CHEMICAL REVIEWS

Review

Cite This: *Chem. Rev.* XXXX, XXX, XXX–XXX

pubs.acs.org/CR

RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview

Jiří Šponer,[*,†] Giovanni Bussi,[*,‡] Miroslav Krepl,[†,§] Pavel Banáš,[§] Sandro Bottaro,[∥] Richard A. Cunha,[‡] Alejandro Gil-Ley,[‡] Giovanni Pinamonti,[‡] Simón Poblete,[‡] Petr Jurečka,[§] Nils G. Walter,[⊥] and Michal Otyepka[§]

[†]Institute of Biophysics of the Czech Academy of Sciences, Kralovopolska 135, Brno 612 65, Czech Republic

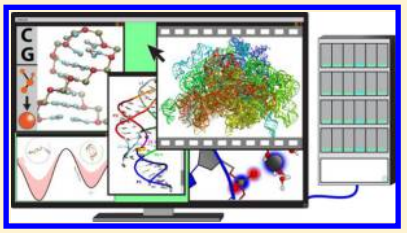
[‡]Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, Trieste 34136, Italy

[§]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacky University Olomouc, 17. listopadu 12, Olomouc 771 46, Czech Republic

[∥]Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Copenhagen 2200, Denmark

[⊥]Single Molecule Analysis Group and Center for RNA Biomedicine, Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

ABSTRACT: With both catalytic and genetic functions, ribonucleic acid (RNA) is perhaps the most pluripotent chemical species in molecular biology, and its functions are intimately linked to its structure and dynamics. Computer simulations, and in particular atomistic molecular dynamics (MD), allow structural dynamics of biomolecular systems to be investigated with unprecedented temporal and spatial resolution. We here provide a comprehensive overview of the fast-developing field of MD simulations of RNA molecules. We begin with an in-depth, evaluatory coverage of the most fundamental methodological challenges that set the basis for the future development of the field, in particular, the current developments and inherent physical limitations of the atomistic force fields and the recent advances in a broad spectrum of enhanced sampling methods. We also survey the closely related field of coarse-grained modeling of RNA systems. After dealing with the methodological aspects, we provide an exhaustive overview of the available RNA simulation literature, ranging from studies of the smallest RNA oligonucleotides to investigations of the entire ribosome. Our review encompasses tetranucleotides, tetraloops, a number of small RNA motifs, A-helix RNA, kissing-loop complexes, the TAR RNA element, the decoding center and other important regions of the ribosome, as well as assorted others systems. Extended sections are devoted to RNA−ion interactions, ribozymes, riboswitches, and protein/RNA complexes. Our overview is written for as broad of an audience as possible, aiming to provide a much-needed interdisciplinary bridge between computation and experiment, together with a perspective on the future of the field.



CONTENTS

Special Issue: RNA: From Single Molecules to Medicine

Received: July 14, 2017

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX–XXX

1. INTRODUCTION

Ribonucleic acid (RNA) is one of the most important and versatile chemical species in molecular biology. According to the original central dogma of molecular biology, messenger RNA (mRNA) is an intermediary carrier of genetic information,[1] whereas transfer RNA (tRNA) is the molecular adaptor used to translate this information into a protein sequence.[2] The translation process is catalyzed by the ribosome, a large multicomponent RNA/protein complex that controls specific pairing of each mRNA with the appropriate sequence of corresponding aminoacylated tRNAs and forms the peptide bonds of the resulting protein.[3−5] Discovery of catalytic RNAs (ribozymes) in the early 1980s demonstrated that RNA can catalyze specific chemical reactions,[6] achieving accelerations comparable to those of many protein enzymes.[7] Determination of the structure of the ribosome,[3,4,8−14] and especially the large ribosomal subunit,[15] revealed that the ribosome itself is an unusually large ribozyme.[16] In addition, these atomistic structures provided a tremendous amount of information about RNA structural motifs and the principles governing RNA evolution. Retrospectively, the discovery of a multitude of catalytic RNAs[6] revolutionized modern biology, demonstrating RNA's unique capacity to simultaneously store genetic

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

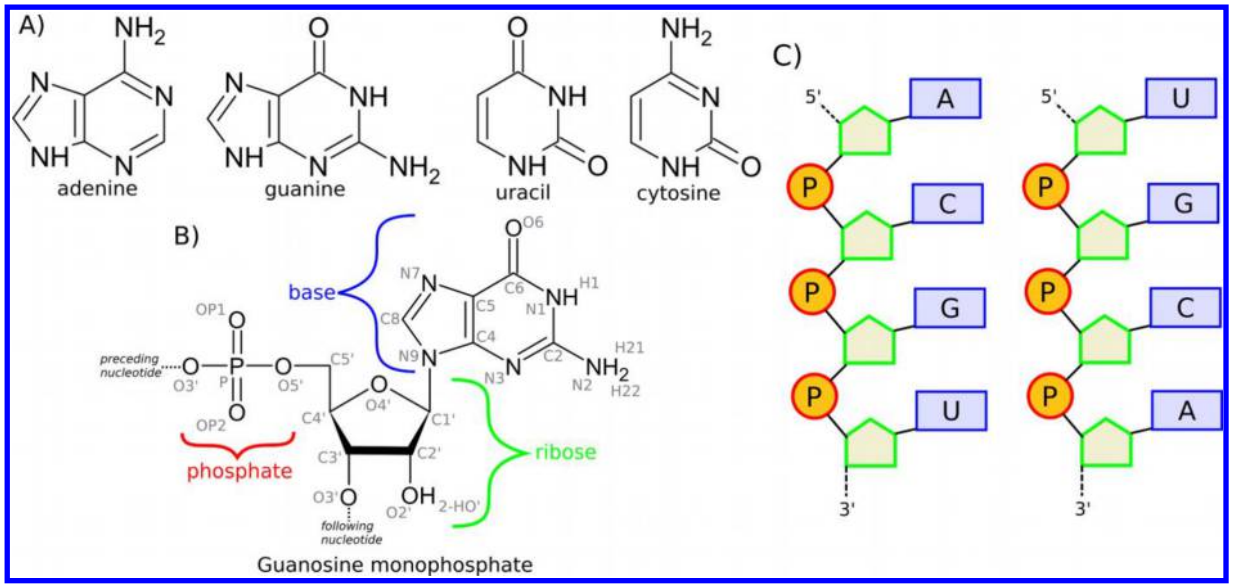


Figure 1. (A) The four aromatic bases found in RNA molecules. (B) Example of a single RNA nucleotide within an RNA chain. The base, ribose sugar, and phosphate are marked by blue, green, and red brackets, respectively. The connectivity to the preceding (upstream) and following (downstream) nucleotides is indicated. The commonly used designations (names) of the individual atoms are written in gray. (C) Schematic representations showing the inherent asymmetry of 5′-ACGU-3′ (left) and 3′-ACGU-5′ (right) RNA sequences. The phosphates, ribose sugars, and bases are stroked with red, green, and blue, respectively.

information (as do DNA genomes) and catalyze chemical reactions (as do protein enzymes), expanding the central dogma of molecular biology and supporting the RNA World hypothesis that RNA-based systems that bridge genotype and phenotype may have been the first forms of life to emerge on Earth.[17−19]

Perhaps unsurprisingly given this functional breadth, as more genomes were sequenced over the last 20 years, many other cellular roles of RNA have been identified. In bacteria, riboswitches ubiquitously control gene expression in response to physiological cues.[20] In eukaryotic and particularly multicellular organisms, RNA is involved in numerous processes essential for the maintenance, regulation, and processing of genetic information, such as RNA silencing.[21−27] RNA can thus be broadly classified into protein-coding and noncoding RNA (ncRNA); the vast majority of RNA transcripts in human cells, numbered at least 80 000 and likely more, belong to the latter category and play fundamental roles in many regulatory processes. Consequently, information on the function of RNA is starting to have impacts well beyond fundamental research. For example, mutations in ncRNAs are involved in cancer and other diseases,[28−32] and riboswitches and other ncRNAs are emerging as attractive targets for novel antibiotics.[33−37]

Because of the rapidly expanding RNA universe, the throughput of experimental tools has quickly become a limitation in our ability to dissect RNA structure−dynamics−function relationships. Computational tools can help overcome this limitation by interpreting the available experimental data, by adding connections between structure, dynamics, and function, and by generating experimentally testable hypotheses. In light of rapidly advancing computational and algorithmic powers, the impact of these tools stands only to further expand. Here, we broadly review the current state-of-the-art of computational treatments of the conformational dynamics and chemical transformations of RNA molecules. After a short introduction to RNA structure and dynamics, we will discuss the methodological constituents of molecular dynamics (MD) simulations. We will start surveying the available force fields, the methods that can be used to access long time scale dynamics, and the available coarse-grained models. We will then discuss how interaction of RNA with ions can be modeled. In the second part of this Review, we will focus on applications of MD to RNA systems. We will start with simple oligonucleotides and continue surveying applications of classical MD to RNA motifs, riboswitches, ribosomes, and RNA/protein complexes. Finally, we will discuss studies of RNA reactivity.

2. PRINCIPLES OF RNA STRUCTURE AND DYNAMICS

RNA is a linear polymer composed of four nucleotide monomers: A, C, G, and U. Each nucleotide consists of a planar aromatic base attached to a ribose unit, a furanose-ring sugar moiety with a 5′-phosphate group (Figure 1). RNA chains are built up by formation of phosphodiester linkages between successive sugar residues, with phosphate groups linking the 3′-carbon of each sugar to the 5′-carbon of the next one, leaving a free (unlinked) 5′-position at one end of the chain (the "5′-end") and a free 3′-position at the other (the "3′-end"). This makes RNA chains asymmetric, so 5′-ACGU-3′ and 3′-ACGU-5′ are different molecules (Figure 1). The 5′-end is considered the beginning of the chain because that is where chain synthesis begins in living organisms.

2.1. Role of the 2′-Hydroxyl Group

The 2′-OH group of the ribose moiety gives rise to profound differences between DNA and RNA. Primarily, this is because it makes RNA less chemically stable than DNA by facilitating self-cleavage reactions. As a result, DNA is better suited for stably encoding large genomes (above ∼10 000 nucleotides in length), whereas RNA is chemically more transient and thus often serves as a dynamic adaptor to changing cellular conditions. In prebiotic chemistry, the balance between RNA cleavage and ligation is assumed to have played a central role in the spontaneous untemplated assembly of the first RNAs, and

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

the subsequent onset of Darwinian evolution based on competition for chemical resources between sequence variants.[38] Because the 2′-OH group is a versatile hydrogen (H−) bond donor and acceptor, it enhances the ability of RNA to form complex architectures unavailable to DNA. The 2′-OH group stabilizes a broad spectrum of non-Watson−Crick (noncanonical) base pairs[39−42] not found in DNA and facilitates compact packing of RNA helices. Evolution has exploited the versatile self-interaction properties of RNA to generate an astonishing diversity of structures capable of countless specific RNA−RNA, RNA−protein, RNA−DNA, and RNA−small molecule or RNA−ion interactions.

2.2. Primary, Secondary, and Tertiary RNA Structures

RNA molecules can be described in a hierarchical manner by analyzing their primary, secondary (2D), and tertiary (3D) structures. Primary structure defines an RNA sequence in terms of its constituent nucleotides, which in addition commonly undergo post-transcriptional chemical modifications (see, e.g., ref 43). 2D structure is defined by the pattern of Watson−Crick (WC) canonical base pairs (A═U and C═G base pairs). These canonical base pairs are particularly stable; they stabilize RNA duplexes by 1−3 kcal/mol per base pair.[44−46] Although RNA molecules are formally single-stranded, most mixed-sequence RNAs form series of short antiparallel canonical double helices by folding back upon themselves to align WC-complementary sequences. In addition to the A═U and G═C base pairs, A-form double helices often contain a significant fraction of G/U "wobble" base pairs.[47]

Canonical double helices alternate with regions of nucleotides that do not form canonical base pairs and so are nominally unpaired. 2D structure is therefore a summary of the adjacent canonical base pairs formed when an RNA molecule folds. Drawings representing the 2D structures often show only the nested canonical base pairs. The remaining nucleotides are frequently shown as unpaired "loops" in 2D plots. However, many of these nominally unpaired nucleotides form non-canonical base pairs. In most naturally occurring RNAs, canonical helices tend to be short, generally containing no more than around 12 consecutive canonical base pairs. Longer stretches of canonical base pairs are probably too monotonous and stable to be useful for the evolution of complex and often conformationally dynamic RNA molecules and RNA-based biomolecular machines; biologically, they are often used to distinguish self from nonself RNA.[48] Canonical helices provide a major thermodynamic contribution to RNA's structural stability, exceeding that associated with secondary structure formation in proteins.[44−46]

Building on the basic architecture defined by the 2D structure, RNA 3D structure is determined by some canonical and many noncanonical base pair interactions that often involve H-bonding through the 2′-OH group. RNA base pairs have been classified into 12 basic families of internucleoside interactions,[39−41,49] with base−sugar and sugar−sugar H-bonds being considered of importance equal to that of base−base H-bonds (Figure 2). This classification was later extended to include sequence-specific base−phosphate interactions,[50] allowing RNA base pairing to be generalized as a network of sequence-specific internucleotide interactions. This broader definition and the classification of base pairs in terms of their isostericity are crucial to understanding the evolution of RNA molecules. The powerful principle of isostericity posits that preserving the exact geometry of base pair interactions (within

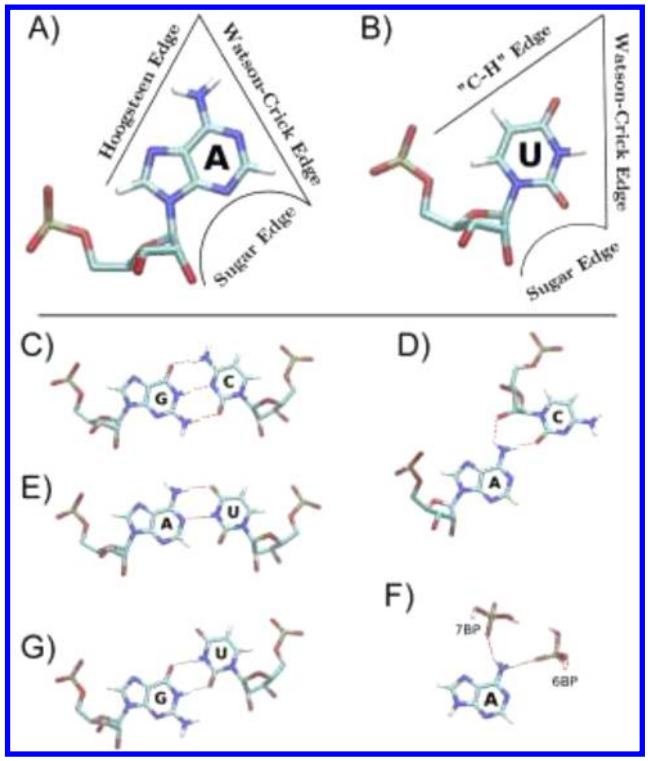


Figure 2. There are 12 basic mutually nonisosteric families of RNA base pairs, created by utilization of the three nucleoside edges and assuming either *cis* or *trans* orientations of the glycosidic bonds. These are complemented by sequence-specific base−phosphate interactions, which involve ∼12% nucleotides in the ribosome. The figure shows the interacting edges of nucleobases (A and B) and examples of interactions between nucleosides (C, D, E, and G) and between bases and phosphate groups (F).[39,50] H-bonds are represented by red dashed lines. (C) and (E) show the canonical base pairs (*cis*-Watson−Crick/Watson Crick, i.e., *c*WW family) formed by guanine and cytosine, and adenine and uracil, respectively. (G) A wobble pair formed by guanine and uracil, which also belongs to the *c*WW family. (D) An example of noncanonical base pair between adenine and cytosine (*trans*-Watson−Crick/Sugar edge, i.e., *t*WS family), which involves H-bonds with the 2′-hydroxyl group of the sugar of pyrimidine. (F) Two possible base−phosphate interactions involving adenine.

double helical regions and in tertiary interactions) is a key requirement governing the evolution and covariation patterns of RNA sequences.[39−42,47,49−56] That is, it embodies the idea that 2D and mainly 3D structures dictate primary structure. In a given sequence position, functionally equivalent RNA molecules from different organisms typically only alternate base pairs of identical (isosteric) geometries; only those sequences that can fully satisfy the isostericity in a given position are selected in the course of evolution.[39,42,49,50,54,55] Although the full set of isosteric possibilities is not always exploited, violations of isostericity are rare. This is because replacing one base pair with another of a different geometry changes the overall topology of the RNA.[39−42,50] Analysis of isostericity is actually a typical application for MD simulation studies, because the technique quite reliably describes shapes of nucleobases and their interactions, allowing realistic studies of base substitutions.

RNA topology is determined by the backbone path, which depends on the position and direction of the backbone segments attached to the bases.[39] This links a local property (base pair geometry) with the global topology, which determines the molecule's biological functionality. Change in

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

the shape of a base pair (i.e., a nonisosteric substitution) within a folded molecule changes the associated backbone path, causing ripple effects across the entire RNA structure. This is the physical reason underlying the isostericity principle.

At first sight, RNA backbone appears sequence-independent and flexible. However, its flexibility is not unlimited. The backbone has a certain number of allowed rotamers (combinations of the consecutive dihedral angles along the chain) that have been classified into approximately 50 suites (Figure 3).[57] Each suite is a set of seven consecutive dihedral

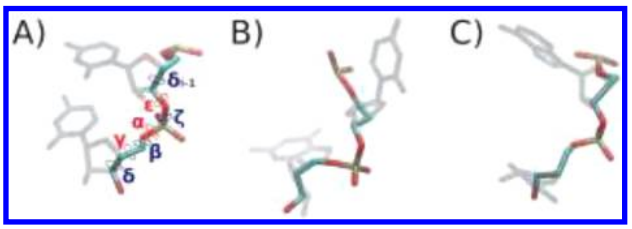


Figure 3. RNA backbone can adopt multiple conformations. Examples of three different conformations (conformational families) of the RNA backbone segments between two consecutive nucleotides (i.e., suites) together with the dihedral angles that define them; note that the numbering of the seven angles defining a suite is phosphate-centered; see the text. Using the notation introduced in ref 57, part (A) shows backbone of canonical A-form (family 1a) with the definition of the torsional angles, while parts (B) and (C) show the structures "found before a bulge" (family 1o) and "between the nucleotides 2 and 3 belonging to an UNCG tetraloop" (family 2[).

angles centered on a phosphate group P_i. It links two nucleotides, i−1 and i, in the covalent chain (the seven angles are δ_{i-1}, ε_{i-1}, ζ_{i-1}, α_i, β_i, γ_i, and δ_i when using the conventional nucleotide-centered numbering or, equivalently, δ_{i-1}, ε_i, ζ_i, α_i, β_i, γ_i, and δ_i when using phosphate-centered numbering). Different suites allow diverse mutual positions of the two linked nucleosides, with the nucleobase positions being further adjusted by the χ glycosidic angles. To a first approximation, 3D structure is thus determined by a complex interplay of sequence-dependent internucleotide interactions that are modulated by intrinsically sequence-independent preferences of RNA backbone. A detailed review of this phenomenon and the associated experimental evidence has been specifically provided elsewhere for those carrying out theoretical studies.[58]

The 12 base pair families are mutually nonisosteric, and, in addition, each family can be further subdivided into several nonisosteric subfamilies.[39] In a folded RNA, every single H-bond matters, and RNA structures are determined much more strictly than typical protein structures. Specifically structured RNA building blocks, or motifs, are characterized by so-called signature H-bonding and stacking interactions that are essential for their formation and dictate their underlying consensus sequences. Consequently, structures are more important and more conserved than sequences. For example, the core of the bacterial ribosome exhibits 72% conservation at the level of sequence identity, but 98% conservation at the level of isostericity.[42] This strictness of RNA structures must be taken into consideration when performing MD studies because failure to account for the isostericity principle may lead to inaccurate interpretation of simulation results (see section 4.2.1). Losses of signature interactions determined by a consensus sequence during MD simulation of a folded RNA can be a sign of structural instability resulting from the approximations of the force field.

As noted above, RNA 2D structures generally consist of short canonical helices punctuated by nominally unpaired segments that are shown as "loops" in planar 2D representations (Figure 4).[41,52−54,59] 2D structure loops consist of one or more strand

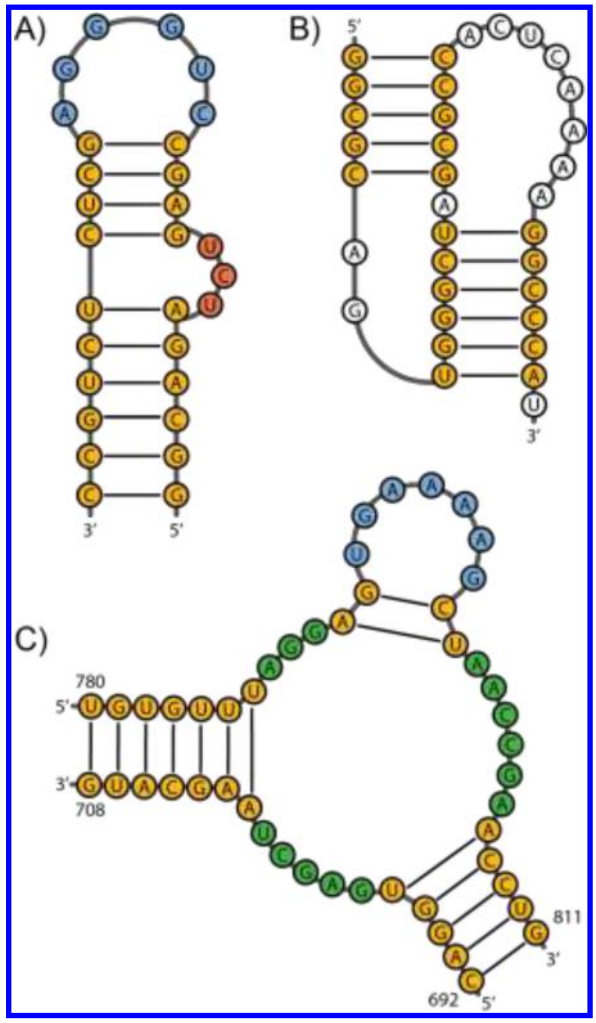


Figure 4. Secondary structure of RNA molecules. Stems (including canonical A═U, G═C, and G/U wobble base pairs; see Figure 2G) are shown in yellow, an internal loop (in this case, a three-nucleotide bulge) is in red, hairpin loops are in blue, and a three-way junction is in green. (A) TAR molecule (PDB 1ANR),[62] (B) MMTV pseudoknot (PDB 1RNK),[63] and (C) fragment of large 50S ribosomal subunit from *Deinococcus radiodurans* (PDB 1NKW).[13] Note that the pseudoknot in B formally consists of two intertwined hairpins.

segments and can be classified into three basic types: (1) hairpin loops formed by single continuous strand segments folded onto themselves at ends of helices; (2) internal loops comprising two strand segments located between two helices; and (3) multihelix junctions from which three or more helices emanate; for more examples, see Figure 3 in ref 41. The term "loop" may evoke the idea of a lack of structure. However, many RNA 2D loops are precisely structured, including most common hairpin loops.[41,52−54,59,60] Nucleotides of structured loops form multiple interactions with each other, and frequently with other (distal) parts of the same RNA or with other molecules. Consequently, the sequences presented as loops in 2D diagrams are often the most interesting and functionally important components of any given RNA molecule. Many of them correspond to recurrent, highly

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

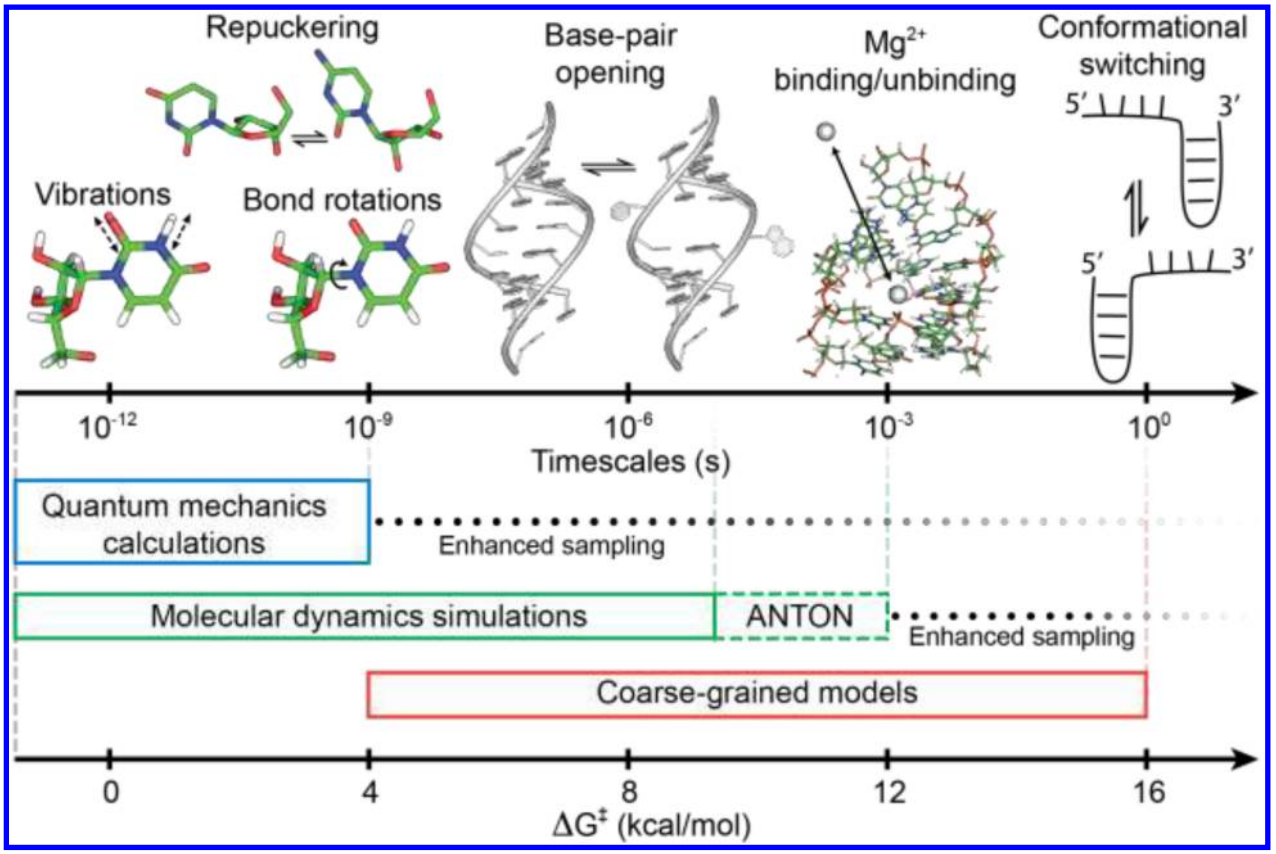


Figure 5. Schematic representation of the time scales that are relevant for RNA structural dynamics. In the upper part, sample structural changes are depicted. In the lower part, the typical time scales nowadays accessible to the techniques discussed in this Review are shown, including quantum mechanical calculations, atomistic explicit-solvent MD simulations, and coarse-grained models (section 3.3). Hardware and software improvements led to an order of magnitude gain every few years in the past. Dedicated hardware such as Anton[73] allows for significantly longer time scales to be accessed in atomistic MD, and massively parallel approaches like the folding@home infrastructure[74,75] allow one to gather large cumulated simulation times (although typically composed of a large number of short trajectories). However, complex conformational changes still remain out of reach of MD simulations. Enhanced sampling techniques (section 3.2) might be used to probe those time scales. The bottom arrow shows typical free-energy barriers involved in these processes. Complex and biologically relevant molecular machines such as the ribosome utilize an endless spectrum of dynamical processes, extending from movements of single nucleotides up to large-scale movements of their whole subunits on a wide range of time scales (section 4.5.3).

specific molecular building blocks and are often called "RNA 3D motifs".[41,52−55,59,61]

The description of the RNA formal structures is completed by pseudoknots that result from (canonical) base pairing between a hairpin loop (or another secondary structure element) and a single distal complementary strand region.[64] In topological terms, pseudoknot formation produces a long base-paired double helix, formed by the combination of two base-paired double helices, where one of the strands is continuous and the other is discontinuous, and two loops that embed the helix within the complete continuous sequence (Figure 4B). Note that by discontinuous strand we mean that the strand lacks continuous covalent bonding along the helix.

As suggested elsewhere,[41] when performing computational chemistry analyses, it is instructive to think of large RNAs as toys composed of intricate, interlocking parts, like jigsaw puzzles, LEGOs, and Russian dolls. Isostericity principle and the dependence of RNA structure on precisely shaped interactions make RNA molecules akin to complex jigsaw puzzles, with natural selection favoring "pieces" that preserve the local shape. Like sophisticated LEGO toys, RNAs then often utilize recurrent modular building blocks that may be, in addition, flexible and dynamic within a given set of topological constraints. Many of the interactions in RNA molecules encode conformational substates that are important for their functional dynamics. Finally, like Russian dolls, RNA architectures are hierarchical. Often, a given structural interaction pattern or motif (with its associated sequence signature) includes subpatterns or submotifs while also being an element of larger RNA molecules.[65] Submotifs are smaller RNA segments that cannot be regarded as independent building blocks that adopt specific 3D folds independently of their context. In other words, a submotif may participate in a number of distinct motifs, and combinations of motifs are used to create larger RNA structures, with astonishing richness of combinatorial possibilities. This may complicate the definition of appropriate model systems for biologically meaningful MD simulation studies.

RNA folding kinetics is hierarchical and sequential, with 2D structure folding occurring usually (but not always[66]) before and independently of 3D structure. This is a consequence of the high thermodynamic stability of 2D with respect to 3D contacts as well as of the fact that the canonical RNA helices are typically able to fold autonomously.[67−69]

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

2.3. RNA Dynamics

As explained above, 3D structures of RNA molecules are largely determined by their sequences. However, as will be discussed in this Review, RNA structures are not static. A single RNA molecule with a well-defined sequence often has multiple accessible 2D and 3D structures that lie within a narrow range of folding free energies and are sampled as a consequence of thermal fluctuations or interactions with proteins and other cofactors that induce or capture specific RNA conformations. The time scales over which these alternate structures form and disappear can range from microseconds for simple base pairing changes to seconds (or even longer) for complex 2D refolding events (Figure 5).[70−72] Because RNA molecules often play multiple roles in biological processes, their native states can be anything from firmly structured folded architectures to intrinsically disordered dynamical single-stranded ensembles.

Despite the tremendous importance of RNA structural dynamics, its experimental characterization is even more challenging than obtaining static structural data. Thus, there have been intense efforts to complement the available experimental techniques by advanced MD simulation approaches. The main goals of atomistic MD simulations are (a) to simulate the structural dynamics of RNA molecules explicitly and in real time to support the interpretation and planning of experimental measurements of such dynamics; (b) to provide reliable and experimentally testable predictions; and (c) to furnish insights that are not obtainable by current experimental methods but which help explain RNA's multitudinous functional roles. As we will discuss, the approximations required in MD sometimes do not allow experimental data to be quantitatively reproduced or predicted. Even in such cases, simulations can still be extremely useful in guiding chemical and physical intuition to design new experiments. Carefully designed MD simulations may often prevent incorrect interpretations of experimental data. MD simulations may be complemented by quantum mechanical (QM) calculations, which can be used to assess the likelihood of specific chemical reaction pathways involving RNA enzymes and to thereby inspire new experimental tests or help interpret existing experimental data on RNA catalysis (section 4.8). Nowadays, QM calculations are often interfaced with molecular mechanical (MM) treatments of distal layers of the RNA that are not directly involved in forming or breaking covalent bonds, to provide context and to capture the impact of conformational dynamics on the reaction probed by QM. QM methods are also indispensable for the parameterization of the MM force fields used in MD simulations (section 3.1). Moreover, larger scale dynamics and conformational changes not accessible via conventional atomistic MD simulations can be studied using various coarse-grained methods, all to be summarized here.

3. MOLECULAR DYNAMICS SIMULATION METHODOLOGIES

3.1. RNA Force Field

The main factor limiting accuracy of MD simulations is the molecular mechanical (MM) force field that defines the relationship between the simulated molecule's exact geometry and its potential energy. Negative gradient of the MM potential energy defines a force for every possible Cartesian configuration of the system. To avoid confusion, we recall that molecular mechanics (MM) here indicates that the forces are computed using some empirical force field, in opposition to quantum mechanics (QM), whereas molecular dynamics (MD) indicates that time-dependence of the atomic positions is simulated.

3.1.1. Basic Force-Field Approximation. When assessing a force field, one must consider its basic approximation separately from its specific parameterization. The basic approximation determines the principal physical limits of a given force-field form. The basic physical limits of MM are very substantial, but their consequences for practical simulations are unknowable a priori and so must be determined by performing simulations of diverse systems. Within these basic limits, one tries to tune the force field via specific parameterizations to obtain the best possible performance for RNA molecules. As one approaches the basic limits of the force-field form, improvements in individual components will increasingly conflict with one another such that improving one aspect of the force field may have adverse effects on other aspects that outweigh the gains. There are a variety of strategies for improving parameterizations (see below), but when applying any such strategy it is important to remember that most critical is the force field's empirical performance in MD simulations. A force field's performance across a set of systems will not be uniform, and will depend on its parameterization and on the extent to which compensation of errors occurs for the systems under consideration. The applicability of a given force-field version may vary between simulated biomolecules, and even between different parts of a single simulated molecule. A textbook example can be found in G-quadruplexes, whose single-stranded loops are described less accurately than the G-stems.[76,77] Improvement of force fields requires the open reporting and analysis of their failures. Unfortunately, many MD publications show a notable reluctance to report force-field failures, and to discuss limitations. This reluctance is rather notorious even in some high-impact journals, especially in short-communication-style papers, because nowadays many such journals seem to strongly prefer short papers that describe "flawless", "hot" research while sidelining conservative works.

3.1.1.1. Force-Field Functional Form. At present, the vast majority of RNA simulations are performed using nonpolarizable force fields whose form is simple and has remained essentially unchanged for several decades (see also ref 78). A notable example is the form used in the force fields of the AMBER simulation package,[79] which are based on the work of Cornell et al. (eq 1).[80] This form describes the covalent connectivity of the investigated system using a set of harmonic springs representing bond lengths and valence angles with force constants k_b and k_a, and equilibrium values r_{eq} and θ_{eq}, respectively, supplemented with sums of dihedral terms characterized by an amplitude V_n, periodicity n, and phase γ. While it would be possible to add higher-order terms to better represent coupled dynamics such as bend−stretch coupling and nonharmonic terms, such a refinement appears unnecessary for biological molecules. Adding higher-order terms would also complicate the parameterization. A consequence of the harmonic covalent bond description is, however, that covalent bond breaking and formation (i.e., chemical reactions) cannot be simulated.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

$$V = \sum_{\text{bonds}} k_b(r - r_{eq})^2 + \sum_{\text{angles}} k_a(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} \times [1 + \cos(n\varphi - \gamma)] + \sum_{i<j} \varepsilon_{ij} \left[\left(\frac{R_{ij}^0}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^0}{R_{ij}} \right)^6 \right] + \sum_{i<j} \frac{q_i q_j}{R_{ij}} \quad (1)$$

Interatomic interactions are approximated by using Lennard-Jones spheres to represent atomic repulsions and dispersion attraction (parameters used for this purpose are the equilibrium interatomic van der Waals (vdW) distance R_{ij}^0 and the potential well depth ε_{ij} in eq 1). Coulombic term is represented by point charges (q_i in eq 1). The charges are localized at the atomic centers and have fixed values; that is, they do not change upon molecule conformational changes and do not respond to external electric fields, including those stemming from solvation. In eq 1, r, θ, φ, and R_{ij} stand for bond lengths, bond angles, dihedral angles, and interatomic distances, respectively. This is the simplest meaningful form for an atomistic empirical potential. Notably, it lacks an explicit term for the polarization energy. Polarizable force fields (i.e., force fields with explicit polarization terms) would have the advantage of being able to reflect the distribution of polarization effects in the studied system. While promising polarizable models are being developed for DNA simulations,[81−83] there are currently no polarizable potentials for RNA simulations in the CHARMM[84] or AMBER[79] force-field families. In addition, despite tremendous efforts, the accuracy of polarizable DNA force fields does not currently surpass that of the best pair-additive alternatives, suggesting that tuning polarizable force fields will be a tedious task whose difficulty is comparable to that of tuning fast QM methods for biomolecular computations.[85,86] We therefore focus on nonpolarizable RNA force fields in the following discussion. Still, we consider the development of polarizable force fields to be of utmost importance because, in our opinion, pair-additive force fields are reaching the limits of their capabilities and are starting to restrict our ability to fully exploit continuing advances in computer power and sampling methods when studying nucleic acids.

The force-field description in nonpolarizable force fields is unphysical and, by definition, neglects many effects (such as all types of polarization and charge-transfer effects) by definition. Because these force fields ignore all nonadditive contributions, they are called "pairwise additive force fields". The neglected contributions are in part effectively (implicitly) accounted for in the parameters of the explicit force-field terms, mainly the partial charges, so that such force fields are also called "effective force fields". A large part of the neglected contributions is compensated for by tuning of dihedral force-field terms.

3.1.1.2. Fitting the Parameters. Bond and angle parameters can be derived from equilibrium distances and angles observed in X-ray diffraction data, while microwave and IR spectroscopies can be used to obtain stretching and bending force constants. The bonded terms can also be derived via high-level QM reference data. vdW radii and well depths can be derived by matching experimental densities and heats of vaporization,[87] benchmark electronic structure calculations, and/or through ad hoc assumptions. There are two commonly used approaches for parameterizing the electrostatic term: (i) parameterizing the atomic charges through fits to QM-derived electrostatic potentials (ESP),[80,88] as ESP fits are typically used for AMBER force fields, or (ii) by fitting the force field to reproduce QM computations of interaction energies for small molecular fragments interacting with ligands such as water molecules, as employed in the parameterization of CHARMM force fields.[89] ESP-based methods originally used simple ESP fits. Later, improved ESP protocols were introduced, including restrained ESP (RESP) fits to eliminate spurious effects such as artificial growth in the charges of buried atoms.[88,90] Artifacts arising from the orientation of rotatable groups may be suppressed by utilization of multiple conformations for fitting.[91]

The most arbitrary task in force-field parameterization is the fitting of the dihedral potentials.[92−98] Note that the dihedral terms, despite often being fitted with the help of QM potential-energy scans, are not related to any real electronic structure effects and thus represent ad hoc functions used for the final tuning of the force-field behavior. Dihedral parameterizations have very large impacts on the simulations, as demonstrated by the RNA glycosidic potential χ: small changes in the shape of the χ potential in the *anti* to high-*anti* regions can have dramatic effects on RNA simulations (Figure 6).[94] An inappropriate parameterization may completely destabilize the native A-RNA structure (giving rise to ladder-like artifacts[99] in some cases, cf., Figure 7) because of cumulative effects in longer RNA sequences.[94] Consequently, most refinements of established force fields for nucleic acids as well as proteins have been based on modifications of the dihedral potentials. Because the parameters of dihedral potentials are fitted after all of the other parameters, any significant change in the other terms would require a subsequent retuning of the existing dihedral potentials.

Presently, dihedral potentials consist of one-dimensional potential energy profiles along individual torsions, so there is no explicit coupling between torsions. Although several parameterization procedures that will be discussed later (see section 3.1.2.2) allow multiple torsions to be simultaneously parameterized, the final potential is still a sum of 1D profiles. However, in real molecules, the neighboring torsions are nontrivially coupled through electron density redistributions upon rotations, which would imply nonadditivity of dihedral parameters. Studies on the importance of these nonadditivities are rare. It has been shown that the magnitude of dihedral nonadditivity in dimethylphosphate can be as large as 1 kcal/mol.[102] An efficient way to overcome the lack of nonadditivity that has been incorporated into the CHARMM protein force field is a grid-based two-dimensional energy correction map (CMAP).[96,103] CMAP was designed to account for nonadditivity of the neighboring torsions observed in QM calculations. However, final CMAP corrections for proteins were obtained by empirically adjusting the map such that the MD simulations reproduce experimental data. CMAP was introduced to better describe protein backbones,[96,103] and proved useful for accurately simulating proteins using CHARMM force fields.[97] Interestingly, CMAP-type corrections are not used to model DNA and RNA even in the latest CHARMM36 and CHARMM polarizable nucleic acid force fields,[81,98] which may indicate that such corrections are either not vitally important in nucleic acid simulations or do not appreciably improve performance or are too difficult to parameterize. The AMBER force fields, which are often considered to provide more accurate descriptions of DNA and RNA molecules than other alternatives, do not include any CMAP-type corrections.

3.1.1.3. Physical Limitations. Although the force-field terms look intuitive, they are in fact unphysical, even though they are

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

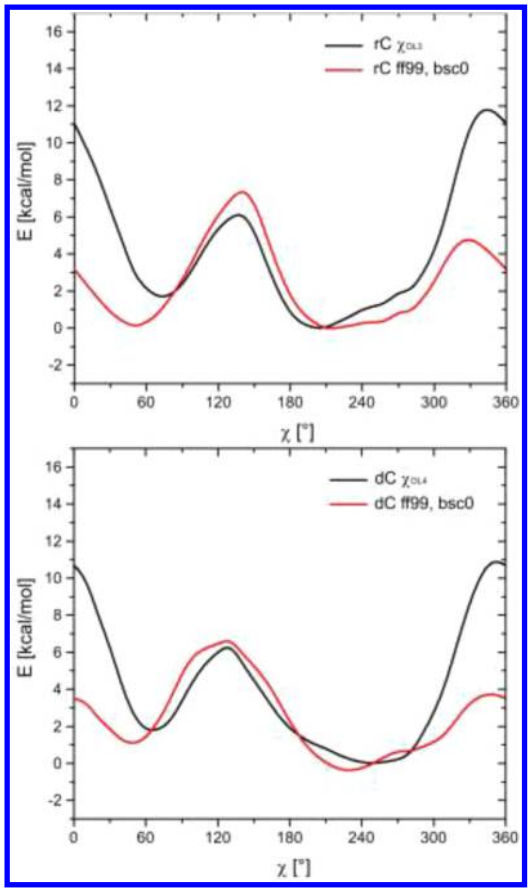


Figure 6. (Top) Total potential-energy scan along the χ dihedral angle (calculated with inclusion of Poisson−Boltzmann continuum solvation energy contribution) obtained for a cytosine nucleoside model with the RNA χ_{OL3} glycosidic potential[94] and the original ff99[90] parameters. Small changes in the *anti* and high-*anti* regions (i.e., tuning of the dihedral profile between ~180° and 260°) have major effects on the simulations. The χ_{OL3} refinement successfully prevents the simulated molecules from adopting spurious ladder-like structures (Figure 7); the ladder is more stable than A-RNA with the original force field. Importantly, χ_{OL3} minimizes adverse flattening of the A-RNA due to reduction of the inclination when refining the *anti* and high-*anti* χ regions, which is a problem of some other χ angle modifications (for more details, see sections 3.1.2.1, 3.1.2.2, and 3.1.2.3). (Bottom) χ_{OL4} DNA potential is shown for comparison.[100] Adapted with permission from ref 101. Copyright 2014 American Chemical Society.

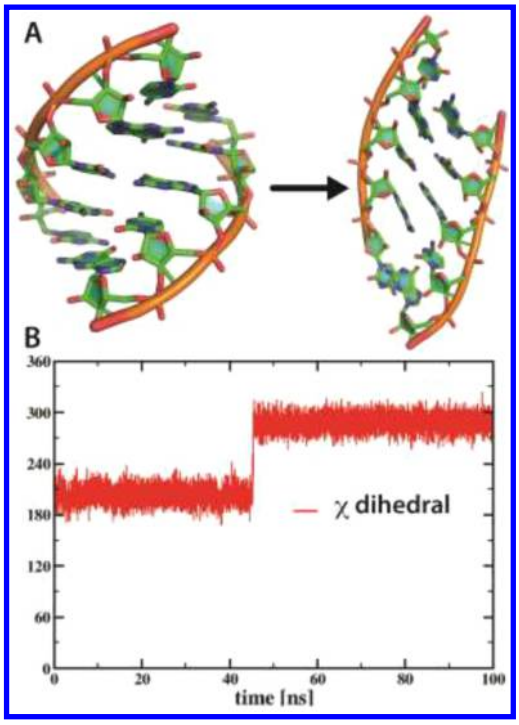


Figure 7. Typical example of a formation of a ladder-like structure in ff99 or bsc0 RNA simulations. The force-field artifact is characterized by (A) loss of helical twist accompanied by (B) collective shift of the glycosidic torsions of all nucleotides from the *anti* to high-*anti* region (see section 3.1.2.1).

motivated by physics. For instance, Coulomb's law is fully physical, but the atomic charges used in the force-field formulas are not observables as there is no QM operator for atomic charges. Thus, they cannot be derived from the first-principles of quantum mechanics, and they are not measurable by any (even hypothetical) experiments. Partial atomic charges are entirely arbitrary quantities that do not exist in nature. It is therefore pointless to discuss the values of individual atomic charges, although such discussions do occasionally appear in the literature. However, the complete set of atomic charges of a given molecular fragment can be tuned to reproduce the ESP around the fragment or some other real physical property of interest. ESP is an unambiguously defined physical property of the molecule that in principle is measurable.

Use of ESP fitting of atomic charges is probably one of the key reasons for the success of the AMBER nucleic acid force fields based on the seminal 1995 parameterization of Cornell et al.[80] ESP charges lead to a very good description of the electrostatic contributions to H-bonding and stacking interactions, as demonstrated by QM benchmark calculations.[104] This is because bases are generally flat and rigid, with the exception of pyramidalization of amino groups, an effect not accounted for in the parameterization.[105,106] However, success of ESP charges for base−base interactions is not transferable to the description of nucleic acid backbone due to its flexibility. In addition, the backbone is a polyanion, which further complicates its description. A given fixed set of backbone point charges cannot simultaneously describe the ESPs for different nucleic acid backbone conformers (backbone families). This problem is especially severe for RNA, which has dozens of biochemically relevant backbone families (Figure 3).[57] Attempts to resolve this issue by basing the ESP fits on weighted averages of multiple conformations have not improved performance,[58,85] and a better description of the RNA backbone will probably require use of a polarizable force field or some kind of conformation-dependent electrostatic term. It thus appears that a balanced description of diverse nucleic acid backbone families perhaps cannot even be obtained simply by tuning dihedral potentials.

Similarly, despite seeming well-defined at first glance, atomic radii are purely arbitrary parameters. The apparent "size" of an atom depends on its chemical environment, which affects its electronic structure. The force-field radius of an atom is an empirical parameter that is selected as the best compromise with which to describe molecular interactions in different chemical situations. For example, a hydrogen atom attached to an H-bond donor looks very small (hidden inside the H-bond donor and having zero radius) when involved in an H-bond but large when pointing toward a nonpolar chemical group.[107]

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Therefore, any parameterization with fixed radii is necessarily a crude compromise; real atoms are not van der Waals spheres.

This unphysicality is why the point charges and atomic radii of a given atom in different force fields can differ significantly, and there is no way of determining a priori which of them is more realistic. Dihedral potentials may be considered the perhaps most unphysical part of a force field. These formally intramolecular terms are used to compensate for major flaws in the description of the intermolecular nonbonded terms.[94]

Limitations of the force-field description can be demonstrated by comparing MM and QM calculations, because the latter can be considered to reflect properties of real molecular fragments. MM and QM (i.e., real) molecules are very different objects. For example, base pair formation at the MM level is a result of complementarity of the unperturbed ESPs, balanced against the vdW Waals term. Real base pairing is a rich phenomenon involving mutual electronic structure redistributions (adaptations) of the interacting bases and communication between their molecular orbitals. Hallmark of H-bonding between bases is the elongation of the X−H covalent bonds in the H-bond donor by up to ∼0.04 Å, clearly demonstrating a change in the electronic structure.[108] Existing force-field terms cannot reproduce such changes in the electronic and geometric structures of the interacting monomers; the MM bases are excessively rigid. Although force fields do allow for some elongation of the X−H bonds, its origin is different from that in the QM description of real H-bonds. In the MM description, X−H elongations arise only from the electrostatic interactions in the point charge approximation, which are counterbalanced by the harmonic springs used to model the X-H bonds. In addition, in the course of MD simulations, the X−H elongations are usually completely eliminated by rigid constraints[109] to enable larger integration time steps, further compromising the MM description of H-bonding; for an illustration of MM and QM/MM descriptions of H-bonding at protein/RNA interface, see ref 110. Further, the monomers in base pairs come into such close proximity that divergence from the MM point charge model may become a significant problem.[111] The lack of physicality in the description of H-bonding is a major problem when balancing the different types of H-bonds that exist in RNAs, the impact of which has only recently been recognized. Another notorious problem in MM simulations of RNA relates to the inclusion of divalent ions, because real divalent ions are hubs for major networks of polarization/charge-transfer effects that significantly influence the electronic structures of all molecules in their first ligand shell and propagate further into the surrounding environment (see section 3.4.2). The currently used force-field approximations neglect all of these effects.

Importantly, the statement that force fields are unphysical says nothing about their ability, after suitable calibration, to mimic certain physical properties of the studied systems. The only thing that matters in this respect is the final performance of the complete force field. However, the lack of a physical basis means that a given type of a force field has certain principal accuracy limits beyond which further tuning becomes impractical. This explains why improvements in force fields have so slowly and intermittently arisen over the last two decades, contrasting the steady and very substantial improvements in QM methodologies.

Simplicity of force fields affects the flexibility of simulated RNA molecules. Their harmonic terms and pair-additivity produce potential energy surfaces with considerably less flexibility than those derived from the wave functions of the QM description. Thus, while we cannot currently directly compare MM and QM simulations of nucleic acids, we would expect MD simulations to underestimate the richness of the local dynamics within a given conformational basin when compared to real (QM) molecules. The previously discussed deformation of monomers upon H-bonding illustrates this point. This is consistent with the available QM calculations on nucleic acid building blocks,[86,112−114] the stiffness of B-DNA in ultralong MD simulations,[115] and the increased flexibility observed in MD simulations of proteins using polarizable force fields.[116] However, despite their lack of local flexibility, classical force fields may paradoxically cause an excess of flexibility in the sampling of different substates on the folding landscape. A hallmark of reported folding simulations on tetranucleotides and tetraloops (see sections 4.1, 4.2.2.3, and 4.2.2.4) is the coexistence of many different conformations with significant populations, which is not generally supported by experiments. This may indicate that the force field is not only unable to find the correct native basins, but also predicts too many simultaneously populated structures; it appears that current force fields tend not to adequately separate the global minimum from the rest of the folding landscape for many systems, although more research will be needed to confirm this observation.

A key question is "can we improve MD simulations by making the force-field form more physical?" Natural candidates for such advancement are polarizable force fields.[81−83] However, the answer is not straightforward. In general, a more complex force field is more difficult to balance. It therefore remains to be seen whether the much-needed polarizable force fields will yield the desired improvements in simulation quality; their complexity may simply increase the occurrence of conflicting imbalances on the potential energy surfaces without adequately improving the physicality of the description. This risk is underscored by the notoriously poor performance of fast low-cost QM methods for biomolecular fragments.[85,86]

Despite their weaknesses, we suggest that the atomistic force-field simulations will remain the only viable computational technique for studying the atomistic structural dynamics of biomolecules for the foreseeable future. It is not even clear that more sophisticated polarizable force fields will surpass the performance of pair-additive force fields. On the other hand, coarse-grained methods (see section 3.3), while able to provide important insights, will always lack the atomistic resolution required in many analyses, especially for RNA with its rich spectrum of functionally and evolutionarily important interactions (see section 2.2). On the other hand, despite the impressive development of QM methods in recent decades, good quality QM methods remain prohibitively slow. Attempts to use QM calculations to study large blocks of nucleic acids (4−11 nucleotides) revealed that series of gradient optimizations in either implicit solvent or explicit solvent modeled using QM/MM methods are generally not sufficient to study the relative energies of different RNA or DNA conformations, even when using high-quality QM methods.[86,113,114] Although new insights were obtained in some cases, the calculations were typically affected by substantial noise in the calculated energies due to the complexity of the potential energy surface.[117] A point that is not always acknowledged in the QM literature is that potential energies cannot be unambiguously linked to free

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Table 1. List of RNA Force Fields Based On the Original Cornell et al. Parameterization

force field (original/recommended name)	some alternative names	modifications introduced	full composition	motivation and outcome
Force Fields Distributed in the AMBER Package at the Time of Writing This Review				
ff94[80]	parm94		ff94	
ff98[130]	parm98	Pucker and χ	ff94 + P,χ	increases twist and pucker, rather small change as compared to ff94
ff99[90]	parm99	Pucker	ff98 + P	improves pucker, rather small change as compared to ff94
parmbsc0[95]	bsc0, ff99bsc0	α/γ_{bsc0}	ff99 + α/γ_{bsc0}	suppresses non-native α/γ conformations but may also penalize native γ-*trans* conformations[132]
AMBER99χ[136]	χ_{YIL}	χ_{YIL}	ff99 + χ_{YIL}	improves χ *syn*/*anti*/high-*anti* balance and incidentally also suppresses "ladder-like" structures, albeit at the cost of excessively reducing A-form inclination[137]
ff99bsc0χ_{OL3}[94a]	χ_{OL3}, χ_{OL}, OL3	χ_{OL3}	ff99 + α/γ_{bsc0} + χ_{OL3}	suppresses "ladder-like" structures without side effects, improves χ *syn*/*anti*/high-*anti* balance
Some Other Cornell et al.-Based RNA Force Fields				
AMBER99TOR[138]		β,ε,ζ	ff99 + χ_{YIL} + $\beta,\varepsilon,\zeta_{YIL}$ + α/γ_{bsc0}	attempted full dihedral reparameterization as compared to ff94, suboptimal performance for canonical A-RNA[58]
Chen–Garcia[120]		vdW of bases, χ; standard vdW combination rules are broken	ff99 + vdW_{CG} + χ_{CG}	weakened stacking (desirable behavior), overly strong hydrogen bonding between bases (undesired side effect),[139,140] excessive reduction of inclination[140]
Aytenfisu–Spasic–Stern–Mathews[141]		$\alpha,\beta,\gamma,\varepsilon,\zeta,\chi$	ff99 + $\alpha,\beta,\gamma,\varepsilon,\zeta,\chi_{ASSM}$	another complete dihedral reparameterization, not fully tested

[a]Note that this force field is often cited as ff10, ff12SB, or ff14SB, but this is a notoriously confusing practice because ff12SB and ff14SB denote protein force fields.[142] Use of these and other (such as ff03) internal AMBER code abbreviations instead of the correct original force-field names is often accompanied by nonsensical citations to the protein force-field parameterization papers containing no nucleic acids parameters and done by groups not parameterizing nucleic acids force fields. Such miscitations make it difficult or even impossible for the reader to determine which nucleic acid force field was used. Especially any usage of the "SB" tag in connection to nucleic acids force fields should be avoided.

energies, as the latter require Boltzmann sampling (averaging); see also section 4.3.3.[118,119]

The sampling problem in large-scale QM calculations is in fact reminiscent of problems with early MM calculations before the arrival of MD. In addition, testing of available fast (low-cost) QM methods has shown that they are in many respects less accurate than well-calibrated force fields for studying nucleic acid fragments.[85,86] Because low-cost QM methods are also still much slower than force field calculations, it is not clear whether any low-cost QM method will be able to achieve satisfactory sampling in the foreseeable future. These issues underscore the point that the different available computational methods each have their own advantages, limitations, and ranges of applicability, and thus complement each other.

An obvious question is "what are the most serious problems affecting the performance of current RNA force fields?" Several answers have been suggested. It is possible that the conceptual underpinnings of fixed-charge models make them incapable of correctly describing different RNA backbone families.[58,85] In addition, the backbone is anionic, and anions are polarizable species because their electron clouds extend far away from their atomic nuclei. This point is supported by several QM benchmark studies.[58,85] Another problem could be over-stabilization of base stacking,[120−122] although an unambiguous comparison with experiments is difficult due to mutual inconsistency in the experimental stacking free-energy data.[122,123] Overstabilization of stacking may be surprising at first glance because QM benchmark studies suggest that the AMBER parameterization offers quite good description for base stacking.[104,119] However, quantitative QM benchmarks indicate that the AMBER force field may indeed somewhat overestimate the energetics of base stacking.[124] It is nevertheless important to emphasize that QM calculations characterize the accuracy of a force field in terms of its ability to reproduce the direct (intrinsic) stacking interaction between two nucleobases, that is, the potential energy of stacking in the absence of other molecules.[118,119] In contrast, the stability of stacking in MD simulations is determined from the populations of stacked and unstacked conformations, and thus the free energies of stacking.[119] The free energies of stacking depend on several factors, including the balance between the intrinsic stacking energies and the solvation energies.

When considering all of the available data, it is becoming evident that force fields have difficulties with properly balancing hydration against the diverse interactions that are important in RNA molecules such as stacking, base pairing, and many other types of H-bonds (such as base−phosphate interactions). As a result, some interactions appear to be understabilized while others are overstabilized. Moreover, the degree of over- or understabilization may be nonuniform and context-dependent. A concomitant effect of the overall lack of balance is sensitivity of nucleic acid simulations to the chosen water model.[125,126] This has prompted suggestions that the key to better nucleic acid simulations could be to develop better parameterizations of water models.[125,126] While we readily acknowledge the importance of good water models, we do not believe that tuning water models alone will be sufficient to significantly improve the performance of nucleic acids force fields because existing water models suffer from all of the inherent limitations of the pair-additive approach. It is difficult to imagine that their refinement could fully compensate for the intrinsic inaccuracies of the solute biomolecular force field. For example, while the OPC water model[127] was shown to somewhat improve the simulations of RNA tetranucleotides[126] and free-energy computations of stacking in nicked B-DNA,[125] the same water model appears to worsen the structural stability of short G-quadruplex stems.[128] The force-field inaccuracies are caused by a number of mutually interrelated problems.

In view of all of the approximations, it was suggested that general force-field refinements could be complemented by

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

structure-specific force-field modifications targeting selected RNA molecules (see the hydrogen-bond fix potential in section 3.1.2.5).[129]

3.1.2. AMBER Nucleic Acid Force Fields. Most widely used nucleic acid force fields are based on the seminal Cornell et al. AMBER 1995 parameterization,[80] which is commonly abbreviated as ff94. The success of this force field stems from its parameterization of the electrostatic term by fitting the charges to reproduce the electrostatic potential around nucleic acid building blocks (section 3.1.1.3). It was later modified, yielding the ff98[130] and ff99[90] variants, by slightly tuning the pucker and χ dihedral profiles (Table 1, see also Figure 3 in ref 78). The prime purpose of these modifications was to eliminate the under-twisting of B-DNA observed with ff94, but the changes were only moderately successful in this respect. Nevertheless, many subsequent reparameterizations took ff99 as their starting point rather than ff94.

Long simulations have shown that the ff94−ff99 force fields do not provide acceptably stable DNA and RNA trajectories. While most older studies were unaffected by this problem because of their short time scales, some papers have inevitably presented results based on corrupted trajectories that were either not noticed or not reported by the authors. Increasing awareness of these problems prompted two key refinements of the AMBER dihedral potentials.

3.1.2.1. Toward Basic Stability of DNA and RNA Simulations: bsc0 and χ_{OL3} Refinements. The 2007 bsc0 version[95] of the AMBER force field reparameterized the α/γ dihedrals to prevent the collapse of B-DNA due to the accumulation of irreversible non-native γ-trans backbone dihedral states (cf., Figure 4 in ref 95). Omission of the bsc0 correction in DNA simulations inevitably leads to formation of major artifacts on a time scale of ∼10 ns. Regarding A-RNA, the γ-trans states formed only reversibly when using the original force field, but upon forming they temporarily and severely reduced the system's helical twist.[131] The bsc0 reparameterization disfavors this behavior and is therefore also beneficial for RNA simulations. While it probably overpenalizes the γ-trans backbone substates in many RNA structures, no better α/γ correction has yet been proposed.[132]

The 2010 χ_{OL3} RNA-specific correction[94,133] reparameterized the χ dihedral to suppress *anti* to high-*anti* χ shifts in RNA, which (among other problems) caused simulated molecules to undergo sudden irreversible transitions into entirely untwisted ladder-like structures (Figures 7 and 8).[94,99,133,134] These RNA ladders usually form more slowly than γ-trans artifacts in B-DNA. However, the χ imbalance becomes detrimental in long RNA simulations and studies on folding. For instance, a folded riboswitch collapsed within ∼0.5 μs[134] (Figure 8), and ladders form even more rapidly in simulated tetraloops.[133] The χ_{OL3} version also improves the description of the RNA *syn* region and the *syn*/*anti* balance. Achieving an accurate balance of the *syn*/*anti*/high-*anti* regions in the force field required consideration of conformation-dependent solvation effects in the course of parameterization;[135] this novel approach was first applied in the derivation of the χ_{OL3} potential.[94]

3.1.2.2. Other RNA Dihedral Potential Reparameterizations. The χ reparameterization of Yildirim et al. (χ_{YIL}),[136] which was originally designed to correct the *syn*/*anti* balance on the basis of nuclear magnetic resonance (NMR) data, has achieved performance similar to that of χ_{OL3}. It was also later shown to eliminate ladder formation.[137] However, it slightly over-destabilizes the high-*anti* region (to a greater degree than

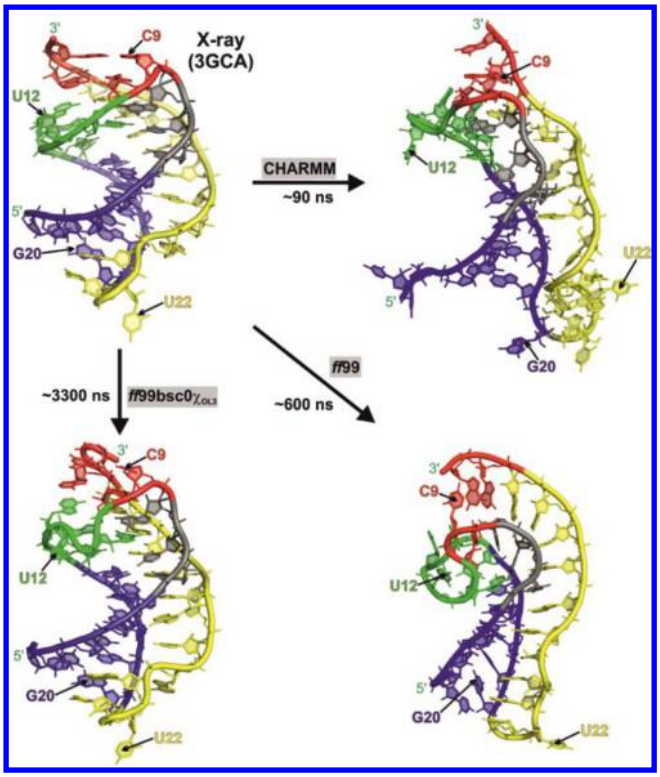


Figure 8. Simulations of preQ$_1$ riboswitch aptamer starting from an X-ray structure upon removal of the ligand, showing the most severe artifacts that may occur in RNA MD simulations. While the χ_{OL3} force field shows a stable trajectory with just local dynamics, the ff99 simulation collapses to the ladder-like structure (the bsc0 correction would not improve this behavior) and the CHARMM27 simulation experiences extended fraying; the latter behavior was only partly corrected by CHARMM36. The difference between the χ_{OL3} simulation and the experimental structure reflects an intricate mixture of genuine dynamics of the system, effects due to removal of crystal packing deformations, and contributions from more subtle force-field imbalances. Reprinted with permission from ref 134. Copyright 2012 American Chemical Society.

χ_{OL3}) and rather considerably reduces the inclination of A-RNA.[137] The original χ_{YIL} force-field version excluded the bsc0 correction, which might be considered suboptimal. Subsequent AMBER99TOR version[138] reparameterized the remaining β, ε, and ζ dihedrals, which (if combined with the bsc0 α/γ correction) would yield a force field with completely reparameterized RNA dihedrals. Unfortunately, AMBER99-TOR performs suboptimally for A-RNA (see Figure 12 in ref 58).

Alternative RNA force-field dihedral potential modifications have been proposed. First, Gil-Ley et al. made an attempt to fit dihedral potentials based on the distribution of dihedral angles taken from a database of crystallographic structures[143] (see section 4.1). Second, Cesari et al. used a maximum entropy approach to fit some of the backbone torsions to solution-phase experimental data,[144] with promising results. In particular, the force field was trained to reproduce NMR data for nucleosides and dinucleoside monophosphates. The number of parameters used in the fit was significantly smaller than the number of available experimental data points to avoid overfitting. The corrections were then validated on tetranucleotides, resulting also in this case in an improved agreement with NMR experimental data. Notably, both the training and the validation

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

were performed using as reference solution NMR data, which is expected to describe structural dynamics of flexible motifs better than an individual crystallographic structure. Lack of NMR data for the phosphate conformation makes it difficult to train and assess the torsional potential for α and ζ angles. Possible problems in these torsions might be related to the occurrence of intercalated structures in simulations of tetranucleotides (see section 4.1). Third, a simultaneous reparameterization of most of the backbone dihedrals (α, β, γ, ε, ζ, χ) was performed by Aytenfisu et al., based on QM calculations on a training set of structures extracted from the PDB. Their modifications were tested on selected tetraloops, tetranucleotides, and RNA duplexes, with promising results,[141] albeit with only limited sampling of the tetranucleotide simulations. A particularly interesting aspect of all of these works is that all dihedral potentials were fitted simultaneously, at variance with the usual procedure where one dihedral angle at a time is modified. None of these new modifications has yet been tested extensively, and so all of these force fields should be considered experimental and only used with this caveat in mind. The same holds also for the subsequent attempt to modify the α/γ potential.[145]

3.1.2.3. The Split between DNA and RNA Force Fields. The bsc0 and χ_{OL3} corrections are presently the most salient modifications of the original ff94 force field, and have positive synergistic effects when applied to RNA simulations. However, the χ_{OL3} refinement cannot be used for DNA because RNA and DNA have conflicting requirements on the χ potential in the most critical *anti* to high-*anti* region. This divergence necessitated the development of separate Cornell et al.-based force-field variants for simulating DNA and RNA. χ_{OL4} parameterization (DNA χ dihedral, published in 2012, Figure 6)[100] is the DNA counterpart of χ_{OL3}, which improves the behavior of canonical and noncanonical DNAs. $\varepsilon\zeta_{OL1}$ (ε and ζ, 2013) parameters[146] were introduced to tune the BI/BII populations in B-DNA, and ultimately corrected the B-DNA helical twist, which was described incorrectly by earlier force fields. $\varepsilon\zeta_{OL1}$ was also reported to improve description of the structural dynamics of the hairpin ribozyme's catalytic center,[147] although other tests have shown that it worsens the description of some other RNAs. We therefore do not recommend its use in RNA simulations.[132] Bsc1 (which combines bsc0 with reparameterized pucker, χ, ε, and ζ dihedrals; 2015)[148] is another DNA-specific dihedral reparameterization that achieves performance similar to that of bsc0$\varepsilon\zeta_{OL1}\chi_{OL4}$. This force field is also not recommended for use with RNA. The final reparameterization of the AMBER dihedrals for DNA was β_{OL1} (β dihedral, 2015),[149] which is especially useful for noncanonical DNAs. The (ff99)-bsc0$\beta_{OL1}\varepsilon\zeta_{OL1}\chi_{OL4}$ force field has been abbreviated as OL15, and represents a complete reparameterization of the DNA dihedral potentials as compared to the original ff94 version.

Among the various dihedral refinements of the original ff94 force field, bsc0 and χ_{OL3} stand out because they are critical for stabilizing DNA and RNA simulations, respectively. Reparameterizations of the remaining dihedrals achieved some additional improvements for DNA, and OL15 probably represents the upper limit of what can be achieved by reparameterizing uncoupled dihedrals for DNA. Despite these improvements, the force field is far from flawless. For RNA, the combination of χ_{OL3} and bsc0 appears to be the best compromise at present because (as noted above) the available reparameterizations of other dihedrals may have undesirable side effects. The fact that state-of-the-art force fields for RNA and DNA require different dihedral parameters is a further confirmation that these parameters are nonphysical and only used to compensate errors arising from other missing interactions. For instance, the atoms involved in the glycosidic torsion are expected to be chemically equivalent in RNA and DNA.

3.1.2.4. Toward Tuning of the Nonbonded Terms. Further tuning of the force field would require modification of the nonbonded terms and consideration of better solvent models. As a first attempt in this direction, Chen and Garcia modified the balance between stacking, H-bonding, and solvation. This involved (i) rescaling the vdW parameters of the nucleobases, and (ii) adjustment of vdW combination rules for base-water interactions (nonbonded fix, NBfix).[120] The main aim was to eliminate the overstabilization of stacking. It was accompanied by a modification of the χ dihedral potential resembling those by χ_{OL3} and χ_{YIL}; the bsc0 α/γ refinement was excluded (Figure 9). The modified force field offered partial improvements in simulations of RNA tetraloops, albeit of a lesser magnitude[129,139] than was originally suggested,[120] and some side effects have been later reported (see section 4.2.2.3).[139,140] While the NBfix procedure may seem unphysical, it is entirely justified because NBfix increases the scope for tuning the force field's overall performance by balancing the treatment of solvation effects.

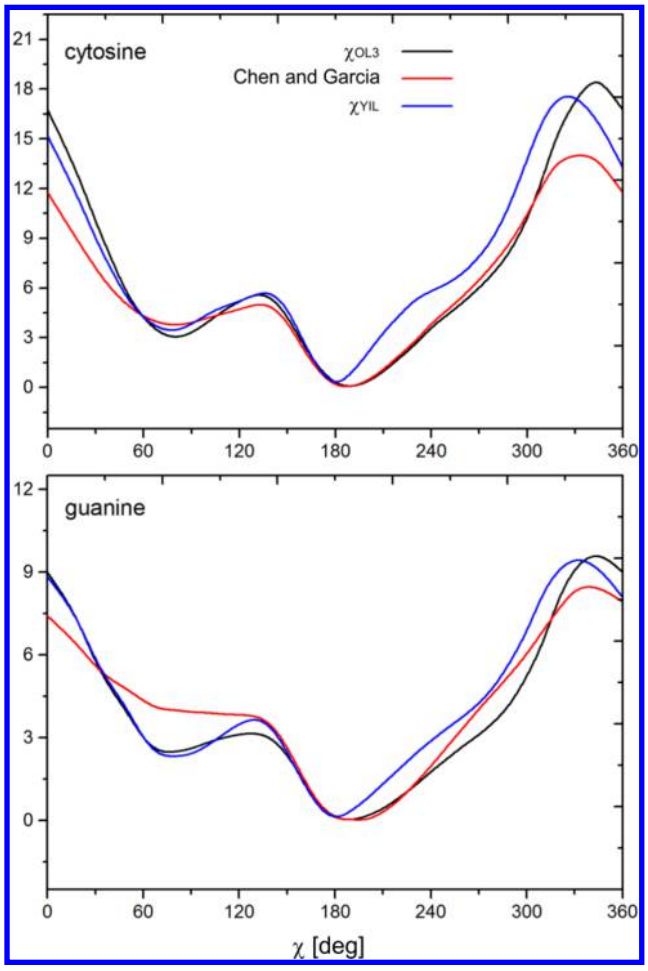


Figure 9. Comparison of full χ potential energy profiles in χ_{OL3}, χ_{YIL}, and Chen and Garcia force fields, obtained for nucleoside models.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Another work considering alternative vdW parameters was published by Bergonzo et al.[126,139] Instead of a new reparameterization, the authors opted to test alternative phosphate parameters proposed by Steinbrecher et al.[150] in combination with the OPC water model.[127] Steinbrecher's parameters were originally derived for bioinorganic phosphates.[150] Bergonzo et al. showed that using the larger, less hydrophilic phosphate groups described by these parameters yielded a better description of tetranucleotides[126] and tetraloops.[139] The Steinbrecher phosphate parameters were also used by Darré et al.[151] in a different way, in conjunction with abandoning the Lorentz−Berthelot combination rules for the Lennard-Jones potential. The simulations showed a slightly better description of tetranucleotides than reported by Bergonzo et al.[139] In addition, Darré et al. modified also the torsional parameters of the C2′−OH group of ribose. However, the latter modification seems to have only marginal effects on the simulations, which become apparent after detailed scrutiny of the published data. Yang et al.[152] also replaced the Lorentz−Berthelot combination rules with a specific phosphate (OP) oxygen parameterization (along with an O2′ adjustment), with similarly good results in tetranucleotide simulations. However, none of these modifications were tested on a broad range of RNA structures, and their general applicability remains to be determined.

In summary, most attempts to modify the vdW terms have required simultaneous modification of the vdW combination rules. It is therefore not clear whether RNA simulations can be improved by tuning the vdW term alone without using additional tricks such as NBfix, which effectively increase the number of parameters that can be tuned. In addition, none of the reported attempts seems to have yielded a real breakthrough in the quality of RNA simulations.

3.1.2.5. Structure-Specific Modifications. In addition to attempts to improve the general parameterization of the RNA force field, simulations of specific systems can be improved by structure-specific force-field modifications.[129] The goal of such modifications is to correct specific weaknesses of the force field while avoiding undesired side effects. The HBfix method (not to be confused with the NBfix-type modifications discussed above) adds a local spherical auxiliary potential supporting native hydrogen bonds.[129] Such local potentials do not affect the ensemble except for a rather narrow part of the configuration space where the heavy-atom distances of hydrogen donor−acceptor pairs lie within a given range (e.g., 3−4 Å in the original work). HBfix only indirectly promotes forward folding (corresponding to k_{on} in experiments) but directly increases the lifetime of folded structures (i.e., reduces k_{off}). In effect, the potential compensates for the unmodified force field's inability to describe the electron-redistribution effects that accompany real H-bonding for a list of a priori chosen interactions that are present in the native structure. Note that when utilizing structure-specific potentials, one is typically limited by the functions that are implemented in the simulation codes; in practice, HBfix is constructed as a linear combination of two standard restraints. For this reason, HBfix has been so far implemented in a way that does not account for the directionality of the H-bonds.[129]

The short-range HBfix bias is much milder than commonly used tools such as restraints, targeted MD, Gō-type potentials,[153] and other brute-force approaches for forcing simulations to sample desired (target) structures. The amount of energy included to stabilize each native H-bond in the HBfix in ref 129 is indeed on the order of the thermal energy k_BT, that is similar to the bias used in Gō model potentials. However, in Gō models all of the native contacts are stabilized, whereas in the HBfix approach only a selected number of contacts are stabilized. For instance, the typical number of hydrogen bonds stabilized in an 8-mer including a GAGA tetraloop in the HBfix approach would be 9.[129] In an atomistic Gō model simulation for the same system, the number of native contacts would be approximately 100.[153] In addition, Gō models are not used within the framework of explicit solvent environment.

Simulations using HBfix with 1 kcal/mol support per H-bond added to the χ_{OL3} force field successfully folded the GAGA tetraloop, retaining all of its signature interactions.[129] HBfix also successfully stabilized simulations of the U1A protein/RNA complex.[154] Note that the HBfix potential can also be used to destabilize selected H-bond interactions if they are overstabilized. Obviously, the use of structure-specific force-field adjustments may seem unsatisfactory. However, pragmatic use of such biases is legitimate, and, due to the persistent performance problems of the general force fields, their use may become increasingly common, or even inevitable. We suggest that the best approach is to first try to achieve the best possible performance with the general force field. Once one has then reached the point at which further tuning is unproductive, the native state(s) can be supported with gentle structure-specific biases rather than continuing with cumbersome force-field refinements that may cause many undesired side effects. In addition, in principle, the HBfix type of potential could be generalized in an interaction-specific manner.

3.1.3. CHARMM RNA Force Fields. CHARMM is another major MM potential that was systematically developed for simulating biomacromolecules, including nucleic acids. The latest version is CHARMM36,[98] and earlier versions are known as CHARMM27[89,155,156] and CHARMM22.[157] Despite its common use to study RNA, its performance has not been as thoroughly and systematically tested as that of the AMBER RNA force fields; literature discussions of limitations of RNA force fields based on the work of Cornell et al. appear to be more systematic than those of the CHARMM derivatives. Several groups have reported that CHARMM27 underestimates base pairing stability in A-RNA and may even lead to uncontrollable helix fraying (Figure 8);[134,158−160] some groups switched from using CHARMM to AMBER during their ongoing RNA projects.[158,161,162] Conversely, other groups have used this force field for RNA simulations without reporting any problems. It is not clear whether these contrasting reports reflect the use of somewhat different force-field implementations (including a possible difference between periodic boundary and solvent sphere computations; see section 4.5.3) or whether some groups used unreported means such as restraints to stabilize their simulations. The instability of A-RNA in CHARMM27 simulations was addressed by the CHARMM36 reparameterization,[98] which modified the 2′-OH ribose dihedral potential because the stability problem was attributed to an inappropriate orientation of the 2′-OH groups. The robustness of this refinement is not yet fully clear; in our opinion, it has reduced but not eliminated the tendency toward fraying. As we noted in several other places of this review (e.g., section 4.2.2.3), we suspect that even the AMBER force field underestimates strength of the base pairing, due to the lack of inclusion of electronic structure redistributions, which actually has been the reason for introducing the HBfix approach (section 3.1.2.5).[129] It should be noted that in some cases the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

lower structural stability of simulations can even be quite useful to speed up conformational changes, by reducing barriers between different parts of the folding landscape. However, there is always a potential risk that such structural changes may not be fully realistic. We reiterate that the CHARMM force-field developers are currently leading the efforts to derive polarizable force fields for nucleic acids.[81−83]

3.1.4. Force-Field Validation by Reweighting. A very important issue closely related to the development of new force fields is their validation. In general, as we discussed, classical force fields contain nonphysical terms that might predict correct relative stabilities of multiple conformers as a consequence of error cancellation. This makes it very difficult to validate the force field from an ab initio perspective, by using for example QM benchmarks.[85] A usual way to assess the quality of a force field is thus testing it within an MD simulation and comparing results directly with experimental data. When possible, this comparison should be made in situations where the MD trajectory is ergodic (see section 3.2.2), so as to separate statistical errors (arising from finite sampling) from systematic errors (due to force-field inaccuracies).

This analysis is computationally demanding when testing several force-field variants due to the fact that a full simulation has to be performed every time a force-field term is changed. An efficient alternative comes from reweighting the already available simulations to take into account changes in the force field. In particular, a trajectory obtained using a given force field with potential energy function $U_0(x)$ can be used to compute ensemble averages corresponding to another force field with potential energy function $U_1(x)$ by weighting each of the visited conformations x proportionally to $w(x) \propto \exp(-(U_1(x) - U_0(x))/k_BT)$. This procedure is analogous to the free-energy perturbation method (see section 3.2.7) and is used to switch between two alternative force fields. Reweighting has been used, for instance, in ref 163 to predict the effect of small perturbations applied on the dihedral angles on tetranucleotides and tetraloops, and in refs 143 and 144 to test full dihedral reparameterizations on tetranucleotides. Unfortunately, this procedure, known as exponential averaging, is only effective when the fluctuations of the difference between the two potential energy functions are small. The statistical efficiency can be quantified computing the Kish's effective sample size,[164] which is defined as $(\sum_x w(x))^2/\sum_x (w(x))^2$ and reports the effective number of conformations with non-negligible weight. Effective sample size will be large only when the ensembles generated by the two force fields are significantly overlapping. If some structure that is stabilized by the force field U_1 is never visited by the force field U_0, its effect on the ensemble averages cannot be estimated without running a new simulation using the force field U_1. In addition, reweighting might be highly inefficient when charges are perturbed, because, due its long-range nature, electrostatic energy can be heavily affected by very small changes in the charges.

3.2. Accessing Long Time Scales, Thermodynamics, and Kinetics

MD simulations allow the equations of motion to be solved and the evolution of the system to be followed in real time. This is achieved using a model empirical potential, the force field, which mimics the real interatomic forces acting on the simulated molecular system. Aside from the approximations inherent in the force field, which are discussed in sections 3.1.1 and 3.4.2, the most important limitation of MD is that the time scales that can be simulated are limited by the available computational resources. The time scales over which the conformational transformations of RNA occur are very heterogeneous.[70] Fast processes such as the internal dynamics of small hairpins[165] occur over microseconds to milliseconds. Slower processes such as ligand-induced riboswitch folding[166,167] occur on scales of seconds to minutes or beyond.[168] In addition, because simulations are performed on a single unique copy of a molecule, it would be necessary to observe multiple individual rare events to obtain statistically converged properties and thus realistic thermodynamic and kinetic properties of the process under investigation. Such comprehensive simulations are clearly beyond the reach of current computers.

3.2.1. Considerations about Convergence. From a theoretical point of view, MD simulations can be seen as Markov chains (see section 3.2.3), which means that the coordinates at a given time step only depend on the coordinates at the previous time step. When used to compute ensemble averages such as populations of individual substates, MD simulations will suffer, as any method based on Markov chains, from a statistical error due to the finite length of the simulation. In addition, if the simulation is not long enough, the population of the initial structure will be artificially enhanced, with an effect known as "initialization bias" in the Markov-chain Monte Carlo literature. The former effect can be decreased by making a simulation longer, and the latter by discarding the initial equilibration part (see, e.g., ref 169 for a discussion on the compromises to be taken in this sense). However, whenever an MD simulation remains stuck in a given conformation, one should try to understand whether this specific conformation corresponds to the global minimum of the free energy of the system or it is just a kinetically trapped local minimum. Rigorously speaking, the only way to answer this question is to run a simulation capable to explore the whole conformational space. In practice, one might try simulations starting from different conformations and see if results are independent of the starting point. Using state-of-the-art hardware and software, the only RNA systems for which a fully converged exploration of the conformational space can be achieved with plain molecular dynamics are probably nucleosides or dinucleotides. However, using highly optimized hardware[73] or with the enhanced sampling techniques discussed in this section, tetranucleotides or even tetraloops (see sections 4.1 and 4.2) can also be explored exhaustively. Probably, neither plain MD nor enhanced sampling methods can completely sample the conformational space of larger systems. In these cases, one might only be able to sample different conformations that are in the vicinity of the initial structure. Still, in some cases, it is possible to obtain relative populations of relevant substates that can be compared to experiments. In other words, whereas full convergence might be impossible to reach, one might be in the situation where multiple transitions between the locally available substates are seen and local exploration is virtually converged. In addition, series of smartly designed simulations initiated in different parts of the conformational space and characterizing properties of different types of conformations present on the free-energy landscape may provide unique insights complementing the available experimental data even without simulating large-scale transitions.[170]

3.2.2. General Considerations about Sampling Enhancement. In general, RNA tends to have multiple metastable states, and its folding landscape is very

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

rugged.[171−174] The time scale distribution of its ruggedness and metastability is continuous, without clear borders. In other words, the metastable states may persist over diverse time scales and may include different backbone conformations, diverse patterns of directly bound ions, differences in base-pair geometries, or even entirely different folds. The definition of metastability depends on the observed time scale and the capability of experiments to detect the ruggedness of the folding landscape. When reconformations are faster than the temporal resolution of the experiments, or in a bulk experiment where a macroscopic number of copies of the same molecule is present in a buffer, an experiment would probe some averaged ensemble property.

MD is a fundamental method for studying the ruggedness of the RNA conformational landscape. In principle, MD simulations are not limited by experimentally detectable properties and temporal resolutions, which is important because many dynamic processes not resolvable by experiments may be critically important for the biochemical and biological functions of RNAs. As extensively discussed elsewhere, MD simulations are analogous to single-molecule experiments performed on extremely short time scales (typically a few μs at most) initiated from a specific starting structure (single geometries).[154,175,176] Therefore, the explored ensembles are usually highly dependent on the initial state (including uncertainties in the initial, usually experimentally determined, geometries).[154,175−177] Using standard simulations, it is difficult if not impossible to accurately evaluate the relative populations of all of the available metastable states on the free-energy landscape. To tackle this issue, several groups have worked over the last few decades to develop methods that allow properties that emerge over long time scales to be investigated using relatively short simulations (see, e.g., refs 178−182 and Figure 10).

It is important to mention that enhanced sampling techniques would be almost wholly unnecessary if they could be replaced by straightforward MD simulations with sufficiently long time scales. Development of optimized MD software[183−186] and, more recently, availability of GPUs[187−190] have made it possible to simulate systems consisting of a few tens to hundreds of nucleotides for time scales on the order of a few microseconds. Even more striking has been the effort made by the D. E. Shaw group, which has developed a dedicated machine for MD[73] that allows access to the millisecond time scale when simulating small proteins[191] and DNA.[115] This machine is called Anton and is based on MD-specific integrated circuits that interact in a tightly coupled manner using a specialized high-speed communication network. Despite this remarkable progress, biologically relevant time scales are currently out of reach and will likely remain so for decades. Therefore, even Anton simulations are sometimes performed using enhanced sampling techniques to exhaustively sample the studied system's conformational space.[192] Even in the future, MD practitioners will certainly still need enhanced sampling methods to bridge the time scale gap between simulations and experiments.

A wide array of methods have been developed over the years for obtaining information about biomolecular systems that emerges over long time scales from short simulations. In general, they have been tested more extensively on proteins than nucleic acids, and fewer studies still have focused on RNA. These methods are also difficult to classify because many of them combine elements of previous methods that were developed on the basis of diverse principles.

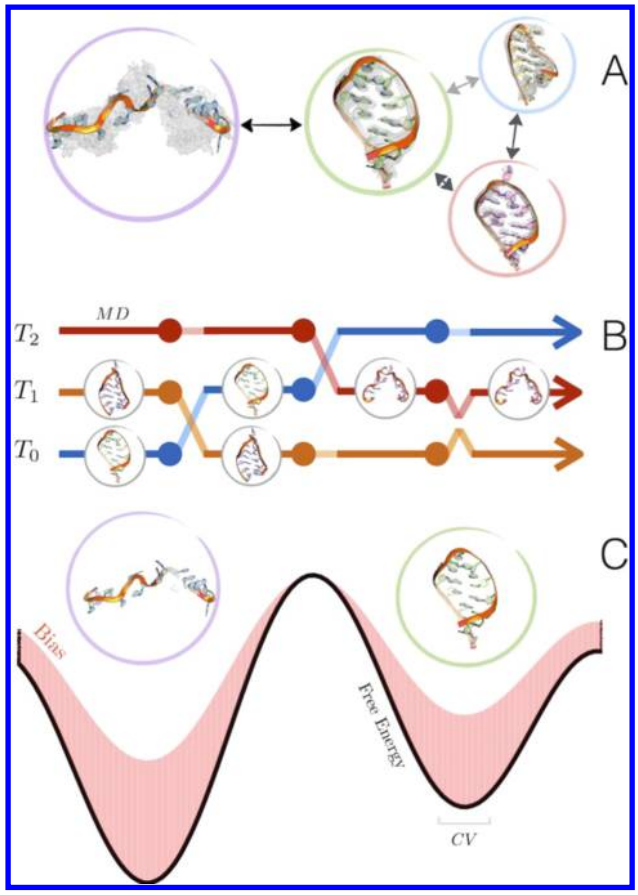


Figure 10. Scheme representing some of the methods discussed in this section. (A) In Markov state models, extensive simulations (usually sets of simulations) are analyzed, and the observed states are clustered. A kinetic matrix is then constructed that provides the probability of observing transitions between pairs of clusters (section 3.2.3). (B) In replica exchange simulations, numerous replicas of the system are simulated in parallel using different parameters (e.g., temperature or potential energy function). From time to time, exchanges are attempted with a Monte Carlo procedure. Sampling in the reference (unmodified) replica is enhanced by the method (section 3.2.4). (C) In metadynamics, a bias potential is added to compensate the underlying free-energy barriers along a preselected collective variable (CV). If the chosen CV is capable of discriminating the transition state, the transition probability is enhanced (section 3.2.5).

The conceptually simplest way to obtain long time scale information is to combine multiple short trajectories (section 3.2.3, Figure 10A). Techniques of this class, such as Markov state models (MSM), mostly provide recipes for initializing the simulations in a way that maximizes the sampling of important events (e.g., crossings of free-energy barriers), and for properly combining the resulting information. MSM methods can also be exceptionally useful for analyzing MD simulation results from huge amounts of simulation data, and visualizing them in a humanly comprehensible way. Indeed, meaningfully analyzing raw MD trajectories is often insurmountably complex due to the vast amount of data they contain.

Alternatively, one could use enhanced sampling methods to accelerate events. The methods of this kind that have been most widely used for exploring the conformational space of RNA molecules are based on the principle of annealing and

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

replica-exchange (section 3.2.4, Figure 10B). In this approach, a set of simulations (replicas) at different temperatures are performed in parallel, and exchanges among the replicas allow free-energy barriers to be crossed by coupling the cold replicas with the more ergodic hot ones. This idea has been extended further, allowing the temperature to be replaced as the control parameter in favor of arbitrary (suitable) modifications of the system's potential energy (i.e., the Hamiltonian).

Other approaches are based on the principle of importance sampling, where an artificially modified ensemble is explored (section 3.2.5, Figure 10C). In general, importance sampling techniques aim to derive properties of a particular ensemble (probability distribution of structures) of interest from samples generated from a different (biased) distribution. In these methods, a biasing force (or a bias potential) is used to accelerate sampling in a reduced (low-dimensional) space consisting of slow degrees of freedom or collective variables (CVs). The CVs should reflect the essence of the studied processes, resembling reaction coordinates in simple chemical reactions. These methods thus assume that the studied process can be described with sufficient realism using just a few degrees of freedom, which can be seen as a coarse graining of the full coordinate space, that is, a low-dimensional projection. The remaining dynamics that is orthogonal to the space defined by the CVs is assumed to be unimportant to the studied process. CVs can be very complex functions of the Cartesian coordinates of the system. Enhanced sampling methods aim to flatten the Boltzmann distribution in the CV space by imposing a bias that allows sampling of the whole CV space. The effect of this external potential is reweighted a posteriori (using procedures similar to that discussed in section 3.1.4) to recover the properties of the unbiased canonical ensemble.

Enhanced sampling methods are based on the idea of guiding the system in some way through unrealistically fast trajectories so as to observe the desired events. This idea can be taken to its logical extreme by considering a case in which the path consists of a transformation in which the chemical identity of the simulated molecules is changed, as it is done in alchemical methods (section 3.2.7). These alchemical transitions may involve the use of chemically unrealistic intermediate states (in which, for example, atoms present in the initial system are simply removed) because the free energy is a state function and the free-energy difference between two states is path-independent.

Finally, it is possible to remove solvent degrees of freedom to accelerate sampling (section 3.2.8). Whereas this procedure introduces some unavoidable approximations that might be critical in RNA systems, it is particularly attractive as it allows free energies to be computed from single conformations, bypassing the need to simulate reactive trajectories.

The following sections discuss these methods and their strengths and limitations in more detail, with a particular focus on those that have been applied to RNA systems.

3.2.3. Markov State Models. In recent years, the availability of massively parallel computing resources has increased exponentially. Modern supercomputing clusters make it possible to run hundreds of parallel simulations. However, advanced analysis techniques must be used to combine the data generated by multiple, out-of-equilibrium, short simulations, and extract relevant information from them. The framework of Markov state models (MSMs) is perfectly suited for this task. In the last decade, MSMs have been exploited to elucidate the slow kinetics of complex biomolecules,[193−199] and the theory behind them has been increasingly refined.[200−203]

In brief, the workflow for constructing a MSM from a set of MD trajectories is as follows:

(1) MD data are clustered into a finite set of microstates. Each microstate consists of a number of structures that can be considered sufficiently similar to be indistinguishable (equivalent) in kinetic terms. A lag-time, τ, is selected for time-discretization of the process, and transitions between the microstates that occur during the simulation are registered with a stride corresponding to the lag-time τ.

(2) Information from MD (the observed transitions at the lag-time window) is used to construct a transition probability matrix, **T**, whose elements, T_{ij}, represent the probability that the system, starting from microstate j, will transition to microstate i after a time τ. (Note that the **T** matrix is often used in its transposed form, and then the meaning of indexes i and j is interchanged.)

(3) The system is modeled as a Markov chain between the aforementioned microstates, governed by the transition probability matrix, so that the probability distribution of the microstates at time $t + \tau$ depends only on the distribution of microstates at time t and not on the system's history. In other words, its dynamics is modeled as a discrete process, in which at each time step, τ, the molecule "jumps" from one microstate (i.e., conformation) to another.

(4) The spectrum of eigenvalues of the matrix **T** is analyzed to extract information about the system's thermodynamics, dynamics, and kinetics.

The resulting discrete MSM aims to approximate the continuous dynamics of the simulated system by a discrete process. Using a MSM, one can evolve either the individual stochastic trajectories (time series of microstates, i.e., a time-discretized counterpart of the MD trajectory) or the probability distributions (of the microstates) of the system with a discrete time step τ. The MSM approximation can be extremely accurate; that is, it can provide the same time-development picture as the MD simulations. As compared to other methods for accessing long time scales that will be discussed below, such as replica exchange or CV-based methods, a MSM can provide information on a system's kinetic properties and transition rates between states of interest, whereas information on these properties is generally lost when the system's dynamics is biased. The interpretation of the eigenvalues and eigenvectors of the transition matrix of an MSM is the following: the first eigenvector of the transition matrix has a unit eigenvalue and corresponds to the stationary distribution vector, that is, the equilibrium probabilities of the microstates; the other eigenvectors describe relaxation processes toward equilibrium, and the corresponding eigenvalues, λ_i, are related to the time scales, t_i, of these processes in descending order, by the relationship $t_i = -\tau/\ln(\lambda_i)$.

MSM can be fruitfully applied to help extracting human-interpretable information from a single long MD trajectory, which repeatedly and spontaneously samples the rare event under investigation. If the simulation samples the process sufficiently well, the MSM can be used to create a coarse-grained model of the process, providing the much-needed insights into the otherwise overwhelming amount of raw simulation data. Alternatively, MSMs can be used to merge together separate MD trajectories, by discretizing the ensemble spanned by all trajectories, and counting the transitions occurring in any of the trajectories. This enables one to

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

rigorously combine the information coming from multiple trajectories in a single quantitative model.

Two main approaches for constructing a reliable MSM are (i) to employ an extremely fine discretization of the phase space into microstates, or, (ii) alternatively, to count transitions between larger basins that correspond to metastable states, in a so-called "coring" approach.[204] We will focus here on the description of the former approach, as it is most commonly used in current research.

The first step in the construction of a MSM is the discretization of the phase space into microstates. This can be done employing different clustering methods as well as different metrics.[200−203] An important advance in this field is the development of time-lagged independent component analysis (TICA).[201−203,205,206] In this method, linear combinations of the input variables are constructed so as to find degrees of freedom that have the largest autocorrelation times and that are thus expected to be more important in the kinetic analysis. This allows the subsequent clustering to be done in a lower dimensionality, in the space of the leading TICA components, by projecting the simulation trajectories onto the largest TICA components. TICA can be viewed as a method analogous to principal component analysis (PCA), which has been conventionally used to process MD trajectories in the past. PCA identifies linear combinations of the input degrees of freedom with the highest variance, while TICA finds those with the highest autocorrelation times, that is, corresponding to the slowest processes occurring in the simulations. TICA can be performed starting from a description based on the Cartesian coordinates of all of the solute atoms, or using a description defined on some internal coordinates, as, for instance, dihedral angles or pairwise distances between relevant atoms. As with other dimensionality reduction methods, one must always be aware that some important pieces of information might be discarded, distorting the description of fast time-scale processes. Because of the remarkable results it has achieved, TICA has been recommended as a standard tool for coordinate transformation and dimensionality reduction of MD trajectories data.[207]

Discretization of the phase space into a finite number of microstates is the main source of systematic error in a MSM. This step breaks the Markovianity of the system, that is, the assumption that the transition probabilities only depend on the current state of the system. Thus, modeling the system as a Markov chain causes deviations from the true dynamics. It has been shown[200] that these deviations can be reduced in two ways: by increasing the lag-time, τ, and by using a finer discretization scheme. In practice, when dealing with real finite-length simulations, both factors affect the quality of the computation. The lag time depends intrinsically on the Markovianity of the system and the desired temporal resolution. Too short of a lag time will make the model non-Markovian. As a rule of thumb, the interconversions among structures within each individual microstates must be fast as compared to the lag time. Typical values of τ used in MSM of proteins and nucleic acids are in the range from 10^{-1} to 10^{1} ns, and are usually selected by looking at the convergence of the implied time scales, as described below. The number of microstates should be large enough to avoid losing resolution due to coarse graining of the phase space but small enough for there to be a reasonable number of transitions between them (i.e., to have a sufficient statistical precision) and for the matrix size to be manageable. These considerations do not apply to "coring" methods,[204] which use much smaller sets of microstates that do not need to cover the system's full conformational space.

For simulations of medium-sized biomolecules with contemporary methods, one typically uses MSMs with at least $10^{2}-10^{4}$ microstates.[207] This number of microstates makes visualization and intuitive analyses of the model impractical. Several methods exist to overcome this problem by exploiting the kinetic information provided by an MSM to construct an even coarser representation of the system, lumping the MSM microstates into a few metastable macrostates. A commonly used approach is Perron-cluster cluster analysis (PCCA),[208] a method that exploits the sign structure of the eigenvectors of the transition matrix to define the optimal metastable partition of the MSM microstates. More advanced versions of the method known as PCCA+[209] and PCCA++[210] assign each microstate a probability of being a member of a given metastable macrostate.

Another possibility is to reduce the model's complexity by constructing a hidden Markov model (HMM) of the dynamics.[211] In this approach, the system is represented as a Markov chain between hidden metastable macrostates. These macrostates are not directly observable but are measured by looking at the microstate, which at every step is extracted from a distribution probability that depends on the hidden macrostate. Thus, one assumes that an additional (hidden) variable can be used to label the states, and its time series is inferred by the time series of the observed variables. The HMM defines states without neat boundaries, and a given conformation has probabilities to be simultaneously participating in multiple macrostates.

A key strength of the MSM approach is the observation that it is not necessary to assume global equilibration in the ensemble of trajectories provided that the MD is in local equilibrium within each microstate. This is what makes MSMs powerful tools for accessing long-time-scale kinetics. In fact, by choosing smart initialization points for the simulations, one can obtain an ensemble of relatively short trajectories, each of them sampling transitions relevant to different steps of a complex and slow configurational change. By combining these trajectories in a MSM, it is, in principle, possible to reconstruct even processes that occur on a time scale longer than the span of any of the individual trajectories. The largest implied time scale can be of the same order of magnitude of the aggregate duration of all MD trajectories used to build the MSM.[212] Obviously, the initialization points must be chosen with care, and one must ensure that no important regions of the phase space are ignored.

There are various ways of selecting the starting points for MD simulations to be used in the construction of a MSM. If available, prior knowledge about the system can be used to initialize simulations in different positions along interesting conformational changes (for example, if experimental structures of multiple conformations are available). Alternatively, new simulations can be initialized whenever a single "exploratory" simulation reaches a new interesting conformation. If this step is repeated recursively on each new simulation, it will produce a cascade of MD trajectories, sampling increasingly larger regions of the available phase space. By changing the criteria for identifying candidate starting points for new trajectories, it is possible to drive the system along the path of the conformational change of interest.[213] This approach is usually referred to as "adaptive sampling", and it can be de facto considered as a

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

subclass of enhanced sampling techniques. Another powerful approach is to extract the initial structures from an ensemble of configurations obtained with some different enhanced sampling technique.

As with all methods designed to reduce the computational cost of MD simulations, a MSM may provide wrong results when used in a way that is inconsistent with its basic approximations and assumptions. It is therefore important to test the validity of the Markovian approximation before drawing any conclusions from a MSM. This is usually done by looking at the convergence of the implied time scales, t_i, as the value of τ increases.[214] The implied time scales should be independent of τ in a Markovian system, so their convergence can be used to select an appropriate lag time (if it exists) that corresponds to a good approximation of the true dynamics. It is useful to point out that the common practice of showing the time scales in logarithmic scale may give a false impression of convergence due to the negative convexity of the logarithm function.

It is important to note that the convergence of the implied time scales is a necessary but not sufficient condition for Markovianity. When convergence is reached, the slowest implied time scale should correspond to the slowest transition mode in the studied system. Its comparison with previous knowledge of the system can provide some hint on whether the full free-energy landscape has been sufficiently sampled or not. A too short implied time scale (as compared to experiment) could indicate that important parts of the free-energy landscape are entirely missing in the simulation data set, and the MSM characterizes only a local segment of the folding landscape. This could happen when simulations are too short or not initialized to cover sufficiently the relevant portions of the phase space. For example, in studies of the folding landscape, series of simulations may be initiated (seeded) from some unfolding pathway, obtained by forced unfolding or high-temperature simulation initiated from the folded state. This may work well for molecules with fast folding via a funnel mechanism. However, when the folding landscape is rugged, inherently obeying kinetic partitioning,[215−218] major parts of the landscape are likely to be unreachable by simulations initiated from a "seeding" unfolding trajectory.[170] Such trajectory can connect the ensemble pertinent to the starting folded structure with the putative unfolded ensemble, but is not sufficient to identify misfolded structures that are present elsewhere on the folding landscape.

A stricter test of the validity of the Markovian approximation can be obtained by performing a so-called "Chapman−Kolmogorov test"[200] in which one examines the time evolution of the probability of the system to be in a certain metastable state, when initialized in another, and compares the model's predictions to the observed time evolution. Note, however, that as in all of the algorithms based on statistical sampling, there is no way to infer information about conformations that were never sampled in the simulated trajectory. Thus, these tests are not a panacea. As mentioned in the previous paragraph, in case of kinetic partitioning, even sophisticated convergence tests would not reveal the lack of convergence because the free-energy basins corresponding to the misfolded states are entirely inaccessible to the simulations. Thus, the investigators must always perform a rational overall appraisal of the studied process and not rely merely on the numbers provided by the computational procedures.[170]

Another important source of uncertainty in the predictions of a MSM is the statistical error due to finite sampling. This can generally be estimated by Markov chain Monte Carlo (MCMC) sampling of the transition matrix.[219] It is vital to take statistical error into account because poorly sampled transitions can easily lead to uncertainties on associated implied time scales of the same order of magnitude as the time scales themselves.

It is only recently that MSMs have been employed to address the kinetics of RNA systems.[220,221] The first works suggested how the millisecond-long folding process of GNRA hairpins can be, in principle, described with the use of MSM and plain MD simulations. The results of these studies with all of their limitations are discussed in section 4.2.2.3. The MSM has also been used to elucidate the kinetics of protein/nucleic acid complexes, particularly the various aspects of RNA polymerase activity. The results of these studies are discussed in section 4.7.10.

MSMs were used to characterize kinetic properties of very short RNA oligonucleotides,[222] dinucleotides composed of combinations of adenine and cytidine as well as tri- and tetranucleotides composed of adenines. The coordinates used were the dihedral angles of the studied systems and the G-vectors introduced in ref 223, which take into account the formation and direction of stacking interactions. The stacking time scales predicted by χ_{OL3} force field were in reasonably good agreement with experimental temperature-jump measurements.[224] These results suggest that the main relaxation modes for short oligonucleotides can be described in terms of transitions between alternative folded states rather than between random coils and native structures. However, because of the known problems of the χ_{OL3} force field, the simulations of the r(A_4) sampled intercalated structures that are inconsistent with nuclear magnetic resonance (NMR) experiments (see section 4.1). Thus, agreement with experiments was only obtained after manually removing the intercalated structures from the trajectory. Results for r$(A)_3$ are summarized in Figure 11, which represents typical output of MSM analysis followed by HMM analysis and reports the obtained clusters as well as the transition rates. MSM modeling has also been used to analyze conformational substates of conformationally restricted single-stranded three-nucleotide loops of DNA quadruplexes;[76] such approaches should be readily applicable also to various single-stranded RNA segments with restrained positions of the strand termini.

Another study investigated the process of pairing and fraying of a terminal base pair of an RNA duplex with methods closely related to MSM.[225] In that paper, the authors limit their analysis to the fraying of a single base of the terminal base pair, freezing the positions of the rest of the double helix considered. Interestingly, they identified a rate-determining trapped state, in which the base is stacked but the backbone assumes a non-native conformation.

MSM methodologies have developed rapidly over the past few years, and many procedures that were favored in the past have been surpassed by better alternatives. Being already successfully applied to the study of many protein systems, they are now starting to cut their space also in the world of RNA. We expect that new applications and protocols will emerge soon, and that the limits of the method's applicability will be more rigorously discussed.

3.2.4. Methods Based on Annealing. The most widely used enhanced sampling method in biomolecular simulations is probably the parallel tempering (PT) method, which is also known as temperature-replica-exchange MD (T-REMD).[226] T-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

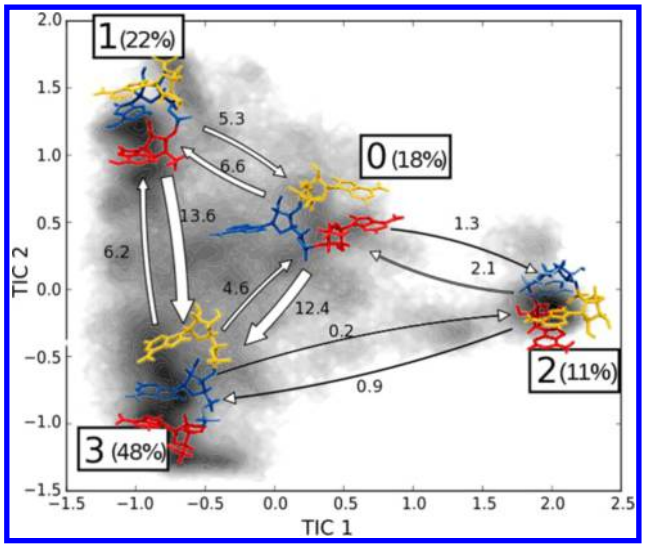


Figure 11. Schematic representation of four-state hidden Markov model for adenine trinucleotide.[222] Nucleotides are colored according to their position in the sequence (red, 1; blue, 2; yellow, 3). Percentages indicate the equilibrium population of each state; the widths of the arrows are proportional to the transition rates between the states, which are also indicated in μs^{-1} units. Shading indicates the distribution of the simulation data projected on the plane defined by the two leading TICA components. Reprinted with permission from ref 222. Copyright 2017 American Chemical Society.

REMD belongs to a wider class of methods based on simulated annealing.[227] The principle of simulated annealing is that a system kept at a higher temperature can more easily cross energy barriers and is thus better able to explore its conformational space than a system at low temperature. This is because the time required to cross a free-energy barrier depends exponentially on the height of the enthalpic part of the barrier divided by the temperature of the system. However, the temperature also affects the equilibrium populations of different conformations. Therefore, running MD simulations at a high temperature would yield faster conformational transitions but might also result in extensive sampling of structures whose population is negligible at lower temperatures (e.g., room temperature). High temperature simulations have sometimes been used qualitatively to enhance sampling of RNA systems (see, e.g., ref 228). However, high temperature simulations alone cannot be used to directly estimate the values of experimental observables at physiological temperatures, and this approach is considered rather obsolete these days. The idea of annealing is, after the transition, to slowly decrease the simulation temperature so as to gradually shift the system to explore a relevant region of its conformational space.

The main problem of simulated annealing is that the results can heavily depend on the schedule used to reduce the temperature. In particular, final conformations might retain properties of the initial high temperature part of the simulation if the cooling is too fast. An important step forward was introduction of the simulated tempering approach.[229] In a simulated tempering simulation, the temperature is not changed on a predefined schedule but evolves under the control of a Monte Carlo (MC) procedure that allows the system to evolve at equilibrium. The goal of the method is to perform a random walk across the temperature space, leading to multiple heating and cooling cycles. Once a set of temperatures is chosen from a given range, a weight is assigned to each temperature state that determines the probability of visiting that state (i.e., how much time the system will spend there). If the weights are not chosen properly, the random walk in the temperature space will be confined to a subspace to some degree rather than fully exploring the entire space. Schemes for automatically adjusting these weights have been proposed.[192,230] Practical solution that avoids the assignment of weights is the T-REMD (i.e., PT) algorithm, in which multiple replicas are simulated simultaneously at different temperatures. From time to time, an exchange of coordinates between two replicas in the temperature ladder is attempted and either accepted or rejected using a MC procedure based on the potential energies of the simulated systems. Because the number of simulations at each temperature is fixed (i.e., there is one at each temperature), it is not necessary to compute weights, and the computational effort expended at each temperature is identical by construction. T-REMD was originally introduced in studies on spin glasses[231] and was subsequently used by the biomolecular research community in conjunction with Monte Carlo methods[232] and then with MD.[226] It is probably the most widely used enhanced sampling method in biomolecular simulations, and several papers have described applications of T-REMD to RNA systems; see, for example, refs 120, 129, 221, and 233−244.

One important choice when setting up a T-REMD study is the temperature range spanned by the replicas. The range typically goes from the reference temperature to a temperature high enough for enthalpic barriers to be easily crossed (usually between 400 and 600 K). These high temperatures are far from the physiological conditions because they are well above the boiling point of water under simulated conditions. However, one should note that the water models used in MD simulations typically have higher boiling points than that of real water. Additionally, most T-REMD simulations are performed at constant volume and so are not subject to this issue. While this is not usually a problem, it should be accounted for properly whenever conformational changes are correlated with changes in the effective volume of the solute, as done by Garcia et al.[237,241] The highest temperature should not be seen as a physical temperature but as a computational tool for enhancing sampling. It should therefore be optimized to enable the observation of the greatest possible number of conformational changes.

T-REMD is certainly a robust tool for overcoming enthalpic barriers. However, folding barriers (often) contain significant entropic contributions, and then the effect of high temperatures to enhance sampling is limited. Paradoxically, the addition of more high temperature replicas could even make the algorithm less computationally efficient because the additional cost of the extra replicas might not be fully compensated by more effective sampling and corresponding shorter folding time. Moreover, the number of states that should be explored increases significantly with temperature, so the dimensionality of the generalized ensemble (i.e., the combination of all temperature ensembles) becomes larger. This makes achieving convergence in a T-REMD simulation more difficult. For these reasons, evaluations of the computational efficiency of a T-REMD procedure could yield very different answers depending on the specific system investigated.[245]

The goal of replica exchange methods is to maximize the number of statistically independent visits at low temperatures. To do this, it is important to maximize the number of successful exchanges and reduce the time a replica needs to visit a high

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

temperature and return. Put another way, one must minimize the time the replica needs to traverse the temperature ladder. The efficiency of parallel tempering simulations is thus often measured by the round-trip time τ_{rt}. This time depends on the average exchange acceptance across the replicas, the number of replicas, and the stride of the replica exchange attempt. If subsequent exchanges are not correlated, one would expect τ_{rt} to be proportional to the stride between consecutive exchange attempts, which should thus be as small as possible.[246,247] The benefit of reducing the stride between consecutive attempts is moderated by the fact that they are correlated, so it is rarely useful to choose a stride that is significantly shorter than the autocorrelation time of the acceptance itself. In practice, the exchange stride in explicit-solvent T-REMD is typically set to a fraction of a picosecond. Recent algorithms have achieved performance improvements by using a procedure that mimics the limit of exchanging at every MD step (the so-called infinite swapping limit).[248]

When setting up a T-REMD simulation, it is also important to properly fix the temperature distribution so that the potential energy distributions of neighboring replicas overlap. This is often done so as to obtain a high and, possibly, uniform acceptance across the replica ladder. Analytical derivations in simple systems have shown that for a uniform acceptance ratio, the difference between adjacent temperatures should be chosen inversely proportional to the square root of the systems heat capacity and should increase with the temperature.[249] Because the heat capacity is proportional to the number of particles, the number of replicas required to span a given temperature range should grow in proportion to the square root of the number of atoms. This makes T-REMD very computationally demanding in large systems such as explicitly solvated biomolecules. It must be also noted that the assumption that specific heat is temperature independent, which leads to a geometric progression of temperatures, is not always realistic. This approximation is less appropriate for systems simulated in an implicit solvent (see section 3.2.8), whose heat capacity is not dominated by the solvent. Many approaches have been developed to optimize the distribution of replicas to maintain a uniform acceptance, to reduce round-trip times, or to adjust the replicas to specific heat capacity (see, e.g., refs 250−252).

After a T-REMD simulation has been performed, unbiased populations of different substates can be obtained from the reference low-temperature replica. The key point is that fully converged T-REMD should provide information on the system's equilibrium thermodynamics (as defined by the force field) at each temperature in the temperature ladder, although the system's kinetics will be obscured by the discontinuity caused by the swaps. Interestingly, there have been some attempts to reconstruct kinetics from T-REMD simulations by making some a priori assumptions on the dependence of the transition rates on the temperature, including one on an RNA system simulated using a native-centric potential.[253] It is also possible to combine information from multiple replicas by performing a weighted-histogram analysis (WHAM, see section 3.2.5).[254,255]

Analysis of errors in T-REMD simulations is not trivial. While a standard blocking analysis[256] could be used to analyze the populations in the reference replica of a T-REMD simulation, the resulting errors might be significantly underestimated because of the correlations hidden in the replica swaps. In an extreme case, one might encounter a large number of apparent transitions that are only a signature of the swaps and do not correspond to a true exploration of the conformational space.[257] This issue can be tackled by reconstructing continuous trajectories that follow the system across the swaps, that is, as it travels across the temperature ladder (also known as "demultiplexed" trajectories). Several works (see, e.g., refs 129, 242, 258, and 259) have examined such demultiplexed trajectories and used differences in the populations of the relevant substates in these trajectories to qualitatively identify convergence problems (Figure 12). Demultiplexed trajectories can also be used in more rigorous error estimations based on an autocorrelation function.[255]

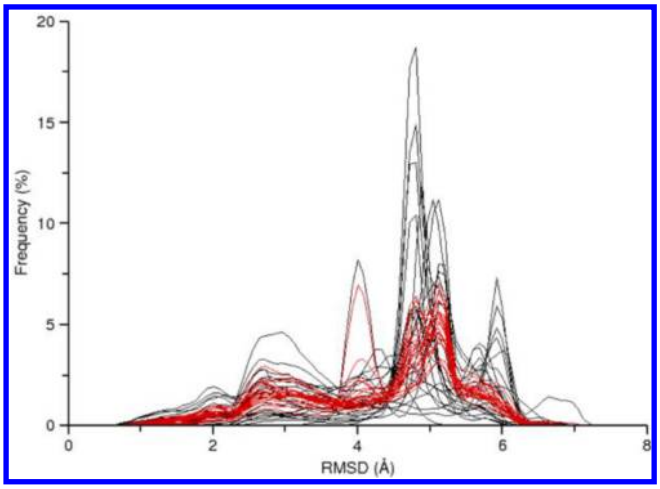


Figure 12. Histograms for RMSD with respect to the native structure for a rGACC tetranucleotide. Twenty-four histograms from a T-REMD simulation after 200 ns (black) and 2010 ns (red) per replica are shown. Each histogram contains data from one "demultiplexed" trajectory (see the text), which is a continuous trajectory traveling through the temperature ladder. In the first part of the simulation (black data), it is evident that different trajectories sample different parts of the space, suggesting the simulation not to be converged. At the end of the simulation (the red data), all of the substates are almost equivalently sampled by all of the replicas. Reprinted with permission from ref 258. Copyright 2013 American Chemical Society.

A common practical problem with all replica exchange methods is that the minimum number of replicas required to achieve reasonable acceptance might be impractically large. A possible solution is to use the simulated tempering scheme discussed above, where a single simulation is performed and the temperature is changed over time. Alternatively, one could use the well-tempered ensemble (WT) scheme[260] in which a bias is imposed on the system's potential energy to partially flatten its probability distribution histogram (and thus also the corresponding free-energy surface). In the WT ensemble, the partition function can be expressed as $Z_\gamma = \int [e^{-\beta U} N(U)]^{1/\gamma}\, dU$, where U is the potential energy, $N(U)$ is the temperature-independent density of microstates expressed as a function of U, $\beta = 1/k_B T$, and γ is a biasing factor (for the unbiased distribution, the partition function is standardly $Z = \int [e^{-\beta U} N(U)]\, dU$). This approach could thus be interpreted as a CV-based method like those discussed in section 3.2.5, in which the potential energy is considered as the biased CV along which the probability distribution is flattened (scaled down). WT ensemble increases the mean square fluctuations of U by a factor of γ while keeping the averages of U essentially unchanged as compared to the unbiased canonical ensemble. The WT ensemble approach thus can be directly combined

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

with T-REMD in a hybrid approach where multiple replicas are still employed but the number of replicas is significantly reduced, thanks to the increased fluctuations of the potential energy. Alternatively, one could increase the acceptance rate by using short nonequilibrium simulations provided that the acceptance rate is properly computed.[261] It must be stressed that these approaches allow the number of simulated replicas to be decreased but do not necessarily decrease the round trip time, which is the time required for a demultiplexed replica to travel from the lowest temperature to the highest and back. Therefore, there is no theoretical guarantee that these methods will accelerate sampling, although they have done so in empirical cases examined to date.[262] A full discussion of these methods is beyond the scope of this Review.

The replica exchange protocol can be generalized to methods where ergodicity is not obtained by increasing temperature but by scaling portions of the force field or adding penalty potentials disfavoring specific structures (e.g., by biasing/flattening potentials along selected dihedral angles). These methods are generally known as Hamiltonian replica exchange (H-REMD) methods because the different replicas use different Hamiltonian functions.[263] Again, an unbiased replica (with the unperturbed original Hamiltonian) is used to gather data to be compared to experiments, like the reference temperature replica in T-REMD. Another way to significantly decrease the number of replicas in a REMD simulation is to use replica exchange with solute tempering (REST),[264,265] a specific variant of H-REMD. Here, instead of changing the temperature, one scales down portions of the potential energy of the system. Typically, it is only the solute's potential energy that is scaled. REST has been applied to the full folding of an RNA tetraloop and provided results compatible with T-REMD.[129] Notably, it is possible to only apply the scaling to a portion of the solute,[266] which significantly reduces the size of the conformational space to be explored. This variant of REST has been used to accelerate the dynamics of an RNA tetraloop without affecting those of the corresponding stem, allowing different conformations of the capping nucleotides to be captured in a very short time.[267]

Other variants of H-REMD are possible. For instance, REMD has been used in combination with accelerated MD (aMD)[268] to sample structures in RNA tetranucleotides.[269] In the aMD method, the dihedral potentials are modified to avoid the system spending too much time in the minima, at the expense of not sampling the Boltzmann distribution. By constructing a ladder of such replicas that are increasingly biased, one can exploit the REMD strategy to obtain an unbiased ensemble in the reference replica. It is worth mentioning that alternative procedures to reconstruct unbiased distributions from aMD simulations have been proposed that do not require the use of replicas but instead make some assumption on the distribution of the potential energy.[270] H-REMD approaches in which the replicas have different scaled dihedral energies have been used to simulate RNA tetranucleotides.[259,271] This approach has been also combined with T-REMD in a multidimensional replica exchange method, that is, M-REMD. In the case of tetranucleotides, this replica exchange framework proved more efficient than using the Hamiltonian or the temperature separately,[271] although it required a very large number (192) of replicas. It should be noted that, although using large numbers of replicas is sometime helpful in exploiting large computing clusters, it makes it more difficult to assess the simulation's convergence. Indeed, it reduces the likelihood that "demultiplexed" replicas (see above) will be able to sample the whole range of temperatures and Hamiltonians. Finally, the binding and unbinding of ions to nucleic acids can be accelerated by scaling down the ions' electrostatic and Lennard-Jones parameters in the higher replicas of the ladder.[272] The idea of using a replica ladder is also employed in other REMD methods that are discussed below.[259,273]

The idea of accelerating a portion of the system (e.g., a particular tetraloop) without affecting the dynamics of another part (e.g., a stem) is very powerful and has been applied to RNA in various ways. For instance, in ref 233, the capping nucleotides of an RNA stem-loop were accelerated by coupling each region to a different thermostat. Historically, the idea of accelerating only a portion of the system was first proposed as an enhanced sampling method per se. In the locally enhanced sampling (LES) technique, multiple copies of a system are simulated.[274] Some of the atoms are kept in common between all of the copies, while atoms in regions selected for enhancement are replicated. The replicated atoms are subjected to a force that is scaled down by a factor corresponding to the number of copies. There is a clear parallel between the copies in a LES simulation and the biased replicas in the solute tempering approach: the number of LES copies is equivalent to the inverse of the potential-energy scaling factor in REST. The LES method has been used by several groups to enhance sampling in RNA systems, for example, in refs 233 and 274−277. A strong point of the LES approach is that the use of multiple copies enables an intrinsically parallel (and thus fast) exploration of the conformational space for the region of interest. However, when used in explicit solvent, all solvent molecules are typically shared among the copies and effectively keep the copies highly correlated. Therefore, it has been suggested that LES is best used with implicit solvent simulations (see section 3.2.8).[275] LES does not allow the unbiased populations to be recovered. Thus, while it significantly improves the speed of conformational sampling, it cannot be used to estimate populations that can be directly compared to experiments. This is probably the main reason why LES has not been used in recent works.

3.2.5. Methods Based on Importance Sampling. Another class of enhanced sampling methods is based on umbrella sampling (US).[278] The main idea of US is that one chooses a priori a set of collective variables (CVs) that describe and distinguish the metastable states of the system and the corresponding transition states. In its original formulation, US works by adding a bias potential that is a function of the selected CVs and that can bring the system as an umbrella through the transition state. Like a catalyst, this bias potential should ideally decrease the energy of the transition state to increase the transition rate.

In contrast to the above-mentioned annealing methods, the important sampling techniques are generally able to address entropically driven processes such as folding events by using properly chosen CVs. In addition, they allow high free-energy barriers on a priori known CVs to be crossed in a short time. On the other hand, the sampling is enhanced only in these few selected CVs, so that the probability of overcoming barriers is increased only in these few dimensions, while all perpendicular degrees of freedom are sampled as (in)efficiently as in the plain unbiased simulations. The entropically rich unfolded state that typically occupies a large part of the multidimensional configuration space will be thus sampled less efficiently than the conformationally well-defined folded state.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Another main obstacle of this approach is that it is usually difficult to know a priori the position of the transition state in the CV space and its stability. For this reason, US is almost invariably used by performing multiple simulations where the CV is restrained to explore the vicinity of a series of given values along the CV joining the initial and final states, or by defining a path along which the free-energy profile is calculated. Note that, in practice, the US simulations are usually seeded using some form of a short pulling simulation, which generates series of starting configurations for the individual US windows between the initial and final states. The effect of this procedure can be nontrivial as discussed later in this section. The US simulations are finally merged using the weighted histogram method (WHAM).[254] As a result, one can obtain a free-energy profile (landscape) as a function of the biased CVs, which is also known as the potential of mean force (PMF). The name PMF derives from the fact that its negative gradient determines the mean force acting on the system in the given point of the CV space. The words PMF and free-energy profile are largely interchangeable when discussing the dependence of the free energy on a CV. The relative probability of finding the system in any two points i and j of the CV space can be calculated as $p_i/p_j = e^{-\Delta w/kT}$, where p_i and p_j are the probabilities, and $\Delta w = w_i - w_j$ is the difference between the corresponding PMF values. This equation relates the populations of different structures with the free-energy landscapes. Note that the relative free energies of two (sub)states from the free-energy landscapes can be obtained by integrating the corresponding populations over the respective parts of the landscape that correspond to the (sub)states.

Alternatively, WHAM can be formulated to provide the full unbiased distribution, allowing one to compute populations of specific states even though these states are not identified by the biased CVs alone.[279] In practice, this allows one to choose biased CVs on the basis of a priori information and then analyze the resulting trajectory by projecting it onto different CVs selected a posteriori.

US with multiple restraints is probably the method of this class that has been most widely applied to RNA systems (see, e.g., refs 98, 141, 159, and 280−285), usually employing one or two CVs. For instance, the unfolding of a small hairpin has been studied by biasing its end-to-end distance.[98,159] The usual requirement of US calculations is that the histograms (i.e., probability distributions) of the CVs in neighboring simulations (the adjacent windows) sufficiently overlap. Overlaps are required for the self-consistent procedure used by WHAM method to converge to a unique solution. This limitation might be to some extent lifted by using an umbrella integration method.[286]

As with all other methods based on importance sampling, the accuracy of US depends on the convergence with respect to the simulated time. Evaluating this convergence is not trivial. Statistical errors on the reported free-energy landscapes (or in the corresponding populations) are sometimes computed by bootstrapping, which could significantly underestimate the error due to sample correlation. A more robust approach is to use blocking analysis[256] or, equivalently, to use block bootstrapping,[287] or at least to compare two independent US simulation sets. Care is also necessary when using this approach: an often underestimated problem of US simulations is that because each simulation is restrained to a small portion of the conformational space, transitions between metastable states are not expected to occur by design. This assumption makes it impossible to determine whether neighboring simulations overlap in the full conformational space rather than only in the CV space; overlap is very difficult to test without observing transitions, and its absence can lead to severe error underestimation.[288,289]

A study on a kissing-loop complex provides a paradigmatic example of this method's application to an RNA system (Figure 13).[283] In this work, the PMF associated with complex formation was evaluated by performing a series of common multiple restraint simulations in which the distance between the two loops was used as a biased CV. The error was estimated using a standard blocking analysis that in principle accounted for the time correlation between samples. At first sight, the calculations appeared to be adequately converged, as is commonly claimed in the literature describing such computations. However, two independent US simulations initialized in two different ways yielded completely different PMF profiles, both of which were incompatible with the statistical error calculated for the other simulation, indicating a drastic but hidden lack of convergence. The difference between the two simulations stemmed from the use of different initial structures: one was performed with an initial structure where the loops

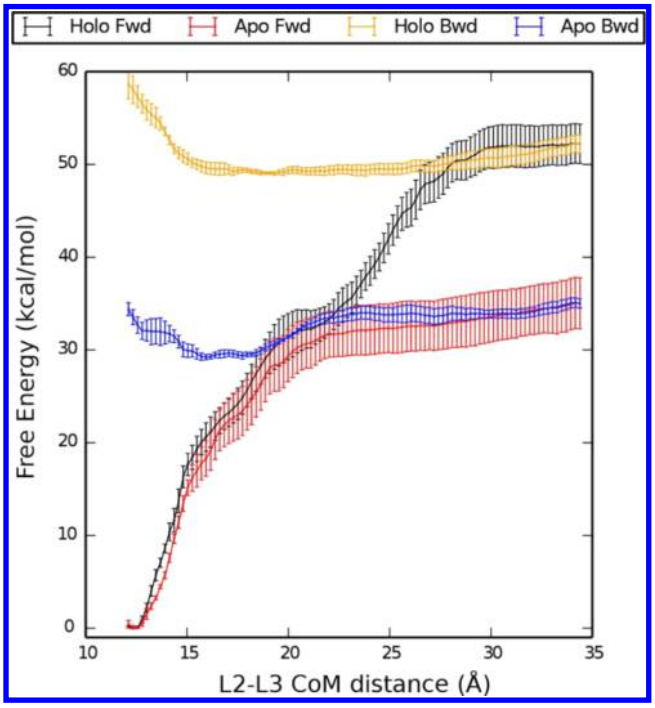


Figure 13. Potential of mean force (PMF) as a function of the distance between the centers of mass of the kissing-loops in an adenine riboswitch. Results for Holo (with ligand) and Apo (without ligand) forms are shown as obtained from two independent US simulations using different protocols to obtain the initial structures (forward, Fwd, where the loops are initially in contact; and backward, Bwd, where the loops are initially separated). Fwd and Bwd profiles are aligned at the maximum distance. The result is strongly dependent on the initialization procedure, indicating that it cannot be quantitatively trusted, even though errors computed with standard block analysis (shown as error bars) would suggest so. Still, the higher stability of the loops in the Holo form when compared to the Apo form is qualitatively captured by the simulations, though the range of calculated values is ∼30 kcal/mol. Reprinted from ref 283; http://creativecommons.org/licenses/by/4.0. Copyright 2015 BioMed Central Ltd.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

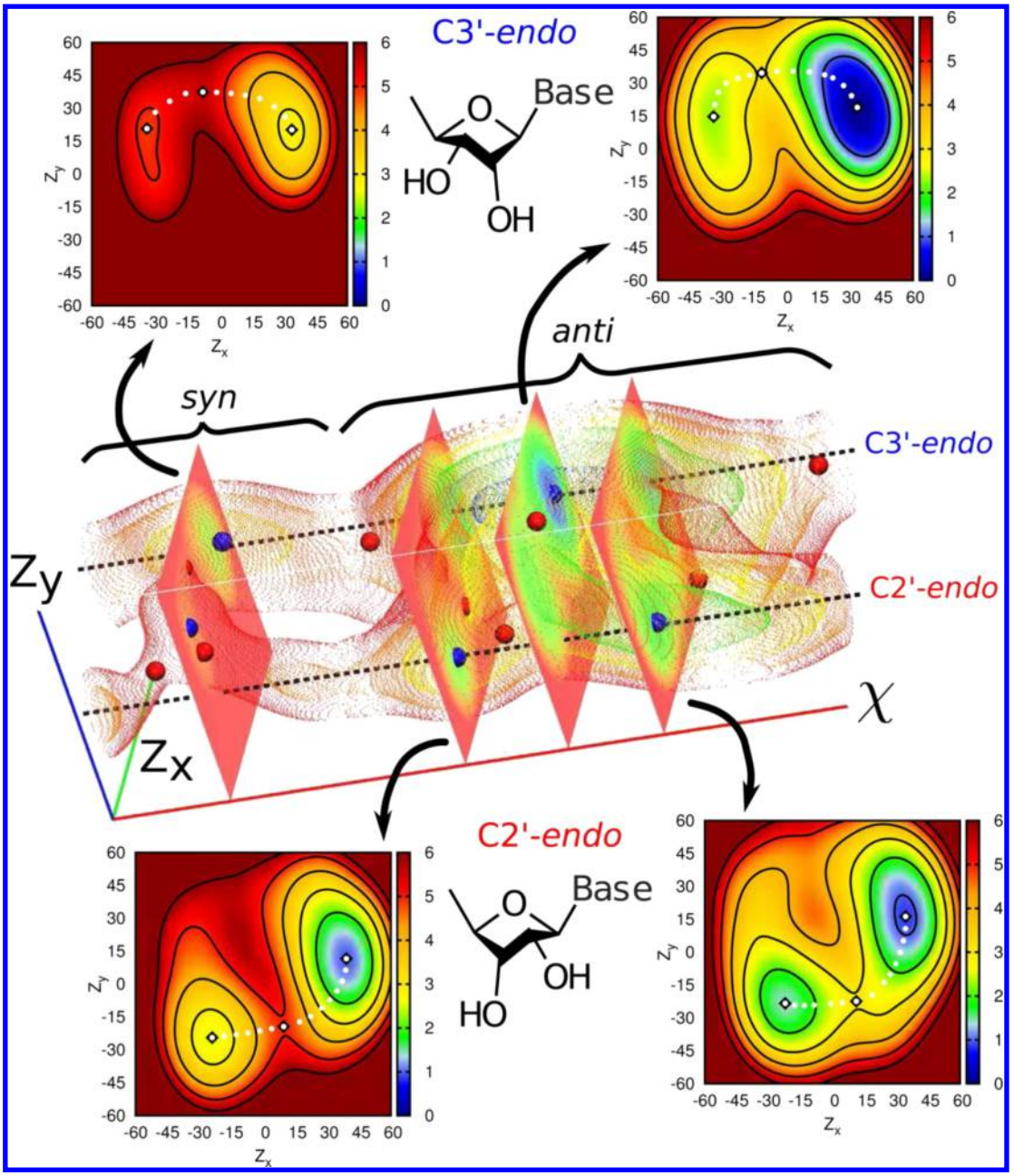


Figure 14. Example of a free-energy landscape (i.e., PMF plot). Three-dimensional free-energy landscape for a solvated uracil ribonucleoside along collective variables describing pucker conformation (Z_x and Z_y)[293] and glycosidic bond angle (χ). Results are obtained using an asynchronous replica exchange US procedure where these three CVs are biased. The system has been simulated using a semiempirical potential in a QM/MM scheme. Free energies are color-coded and are in kcal/mol (the bars right to the plots), and axes are in degrees. White dots represent lowest free-energy pathways connecting different pucker conformations. Local minima and saddle points are shown as blue and red spheres, respectively. Reprinted with permission from ref 291. Copyright 2015 American Chemical Society.

were docked and the other with a structure in which they were not. The selected CV was not sufficient to sample transitions between docked and undocked state likely due to the presence of other kinetically relevant CVs, so-called hidden variables, that were not biased. The calculated free energy thus depended strongly on the chosen computational protocol. In other words, the study's objective proved to be outside the applicability range of the chosen US protocol using the selected CVs. Note that many published papers do not provide such error analyses, and in many cases their authors are even unaware of the danger of a hidden lack of convergence. Common statistical methods to analyze convergence are usually not sufficient to identify such problems.

A much more robust approach to US is to include it in an REMD scheme where swaps in the conformations of neighboring replicas are periodically attempted and either accepted or rejected on the basis of an MC scheme.[290] The acceptance calculation should reflect the fact that different replicas feel different bias potentials, as in the H-REMD methods discussed above. Example application of replica exchange US to the conformational sampling of an uracil ribonucleoside has been reported in ref 291 (Figure 14). The advantage of this formulation is that demultiplexed trajectories

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

can be constructed (see section 3.2.4) and inspected to see if transitions between relevant metastable states are observed. In contrast to classical US, which might lead to large differences between PMFs for association and dissociations pathways as highlighted in the preceding paragraph (see, e.g., Figure 13), the combination of US with REMD scheme typically provides significantly improved convergence as demonstrated, for example, for minor-groove binding of a ligand to DNA.[292]

It is important to note that combining simulations with WHAM to obtain a free-energy profile essentially prevents access to any information about the system's dynamics. Alternative approaches to WHAM that retain the dynamics information have been proposed (see, e.g., the TRAM method in section 3.2.6). The advantage of these approaches is that they are less sensitive to the initial structures.

If one considers a limiting case in which the restraints used in US are infinitely strong (i.e., the value of the CVs is fixed or constrained instead of restrained), multiple restraints US becomes equivalent to a thermodynamic integration (TI) calculation (see section 3.2.7).[286] TI applied to CVs has been used to investigate nucleic acid systems in multiple studies.[240,294] A very similar principle is used in steered MD,[295] where the restraint is slowly moved so as to steer a system across the transition state. This allows one to explicitly account for the nonequilibrium work using the Jarzynski equality.[180,296] With this approach, many trajectories are simulated following the same steering protocol, and those for which less work is performed (i.e., those in which the system achieves the transitions more easily, with less force) are weighted more.[180] The Jarzynski equality allows one to obtain an exact free-energy difference even though the transition happened out of equilibrium, that is, via series of non-equilibrium simulations. Steered MD has been combined with WHAM analysis to compute the free-energy change associated with base pair fraying.[297] The strength of direct stacking interactions between a ligand and a canonical stem in a riboswitch has also been estimated using this approach.[298] The principle of steered MD and TI can be applied to any CV. When applied to root-mean-square deviation (RMSD) from a native (target) structure, it is known as targeted MD.[299] This method has been used to study dimerization of stem loop SL1 from the HIV-1 genomic RNA.[300] However, it is usually very difficult to recover quantitative free-energy differences from targeted MD simulations, especially when large conformational rearrangements are involved, because the change in the number of available states when steering the system to a low-RMSD ensemble produces an entropic contribution that largely dominates the free-energy change.

A common problem of all of the unidirectional approaches is that there is an intrinsic bias in the final result. More precisely, whenever the transition from state A to state B is enforced, the measured free-energy change $\Delta G = G^{B} - G^{A}$ is systematically overestimated. Whereas this might not be a problem when qualitative results are desired (e.g., when ranking PMFs obtained under different conditions as in ref 297), it is certainly a problem when results are interpreted quantitatively. This is also true when Jarzynski-based methods are employed with a finite number of simulations, although the bias decreases as the number of simulations increases.[301] A practical solution to this problem is to combine simulations going from A to B with simulations going from B to A so as to exploit some error cancellation. In the simplest formulation of this approach, forward and backward calculations might be compared. The result might be trusted only if it is independent of the steering direction. There is a clear analogy between this idea and the validation of US simulations shown in the Figure 13. It is also possible to rigorously combine forward and backward calculations using a maximum likelihood approach.[302] This approach was used to compute the binding affinity of the TAR RNA to a cyclic peptide.[303] CV chosen to describe the binding process in this work was the strength of the electrostatic interaction between the RNA and the peptide, which was evaluated using a Debye−Hückel model. This illustrates the richness of the possibilities that one can consider when choosing CVs.

Another notable idea derived from the original US method is that of constructing a potential adaptively during the simulation. It is important to note that this is not usually done by restraining the simulation to a given region in the CV space as is common in multiple-restraint US. Rather, the approach adheres to the original concept of US in that it involves constructing a potential that stabilizes the transition state and accelerates the system's conformational dynamics. Several methods of this type have been proposed.[304−310]

One of the most popular adaptively biased methods is metadynamics, where a history-dependent procedure is used to disfavor already visited states.[307] Penalty potential is constructed step-by-step as a sum of Gaussians, in a way that is designed to flatten the distribution of the sampled points in the low-dimensional space of the CV coordinates such that, ideally, each value of the CV is equally likely to be realized. In other words, the free-energy landscape (and thus also the probability histograms of the biased simulation) is flattened in the CV space. Unbiased free-energy surface (as defined in the low-dimensional CV space) can then be estimated by inverting the bias potential. In a more recent well-tempered variant of this approach (WT-metadynamics),[308] the penalty potential grows more slowly as the simulation progresses, eventually reaching a quasi-equilibrium state. This is achieved by a bias growth control factor γ. As $\gamma \rightarrow$ infinity, the method becomes equivalent to the original metadynamics approach, while for $\gamma = 1$ the method corresponds to the unbiased ensemble. WT-metadynamics reduces fluctuations in estimation of the free-energy differences between the metastable states of the system. Moreover, in WT-metadynamics, one can easily tune the parameters of the simulation to obtain a controllable partial flattening of the free-energy landscape, avoiding the exploration of unnecessary high-energy states. This is because WT-metadynamics yields the relationship $V = -(1 - 1/\gamma)F$ between the constructed bias potential V and the target unbiased free-energy landscape F (both defined as low-dimensional functions of the CVs). Another important parameter controlling flattening of the free-energy landscape is the shape of the Gaussian functions used as a bias. Nonetheless, a method for setting their width automatically was also developed.[311] Excellent reviews of both flavors of metadynamics are available.[182,312]

WT-metadynamics has been used to study RNA tetraloops (TLs).[129,313] However, full convergence was not achieved in these simulations likely due to a too complex nature of the TLs conformational dynamics that was excessively oversimplified by the selected CVs. Nevertheless, even simulations that did not achieve statistical convergence have provided useful insights into the studied systems. For instance, WT-metadynamics results complemented those obtained with other methods indicating that the native structure of a TL was not correctly

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

predicted by the tested force fields (see sections 4.2.2.3, 4.2.2.4, and 3.1.2.5).[129]

The main advantage of methods based on construction of an adaptive bias potential when compared to multiple restraint US is that they only converge when explicit conformational transitions are observed, and so their reliability is significantly easier to assess. In other words, when multiple recrossing events are reported in metadynamics simulations, the result is statistically trustable. Simple tests can be performed to verify whether important CVs are missing.[314,315] A method was also proposed that allows improved CVs to be chosen after a metadynamics simulation has been performed, based on the rule that better CVs should find higher free-energy barriers.[316] In principle, this method enables the use of iterative optimization procedures. Note that a basic hysteresis in metadynamics typically leads to some overestimation of the barriers. However, use of entirely inappropriate CVs that create an overlap of different metastable states in the low-dimensional CV projection typically causes barriers to be severely underestimated. In this case, finding more appropriate CVs that lift the overlap of the metastable states reveals the true free-energy barriers; the barriers in the mentioned work were estimated using a reweighting procedure (see section 3.1.4).[317]

Easy-to-detect unstable behavior of metadynamics when the chosen CVs are not sufficient is perhaps one reason why the community using this method has developed a large number of CVs. PLUMED[318,319] and colvars[320] are software plugins that implement a very large number of possible CVs and can be used both with metadynamics and with other CV-based enhanced sampling methods. Having a large database of implemented CVs is very useful, as finding the right CVs is essential for all CV-based enhanced sampling methods. Additionally, algorithms have been proposed to take advantage of multiple simulations that can be run in parallel on large computer clusters, such as multiple-walkers[321] and altruistic[322] metadynamics. Multiple-walkers metadynamics allowed free-energy profiles associated with phosphodiester backbone cleavage in small RNA motifs to be computed with high statistical accuracy using QM/MM molecular dynamics with semiempirical potentials (see also section 4.8.2).[323] Interestingly, although metadynamics was originally conceived as a method for flattening (either completely or, in the WT version, partially) the histogram of a selected CV, it has been extended to allow sampling of arbitrary preassigned probability distributions in the CV space.[324,325] In particular, it has been used to enforce dihedral angle distributions on dinucleotide systems corresponding to the populations observed in structural databases.[143]

The key complicating factor of all CV-based methods stems from the difficulty of choosing appropriate CVs. Even more importantly, their applicability is critically dependent on the properties of the studied process, that is, the extent to which it can be acceptably described using a low-dimensional set of suitable CVs. On the one hand, processes that can be properly (i.e., without a loss of information) described by reaction coordinates with well-defined transition states and intermediates are ideally suited for CV-based descriptions. On the other hand, complex folding landscapes possessing a kinetic partitioning[215−218] (multiple competing metastable states, i.e., free energy basins) might be not reducible to CV-based descriptions (see also the cautionary comment in section 3.2.3).[170] To be effective in accelerating sampling, CVs should have two properties. First, they should be able to distinguish between relevant metastable states. If they lack this property, any bias potential applied to these CVs would affect the nondistinguished metastable states by an equal amount of bias as compared to the unbiased simulation. Indeed, such states would effectively merge (overlap) in the reduced (low-dimensional) space of the CVs despite being well separated in the full space of the Cartesian atomic coordinates. Thus, if these states have significantly different unbiased stabilities, some of them (the less stable ones) might be completely masked by other states in the CV-based simulation and thus not observed. Additionally, the CVs should properly describe the transition state to allow for its stabilization (i.e., to increase its accessibility) and the acceleration of the relevant transitions (Figure 15). This is usually even more difficult to achieve. In other words, the construction of CVs is the most fundamental issue when using any CV-based method, irrespective of the specifics of its implementation. One should therefore try to assess the sufficiency (completeness) of the chosen set of CVs for the intended computation before conducting simulations. Common sense can often indicate whether a system or phenomenon of interest is likely to be outside the method's limits of applicability. The easiest way to conduct such an assessment is by checking whether a CV is truly capable of describing and distinguishing between the relevant stable or metastable conformations, for instance, by performing short unbiased simulations starting from those metastable conformations that are already known and projecting them on the putative CVs. Note that, especially in the biomolecular world, some processes and free-energy landscapes may be too complex to be reducible to a few collective degrees of freedom, no matter how smartly constructed. In such cases, a meaningful coarse graining of the dynamics to a few CVs might be fundamentally unachievable. This issue has been extensively discussed in the context of an intricate problem involving folding of guanine quadruplex DNA molecules that provides clear examples of processes not reducible to a few CVs.[170,326] In addition, even CVs that appear chemically intuitive might be finally found as inappropriate for correct description of the studied process. A textbook example is proton transfer reaction in protonated water trimer.[327]

A simple empirical consistency check can be performed by looking at the reported free-energy landscapes. In a CV-based method, the maximum boost is on the order of $\exp(\Delta G^{\ddagger}/k_{\mathrm{B}}T)$, where $\Delta G^{\ddagger}$ is the barrier reported in the free-energy landscape. If this boost is not larger than the ratio between the real (experimental) time scale and the simulated time scale, it is unlikely that the correct event has been simulated, unless the transition time predicted by the force field is also much faster than the experimental one, due to (in this particular case potentially fortunate) force-field errors.

Most commonly used CVs are based on geometric parameters such as distances, angles, or dihedral angles (Figures 13 and 14). The dihedral angles of chemically bonded atoms are often used because they make it possible to directly monitor isomerization processes. Distances are often used to describe binding events. When considering the formation of hydrogen bonds, one could use either distances or combinations of distances and angles. However, CVs can be arbitrary functions of the Cartesian coordinates of all of the atoms of the system, although typically they depend only on a fraction of them. At every time step, the CVs are computed from the atomic coordinates and used to compute the bias potential, which predicts how much every conformation in the CV space should

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

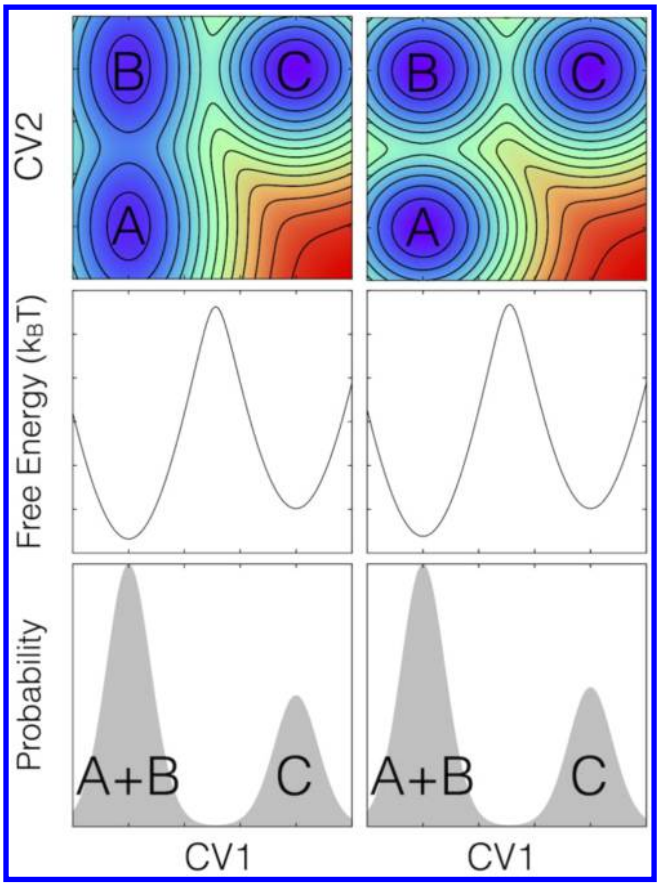


Figure 15. Free-energy landscapes for a model three-state system. Upper panels: Two-dimensional landscapes, with spacing of contour lines 1 k_BT, as a function of two collective variables, CV1 and CV2. Lower and higher free-energy levels are colored in blue and red, respectively. Middle panels: Free-energy profiles analytically projected on the collective variable CV1. Lower panels: Probability distributions as a function of CV1. Notice that states A and B correspond to the same value of CV1 and are thus seen as a single minimum in the one-dimensional CV1 projection. Left panels: In this case, the free-energy barrier separating states A and B is only a few k_BT high, and transitions from A to B and vice versa happen spontaneously due to thermal fluctuations, without any need to enhance sampling. As a consequence, an enhanced sampling method biasing only CV1 is expected to properly capture the transitions between all three states. Right panels: In this case, the free-energy barrier separating states A and B is large, and CV2 is an important slow degree of freedom. Thus, enhanced sampling simulation biasing only CV1 would not be able to enhance transitions from state A to state B or C. If the system reaches A, it will be trapped there for a long time. If the simulation starts at B or C, biasing the CV1 will facilitate transitions between these states while state A will remain invisible.

be penalized. Knowing the derivative of the bias potential with respect to the CVs and the derivatives of the CVs with respect to the atomic coordinates, one can compute the extra force that should be applied on the system during the simulation. Some CVs can be very intricate functions of the atomic coordinates, such as the number of native H-bonds.[313] Even electrostatic interaction energy,[303] full potential energy,[260] or conformational entropy[328] can be chosen as CVs. Interestingly, TICA projections (see section 3.2.3) were also proposed as CVs for metadynamics simulations.[329] Notice that all of the conformations having the same values of CVs merge in CV space, and thus are biased by an identical bias potential. Thus, when the number of native H-bonds is used as a CV, all geometries with the same number of these H-bonds are biased by the same amount, irrespective of which native H-bonds are present and which are absent. This dimensionality reduction obviously brings a loss of information.[170] From a practical point of view, CVs should be computed on the fly on every step of the MD simulations, so a CV that is particularly expensive to compute might slow the simulation. However, this problem can be alleviated using a multiple time step framework.[330]

Structural deviation from a starting conformation chosen a priori might also be used as a CV. However, despite its popularity, standard RMSD after alignment[331] is a very poor CV or parameter for monitoring MD simulations (see section 4.2.2.3). RMSD between torsion angles usually provides a better discrimination of alternative structures, but its ability to describe the formation of contacts related to atoms that are far from each other in the molecular topology is low. As an extreme example, the torsional angles in a RNA duplex and in two separated A-form single stranded RNA molecules are identical, so torsional-RMSD would be unable to distinguish between these two structures. A promising RNA-specific variable is εRMSD,[223] which provides a three-dimensional contact map of base−base interactions. Given the importance of base−base interaction patterns in RNA (section 2.2),[40] this structural deviation is capable of summarizing in a single, continuous function of the coordinates the relevant disparities between two structures. εRMSD has a high correlation with interaction network fidelity,[332] which is often used to compare RNA structures, but has the advantage of being a continuous function of the atomic coordinates. It is thus usable as an RNA-specific CV for enhanced sampling methods.

Structural deviations from multiple intermediates along a pathway can be combined to create path-associated CVs.[333] In such cases, one defines a path between two states in analogy to a reaction coordinate, and the two associated CVs take into account the degree of progression along this path and orthogonal deviations from the path. Another possible route is to combine structural deviations with respect to known templates so as to measure the amount of structure present in a molecule. This idea was proposed for studying the folding of β-sheets in proteins[334] and has not yet been applied to RNA. It must also be noted that sometimes water plays a very important role and should be included in the CVs used to describe a reaction. Perhaps the simplest example is the association of an anion and a cation in water: because desolvation of the ions is an important intermediate step leading to association, good CVs for describing NaCl dissociation could be, for instance, the distance between the two ions as well as the number of water molecules coordinated to each ion. In one study, the electric field generated by neighboring water molecules on the two ions was used in a similar manner.[307] This effect is even more marked when interaction with water is stronger, such as in the case of divalent cations. In another work, the biased CV chosen by the authors was the number of waters coordinated to an Mg^{2+} ion.[335] This significantly reduced the water exchange time in the simulation, enabling fast RNA-ion binding and unbinding events to be observed.

Note that CV-based methods can also be used to compute affinities between two molecules, such as binding of a ligand to RNA. As discussed above, one should bias CVs that are capable to properly distinguish the bound and unbound states, as well as the transition state. In addition, when ligand binding is calculated, one needs to take into consideration that this is an

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

association/dissociation process, which is concentration-dependent. To properly compute affinities at the standard concentration, one should take into account any applied constraints and the relation between the CV space and the corresponding volume in the space of the Cartesian coordinates. For example, when simulating unbinding of a ligand, the distance r between the ligand and the binding site can be taken as a CV. However, the available volume for the movement of the ligand in the solvent in the Cartesian coordinate space increases proportionally to r^2, because it corresponds to a surface area of a sphere with radius r. The larger amount of space that should be explored at larger values of r might slow the sampling. This problem can be circumvented, for example, by applying a cylindrical restraint/constraint fixing the direction in which the ligand is transferred into the solvent. This can be done because the free energy (PMF) of the ligand that is already sufficiently far from the binding site in the bulk solvent is independent of its position (uniform); thus, it is not necessary to sample the whole bulk. However, the free energy associated with such constraints then needs to be included in the calculation. In other words, correct execution of binding free-energy calculations is not easy and requires one to respect certain basic methodological rules; further details are beyond the scope of this Review and can be found in the literature.[336−339]

3.2.6. Combinations of Enhanced Sampling Methods. Most of the methods used to access long time scales in biomolecular simulations can be assigned to one of the three categories reviewed above. However, there are also important ways of combining ideas from methods belonging to different classes.

Metadynamics can be combined with T-REMD.[340] Here, the main role of T-REMD is to accelerate the exploration of all of the system's degrees of freedom, while metadynamics is used to change the balance between different regions in the CV space. This might be useful to stabilize transition states so as to accelerate some a priori known transitions. In addition, the combination can also be exploited to stabilize structures that are only metastable but for which one would like to compute the thermodynamic stabilities. Ideally, this is applied for unfolded or incorrectly folded structures. However, due to the poor quality of current force fields (see section 3.1), this idea can also be used to induce formation of the (otherwise metastable) native structure in a biased simulation. If the force field predicts a spurious structure to be the global minimum, the native structure may have a very low population in unbiased simulations. However, flattening of the free-energy landscape using an appropriate set of CVs may enable sampling of the native structure in biased simulations. This idea has been exploited to quantify the change in the thermodynamic stability of the native structure of RNA TLs in response to modifications in the force field (section 4.2.2.3).[163] This was the first work in which the εRMSD[223] variable was used as a biased CV (Figure 16). An identical protocol was later used by another group to test force-field modifications[152] (see also section 3.1.2.4) and to investigate the folding of a DNA quadruplex.[341]

Another idea is bias-exchange (BE) metadynamics,[342] where different replicas are used in metadynamics simulations acting on different CVs, and coordinates are swapped periodically using a standard H-REMD procedure. Thus, several replicas are simulated, each with an independent metadynamics simulation using a different CV. The difference from conventional

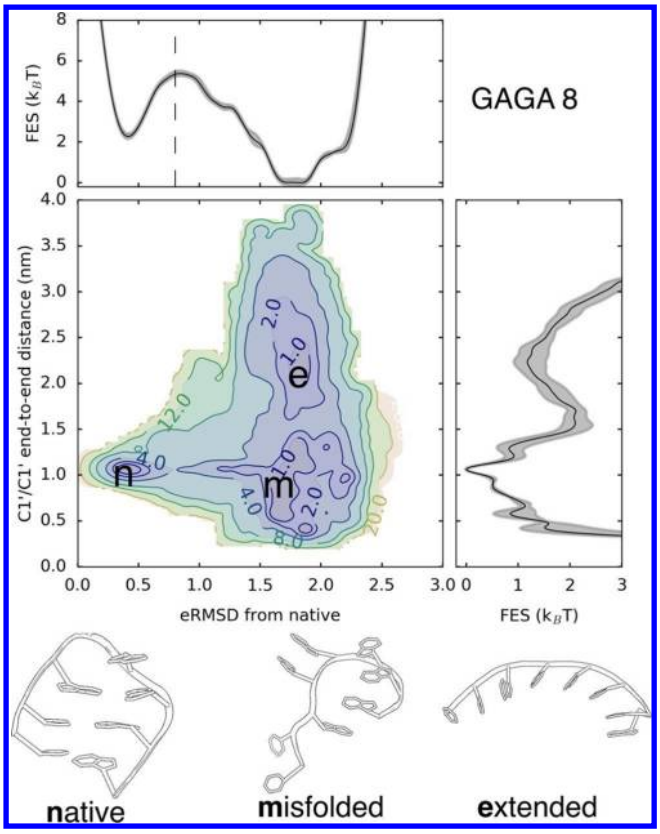


Figure 16. Free-energy landscape of a GAGA tetraloop using the χ_{OL3} force field. The landscape was computed using WT-metadynamics on the εRMSD from the native structure coupled with T-REMD, but is here reported as a function of a posteriori chosen variables, εRMSD (upper panel), end-to-end distance (right panel), and both variables (central panel). Free energies and isoline labels are reported in k_BT units. In the lower panel, typical folded, extended, and misfolded structures are presented. WT-metadynamics allowed one to sample the native structure despite that it is not the most stable structure according to the force field (see section 4.2). It is also instructive to notice that native and misfolded structures have a very similar end-to-end distance. Any enhanced sampling method applied on the end-to-end distance alone would not be capable of distinguishing native and misfolded structures and would thus likely lead to results highly dependent on the initial structure. Reprinted with permission from ref 163. Copyright 2016 American Chemical Society.

metadynamics is that conventional metadynamics combines several CVs to construct a true multidimensional description of the system in the space of the used CVs. However, the computational requirements of such description increase sharply with the number of CVs, so typically only two or three CVs are used. In BE metadynamics, the individual simulations with different CVs run independently, and coordinate swaps are proposed with a procedure analogous to the H-REMD approaches discussed in section 3.2.4. This allows one to use more CVs. However, the coupling (interdependence) between CVs is not explicitly taken into consideration, and the penalty potentials act on one CV at a time. Nevertheless, if such one-dimensional penalty potentials are sufficient to induce an exhaustive exploration of the conformational space, the free-energy landscape can be reconstructed, usually as a function of a few of the biased CVs. The parallel bias method allows similar results to be obtained without the need to use multiple replicas.[343] Simulations using BE metadynamics can be analyzed by using WHAM so as to

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

combine results from different replicas.[344] This approach permits significant flexibility in the setup of the individual replicas. In practice, any change to the Hamiltonian that is believed to improve the quality of sampling can be included and discounted a posteriori. For instance, a variant of BE was used to compute the relative stability of different Mg^{2+} binding sites on RNA molecules (Figures 17 and 18).[335] Here, in each

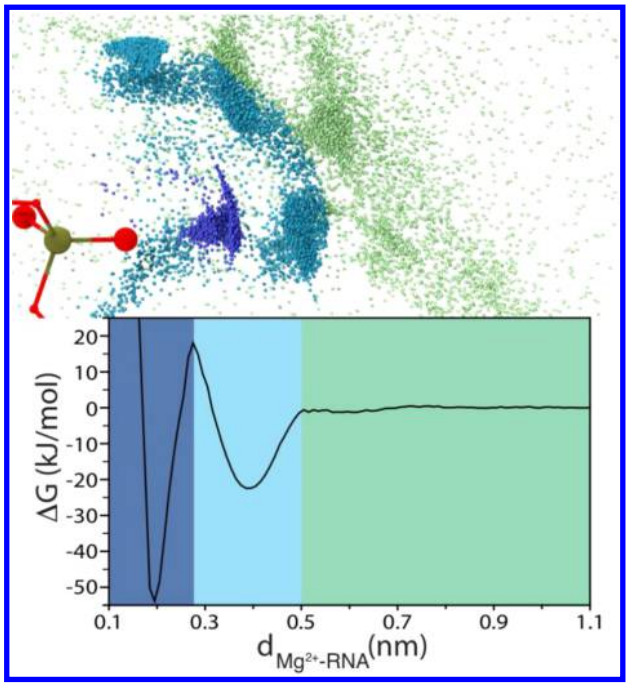


Figure 17. Bottom panel shows the free-energy profile as a function of the distance between Mg^{2+} and the closest RNA atom. The free-energy profile was obtained by a 9 μs long bias-exchange metadynamics simulation, whose details are presented in ref 335. Top panel displays positions of Mg^{2+} ions that are interacting with the central phosphate moiety on the guanine side of an A-form RNA duplex with four G═C base pairs. The phosphorus atom and the nonbridging oxygens are represented as spheres (ochre and red, respectively). The blue dots represent the Mg^{2+} ions that are directly bound to RNA, the cyan dots are those that interact through at least one water molecule of the first coordination shell of the cation, and the green dots correspond to the ions that do not bind in any way to the RNA.

replica, the binding of an Mg^{2+} ion at a different putative binding site was enhanced by performing metadynamics simulations using the distance between the ion and the corresponding binding site as well as the coordination number of the ion with water as the chosen CVs. At the same time, binding to competitive sites was hindered by artificial restraints. These restraints prevented ions in individual replicas from becoming stuck in binding sites that were pertinent to different replicas and were automatically taken into account during the following WHAM procedure. See, however, section 3.4.2 for a discussion of the force-field limitations of Mg^{2+} description.

Another interesting possibility is to combine US with replica exchange. This approach was used in a study where US was used not to progressively steer the system from one structure to another but to penalize stable rotamers.[273] A ladder of replicas was then constructed in which the degree of destabilization was progressively increased. While this method is technically implemented by imposing a bias potential on selected degrees of freedom, its spirit closely resembles that of T-REMD, solute

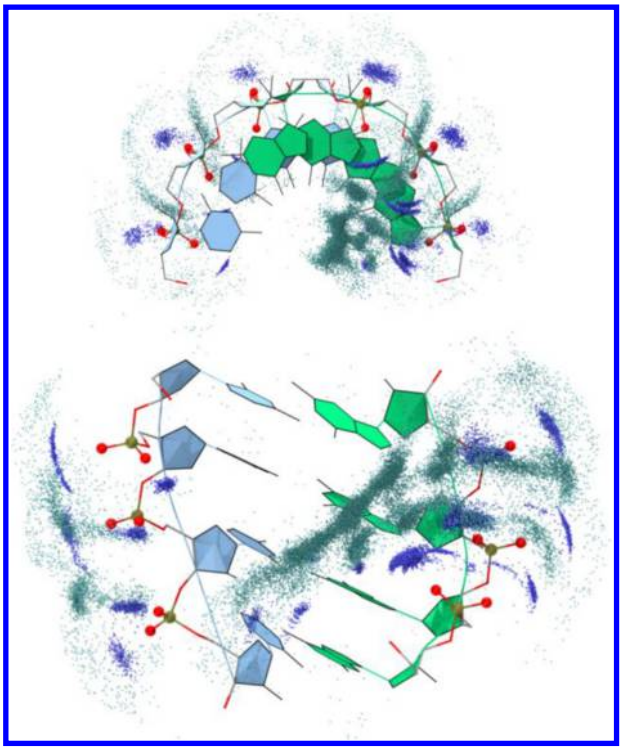


Figure 18. Different angles of an overall view of the Mg^{2+} ion distribution shown in the top panel of Figure 17. The colors correspond to those in Figure 17; that is, blue is used for directly bound ions and green for indirectly bound ones.[335]

tempering, and, in general, H-REMD. Indeed, the number of concurrently biased CVs can be very large, and very little a priori knowledge about the transitions to be enhanced is required.

An extension of this idea in which metadynamics is used to adaptively construct penalizing potentials has been called replica exchange with collective variable tempering (RECT),[259] and has been applied to the conformational sampling of oligonucleotides.[143,144,152,259] Here, a large number of CVs is chosen, and a bias potential designed to flatten their histogram is iteratively constructed. To avoid the difficulties of constructing a high dimensional histogram, the CVs are assumed to be independent when constructing this potential. This potential thus typically disfavors sampling of minima. To reconstruct an unbiased ensemble, the strength of this biasing potential is gradually modified along a ladder of replicas. The ladder of replicas is constructed by modulating the bias factor γ (see above) of WT-metadynamics, which allows for a progressively increasing partial flattening of the individual histograms across the ladder.[308] The use of metadynamics in the construction of penalizing potentials allows simulations to be performed with very limited a priori knowledge. The metadynamics principle is here used to increase dynamics in the high-bias replicas, while the unbiased replica is again used to collect the unbiased populations, that is, the thermodynamic data, as in most REMD schemes.

Metadynamics can be also combined with replica exchange in a framework where the value of some observable averaged across the replicas is restrained so as to agree with experimental data (replica averaged metadynamics, or RAM).[345] This approach can be used to partially compensate for force-field inaccuracies by forcing the system to reproduce experimental,

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

for example, NMR, data and simultaneously enhance sampling. RAM has been used to characterize low-population states of a TAR RNA−peptide complex[346] and a UUCG tetraloop[347] (see section 4.7.5). In both applications, experimental residual dipolar couplings were enforced using a tensor-free method.[348] The problem of such applications is that available primary experimental data might be, in some cases, not sufficient to overcome or suppress the limitations related to the force-field inadequacies.

Importance sampling techniques can also be combined with MSM through the transition-based reweighting analysis method (TRAM). Recently proposed in its most general version,[349] TRAM makes it possible to combine multiple biased simulations (such as those usually analyzed with weighted histogram methods) with unbiased simulations to reconstruct the full thermodynamics and kinetics of the system. Briefly, the idea of the method is to define different "thermodynamic ensembles", one for each bias condition, including the unbiased ensemble. Each of these ensembles is divided into microstates and treated as a single MSM. The parameters of these MSMs are computed from likelihood maximization by solving a system of equations that couples the different ensembles. Under certain conditions, it is possible to obtain unbiased bidirectional rates of a transition even if the transition in the unbiased simulations was observed only in one direction. This approach is thus a promising tool for studying kinetics of conformational changes between states with very different stabilities for which transitions in one of the two directions cannot be sampled in a feasible time with unbiased MD simulations. A closely related method was subsequently introduced and applied to study RNA duplex formation[350] using the oxRNA coarse-grained model[351,352] (see also section 3.3).

3.2.7. Alchemical Methods. All of the enhanced sampling methods discussed above are designed to allow MD simulations to realize conformational changes that would require a very long time to be observed in unbiased simulations. These methods essentially induce transitions by forcing the system through unlikely paths and allow the effect of this forcing to be discounted a posteriori. However, similar methods can be used to compute free-energy differences between systems with different numbers or types of particles. Here, the role of the CV is played by some control parameter λ that continuously converts a given potential energy function to another one. One classic historical example is conversion between an ion and a neutral atom (neon),[353] which can ultimately allow the solvation free energy of an ion to be evaluated. Simulations of this kind are called alchemical simulations because they allow the user to interconvert between atoms of different types and annihilate or create atoms, that is, to perform chemical transformations. An excellent introduction to alchemical methods has been published.[354] Note that in modern computations of ion hydration energies, the ions are typically completely annihilated (alchemical ion to nothing transformation) in a two-stage process. In the first stage (charge neutralization), the ion's charge is slowly neutralized, and in the second stage (disappearing), the vdW potentials are slowly removed. This approach has been used, for example, to derive vdW force-field parameters of ions.[355]

A prototypical method of this class is thermodynamic integration (TI), in which simulations are performed for a set of values of the control parameter conventionally ranging between 0 and 1, where the end points correspond to the chemical species considered. For each simulation, the derivative of the total free energy of the system with respect to the control parameter is computed, and, by integration, one can obtain the free-energy difference. The TI method has a clear conceptual relationship to umbrella sampling (US) simulations that are performed with multiple restraints and combined using WHAM (see section 3.2.5). In both cases, multiple simulations are performed at intermediate steps along the transition (or transformation). The TI method could thus suffer from problems similar to those of the US method. For instance, if simulations at different values of the control parameter are initialized (seeded) by unidirectionally increasing the parameter's value, the estimate of the free-energy difference might be affected by a systematic error. Simulations initialized with an inverted protocol should have a systematic error of the opposite sign, and could thus be used for validation. In the context of TI, this effect is known as "Hamiltonian lagging".[356]

Another alchemical method that has applicability similar to that of TI is the free-energy perturbation method (FEP). Instead of performing the alchemical change in small steps as in the TI protocol, one might perform it in a single step by running a single simulation. This would be made using the potential energy function (Hamiltonian) corresponding to the initial state (A) and virtually replacing it with that of the final state (B). In other words, one runs a normal simulation for state (Hamiltonian) A, but for each counted configuration the energy corresponding to state B is also computed. This can be done in the formally exact framework of FEP.[357,358] However, the statistical efficiency of this procedure is limited by the fact that initial and final states should be significantly overlapping.[359] For this reason, it is more common to see FEP applications where the change is performed in stages, using series of intermediate states between A and B, as in the TI protocol. In addition, similarly to the other "directional" methods, using simulations in both directions of the transformation (from A to B and from B to A) leads to a significant cancellation of error that makes the procedure statistically more reliable.[360]

As with multiple-restraint US, the safest approach to TI and FEP may be to use a replica-exchange framework in which simulations at different values of the control parameter are performed simultaneously and coordinate swaps are attempted from time to time.[361] For instance, Sakuraba et al. used this approach to compute the free-energy impact of base pair substitution in canonical A-RNA[362] (see section 4.3.3). In the case of TI or FEP, this approach actually makes it possible to verify the existence of continuous trajectories through the space defined by the control parameter.

In any transformations of atoms or chemical groups, the control parameters should affect both the charges on the atoms and their vdW interactions. As mentioned above, it is strongly recommended to nullify the partial charges on the atoms before modifying their Lennard-Jones parameters to prevent a singularity in the electrostatic term when the atoms start to overlap. This procedure is known as a two-step transformation. The vdW term also presents some technical difficulties related to the divergence of the $1/r^{12}$ potential, which makes the integral numerically difficult to solve. In other words, there is a visible effect of the singularity that occurs at interatomic distances close to $r = 0$, which are encountered (sampled) at the very end of any disappearing process and at the beginning of any creation process. This problem is nowadays usually addressed by employing so-called soft-core potentials while decoupling or annihilating Lennard-Jones interactions. Soft-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

core potentials guarantee finite pairwise interaction energies for all configurations and thus prevent singularities and instabilities that would otherwise occur at the end points of the TI path;[354,363,364] for more information, see also http://www.alchemistry.org/wiki/Best_Practices.

Note that the potential energy functions used in MD are always defined up to an arbitrary constant, so that the physical quantities that can be compared to experiments are always differences between states with the same number and types of particles. Figure 19 shows a thermodynamics cycle, which

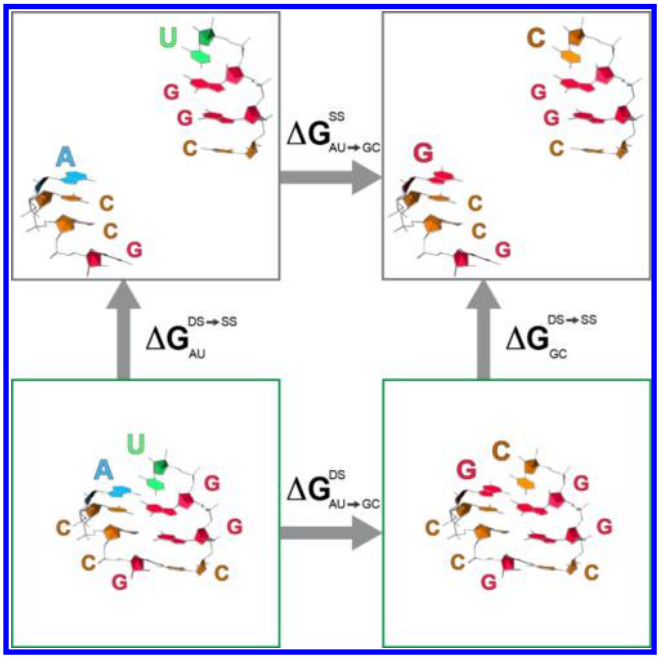


Figure 19. A scheme representing the procedure required to compute the effect of substitution of an A═U base pair with a G═C base pair on the stability of a short duplex. Alchemical transformations can be used to compute $\Delta G^{DS}_{AU\rightarrow GC}$ (lower panels) by simulating the dsRNA. Similarly, an alchemical transformation can be used to compute $\Delta G^{SS}_{AU\rightarrow GC}$ (upper panels) within the ssRNAs. Because free energy is a state function, the difference between these two numbers is equal to the difference in the stability of duplexes with A═U and G═C base pairs in the mutated position. In other words, one computes the processes along the horizontal arrows to obtain the free-energy difference for processes marked by the vertical arrows exploiting the relationship $\Delta G^{DS}_{AU\rightarrow GC} + \Delta G^{DS\rightarrow SS}_{GC} = \Delta G^{DS\rightarrow SS}_{AU} + \Delta G^{SS}_{AU\rightarrow GC}$; that is, $\Delta\Delta G^{DS\rightarrow SS}_{AU\rightarrow GC} = \Delta G^{DS\rightarrow SS}_{AU} - \Delta G^{DS\rightarrow SS}_{GC} = \Delta G^{DS}_{AU\rightarrow GC} - \Delta G^{SS}_{AU\rightarrow GC}$. This protocol allows one to circumvent the explicit simulation of the transition between a dsRNA and two ssRNAs, that is, of the duplex formation, which would be very challenging. However, even when simulating the alchemical changes along the horizontal arrows, one must make sure that both the dsRNA and the ssRNA states are extensively sampled. As we will see in section 4.1, sampling a flexible single-stranded oligonucleotide is far from trivial.

allows one to analyze the effect of base substitutions on the stability of an RNA duplex. The free-energy effect of substituting an A═U base pair by a G═C base pair in a double helix is defined as a difference ($\Delta\Delta G$) between duplex formation free energies (ΔG) with the original and mutated sequence. Calculations of free energies of duplex formation processes are very challenging because they require hybridization and dehybridization to be sampled multiple times to provide a converged result. In contrast, it is much more convenient to make alchemical calculations of the mutations within the duplex and single-strand contexts, as shown in Figure 19. The resulting $\Delta\Delta G$ values can then be compared to experiment. In some cases, equivalent $\Delta\Delta G$ values can be obtained using enhanced sampling methods such as those described in the previous sections, where a physical trajectory is simulated between the ending states. However, there are cases where the physical process of interest is difficult to simulate, and it is more convenient to simulate the "alchemical legs" of the thermodynamics cycle. This is also the case of the ion hydration free energies. We cannot easily study the physical process of the immersion of an ion into water from gas phase, as it could only be directly calculated by simulating a water/vacuum interface, which would require a huge simulation box. However, one can relatively easily calculate the free energy associated with the creation of the same ion in water (and in the gas phase) from nothing. The thermodynamics cycle is then trivially completed by immersing "nothing" into water.

TI has been used to compute free-energy changes associated with chemical modification of one or more nucleobases in RNA systems, for instance, to quantify the effect of replacing a guanine−cytosine canonical base pair proximal to a non-canonical A/G base pair with an isoguanine−isocytosine (isoG═isoC) base pair.[365] Isoguanine and isocytosine have their amino and carbonyl groups transposed relative to guanine and cytosine, so the substitution is entirely isosteric (see section 2). The work compared the stabilities of duplexes with two different arrangements of the A/G base pairs. In this case, all of the TI computations for the thermodynamics cycle were performed on duplexes, with no need for any single-strand computations. This is because the calculations studied the response of the A/G geometry to the substitution. A comparison with NMR data revealed only modest agreement between theory and experiment. This could be perhaps related to the use of the ff99 AMBER force field. Further, the authors attempted a modification of the force field to account for amino group nonplanarity,[365] to capture out of plane H-bonds involving unpaired guanine amino group in the *anti*−*anti* conformation of the A/G base pair with the adjacent base pairs.[106] This work is in more detail described in section 4.3.3.

TI computations can also be used to quantify how a mutation would affect stability of a folded structure of a biomolecule. In this case, one might use TI to introduce the mutation in both the folded and the unfolded structures. This means to simulate conformationally very different states, similarly to the case shown in Figure 19. In an example of this approach, mutations from guanine to inosine were alchemically induced in a sequence containing a GCAA tetraloop in its native structure.[366] To quantify the mutation's effect on the system's stability, the authors also computed the free-energy change associated with the mutation, considering the guanine in the context of small fragments that were used as models for the unfolded state (either a single nucleoside, a trinucleotide, or a pentanucleotide). Use of these small models allowed one to avoid performing difficult calculations on a longer flexible ssRNA. The authors reported both forward and backward protocols so as to provide lower and upper bounds for the results, which were in reasonable agreement with available thermodynamic data.

Another interesting case is the use of TI to quantify the effect of a guanine-to-inosine substitution on the stability of a canonical duplex, which was performed with the χ_{OL3} force field.[367] The free-energy differences of duplex dimerizations caused by the guanine-to-inosine substitutions were derived via

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

thermodynamic cycles using alchemical TI guanine-to-inosine mutations in the context of duplex and single-strand RNAs (Figure 19). ssRNAs are flexible in solution and, as discussed in section 4.1, are very challenging systems on their own. However, in this work, possibly because the simulations were relatively short (~20 ns), the simulated single strands remained structurally stable and allowed the authors to calculate free-energy changes with small statistical errors. The calculated $\Delta\Delta G$ values were within ~1 kcal/mol of the experimental results, which is an excellent agreement given the approximations used. A similar calculation was reported for the C═G → isoC═isoG substitution.[138] In this case, the difference between ff99 simulations and experimental data was on the order of 12 kcal/mol. This difference was reduced to an average of 2 kcal/mol using the Amber99TOR reparameterization designed by the authors. An important difference between the Amber99 and Amber99TOR force field relevant in this context could be the parameterization of the glycosidic torsion, similarly to the χ_{OL3} force field (see section 3.1.2.1). The Amber99TOR potential, which modifies also several other dihedrals, was later shown to be suboptimal for A-RNA double helices.[58]

TI was also used to examine the effect of the Arg142Ala substitution on RNA binding in the SRSF1 RNA recognition motif 2 (RRM2) protein. The computed result was in good agreement with data from isothermal titration calorimetry (ITC) experiments.[368] In yet another study, the same authors used TI to evaluate the influence of two hydration sites reported in simulations of the Fox-1 RRM on protein/RNA binding affinity. The hydration sites were abolished by alanine mutations of the water-coordinating residues Ser155 and Ser122. Good agreement with experiment was obtained for the Ser155Ala mutant but not for Ser122Ala.[369] More details on these results are reported in section 4.7.2.2. Note that in both studies, the effect of mutations in the unbound system was evaluated in the protein. This might have contributed to the reported agreement between experiments and TI calculations because the unbound proteins had a well-defined folded structure that could be properly sampled in the simulation's time scale. This is often very difficult to achieve for flexible single-stranded RNA molecules (see section 4.1).

As a closing comment about alchemical methods, it should be noted that the control variable used to induce the transition is not always capable of properly describing the different possible substates of the system. In other words, using the language of the enhanced sampling methods discussed above, it would not be necessarily a good CV. For instance, if the mutation of a nucleobase is coupled with a *syn*/*anti* flip in the glycosidic bond, it is very unlikely that a TI procedure that only alchemically converts one base into another would capture the correct free-energy change. In general, one can assume that accuracy of the TI method is optimal in case of simple modifications, such as purine to purine or pyrimidine to pyrimidine transformations, which are not associated with significant conformational changes. This means, for example, a substitution of one canonical base pair by another in an A-RNA helix without interchanging the purine and pyrimidine positions. Reliability of alchemical computations is expected to deteriorate sharply if more challenging substitutions are attempted. It may be possible to partially alleviate these problems by combining alchemical methods with enhanced sampling techniques such as those discussed earlier. For more studies using alchemical approaches, see also sections 4.3.3, 4.3.9, 4.5.3, and 4.7.2.

3.2.8. Continuum Solvent Methods, MM-PBSA and MM-GBSA. In all of the methods that we discussed in the preceding sections, free-energy difference is computed by enforcing real or "alchemical" transitions between one conformation (state) and the other. To get converged results with methods such as metadynamics, multiple transitions need to be realized. Similarly, when using methods such as US simulations with multiple restraints, the separate simulations that one performs must bridge in a continuous manner from one state to another. This is required to avoid explicit calculation of the partition function of the system, which, in turn, would require averaging over all degrees of freedom of the system. However, if one were able to average the Boltzmann weights on a significant number of degrees of freedom, a direct calculation of the free-energy difference might be feasible. For example, because a large fraction of the binding free energy between two molecules depends on their interaction with the solvent, an explicit calculation might be possible by integrating over solvent degrees of freedom. This can be done using implicit solvent models. Thus, free-energy differences can also be calculated with the help of continuum solvent approaches. A typical application is computation of affinity between two molecules, for example, between nucleic acid molecule and protein or ligand. The approach can also be used to compute free-energy differences between different conformations (states) of the same molecule or to monitor free energies along MD simulation trajectories. The main advantage of these methods is that it is sufficient to simulate only the end states.

Most common implicit solvent models used in this context evaluate polar and nonpolar parts to the solvation free energy separately. The polar part is computed by assigning to each atom a radius within which the dielectric constant is assumed to be different from that of the solvent. The system is then described by Poisson−Boltzmann equation that is solved either numerically on a grid (Poisson−Boltzmann method, PB) or analytically in an approximate fashion (generalized Born method, GB). The nonpolar part of the solvation free energy is estimated from a surface area (SA), resulting in so-called GBSA and PBSA implicit solvation models. Within their approximations, PBSA and GBSA allow one to compute the absolute solvation free energy of a molecule. For a review of implicit solvation methods for biomolecular systems, see ref 370.

A particularly popular approach based on the combination of PBSA (or GBSA) methods with explicit solvent MD is called MM-PBSA (or MM-GBSA).[371] When calculating free energies of two different states of a molecule, separate MD simulations in explicit solvent are performed for the two investigated conformations, without the need to simulate any transition between them. For each trajectory, one then extracts a number of snapshots including only the coordinates of the solute. For all of these frames, the solvation free energy is then computed using either PBSA or GBSA, whereas the solute energy is computed with the used atomistic force field. Free energy is thus estimated combining the solute force field energy and the solvation free energy. Finally, the free-energy difference between the two states is estimated by means of averaging the energies over all of the snapshots representing each state. Some applications of this technique to RNA systems are described in other parts of this Review. For binding free energies, the continuum solvent calculations are executed

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Table 2. List of Main CG Models Described in the Text, with the Number of Beads per Nucleotide (N_B), Number of Nucleotides of the Largest Structure Reported (Nt_{max}), the Properties Studied with the Model, and a Brief Description of the Interactions Included[a]

model	N_B	Nt_{max}	properties studied	interactions
Vfold[393−397]	2,3	200	thermodynamics	ST, stacked pairs with at least one cBP
TIS[390,398]	3	34	thermodynamics	ST and ES, hydrogen-bond network obtained from native structure
DMD399	3	100, 161 with constraints	structure, specific heat	ST, cBP
HiRe-RNA[400−402]	6,7	76	structure, FES (reduced units)	ST, ES, cBP, some ncBP
Jost and Everaers[403]	1	80	thermodynamics	lattice model, ST, BPs as in NN model
naRES 2P[404,405]	2	44	thermodynamics	ST, cBP, ES
TOPRNA[406]	3	bulges of 7 nts	thermodynamics	connectivity between stacked and canonically paired bases, requires secondary structure
oxRNA[351,352]	1 (anisotropic)	20 for thermodynamics, 1200 for mechanical properties	thermodynamics, mechanical	ST, cBP, ES
RACER[407−409]	5	52, 122 (for structure prediction with restraints)	thermodynamics	ST, cBP, ncBP, and ES
MARTINI[410,411]	6,7	4793	thermodynamics, persistence length of duplexes	ES, requires 3D structure for defining connectivity, interactions with proteins allowed
YUP[412,413]	1	76	structure	connectivity between stacked and canonically paired bases, requires secondary structure and tertiary contacts
NAST[414]	1	388	structure	connectivity between stacked and canonically paired bases, requires secondary structure and tertiary contacts
Ernwin[415]	helices and internal, hairpin, and multiloops	298	structure	A-minor, clashes, and radius of gyration-dependent
RAGTOP[416]	helices, internal loops, and junctions	158	structure	angles and dihedrals between helices; pseudoknots
RNAkb[417]	5	76	structure	implicitly contains cBP, ncBP, and ST
SimRNA[391]	5	189	structure	ST, cBP, ncBP
RNApps[418]	5	26	structure	ST, BPs according to ref 419
SPQR[420]	4	27	structure	ST, cBP, ncBP, BPh
ENMs[158,421−435]	1−3	4793	domain motion	connectivity between neighbors, 3D structure required, interactions with proteins allowed

[a]In the properties, thermodynamics stands for the calculation of free energies, specific heats, and melting curves. Structure stands for structure prediction or refinement applications. In the interactions, ST, cBP, ncBP, BPh, and ES mean stacking, canonical base pairs, noncanonical base pairs, base−phosphate, and electrostatic interactions, respectively.

separately for the complex and the two monomers, and can be performed processing either a single MD simulation of the whole complex (single-trajectory approach) or using separate simulations of the complex and the monomers (multiple-trajectory approach).[371,372]

The MM-PBSA and MM-GBSA calculations are exceptionally popular because they are easy to execute. One just needs to run a standard explicit-solvent all-atom MD simulation, then strip off the solvent, and, finally, add the solvation free-energy terms using the continuum solvent model. The method is essentially an a posteriori treatment of explicit-solvent simulations with a continuum solvent model. A key advantage is that one simulates only the end states of the studied processes and not the actual transitions. However, the use of continuum solvent models brings additional major approximations into the computations, which stem from their fundamental physical limitations. A cardinal problem of all continuum solvation models is the enormous sensitivity of computed solvation free energies to various parameters of the calculations such as the assumed atomic radii. In reality, these methods can be trained for rather narrow sets of molecules, but there is no generally valid (universal) set of parameters. This sensitivity to parameterization applies to both classical and QM-based continuum solvent models, irrespective of their specific formulation.[114,373−376] For studies demonstrating the limitations, see refs 372−374, 377, and 378. Note also that, due to the limitations of continuum solvent models, we do not recommend MD simulations of highly charged nucleic acids using continuum solvent approaches.

3.3. Coarse-Grained Models of RNA

Although the feasible time scales of all-atom MD simulations of biomolecules have increased impressively in recent decades, and enhanced sampling methods allow even longer time scales to be accessed at least in some cases, there are many phenomena that still cannot be efficiently studied by atomistic MD. Biologically interesting processes such as the folding of RNA structures or the domain motions of ribosomal subunits during translocation can develop over time scales of the order of milliseconds or longer, and might also involve millions of atoms or more. Atomistic simulations of these systems therefore usually demand a rather large or even prohibitive computational effort, necessitating the use of alternative approaches. One such alternative is to use simplified, coarse-grained (CG) descriptions. These procedures usually involve reducing the degrees of freedom of the original atomistic system by grouping selected sets of atoms and representing them as a smaller set of CG particles (also called beads) that interact through effective energy functions. This reduction reduces the complexity of the calculation, and, in general, defines a smoother energy surface, which can be explored more

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

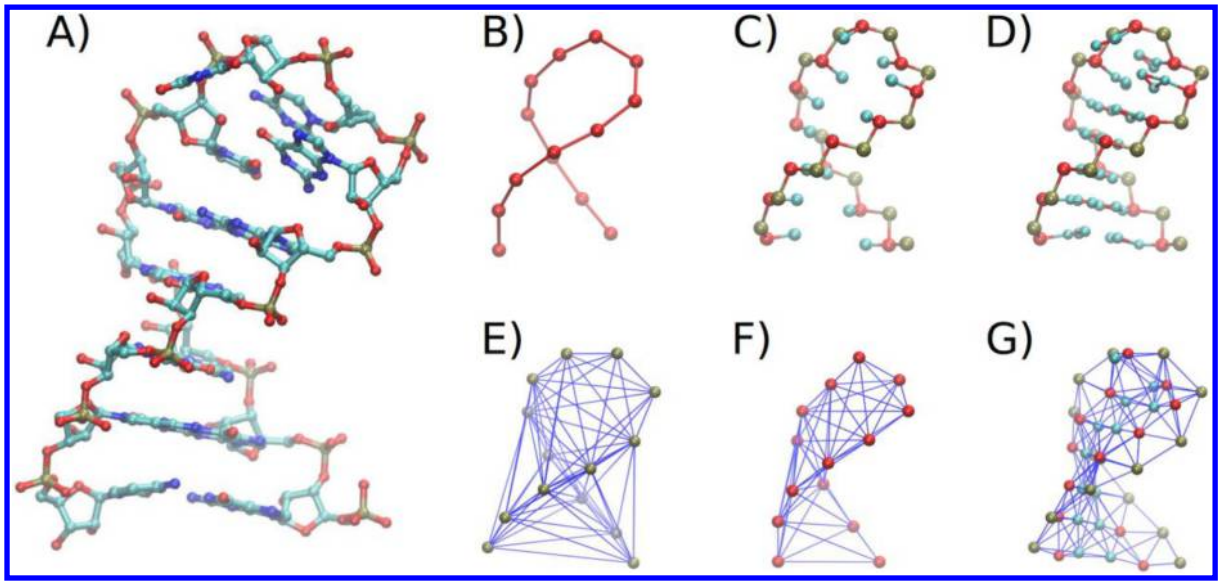


Figure 20. Different representations of the structure of the UUCG tetraloop capping a 5 base-pair canonical stem (PDB ID: 2KOC): (A) atomistic representation; (B) CG representation with one particle per nucleotide as in NAST (one bead situated on C3′ atoms);[414] (C) CG representation with three beads per nucleotide as in Vfold (one bead for N1 (pyrimidines) or N9 (purines),[393] together with C4′ and P atoms along the backbone); and (D) CG representation including three particles per base and two along the backbone, which opens the possibility of formation of noncanonical base-pairs, as in SimRNA.[391] The last three images depict ENMs representing (E) only phosphorus atoms (cutoff 20 Å), (F) the centers of mass of the sugars (as used by Setny and Zacharias,[421] cutoff 16 Å), and (G) the phosphorus atoms, sugar rings (C1′), and bases (C2), as used by Pinamonti et al. (cutoff 8 Å).[422] Sugars are shown in red, phosphate groups in tan, and bases in cyan in the CG representations.

quickly and using a larger integration time step. In consequence, larger systems and longer times can be simulated.

3.3.1. General Considerations about Coarse Graining. There is no unique way of defining a CG representation and its interactions. Accordingly, they are usually tailored to suit the characteristics of the phenomena under investigation. For example, some parameterizations focus on reproducing the distribution functions between the CG sites sampled during atomistic MD simulations.[379−384] These approaches can be practical both for enhancing the exploration of the conformational space and for recovering atomistic resolution in a later step. The quality of their approximations of course depends on the reliability of the atomistic simulations used in their parameterizations. Alternative versions of these methods use sets of experimental structures to obtain such distributions; the resulting models are said to be "knowledge-based".[385,386] This option is in some ways less rigorous, but is particularly useful when atomistic MD sampling is cumbersome or simply unreliable. Models may also be constructed by adjusting selected parameters of a predefined set of energy functions, with the aim of reproducing experimental quantities such as free energies.[387,388]

Because of the used approximations, CG models will generally only faithfully reproduce a specific set of observables in a limited region around the conditions (thermodynamic and nonthermodynamic) assumed in their parameterization. Therefore, these approaches must be used with care, and may sometimes require a redefinition of the observables of interest based on the model's resolution.[389] However, if the model's limitations are understood and respected, CG simulations can provide important insights that complement atomistic MD simulations or help explain experimental observations.

Because CG models allow much broader sampling of the conformational space than all-atom simulations, it is attractive to combine the CG and all-atom methods, with the CG methods scouting the free-energy landscape and the atomistic methods subsequently refining the CG exploration. A prerequisite of such combined approaches is the capability to reconstruct atomistic models from the CG explorations, which could be used to seed atomistic simulations.

The relation between various CG energy functions and atomistic MD force fields is not straightforward. The empirical force fields used in atomistic simulations are sometimes referred to as "physics-based" models. However, the term "physics-based" is quite arbitrary: many quantum chemists would vigorously dispute the claim that atomistic force fields offer a physics-based description because they are entirely empirical. An atomistic force field can be seen as an "unphysical" coarse graining of the true electronic structure. However, we usually expect atomistic MD simulations to be capable of simultaneously describing the structural dynamics, thermodynamics, and kinetics of the studied biopolymers, as long as no chemical reactions are involved. Conversely, when using CG approaches, we can only expect predictions of a limited set of properties of the studied RNA systems under a narrow range of conditions. Despite their apparent primitiveness, CG models may benefit substantially from the compensation of errors, to the extent that a given CG force field may, in certain cases, be more accurate than a "multipurpose" atomistic force field.

Below, we provide an overview of the principal CG descriptions that have been proposed for simulating RNA. These models can be used for various purposes including predicting the thermodynamics or native structure of an RNA system, or modeling large domain motions using a so-called Elastic Network Model (ENM). A brief summary of these features is shown in Table 2 where each method is listed with the number of CG beads per nucleotide in its representation (when applicable), the number of nucleotides of the largest structure studied under the CG representation, the properties

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

that the model has been designed to reproduce or have been studied with it, and the interactions considered in the model. CG methods use diverse representations, as illustrated in Figures 20 and 21, depending on the level of resolution

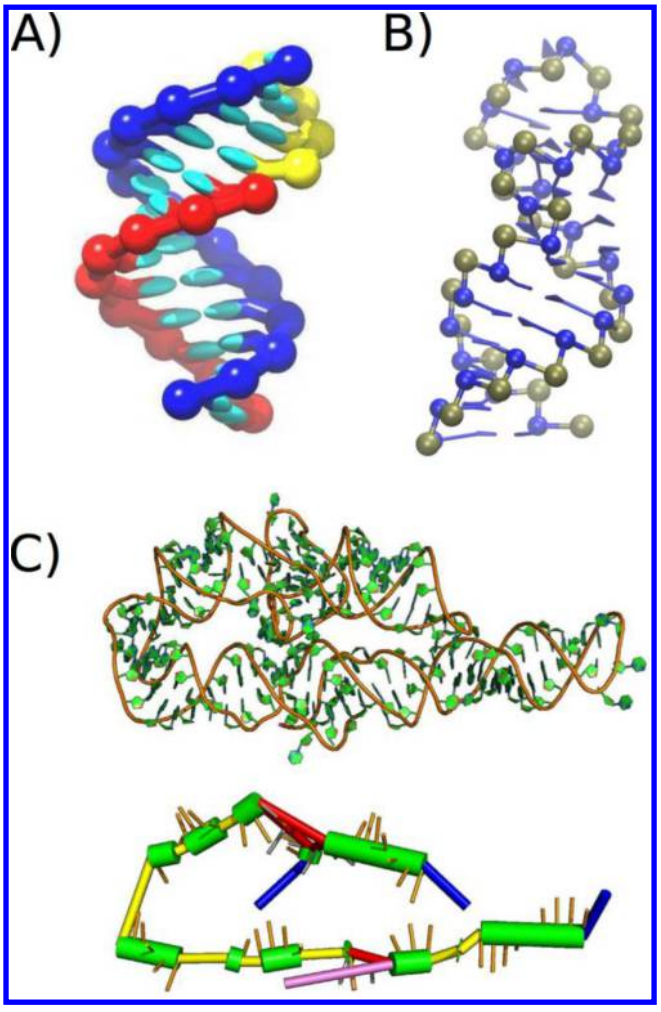


Figure 21. Building blocks of the RNA representation in a simulation do not necessarily have to be spherical beads: (A) in oxRNA, each nucleotide is a rigid object with a well-defined orientation and interaction mapping sites. Reprinted with permission from ref 351. Copyright 2014 AIP Publishing. In (B), SPQR[420] treats each nucleoside as a rigid object, in which the sugar is a spherical virtual site attached to the base, in a place that depends on the glycosidic bond angle conformation, while the phosphate group bead is an independent particle. In (C), Ernwin[415] treats each stem as a cylinder, while keeping trace of the orientations of the loops. Atomistic representation is included for clarity. Courtesy of Dr. Peter Kerpedjiev.

required for the properties they are designed to address. For example, they can represent a nucleotide by a spherical bead for a low-resolution treatment of a large structure, or replace it by a small set of bonded particles, which allow the definition of its internal structure and the introduction of more complex interactions. In the same manner, the interactions between these elements can be of different nature, although they might be described with similar representations, as in the case of structure-prediction models, which allow one to explore a large number of conformations, and ENMs, which have fixed secondary and tertiary structures by construction. We will focus on both aspects when describing a CG model. It is worth noting that, although the thermodynamics of folding is closely related to the problem of structure prediction, the latter is generally tackled using parameterizations and methods that do not necessarily have a direct physical meaning. As an example, Figure 22 shows the free-energy and energy landscape of two pseudoknot structures obtained with different methods: the

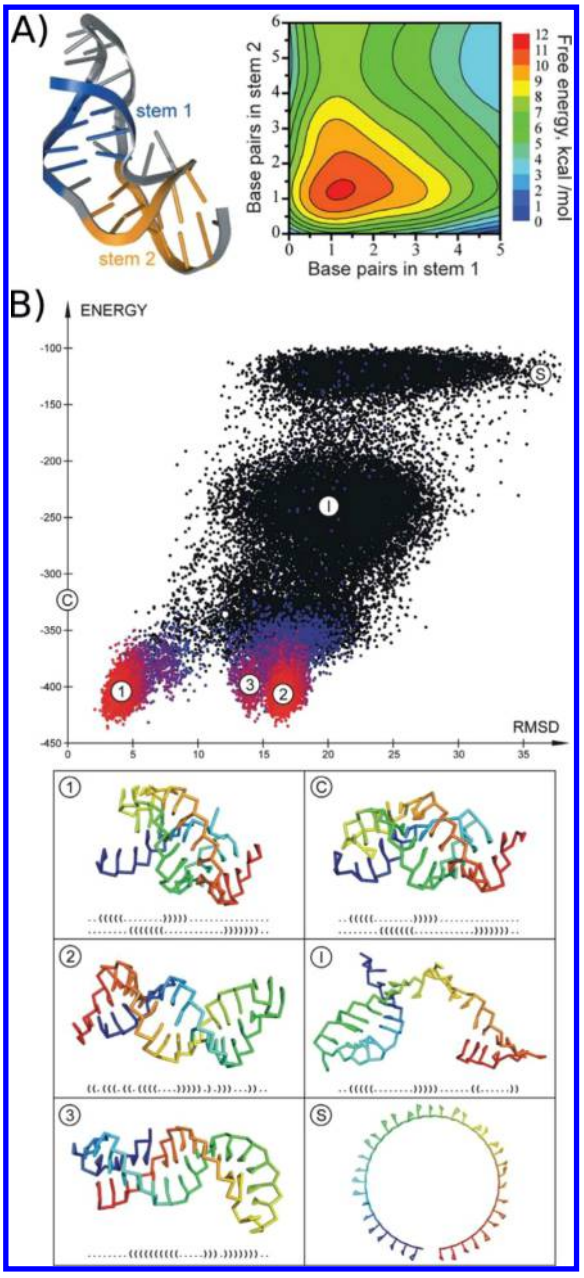


Figure 22. (A) Free-energy landscape of a pseudoknot calculated using the TIS model. The free energy is shown as a function of the number of native contacts in the stems present in the pseudoknot. Adapted with permission from ref 390. Copyright 2013 American Chemical Society. (B) Energy landscape obtained from SimRNA simulations for the gene 32 mRNA pseudoknot of bacteriophage T2 (PDB ID: 2TPK). The scheme shows the relation between the energy and the RMSD from the crystal structure (denoted by C), including an intermediate state (I), the top clusters (1, 2, and 3), and the starting conformation (S), which are also depicted in the lower figures with their secondary structure. Adapted with permission from ref 391. Copyright 2015 Oxford University Press.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

TIS,[390] parameterized with experimental free energies, and a knowledge-based approach, SimRNA.[391] The classification presented here is intended to shed light on the properties that can be faithfully reproduced and/or have been addressed using these approaches, and to guide readers who want to focus on a particular problem but lack knowledge of the field. A similar approach was adopted in the excellent review by Dawson et al.,[392] but their classification was based on the types of input parameters used to design the interactions rather than the applicability of the resulting CG model.

3.3.2. Models Reporting Thermodynamic and Physical Properties. Several models have been parameterized to reproduce the thermodynamics of folding and the free-energy landscape. These works are often built on earlier efforts to model secondary RNA structures, which provided a foundation for the parameterization or validation of the three-dimensional models. An important approach of this type is the nearest-neighbor (NN) model,[44,436−442] which approximates the free energy of the structure as a sum of contributions from steps of consecutive stacked base pairs, including separate terms for dangling ends, bulges, and hairpins. Its parameterization is based on experimental thermodynamic measurements of a large number of oligonucleotides. However, as a tool for evaluating the free energy of a given secondary structure and its melting temperature, it does not provide any information on the system's three-dimensional structure. To obtain a complete description of folding, it would be necessary to use a three-dimensional representation of the nucleotides. In addition, the interactions between the nucleotides should be defined in terms of stacking and base-pairing contributions, which are not fully distinguished in the NN model. Many of the models reviewed here use similar arguments or are parameterized using a consistent set of experimental data, and focus on reproducing the thermodynamics, melting curves, and possibly the salt dependency of the stability of hairpins, duplexes, and pseudoknots, among other types of RNA structures. In some cases, certain mechanical properties have been reported as well.

The Vfold model proposed by Cao and Chen[394,395] uses an estimate of the free energy based on NN parameters. It decomposes a secondary structure into arrays of stacked pairs and unstacked loops. The free energy of the stacks is estimated using the NN parameters, while the free energy of the loops is derived from their entropic contribution, which is calculated on the basis of an exhaustive exploration of self-avoiding random walks in a diamond lattice representation. The RNA chains were initially modelled using beads situated at the positions of the C4′ and P atoms, while the conformations were defined by the dihedral angles along the chains in this reduced description. In a newer version of this approach,[393] beads located at the positions of the N9 atom for purines and N1 for pyrimidines were added to account for the effect of base orientation on loop entropy. From the estimation of the partition function and the free energy of the allowed conformations, it is possible to predict the lowest free-energy secondary structure, and to obtain properties such as the specific heat. Using this approach and further extensions, it was possible to study the thermodynamics of specific regions in ribozymes,[394,395,397] several kinds of pseudoknots,[393] and kissing-loops.[396] The three-dimensional representation is obtained a posteriori by assigning the aforementioned CG representation to each of the predicted secondary structures with the help of a motif template database. Finally, atomistic details are introduced into this low-resolution scaffold and refined by energy minimization using the AMBER force field[80] in explicit solvent. Free energies can also be obtained using the MFOLD parameters.[443]

Gō models[444] are descriptions that favor a particular reference (target) structure (see also section 3.1.2.5). They have been widely used in the study of protein folding.[445] They typically make use of CG descriptions for the residues of a macromolecule, including bonded interactions along the chains and other nonbonded contributions of physical origin. The reference structure is favored by the inclusion of specific interactions associated with native contacts, which are pairs of CG particles separated by a distance smaller than a preassigned cutoff in the reference molecule. The folding pathway of RNA hairpins has been studied using an approach of this kind in which each nucleotide was represented by three particles (one each for the base, sugar, and phosphate group) by Hyeon and Thirumalai.[446] The interaction terms of these models include bonds, angles, and dihedrals to preserve the connectivity of stacked base pairs and the energetics of their interaction; the parameters used for these terms are directly adjusted on the basis of experimental free-energy data.[44,441] In addition, electrostatic interactions are included using Debye−Hückel terms, and there are additional nonbonded interactions including attractive contributions for native contacts and purely repulsive terms for non-native ones. This model has been used to study the nature of the folding transitions in terms of the temperature and force applied to the end sugars of the hairpin, calculate hopping and refolding times for different temperatures, forces, and initial conditions, and to analyze in detail the refolding by force-quenching.[446,447] In a subsequent work, the same group proposed a simpler representation, the self-organized polymer (SOP) model,[448] for studying the unfolding and refolding pathways of hairpins, ribozymes, and riboswitches.[448,449] This representation features one particle per nucleotide, with a set of interactions consisting uniquely of nonlinear bonds along the backbone, attractive and repulsive energy terms for native contacts, and exclusively repulsive terms for non-native ones. Some of the parameters that define these energy functions are determined from experimental structures, while others were adjusted to prevent interchain crossing.

An approach with more rigorous interactions was introduced in a later contribution focusing on the study of RNA folding. The three interaction sites (TIS) model[390] proposed by Denesyuk and Thirumalai represents each nucleotide with three particles, one each for the sugar ring, phosphate group, and base. The interaction energies are defined in terms of bonds, angles, excluded volume, stacking, base-pairing, and electrostatic forces. The terms used to describe these interactions are based on experimental measurements of the free energies of interactions for a large set of pairs of stacked base pairs.[438] A series of assumptions is used to decompose these measurements into independent enthalpic and entropic contributions for canonical base pairing and nucleobase-specific stacking interactions. This decomposition can be used to estimate the melting temperatures for every possible combination of stacked dimers, which are used to parameterize the interaction energies together with the experimental melting curve of the mouse mammary tumor virus (MMTV) pseudoknot and the stability of two hairpin structures. In this representation, all hydrogen bonds are assumed to make identical energetic contributions, and the bases have no internal structure. A hydrogen bond can be formed only if it is present in the hydrogen-bond network of the native structure. A later

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

version of the model allows the formation of canonical base pairs independently of the native contacts. Electrostatic repulsion is accounted for using Debye−Hückel theory, although the effective charge is based on counterion condensation theory.[450] The TIS model has been used to study the folding thermodynamics of the aforementioned pseudoknot (shown in Figure 22A) and two hairpins as a function of temperature and monovalent salt concentration, and for studying the relative stability of the hairpin and pseudoknot structures of the pseudoknot domain of telomerase RNA under molecular crowding.[398]

The CG representation implemented in the discrete molecular dynamics (DMD) algorithm by Ding et al.[399] also uses the NN model to directly calibrate some of its interactions, and has been used to study the specific heat profiles of tRNAs, pseudoknots, and fragments of ribosomal RNA (rRNA) and mRNA. This representation describes each nucleotide using three particles as well, corresponding to the sugar, base, and phosphate group. There are specific terms for the electrostatic repulsion between phosphate groups, and hydrophobic, stacking, and hydrogen-bond interactions between bases. In addition, there is an explicit correction for the entropy of loop formation because it was found to be underestimated by this type of CG representation.

HiRe-RNA[400−402] is a high-resolution CG model in which nucleotides containing purines and pyrimidines are modeled using seven and six beads, respectively: two corresponding to the phosphate group, three to the sugar ring, and the rest to the base. The authors reported specific heat curves for RNA duplexes, some of which included noncanonical base pairs. The force field of the HiRe-RNA model is quite sophisticated and includes local interactions accounting for local stereochemistry, an excluded volume interaction that gives the beads a physical size, and nonlocal interactions accounting for base pairing, base stacking, and electrostatics. Many (although not all) non-canonical RNA base pairs are represented. The set of interaction parameters was derived by analyzing around 200 structures from the PDB database, and refined on the basis of MD simulations. However, no quantitative correspondence between the simulated and real melting temperatures has yet been established. In the latest paper, the authors reported the successful folding of several tetraloops, pseudoknots, and quadruplexes. The representation also allows the reintroduction of atomistic details in a simple way, making possible a multiscale exploration of the conformational space. This is the case, for example, of the work of Stadlbauer et al.,[326] where the DNA variant of HiRe-RNA has been used for sampling configurations of DNA G-quadruplexes, while atomistic MD simulations have been used for obtaining further insight.

Jost and Everaers[403] proposed a one-site-per-nucleotide representation on a face-centered-cubic lattice, using the NN model for the parameterization of its interactions. This limits the possible interactions between the nucleotides to canonical base pairs (including G/U wobble pairs, Figure 2G) and stacking interactions. The comparatively low resolution of the configuration space offered by this description is compensated for by a rigorous treatment of chain conformations and the expressions of its entropy. Their model has been used to calculate melting temperatures and estimate the specific heats of systems such as small hairpins, pseudoknots, and tRNAs with up to 80 nucleotides.

Another representation is naRES-2P,[404] which uses interaction sites for the sugar ring, phosphate, and a base together with its dipolar momentum. The phosphate's position is defined in relation to the two closest sugars, which are directly connected by a bonded potential. In this model, base−base interactions are described using a combination of electric dipole and anisotropic Gay−Berne contributions. These interactions have an analytical form that was parameterized using the AMBER energy function[451] and experimental NN parameters.[452] Backbone conformations are governed by knowledge-based expressions taken from a structural database. The model also includes phosphate−phosphate electrostatic repulsions. Although the original version of this model successfully reproduced the double-helix structures of DNA and RNA, an optimized version was proposed to improve the agreement of the predicted melting curves with experiment.[405]

Starting from a known secondary structure, TOPRNA[406] is a model composed of nucleotides represented by three particles standing for base, phosphate, and sugar groups. It uses a pseudotorsional description for the backbone conformations and takes into account canonical and wobble base pairs through knowledge-based potentials. Stacking interactions are parameterized on the basis of experimental free-energy data. All nucleotides belonging to stems are bonded with their neighbors to reproduce an A-form helix, while the rest of the unpaired nucleotides are allowed to rotate freely around the backbone, without feeling stacking or base-pairing interactions. The model is typically used to explore topological constraints on folding and dynamics, especially in bulges, and their contribution to the free energy. Calculations using this model have produced results in good agreement with low-salt experiments, suggesting that the topology plays an important role in determining the free energy of bulges and junctions under these conditions.[453]

The oxRNA[351,352] model (Figure 21A) describes each nucleotide as a rigid particle with multiple interaction centers, and allows formation of stacking and base pairs. In addition, it incorporates specific interactions for topologically linked nucleotides. Its interactions are parameterized by adjusting the melting temperatures of a set of short duplexes and hairpins using the predictions of the NN model. It has been used to study melting curves of duplexes and hairpins, and the effect of mismatched base pairs and bulges on their thermodynamics. In addition, it was used to study some important mechanical properties of duplexes such as their persistence length, elastic responses under the action of forces and overstretching, and their mechanical behavior under supercoiling conditions. Despite not being parameterized for these properties, the results obtained agree reasonably well with experimental data in general. Electrostatic interactions have been incorporated into the model using the Debye−Hückel theory, and effective charges for the phosphate groups were fitted to reproduce the experimental melting temperatures of selected RNA duplexes.

Ren and co-workers[407−409] proposed the RACER model, which uses a five-bead representation of nucleotides. This mapping assigns two particles to the backbone (representing sugar and phosphate groups) and three to the bases, making it possible to define their orientations and the positions of interaction sites. Interactions are modeled using knowledge-based potentials that also incorporate excluded-volume interactions, electrostatics (in the form of a repulsive Debye−Hückel term between phosphate groups), and explicit hydrogen bonds between bases. Originally designed for structure prediction and modeling, the latest work on this method presented energy landscapes for 14 structures together with equilibrium pulling simulations on RNA duplexes (with

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

between 6 and 10 base pairs), hairpins (10−18 nucleobases), and the HIV TAR RNA, showing good agreement with experimental free energies.

Finally, the MARTINI force field[454] is a systematic CG model designed for biomolecules. Its general approach is based on representing a set of four heavy atoms with their associated hydrogens by a single bead. The bonded interactions between these beads are adjusted to distributions sampled in all-atom MD simulations, while the nonbonded ones are parameterized using experimental values of partitioning free energies. This procedure allowed CG simulations of proteins,[455] carbohydrates,[456] and lipids.[387] The parameterization has been extended for modeling DNA and RNA molecules,[410,411] making possible studies of interactions between RNA and proteins in a common framework. It is important to mention that the tertiary structure of the simulated RNA molecules is kept fixed by additional harmonic springs, so this approach is not intended for the study of folding properties. The authors have reported flexibilities of single- and double-stranded RNAs, which are in good agreement with all-atom MD results. They also evaluated the stability of RNA loops and systems involving the HuD protein, the protein 19 of the human signal recognition particle, and even the whole *Thermus thermophilus* 70S ribosome. Also, they tested formation of protein/RNA complexes by performing simulations starting from an initial condition in which the protein and the RNA were separated by 1 nm.

The aforementioned methods provide descriptions that in general take care of canonical base pairs, stacking, and electrostatics. Unfortunately, the parameterization of interactions in CG models on the basis of free energies requires a large amount of work, and given the paucity of data available involving noncanonical base pairs or base-phosphate interactions, these models are only capable of describing a few kinds of base pairs. They also impose restrictions on the backbone's conformational space and ignore sugar pucker (in most cases) and glycosidic bond angles, which may significantly affect the predictability of the native fold for important motifs.

3.3.3. Models Designed Exclusively for Structure Prediction or Refinement. Several methods discussed in the previous section can be used to predict native 3D structures on the basis of sequence data alone. However, in folded RNAs, the interactions between bases, or those between bases, phosphates, and sugar groups, can be much more variable than those found in typical A-form RNA helices. In addition, an effective structure prediction tool must be able to sample the conformational space in an efficient manner, which can be easier to implement in low-resolution models of RNA, where the building blocks that compose a molecule are constituted by many nucleotides. Hence, it is sometimes necessary to use less rigorous approaches when parameterizing interactions to capture or refine more complex structural features. The methods described in this section generally aim to predict a "best energy" or most populated cluster structure according to a particular energy function, which accounts for various real interactions in a reasonably accurate and computationally affordable manner. In general, the temperatures used in these methods have no physical meaning, and so the conformational space sampled during a constant temperature simulation does not either, unlike the results of an atomistic MD simulation. Parameterizations of most methods in this class rely on extracting data from large data sets of 3D structures instead of being defined on the basis of experimental free energies or melting temperatures.

The early method developed by Malhotra et al.[412] was proposed as a structure refinement tool (to be used in conjunction with experimental data) rather than a truly predictive algorithm. These authors presented a protocol for introducing details at three levels of resolution based on a previously known secondary structure and a set of experimental constraints. The high-resolution level represents a nucleotide by a single particle located at the position of the corresponding phosphorus atom. The next level represents hairpins and duplexes by five or more particles, while the lowest resolution level uses large spherical beads to represent these motifs. Because of the limited amount of information used in the design of the modeled topology, additional experimental data are usually required to guide the selection of the best structure from a set of candidates. The computation of energies in this model includes harmonic potentials for bonds, angles, and dihedrals, with the possibility of introducing constraints for shape and positional information from electron microscopy experiments through specific terms. These potentials are then used for enforcing 2D structure starting from several random walk chains. The method was designed for dealing with relatively large structures, such as tRNA and components of ribosomal subunits. Its features were implemented later in the YUP package.[413]

NAST[414] is another one-bead-per-site model in which each nucleotide is represented by a bead located at the coordinate of its C3′ atom. Like Malhotra's method, it requires data on secondary structure and tertiary contacts as inputs. Its interactions are bond, angle, and dihedral potentials derived from knowledge-based approaches, with additional terms for fixing secondary and tertiary contacts. A large number of decoys is therefore created starting from random coils and circular conformations, which is later clustered and ranked according to the energy function. The method has been tested on yeast phenylananine tRNA and the 388 residue *T. thermophilus* group I intron. This model can also be complemented with small-angle X-ray scattering (SAXS) or solvent accessibility data to refine the results. In addition, it is also been used for proposing models of missing loops in crystal structures.

Another method that makes use of the secondary structure is Ernwin,[415] a helix-centered model shown in Figure 21C. In this approach, stems are represented by cylinders, which are connected to each other by interior loops or multiloops, or simply end in a hairpin loop or the terminal residues. 2D structure is kept fixed during Markov chain Monte Carlo simulations, while energy functions avoid clashes between stems. Tertiary contacts can be formed through stem−loop and loop−loop knowledge-based interactions, and an additional energy term dependent on the radius of gyration can guide the simulation toward conformations with realistic compactness. The method has been tested on several structures, with a number of nucleotides ranging from 63 to 298. An important part of its capabilities relies on its sampling scheme, which allows displacements of large RNA domains, motions that cannot be efficiently sampled under atomistic approaches. At a similar level of resolution, RAGTOP[416] builds 3D graphs from a given secondary structure. Junctions between helices are predicted first, and a subsequent simulated annealing optimizes the structure subjected to knowledge-based potentials. The scoring function contains terms for torsions and angles around internal loops, the total radius of gyration, and a specific treatment of pseudoknots. The method has been tested on a group of 10 riboswitches containing three- and four-way

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

junctions. In addition, kink turns can also be identified and treated with a specific set of parameters.[457]

RNAkb[417] is a knowledge-based approach that models nucleotides using five particles located at the positions of the P, C4′, C2, C4, and C6 atoms, giving the bases an appreciable degree of internal structure. Its interactions are statistical potentials taken from histograms between these mapping points over a set of experimental structures, and have been carefully designed to take a smooth, differentiable form that can be used for MD simulations. The same approach has also been used to parameterize an all-atom knowledge-based model that reflects distributions of all atom types present in the RNA molecule. The performance of these representations was evaluated by using them to define scoring functions for identifying the best candidate structure from a set including multiple decoys, on structures with a maximum of 76 nucleotides. Results obtained with the CG energy function were of a quality similar to those obtained with established all-atom scoring functions such as Rosetta.[458] The all-atom version was better able to discriminate the native structure from a decoy set than the CG representation.

SimRNA[391] is a model in which each nucleotide is represented by five beads corresponding to the sugar, the phosphate group, and a planar base. Its interactions are statistical potentials derived by sampling a curated set of experimental structures. The backbone representation utilizes the pseudodihedral conformational space, while the three-particle representation of bases allows them to form directional interactions including stacking and both canonical and noncanonical base pairing. The simulation procedure involves exploration of the conformational space by means of Hamiltonian Replica Exchange or simple Monte Carlo simulations, selection of the lowest energy structures, and clustering, as shown in Figure 22B. The model allows blind prediction of large numbers of structures, having been tested on a benchmark set of more than 200 structures with a maximum of 189 nucleotides and usually recovering the system's true secondary structure. In addition, if 2D structure obtained by some other means is available, it can be used to impose constraints to accelerate the simulation or guide the simulation toward a structure consistent with the available 2D structure data. This method has been particularly successful for folding of pseudoknots, correctly predicting both their 2D and 3D structures in the vast majority of cases. The method has been implemented in the SimRNAweb server.[459]

More recently, the RNApps model presented by Li et al.[418] introduced another five-particle-per-nucleotide representation, using two particles for the backbone (situated on the P and C4′ atoms) and three that represent a base with its orientation. In this simpler representation, the conformation of a chain is given by a set of torsion angles, which are obtained according to a knowledge-based CG probability distribution taking into account correlations between the consecutive torsions. This method was inspired by the BARNACLE[460] algorithm, and makes use of the RASP[419] scoring function to evaluate the energy during sampling and posterior refinement steps. It is intended to be used as a refinement tool, having been tested in a large number of small loops and junctions (with 20 or less nucleotides) belonging to ribosomal structures and 876 nonredundant structures, among others. One of its most remarkable features is its sampling efficiency.

Finally, the SPQR model[420] uses a point particle to represent the phosphate and an anisotropic particle for the nucleoside, as shown in Figure 20B. Its knowledge-based interactions are inspired by the εSCORE scoring function,[223] which deals exclusively with the geometrical arrangement of bases in the RNA, exploiting the unambiguous relationship between base-to-base positions and the overall arrangement of RNA structures. In this case, the base−base interactions are divided across the sugar, Watson−Crick, and Hoogsteen faces, allowing formation of both canonical and noncanonical planar base pairs,[39] and base−phosphate interactions.[50] The interactions are also reweighted on the basis of experimental data.[39,50] In the same spirit, backbone is defined by a minimal set of interactions to avoid favoring the most frequent RNA conformations found in the PDB database. To this end, the model uses a specific treatment of sugar puckers and glycosidic bond torsion conformations, allowing it to fold several hairpins and tetraloops with high accuracy. The representation also permits construction of structures with atomistic detail by making use of εRMSD,[223] an advanced structural deviation measure that only considers the arrangement of the bases and their relative orientations. εRMSD is more suitable for describing similarity between RNA structures than standard RMSD after optimal superposition.[331] By minimizing εRMSD between the CG-predicted structure and the corresponding all-atom strands in a structure obtained by explicit solvent steered MD, one obtains an atomistic consistent structure that usually also has correct sugar pucker and glycosidic bond angle conformations. This procedure is closely related to the targeted MD approach (see section 3.2.5), and the target structure is here given by the CG model prediction.

The high-resolution HiRe-RNA model[400−402] introduced in section 3.3.2 is also being developed with the aim of creating a tool for predicting 3D RNA structures and folding pathways by exploiting its ability to mimic key noncanonical base pairs. However, some base pairs and base−phosphate interactions remain to be incorporated into the model. Similarly, the DMD model has also been used as a structure prediction tool, correctly predicting 150 of 153 nontrivial structures. This method is also available online through the iFold web server.[461] The RACER[407−409] model discussed in the preceding section has also been used for the blind prediction of 14 structures with 30 nucleotides or less, achieving reasonable accuracy. In addition, the authors were able to introduce restraints reflecting known 2D structures and the presence of coordinated Mg^{2+} ions, and proposed a refinement method for structures of more than 100 nucleotides that can either use distance constraints from NMR data or comparisons between experimental data and calculated SAXS profiles for a set of candidate structures obtained by simulated annealing of a random coil. Vfold can also be used as a structure prediction tool, in particular through its web server implementation.[462]

Capabilities of several of the above-described methods have been tested against other approaches in the RNA-Puzzles challenge,[463−465] a competition involving blind predictions of the three-dimensional structures of given sequences. Authors of the Vfold, SimRNA, and DMD models have participated very actively in this challenge, as have the creators of other all-atom methods such as FARFAR,[458] RNABuilder,[466] MC-Fold/MC-Sym pipeline,[467] and RNA123.[468] Although most of the predictions are performed using hybrid approaches that involve experimental inputs, constraints for 2D structures, and all-atom refinements among other things, the CG models have made important contributions to the predictions when they have been used, and in some cases they have outperformed the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

available alternatives. For example, SimRNA was used effectively as a refinement tool in combination with ModeRNA[469] and other tools, and was particularly successful in predicting the structure of a self-assembling RNA square and the bound state of the L-glutamine riboswitch. In a similar manner, Vfold[394,395] provided the best result for folding of the SAM-I riboswitch and an 84 nt glycine riboswitch domain, for which the DMD method also yielded good results. However, despite some notable successes, predicting RNA 3D structures remains far from trivial, even though known native structures are often used while training the parameterization of the CG models. Particularly, it is very difficult to capture rare backbone conformations, noncanonical base pairs, base−phosphate, and base−sugar interactions, and other features that do not appear frequently in structural databases. Further, many RNA molecules may only adopt their native structures upon interacting with appropriate oligopeptides or proteins. Additionally, it is not trivial to reintroduce atomistic details into the predicted structures, that is, to convert the structure predicted by the CG approach into an atomistic structure. This is required if atomistic insight is needed or if the predicted structures are to be used as starting points for further atomistic MD simulations.

3.3.4. Elastic Network Models. If a reference structure is available, Elastic Network Models (ENMs)[470] are simple tools for studying RNA dynamics based on the interconnectivity of the residues of a macromolecule. They map nucleotides onto a reduced set of particles that are bonded to all of their neighbors within a predefined cutoff radius by elastic potentials. The contact map associated with the native structure is thus enforced in the model so that the molecule cannot depart from it (Figure 20). Despite their simplicity, calculations of the normal modes of these representations can be useful for distinguishing between flexible and rigid domains inside a macromolecule, and even for inferring its catalytic activity using this information.[471,472] Given their great efficiency and extensive application to proteins,[473−476] ENMs are also applicable to protein/RNA complexes, and can generally deal with structures containing thousands of nucleotides. They also provide room for variability in terms of the cutoff of the interactions, the values of the spring constants, their dependence on separation of the particles in the equilibrium structure, and (of course) the choice of mapping sites.

One of the first applications of ENM methods to RNA was reported by Bahar and Jernigan.[423] In their approach, one or two beads per nucleotide (standing for P or P and O4′ atoms, respectively) were used to identify hinge regions in tRNA and in the complex formed by tRNA and glutaminyl-tRNA synthetase. ENMs using the same mapping were also used to study large-scale motions relevant to ribosomal function by Wang et al. (one or two beads),[424] Kurkcuoglu et al. (one bead),[425] and by Tama et al. (one bead),[426] who complemented the ENM analysis with cryo-electron microscopy (cryo-EM) data. This method has also been used to study the dynamics of tRNA in isolation and when attached to the ribosome or a synthase.[427]

Zimmermann and Jernigan[428] investigated the correspondence between the normal modes of ENMs and the principal components (PCs) of the apparent motion extracted from ensembles of experimental structures. Their study used a set of 16 ensembles of at least 5 structures having between 22 and 170 nucleotides, with particular emphasis on tRNA structures. The performance of each model was analyzed using different representations with between one and three particles per nucleotide, located at the positions of one or more of the P, C2′, and O4′ positions. Additionally, they analyzed several options for representing the dependency of the spring constant between two particles and their equilibrium distance, finding that the results obtained were less sensitive to this parameter than is the case when simulating proteins. They also observed good agreement between the relative amplitude of motion for each atom and metrics that compare how the particles fluctuate in each direction as the cutoff distance in the CG representation is increased. In a later step, they compared different approaches for performing the alignment of the structures, which is a fundamental aspect in the determination of the PCs. Finally, the authors proposed a method for RMSD minimization and structure alignment that uses only those atoms that exhibit small fluctuations in the ENMs. It was found that this approach yielded more consistent results than the available alternatives when comparing the overlap between PCs and normal modes for different nucleotide representations.

ENM approaches have also been used in smaller systems, for example, to study unfolding pathways of riboswitches,[429] and to analyze conformational space of a specific sequence, the CUG repeat, at different resolutions, including all-atom networks.[430] The performance of ENMs has also been compared to that of all-atom simulations and normal-mode analysis (NMA). Van Wynsberghe and Cui[158] found that ENMs describe the dynamics of the hammerhead ribozyme less well than higher resolution approaches. A more compact structure, the guanine riboswitch, exhibits better results under the ENM representation. These authors performed a mapping over phosphorus atoms, and ran their MD simulations for rather short times (12.5 ns). Similar findings were reported by Besseova et al.,[431] who studied three three-way junctions extracted from the large ribosomal subunits using all-atom MD and compared the results obtained to data from ENM calculations and a block analysis using the ElNemo software package.[432] They found that for the GTPase associated center part of 23S rRNA ENM failed to predict the hinge region observed by all-atom MD at the three-way junction. Further, the all-atom NMA did not properly describe the system's motions and was in addition computationally inefficient. The authors argued that the failure might be due to the noncompactness of the studied domain. The physical nature of ENM models may make them more efficient at depicting the dynamics of globular molecules because their interaction potentials reflect the density of particles around each residue.[158,431]

A few works have sought to optimize the representations used in ENMs, mostly comparing the resulting fluctuations to those obtained from MD simulations.[158,421,422] In particular, the work of Setny and Zacharias[421] suggested that an optimal model with a single bead per nucleotide could be obtained by placing the bead on the geometric center of the sugar. This result was confirmed by the analysis of Pinamonti et al.,[422] who compared MD simulations of a set of structures with a number noncanonical pairs to ENM simulations. These authors showed that more accurate representations can be obtained by increasing the number of beads, and suggested that the optimal representation has three beads per nucleotide, corresponding to the sugar, the base, and the phosphate.

It should be noted that fluctuations obtained from ENMs can also be compared to experimental results. Comparisons with the *B*-factors obtained from X-ray structures are common, although this approach has known limitations.[433] It has also

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

been suggested[422] that one could compare the nucleotide-resolution flexibility obtained using ENMs to the results of selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) experiments.[477]

In an approach similar to ENMs, the Gō-like model of Hori and Takada[434] incorporates structural information from atomistic simulations for parameterization of its interactions by using the procedure introduced in ref 478. It represents each nucleotide by three beads standing for sugar, phosphate group, and base. In this manner, the interactions are explicitly adjusted to reproduce the mean square fluctuations of the CG sites sampled from the atomistic simulations of a set of RNA and RNA/protein structures. Although the interactions do not correspond to an ENM, the authors use them in a similar manner for studying fluctuations and deformations of tRNAs and the ribosome. The results compare well with *B*-factors and substates seen in the X-ray structures.

Finally, we note the work of Benítez et al.,[435] who used an ENM parameterization to describe the diffusion properties of RNA structures (tRNA and rRNA). In this case, the elastic networks were simulated using a Brownian dynamics[479,480] approach that introduces hydrodynamic interactions and thus realistic solvent friction. Including only secondary structure constraints, the authors reported good agreement between the diffusion coefficients and relaxation times of the interdomain motions of the studied structures.

3.3.5. Final Considerations about Coarse-Grained Models. We have reviewed some of the most relevant CG models of RNA. Their proper use can provide useful information about thermodynamics, mechanical properties, relevant large domain motions, and possible native structures from the knowledge of the sequence and other information. They can also be employed to refine experimental data or complete missing fragments, or to explore conformational changes beyond typical all-atom MD scales, filling the gap between different time scales and experimental and computational techniques. Because they are usually parameterized with very specific tasks in mind, one should be aware of their advantages and limitations and reported results before applying them.

It is important to understand which are the features that the authors have prioritized in their description according to their specific criteria. For example, when addressing thermodynamics, some authors have opted for developing rigorous expressions for the entropy of loops while treating them as structureless objects. These representations in general lack an explicit description of noncanonical base pairs and other interactions, which are critical for RNA structures (section 2). A similar issue must also be considered in the opposite way; given that knowledge-based approaches lack a rigorous interpretation of the conformational space they explore, the definition of temperature and other thermodynamic quantities is not straightforward, and could also depend on the studied structure. Simplifications of systems may lead to problems of choosing optimal sets of degrees of freedom and the possibility of overfitting the model with a reduced representation.

3.4. Interactions of RNA with Monovalent and Divalent Ions

Effects of ions on the structure and function of RNA can be profound, and accurately describing these effects is a challenging but important problem in MD simulations.[175,177,335,481−492] Despite tremendous progress in MD methods, there are several fundamental challenges that make it very difficult to accurately describe interactions between RNA and ions, especially divalent cations. In addition, there are a number of myths and controversies relating to the incorporation of ions into MD simulations; simulation studies have been excessively criticized over rather marginal problems, and some major uncertainties and approximations are routinely overlooked. This section therefore focuses on some of the major problems with modeling of ions in RNA MD simulations. Because of the important effects of ions on RNA structure and function, specific aspects of their modeling are also discussed in some other parts of this Review.

3.4.1. Basic Considerations about RNA−Cation Interactions. Experiments have clearly shown that cations are required to stabilize RNA structures because of RNA's negative charge and compactness.[69,493−501] Many RNAs require divalent ions such as Mg^{2+} to fold properly; if only monovalent ions are present, these RNAs become only compacted. However, some RNAs can also fold correctly in the presence of excess monovalent cations. It should be noted that while diverse experimental techniques have been used to study interactions between RNA and ions, we lack detailed high-resolution data on the dynamic interactions of ions with biomolecules. Individual experimental techniques reveal only specific aspects of the ion-binding processes, and it is not straightforward to compare results obtained using different experimental methods, or to compare experimental and computational results.[502]

It is widely assumed that most cations (both monovalent and divalent) interact with RNA as highly mobile species, creating a diffusive ion atmosphere.[493−496,498,502] These ions are thus expected to interact with the RNA molecule transiently, via the ion-hydration shell in the case of divalent ions. Indirect binding can be defined either on the basis of the distance between the ion and the nearest atoms of the RNA molecule, or by explicitly considering the ion's hydration shell. If the water oxygens in the first coordination shell of an ion are not resolved in a structure obtained by X-ray crystallography, one might use the distance between the ion and the closest electronegative RNA atoms to identify an indirectly bound state.

A small subset of ions can be considered structurally bound; that is, they occupy a specific binding site and typically form one or more direct interactions with the RNA molecule (Figure 23). This is known as inner-shell binding, site binding, or chelation.[69,493−496,498,502] Such interactions are characterized by the presence of an electronegative RNA atom in the ion's first coordination shell; that is, the RNA replaces one of the water molecules normally found in a solvated ion's inner shell. Note that monovalent ions can form inner-shell contacts that are only transient.[485,487,489,503] This is unlikely the case with divalent ions due to the slow exchange kinetics of their directly bound ligands; the energy barrier for displacing a water molecule from the first shell of Mg^{2+} is estimated to be around 9 kcal/mol, with microsecond-scale ligand-exchange times.[504,505] For RNA phosphates, the barrier could be even ∼12 kcal/mol.[506] So, once a divalent ion binds directly to RNA, it stays bound for some time. For monovalent ions, comparison of various MD studies suggests that binding times of directly bound ions can be anywhere between transient subnanosecond time scale and permanent binding, depending on the properties of the ion binding sites.[484,485,487,489,503,507−509]

The importance of structurally bound ions is, however, sometimes overstated; it is assumed that medium-sized RNAs (i.e., RNAs with up to 100 nucleotides, or perhaps a few

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

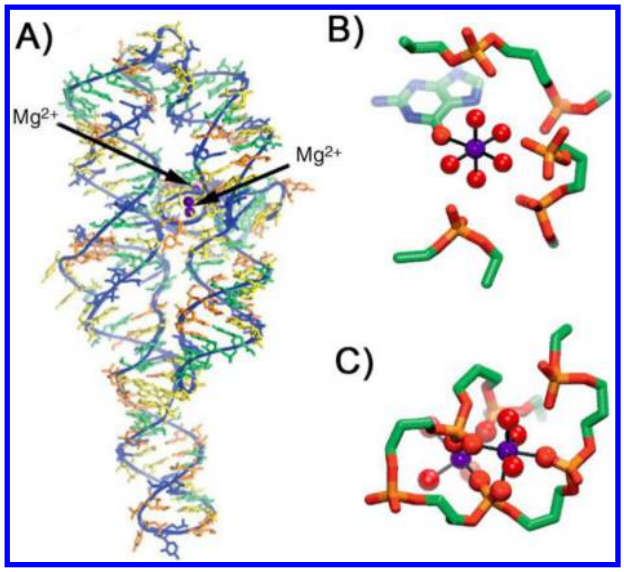


Figure 23. Examples of structural binding of Mg^{2+} to RNA. (A) The structure of the P4−P6 domain of the Tetrahymena Group I intron RNA [PDB entry 1HR2]. Three Mg^{2+} ions are indicated by purple spheres. The coordination of these ions is shown in detail in panels B and C. (B) Coordination of an inner-shell-bound Mg^{2+} ion. This ion is coordinated by RNA and five water molecules. Oxygen atoms of first-shell water molecules are shown [Mg^{2+} 6766 of 1HR2]. (C) Highly coordinated Mg^{2+} ions induce specific conformational states of RNA. Two trichelate Mg^{2+} ions contain tightly packed RNA phosphate oxygen atoms in their first coordination shells [Mg^{2+} ions 6756 and 6758 of 1HR2]. Reprinted with permission from ref 498. Copyright 2012 Elsevier Ltd.

hundred) only contain a few Mg^{2+} ions that are truly structurally bound.[69,493−496,498,502] This conclusion is primarily based on various solution-phase experiments. For example, the Hepatitis delta virus (HDV) ribozyme utilizes one Mg^{2+} ion for folding and another one for catalysis (see section 4.8.3.2).[510,511] However, this ribozyme can fold and retains catalytic activity even in the absence of divalent ions. Moreover, experiments on Tetrahymena and Azoarcus group I intron ribozymes showed that the initial stages of their folding required only the presence of some kind of cation (even polyamines were sufficient), and only the active site was specific for bound Mg^{2+} ions.[69] It should be noted that unambiguous experimental counting of structurally bound divalent cations in solution is not straightforward, and the number of such cations remains a contentious issue.[502]

Because of their immense size and complexity, ribosomal structures provide interesting examples of specific ion binding. For instance, the whole large ribosomal subunit (~3000 nts) contains approximately 120 chelated Mg^{2+} ions.[512] However, this count is based on X-ray crystallography data, and the excessive presence of divalent ions in some X-ray structures may be responsible for the occasional overestimation of their role and number in RNA. In general, characterizing bound ions by X-ray crystallography is far from trivial due to difficulties in interpreting the associated electron density maps and the roles of the ions in crystallization, as they may be present in the crystallization media at rather high concentrations, and may mediate crystal packing contacts (see section 3.4.6).[497,513,514]

Some divalent ion binding sites, for example, in some small and all large ribozymes, are considered functional (catalytic), although many small ribozymes can function even in the absence of divalent ions.[510,511,515−517] For example, hairpin ribozyme does not utilize divalent ions for cleavage,[516] and while the HDV ribozyme does probably use an Mg^{2+}-bound OH^- group as the general base, it can also catalyze cleavage via an Mg^{2+}-independent pathway (see section 4.8.3.2).[510,511] By contrast, self-splicing introns and RNase P directly employ two Mg^{2+} ions for catalysis.[483,518,519] As said before, the 2.4 Å resolution X-ray structure of the large ribosomal subunit revealed 116 bound Mg^{2+} ions.[512] While some binding patterns may be affected by the natural resolution limits of the ribosomal X-ray structures (see section 3.4.6), the analysis disclosed enormous variability in the specific Mg^{2+} binding patterns: the number of inner-sphere binding contacts between individual Mg^{2+} cations and ribosomal RNA (rRNA) ranged from zero to five. The Mg^{2+} cations help to compensate the negative charge of rRNA and can stabilize specific structures, for example, by mediating interactions between autonomously folded rRNA secondary structure elements or between rRNA and ribosomal proteins. Mg^{2+} ions are especially abundant around the peptidyl-transferase center, which is, in an evolutionary sense, the oldest part of the ribosome that shows a notable lack of cationic tails of ribosomal proteins.[520]

Although most of the known Mg^{2+} binding sites revealed by X-ray structures are in large RNA structures, ion binding is also important for small RNA motifs. A textbook example is the RNA kink-turn (see section 4.3.7),[521,522] which critically requires the presence of divalent ions to fold properly in the absence of proteins.[523,524] Ions can also be involved in interesting interplays with noncanonical interactions. For example, a recurrent RNA motif bacterial 5S rRNA loop E (LE)[51,525,526] exhibits an anomalous salt dependence that complicates experimental studies of its thermodynamics. Indeed the X-ray structures of this motif revealed clear structured binding of several Mg^{2+} ions in its widened deep (major) groove (see sections 4.3.2 and 4.3.4).[527] LE is also associated with deep electrostatic potential (ESP) minima.[484−487] MD studies have suggested that two of the closely spaced Mg^{2+} binding sites in LE probably only bind a single ion, with each site fractionally occupied.[485,486] MD simulations also suggested that, in the absence of divalent cations, multiple monovalent ions bind in the LE deep grooves.[484−487] This and the HIV-1 DIS kissing-loop complex (see section 4.3.10)[489] provided two of the first examples of competition between divalent and monovalent ions for the same ion-binding sites documented by MD simulations. In contrast, the Sarcin-Ricin loop (SRL, see section 4.3.6),[528,529] another recurrent RNA motif with a size and base composition similar to those of LE, essentially does not bind ions and forms only a weak ESP.[530] However, the SRL motif is stabilized by several highly conserved base−phosphate interactions that are absent in LE.[50] The deep ESP valley formed along the noncanonical LE helix arises primarily from the disposition of its phosphate groups rather than the ESP of the nucleobases; in fact, the ESP of this motif is insensitive to the four G to A substitutions that distinguish bacterial and spinach chloroplast LE sequences, which are structurally identical.[484] This conservation of structure is a textbook example of the power of the isostericity principle of RNA evolution; for more details, see sections 2.2, 4.3.1, and 4.3.2. In contrast, ESP of the SRL motif is dampened because several of its phosphates form hydrogen bonds with nucleobases. This contributes to the striking difference in the ion binding properties of LE and SRL.[50]

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

The effects of ions on RNA structure and dynamics may be affected by their concentration. The intracellular ion concentrations depend on the specific ion and cellular conditions. Usually, MD simulation protocols aim to utilize either the concentration of ions used in the experiments, to which results are compared, or some "cellular" or "physiological" concentrations. Typical physiological cellular concentrations of K^+ ions are of the order of 100 mM.[531,532] The concentration of Mg^{2+} ions may range from millimolar to micromolar, and cells can survive at very low Mg^{2+} concentration if Mg^{2+} is replaced by polyamines.[533,534] However, one also has to take into account that conditions inside simulation solvent boxes are certainly not the same as those in the crowded molecular environments in living cells; see further comments in section 3.4.5.

In summary, it is evident that the modeling of ions in MD simulations and their proper placement are critically important. However, the treatment of ions is one of Achilles heels of atomistic MD simulations of RNA. Difficulties of simulating ion binding stem from force-field approximations, insufficient sampling of ions, use of small sizes of the simulation boxes, and issues with the identification and experimental characterization of ions' positions in RNA structures. We address each of these issues in detail below. For the sake of completeness, we note that there also exist computational approaches that can provide insights into the binding of multivalent ions to RNA without relying on atomistic simulations.[174,535,536]

3.4.2. Force-Field Approximations Related to Ion Binding. The first difficulty in describing ion binding by MD stems from the approximations of the force-field model. Contemporary MD simulations use pair-additive force fields that disregard polarization effects and thus any nonadditivity of interactions (Figure 24). However, ions are charged species and

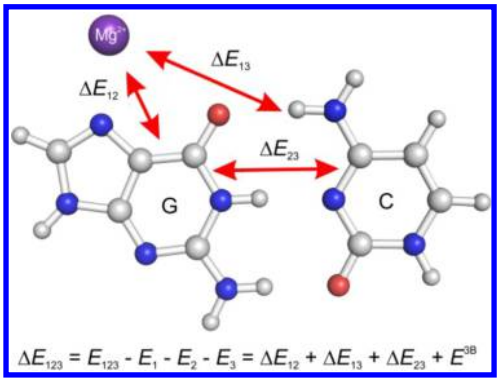


Figure 24. Interaction energy of a trimer $\Delta E_{123} = E_{123} - E_1 - E_2 - E_3$ can be expressed as a sum of pairwise interaction energies ΔE_{12}, ΔE_{13}, and ΔE_{23} plus the three-body term E^{3B}; the latter term includes the polarization and charge-transfer effects and is by definition 0 when using contemporary simulation force fields. G═C base pair with divalent ion bound to N7 of G is shown; in this case, the E^{3B} term brings a significant strengthening of the H-bonds. All three participating particles electronically communicate with each other.[545] E_{123} stands for total electronic energy of the trimer, while E_1, E_2, and E_3 stand for the electronic energies of the isolated monomers.[118,537,545]

thus communicate with other molecules also via polarization.[175,488,498,537−543] Moreover, charge-transfer effects may be important in the case of divalent cations. Note that the distinction between polarization and charge-transfer effects is somewhat arbitrary because there is no unambiguous QM definition of these effects.[537,542,544] However, in imprecise terms, polarization by a positively charged cation means that the cation distorts the electronic density of the ligands in its coordination sphere such that an excess of negative electronic density appears on the ligand in the vicinity of the cation (Figure 25). If the polarization is strong enough, it can lead to a

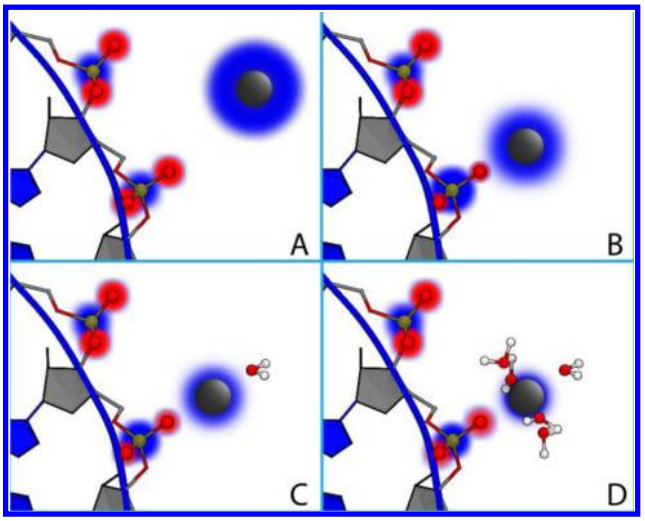


Figure 25. Schematic representation of charge-transfer effects upon Mg^{2+} binding to the phosphate moiety of RNA backbone. Blue and red shadings depict the positive and negative atomic charges (or charge densities), respectively. The red, white, ochre, and black spheres are representations of the oxygen, hydrogen, phosphorus, and magnesium atoms, respectively. The different intensity of the shadings between the four snapshots is a qualitative indication of how charges will be distributed in each scenario. With the increase of the number of ligands in the inner coordination shell of Mg^{2+}, going from zero ligands in panel A, one in panel B, two in panel C, to six in panel D, the nominal charge of the cation is reduced. Notice that also the charge distribution of the ligands is changed upon Mg^{2+} binding. The phosphate oxygens donate charge to the Mg^{2+} ion, while the phosphorus atom itself becomes slightly more positive. The changes in the charge on the phosphate oxygens in panels B, C, and D correlate with the increase of the number of charge donors bound to the Mg^{2+}. Note that the figure needs to be taken only as a qualitative visualization, because atomic charges are not unambiguously defined; see section 3.1.1.3.

donor−acceptor interaction between occupied electronic orbitals of the ligands and unoccupied orbitals of the cation, which is essentially equivalent to a ligand-to-cation charge transfer. Thus, QM-based publications usually refer to charge-transfer effects when changes in the electronic structure (i.e., the molecular orbitals) of the ligand and ion are sufficiently pronounced. Polarization and charge-transfer effects associated with ion binding are sometimes jointly referred to as nonelectrostatic contributions,[544] to separate them from simple long-range electrostatics originating from unperturbed ESPs. However, despite being quite intuitive, the term "nonelectrostatic contribution" is not rigorously defined in terms of QM theory.

Polarization and charge transfer are contributions to potential energy that are not directly accounted for by pair-additive force fields because they have no corresponding force-field terms. To explicitly account for these effects, it would be necessary to use variable electrostatic terms and parameters that respond to external electric fields. In the simplest case, the atomic charges of molecules would depend on their environment. As was already discussed in section 3.1.1.3, rearrangements of the sugar−phosphate backbone conformation could

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

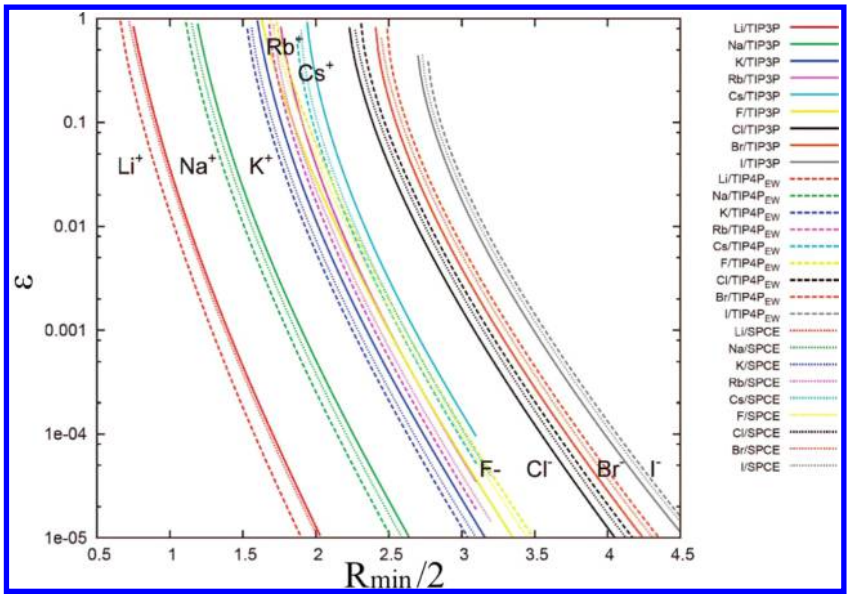


Figure 26. Example of correlation between well depth (kcal/mol) of the potential energy and radius (Å) at the minimum Lennard-Jones potential energy giving constant hydration energies calculated for diverse cations and water models; for more details, see ref 355. Reprinted with permission from ref 355. Copyright 2008 American Chemical Society.

also change the electronic structure of the RNA fragments, which is another reason why current fixed-charge models provide only a very crude approximation of real molecules. The electrostatic term should thus be also conformation-dependent. Polarizable force fields that are currently under development for nucleic acids attempt to include polarization effects using more sophisticated terms, for example, by using the formalism of classical Drude oscillators,[81−83,546,547] but it remains to be seen whether these approximations will be sufficiently robust in practical applications. In conclusion, polarization/charge-transfer effects associated with ion binding are largely orthogonal to the energy terms that are tunable with current standard force-field formulations and parameters sets, and cannot be directly affected by them.[77,538] Consequently, the only way to mimic these effects is through effective cancellation of errors in other force-field terms. In other words, specific parameters might work correctly if used in conditions that closely match those on which their empirical fitting is based, but their transferability is limited. Cations are prominent sources of polarization and charge-transfer effects.

Polarization and charge transfer lead to two salient effects associated with ion binding.[537,544,548] First, there is a polarization/charge transfer from the ion to the ligands in its first coordination shell (see Figure 25). This modulates the ligands' electronic structures in a way that induces ligand−ligand repulsion within the inner shell, partially counteracting the potential energy gain from ligand−ion binding. Consequently, binding of water molecules to an initially "naked" ion becomes progressively less favorable: energy gain resulting from binding of the first water molecule in the inner shell is greater than that for binding the second, which is greater than that for binding the third, and so on until no further water molecules can be added to the first shell. This behavior is partly due to steric factors and unperturbed electrostatics, which are more or less well described by properly calibrated pair-additive force fields. However, some of the anticooperativity of first-shell ligand binding is due to ligand polarization: the binding-associated changes in the molecular induced dipoles within the first ligand shell are repulsive. The second major effect of polarization/charge transfer on ligand binding is that H-bonds between polarized water molecules (ligands) in the first coordination shell and water molecules (or other H-bond acceptors) in the second shell are much stronger than those in bulk water. Therefore, polarization creates anticooperativity in ligand-ion binding within a given shell, and cooperativity in ligand binding between consecutive ligand shells. Neither effect is explicitly accounted for in the pair-additive force-field approximation.

Because the force-field parameters are set with the aim of reproducing observable physicochemical properties of ions, one might still obtain reasonable overall solvation energies, but breaking the overall interactions down into individual terms is unrealistic. In addition, it is difficult to obtain unambiguous experimental ion solvation free energies, and to compare such data to free energies computed in MD simulations on the basis of the periodic boundary condition, which complicates parameterization efforts.[549,550] Moreover, when ion-hydration free energies are used as target values for parameterization, the ion parameters necessarily become coupled to those of the chosen water model.[355,542] In addition, the ion hydration energy alone is not sufficient to unambiguously parameterize the ion potential because a given target hydration energy can be obtained using very different combinations of the two tunable Lennard-Jones parameters, which are the optimal ion−water oxygen vdW distance (minimum on the vdW potential energy curve; summed "atomic radii") and the vdW energy (potential energy-well depth) (Figure 26). Indeed, parameter values used in different ion parameter sets reported in the literature vary widely. Unfortunately, if the ion radius lies outside a certain meaningful range, it can create problems when describing interactions with simulated biopolymers. This ambiguity is only partially removed by simultaneously targeting the equilibrium ion−water oxygen distance, because the latter parameter correlates with the hydration energy; that is, both the ion vdW radius and the energy-well depth affect the ion−water oxygen distance.[355,542] One way to alleviate this issue is to consider additional properties such as crystal lattice parameters

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

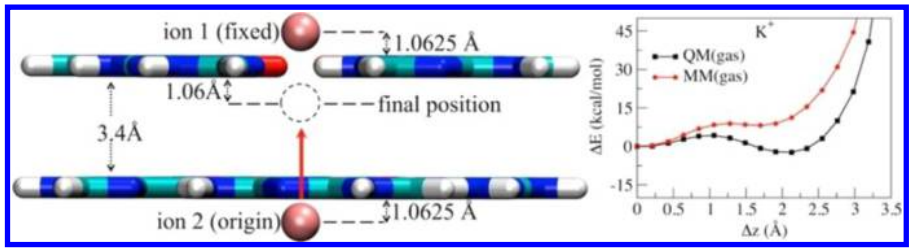


Figure 27. Interaction between two K^+ cations coordinated to two rigid guanine quartets. The first ion is fixed while the second ion moves from the starting (origin) to the final position. The *y*-axis shows relative energy of the system with respect to the starting configuration calculated using QM benchmark (black curve) and AMBER force field (red curve). The force field strongly overestimates the cation−cation repulsion due to a lack of polarization. Reprinted with permission from ref 538. Copyright 2014 American Chemical Society.

to narrow the parameter ranges. A further complication that must be accounted for is that the results of simulations depend not only on the values of the Lennard-Jones parameters but also on the combination rules used to calculate the vdW energies between different atoms. All of these problems can be seen as consequences of the oversimplified and unphysical force-field description of the ions.[542]

Polarization effects are relatively mild (although not negligible) for monovalent ions. Alkaline earth cations such as Na^+ and K^+ are described reasonably well. Note that force fields are not designed to describe monovalent cations whose binding is partially covalent, such as Ag^+ or Tl^+. Cl^- is likely described a little less accurately than the monovalent cations because the electron clouds of anions extend further from the nucleus, and so anions are inherently more polarizable.[551]

A textbook example illustrating limitations of the monovalent ion description is binding of Na^+ and K^+ to G-quartets in RNA and DNA G-quadruplexes.[538,552,553] Force fields qualitatively describe the basic electrostatic stabilization of the G-quadruplex stems by monovalent ions because they capture long-range electrostatics well. However, there are difficulties in describing with a force field the details of binding of desolvated ions inside the G-quadruplexes, due to profound polarization effects. A notorious problem is the occurrence of bifurcated G-quartets in simulations of G-quadruplexes.[128,554] Comparisons with benchmark QM calculations further revealed that force field underestimated the direct interaction energy between the ion and G-quartets;[552] for an explanation of the term "interaction energy", see ref 118. In addition, the ions inside the G-quadruplex often behave as having excessively large atomic radii; that is, they fluctuate as though they do not fit properly in the cavity. This is because it is difficult to simultaneously fully balance all of the relevant energy contributions; this particular problem might arise from a parameterization mismatch between ion hydration energies and ion−quartet interactions. Full extent of the force fields' limitations in this case was demonstrated by a benchmark QM study that scrutinized binding of multiple monovalent ions inside G-quadruplex stems.[538,553] When two or more ions are present inside the stem, their effective mutual repulsion is significantly overestimated by force fields (Figure 27). Real ions communicate with each other electronically through polarization of the G-quartets in the stem. This typically softens their mutual electrostatic repulsion relatively to that present between two +1 point charges. To correctly describe the ion−ion interaction inside the G-stem, it would be needed to somehow reduce the ion charges.[538] However, this would have to be done in a geometry-dependent manner. The charge distributions of both the cations and the G-quartets would have to be recalculated for each snapshot realized in the simulation, rendering the fixed-charge force-field approximation inadequate. In reality, any change in the coordinates of the ions or the quartet geometries will modify the system's electronic structure.

Effects of polarization and charge transfer on the binding and modeling of divalent ions are far greater than for monovalent ions, so the failure of force fields to adequately describe these effects can severely affect the quality of modeling of divalent ions. Therefore, while publications on MD simulation studies often state that the simulations were performed "in the presence of Mg^{2+}", such claims should be treated with care, because force-field descriptions of divalent ions are well outside the limits of approximations of contemporary multipurpose biomolecular force fields. They involve far more severe approximations than are used to describe the organic components of the RNA molecule or the monovalent ions. Total contribution of nonadditive effects in the first ligand shell of Mg^{2+} amounts to ∼70 kcal/mol in absolute terms, which is equivalent to the interaction potential energy of five A═U base pairs.[537,545]

Fortunately, there are large mutual error compensations between the different missing energy terms. For example, the absence of polarization means that force fields underestimate both direct ion−ligand attraction and interligand repulsion. This attenuates the error in description of the difference between hexa-hydrated Mg^{2+} and partially dehydrated Mg^{2+} coordinated with RNA. However, the main problem is that even if a force field can be adjusted to reproduce one target property of a system (for example, the hydration energy), it is still very difficult to simultaneously obtain a correct description of its full range of interactions.[542] One should thus be careful when choosing ion parameters, and perform simulations using parameters designed to reproduce physical properties most important for the problem at hand. For example, one might be interested in the effect of Mg^{2+} on an RNA structure in which direct binding might not be relevant. In such cases, a force field that accurately reproduces the behavior of hexahydrated Mg^{2+} would suffice.

It is difficult to quantitatively estimate the magnitude of errors, or which phenomena one might be missing when using a given force field to simulate the binding dynamics of Mg^{2+}. The only benchmark method that would provide a true picture of ion binding would be a sufficiently long QM dynamics simulation using a sufficiently accurate QM description. Such calculations are far beyond the capabilities of modern hardware. However, simpler QM studies (involving gradient optimizations and energy calculations on small molecular clusters) have yielded results that differ markedly from equivalent force field-based predictions.[537,540,544] For example, the physical proper-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

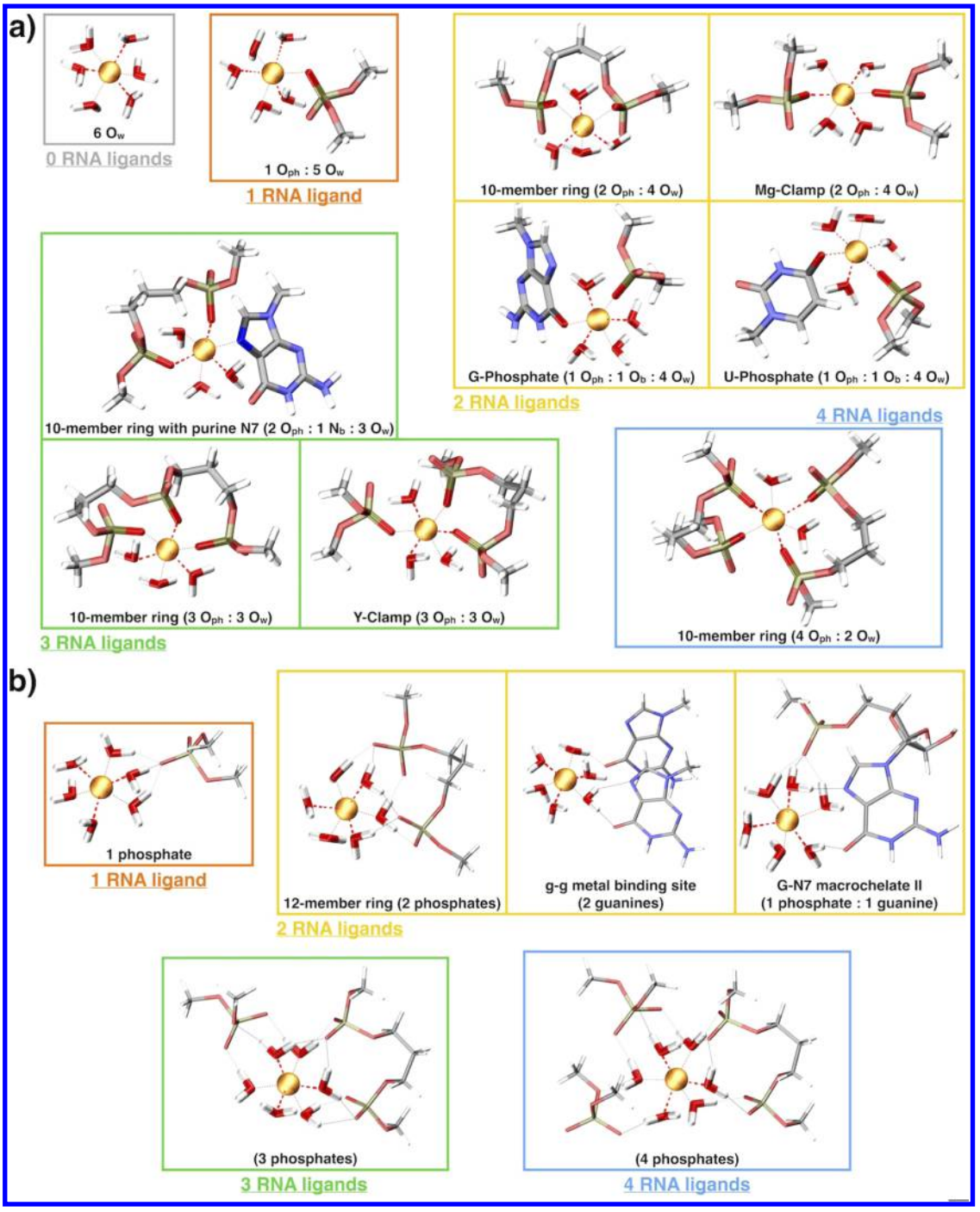


Figure 28. Different RNA–Mg^{2+} binding patterns investigated in ref 488. Reprinted with permission from ref 488. Copyright 2017 American Chemical Society.

ties of water molecules in the first ligand shell of Mg^{2+} are very different from those of bulk water: the ion-bound waters can form much stronger H-bonds, which can be inferred from the large stretch (elongation) of the O_w–H covalent bond.[537] Binding of divalent ions to the guanine N7 may also affect the strength of its base pairing via guanine polarization (Figure 24).[539,545,555−558] There is an intricate network of polarization/charge-transfer effects in the vicinity of any divalent ion bound to RNA that is specifically modulated by the type of the divalent ion and the structure of the binding site.[488,537,544] However, it is important to note that most reported QM results relate to gas-phase clusters or clusters surrounded by a continuum solvent, which may somewhat overemphasize the role of polarization/charge-transfer effects. These calculations reveal the upper limit of these effects as compared to fully solvated systems.

QM calculations of hydrated Mg^{2+} ions interacting with different RNA binding sites clearly demonstrated the magnitude of the complex charge-transfer and polarization effects in these systems using Natural Bond Orbital and Bader analyses (Figure 28).[488] Nevertheless, the authors suggested that the differences in charge transfer and polarization among common Mg^{2+}–RNA coordination patterns are surprisingly small. In particular, charge of the Mg^{2+} ion does not vary greatly

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

between coordination sites, suggesting that approaches based on developing site-specific Mg^{2+} ion parameters (with different vdW parameters for different RNA binding patterns) may not be consistent with the physical origin of the inaccuracies in the MM description of Mg^{2+}−RNA interactions in the MD simulations. However, the results in no way say that the MM description of Mg^{2+} is satisfactory. For example, the authors reported highly variable levels of charge-transfer involving the first-shell water molecules.[488] More specifically, the charge rearrangement of the water hydrogen atoms decreased markedly when RNA atom ligands were included in the coordination sphere. Consequently, the physical properties of the bound water molecules depended on the nature of the cation's other ligands (for discussions of a similar effect, see ref 544). QM calculations also clearly confirmed the existence of sizable charge-transfer/polarization effects between the first and second ligand shells of Mg^{2+}. To complement the QM calculations, the authors presented a comparative analysis of five Mg^{2+} parameterizations based on MD simulations of the Group IIC intron and HDV ribozyme.[488]

The severity of the pair-additive approximation is clearly illustrated also by the difference between Mg^{2+} and Zn^{2+}. Both ions have identical charges and very similar radii, so their simplest force-field models are inevitably similar. In reality, Mg^{2+} and Zn^{2+} are quite different ions due to factors such as the presence of 3d orbitals in the valence shell of Zn^{2+}. While Mg^{2+} is almost always hexa-coordinated, Zn^{2+} can have anywhere between four and six ligands.[548] Moreover, Mg^{2+} preferentially interacts with oxygen atoms, whereas Zn^{2+} prefers nitrogen atoms such as the N7 centers of purine bases due to its ability to accept electron density from nitrogen lone pairs (which are delocalized in the case of nucleobases) into its vacant orbitals. This is a clear example of a charge-transfer interaction. All of these effects are missing in the force-field description, so the differences between Zn^{2+} and Mg^{2+} can only be captured by QM calculations or by using specialized force fields with charge-transfer terms such as Sum of Interactions Between Fragments Ab initio computed (SIBFA).[540,541] However, these methods are not yet suitable for MD simulations of biomolecules.

We suggest that polarization/charge-transfer effects associated with divalent ions are so severe that they may not be adequately described using the formalism of the multipurpose polarizable biomolecular force fields currently in development[559] without significantly compromising their portability. Unlike SIBFA, these polarizable force fields seem to lack terms suitable for explicitly describing charge-transfer effects. Going back to the simplest force fields, the only way to respect the existence of these effects in a pairwise additive force field is to optimize the few available parameters to obtain the correct overall behavior. However, any such optimization will inevitably have limited transferability.

There are ongoing attempts to improve the description of ions by adding additional ad hoc terms to the usual point-charge plus vdW Lennard-Jones 12-6 description, as reviewed by Li and Merz.[542] For instance, 12-6-4 vdW terms have been suggested to improve the behavior of Mg^{2+} in simulations.[560,561] A physical reason for including the r^{-4} term in the parameterization is that it offers a way to mimic charge−induced dipole and dipole−induced dipole energy terms. An attempt to further tune the r^{-4} terms (the C_4 parameters) for RNA has been reported, targeting the balance between binding to phosphates and to N7 atoms of purine nucleobases.[562] Completely different models have also been proposed, in which the Mg^{2+} ion is modeled as a rigid body that includes displaced charges (using dummy atoms) to mimic the electronic structure around the ion.[563−565] However, none of these models directly includes polarization and charge-transfer effects. Essentially, one has just two (or a few when using more sophisticated potentials) parameters available for tuning: the atomic radii and the depth of the vdW potential well. Moreover, as noted above, these two parameters are interrelated to a certain extent.[355,542] Thus, the common practice of training the ion parameters to reproduce hydration energies could create flaws in the description of direct solute−ion interactions, and vice versa. An interesting alternative attempt to obtain ion parameters has been parameterization of Mg^{2+}, Li^+, Na^+, and K^+ Lennard-Jones potentials using experimental osmotic coefficients.[566,567]

Some early MD simulations of nucleic acids suffered from another specific problem that may be indirectly related to the simplicity of the force field. When excess salt (NaCl or KCl) was added to the net-neutralizing ion atmosphere, spurious cation−anion pairing, clustering, and even crystallization occurred when the simulated system was well below the solubility limits of the real system.[481,568] This problem has been severe for some older parameterizations (or combinations of parameters) of the ions, including at that time standard combination of parameters and combination rules in the AMBER code. This could have affected some older simulations when the authors did not notice the problem. The salt crystallization effect has been eliminated or at least efficiently suppressed by newer parameterizations.[355]

3.4.3. Difficulties in Sampling Ions. A second source of difficulties in properly describing ions is sampling. For monovalent ions, good and almost converged sampling of the simulation box can be achieved on a time scale of 100 ns.[569,570] Exceptions include tightly bound ions and especially poorly accessible ion-binding inner pockets that do not easily exchange with the bulk solvent. Conversely, easily accessible monovalent ion binding sites are usually swiftly found by the ions in the course of the simulations. Two examples that illustrate the diversity of ions' kinetic behavior in MD simulations are the open monovalent ion binding site in the catalytic pocket of HDV ribozyme[503] and the buried ion binding site at the quadruplex-duplex junction of the fragile X mental retardation protein (FMRP) RGG peptide−RNA complex;[154,508] see sections 4.8.3.2 and 4.7.10.

Sampling of hexahydrated Mg^{2+} ions that bind via outer-shell (through-water) interactions can be quite tricky. If simulations are performed using typical experimental ion concentrations, the number of included divalent ions will be rather small, requiring more simulation time to converge the exploration of all possible binding sites. A different approach should be adopted if one is interested in modeling the direct binding of Mg^{2+} to RNA, which is a completely different story in terms of sampling. Modeling could be straightforward if the exact position and coordination of the bound ion are known from experiments. However, spontaneously sampling first-shell ligand exchanges for Mg^{2+} is a complex problem. Experiments suggest that the water molecules from the first shell of the Mg^{2+} exchange on a microsecond time scale.[505] Suitable methods should thus be used to explore all of the putative binding sites.[335,571] Sample results obtained with enhanced sampling techniques are reported in Figure 17. In addition, the force fields may not be sufficiently accurate to properly capture the kinetics and thermodynamics of such events (section 3.4.2).

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Thus, in earlier simulations with the at the time popular Aqvist's Mg^{2+} parameters,[572] the formation of spurious inner-shell (direct) binding to the RNA atoms has been often observed even during equilibration.[131] This problem might have been eliminated by newer parameterizations, in which Mg^{2+} is practically maintained hexahydrate by increasing the dehydration free-energy barrier.[504] Such parameterization should improve kinetic properties of the cation ligand shell, albeit possibly at the expense of accurately modeling the ion's thermodynamics properties. Such parameterizations may run into troubles when simulating real cases of inner-shell Mg^{2+} binding, for example, in the ligand binding pocket of the quaternary HutP protein/RNA complex.[573] Here, the experimentally observed Mg^{2+}−nitrogen atom interactions are irreversibly lost on a time scale of few hundreds nanoseconds with the Allner et al. parameters,[504] while the Aqvist's parameters,[572] due to their specific compensation of errors, neatly keep the experimentally observed ligand binding shell unperturbed in multiple microsecond time scale simulations.[110] Interestingly, when testing the consensus 12-6 parameters by Li and Merz,[574] the authors have obtained a quite stable HutP ligand binding pocket with SPC/E[575] but not with TIP3P[576] water model.[110] This is a reminder of the complex dependence of the ion binding processes at different binding sites on ion as well as water model parameters. This is why numerous parameterizations of ions for biomolecular simulations can be found in the literature.[355,542,561,574,577,578] However, as noted above, all of them suffer from the common limitations of pair additive force fields, irrespective of which property is used as the target in the parameterization. It is inevitable that improving one target property may worsen others; tweaking parameters cannot fully compensate for the inadequacy of the underlying physics.

3.4.4. Importance of the Simulation Box Size. A third problem in ion treatment relates to the second. Proper sampling of ions in solution requires appropriately sized boxes.

Most published simulations have used relatively small solvent boxes; the typical recommended distance between the solute and the edge of the water box is 10−15 Å, and simulations are usually performed using periodic boundary conditions and the particle-mesh Ewald (PME) formalism.[579] Small boxes prevent the ion distribution at the box border from matching that in bulk solvent, so the simulations may suffer from finite-size artifacts (Figure 29).[580,581] Large boxes can be used but significantly increase computational costs, effectively reducing the amount of simulation time and thus the sampling of the system.[582,583]

When an ion leaves the box, it does not enter the bulk solvent but instead enters the box on the opposite side. Thus, even if ion binding sites around the solute are well sampled (within the force-field limitations), the exchange between different binding sites may be affected by a small box size. A box that is too small cannot provide a realistic partitioning between local and bulk ions because it is lacking a true bulk. These effects are discussed extensively in refs 580 and 581. In real systems, the presence of the highly charged RNA macro-anion causes counterion accumulation and co-ion depletion in the RNA local atmosphere, so the RNA charge is completely neutralized by the ion atmosphere.[496,502] However, the typically used box sizes cause spurious truncation of the ionic atmosphere around the simulated RNA. This, when performing excess-salt simulations, leads to an unrealistic accumulation of anionic species near the edges of the box.[580,581] A similar effect

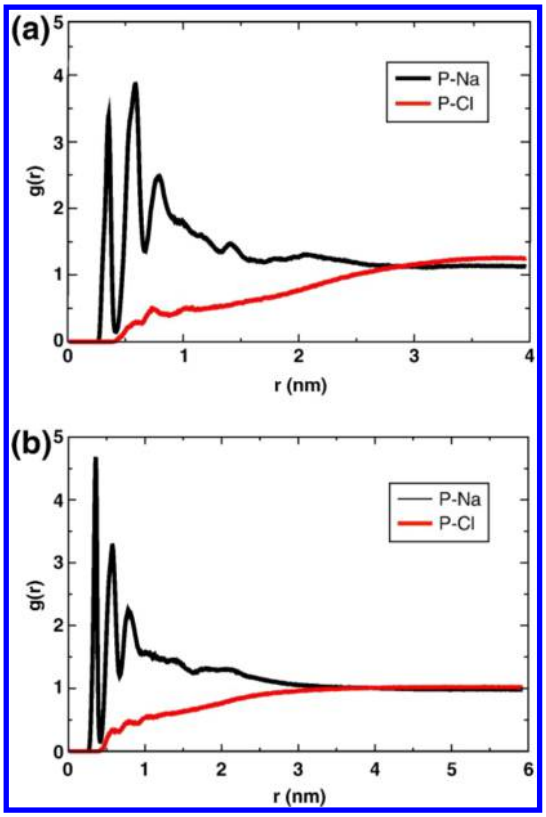


Figure 29. Illustration of how finite-size artifacts lead to concentration mismatches in the bulk. (a) RNA phosphate-counterion radial distribution functions for a large RNA complex in an 80 Å-side box. The anion and cation concentrations do not match at large distance from RNA. (b) The same system simulated in a 120 Å-side box. The anion and cation concentrations at large distance from RNA agree with each other to within 1%. Reprinted with permission from ref 580. Copyright 2009 Elsevier Ltd.

was also observed in ref 335, where the computed affinity between Mg^{2+} and RNA displayed a dependence on box size consistent with this trend. In addition to this, the limited amount of space available in a small simulation box might prevent the system from properly equilibrating on the simulation times-scale; that is, one can expect that the equilibration of the ion distribution is slowed.[580,581] Moreover, because the equilibrium of the ion distribution is substantially affected, also the kinetics of ion diffusion processes is inevitably distorted, even in simulations that reach equilibrium. Simulations with standard boxes thus should not be used to make quantitative suggestions about the nature of ion diffusion processes around the RNA molecule. One MD study[367] correctly reproduced the different free-energy effect of guanine to inosine substitution in A-RNA and B-DNA[584] using the TI method (see sections 3.2.7 and 4.3.3). However, the salt-dependence of this phenomenon could not be correctly reproduced.[367] This may have been due to the use of an insufficiently large box. The consequences of a small simulation box are thus certainly significant but have yet to be fully quantified. Some authors also argued that there could be errors stemming from the PME method,[585] although, again, these have not been clearly documented yet.

Many past nucleic acid simulations were performed using a net-neutral cation atmosphere, which was usually achieved by adding Na^+. While this approach is widely accepted by most in the simulation community, it has been occasionally questioned

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

because it does not correspond to experimental salt conditions. However, while net neutral conditions may look non-natural at first glance, they can be justified and should not cause major problems, at least in comparison to those arising from other approximations. Because RNA has a charge of −1 per phosphate, net-neutralization (one monovalent cation per phosphate) with commonly used solvent boxes corresponds to fairly relevant cation concentrations of ∼0.2−0.3 M. This estimate is obviously not valid for net-neutralization of protein−RNA complexes if the bound protein provides substantial compensation for the net charge of the RNA. The cation concentration can be calculated by simply comparing the number of ions to the number of waters in the simulation box. Adding excess salt would further increase the cation concentration. For completeness, it should be noted that the formal solute concentration in simulations is relatively large (a concentration of 0.01 M for the solute corresponds to ∼5500 water molecules), but solute−solute contacts are prevented by the use of sufficiently large boxes in connection with periodic boundary condition.

3.4.5. Should Ion Conditions in Simulations Match Those in Experiments? Some in the research community have suggested that simulations should exactly mimic the ion conditions of experiments, arguing that certain calculations are flawed because the authors used Na^+ instead of K^+, or vice versa, or did not use excess salt, or did not use divalent ions, etc. Below we explain why such criticisms are usually excessive.

There are several reasons for not requiring the salt conditions in MD simulations to exactly match those of specific experiments. First, it is, in fact, not possible. There are unavoidable intrinsic differences between MD and experiments that should be taken into account. For example, the infinite solvent is mimicked by the periodic boundary condition approximation, and the box size is always much smaller than any experimental buffer. This factor alone means that the conditions in which the MD simulation and experiment are performed can hardly be entirely equivalent. Second, biological relevance of salt conditions used in some experiments can be questionable, for instance, in crystallization experiments (see section 3.4.6). There are indeed works in which the concentration of ions is exaggerated with respect to the intracellular one to help crystallization or to make sure that the folded RNA structure is kept intact throughout the experiment. RNA folding is usually accelerated at higher Mg^{2+} concentration. Nonetheless, there are counterexamples in which the increase of ion concentration does not accelerate folding but instead creates kinetic traps or induces misfolding.[586]

Most importantly, MD simulations are frequently performed to investigate properties of the molecules that may be largely insensitive to small changes in the salt conditions if one accounts for the starting structure and simulation time. MD simulations are relatively short and typically stay within one conformational substate, so they are probably less sensitive to ion treatment than real nucleic acids in folding experiments. The ion-dependence of the simulations should therefore not be compared to known equilibrium experimental thermodynamics salt dependences or experimental conditions with excessive strictness. Experimental results reflect equilibrium thermodynamics, whereas simulations usually sample around the starting structure. Therefore, if the experimental ion-dependence correlates with the relative stability of two ensembles (substates) that are not simultaneously sampled in the simulation, the simulation's behavior may be fairly insensitive to the type and concentration of the ions used. For example, at equilibrium, an RNA kink-turn should unfold in the absence of proteins and divalent ions.[523,524] However, when one extracts the structure of the folded kink-turn from the ribosome, the kink-turn can be simulated in its folded (kinked) conformation in the presence of only a net-neutralizing cation atmosphere (see section 4.3.7).[522] Such simulations are entirely relevant when studying the effects of base substitutions, internal flexibility of the kink-turn in its folded states, molecular contacts at the protein/RNA interfaces, and many other properties of the folded kink-turn. The transition to the unkinked ensemble is not observable on the usual simulation time scales due to the presence of a high kinetic barrier. Such simulations thus cannot reproduce and are not sensitive to any phenomena that result from communication between the folded and unfolded ensembles. Another textbook example of the lack of any profound ion identity effect on MD is the simulation of G-quadruplexes.[170] G-quadruplexes are prominent monovalent ion binders, and changing the ions can affect the population of their different folds in thermodynamics equilibrium in experiments. Despite this, one can safely simulate any folded G-quadruplex using either Na^+ or K^+, without noticing any systematic differences. This is thanks to the fact that the lifetime (i.e., the reciprocal of the k_{off} rate constant) of a folded quadruplex (i.e., a single molecule) is typically on the order of minutes in real time no matter what cations are present. These time scales are orders of magnitude beyond those accessible in simulations with modern hardware. Therefore, while swapping from Na^+ to K^+ may have subtle effects on ion dynamics in simulations of specific single folds (conformations), these effects are usually inseparable from the sampling uncertainties. Furthermore, the effects of different ions on the relative stability of different folds are not sensed by the simulations. All known G-quadruplexes simulated with ions in their channel retain their starting fold independently of the ionic conditions, as discussed at length elsewhere.[128,170,587]

A rather unexplored topic and a potential source of errors is that virtually all MD simulations are performed at a fixed number of atoms. Thus, the most reliable way to know the effective ionic concentration that is comparable to experiments is to measure the ion−water ratio (concentration) as far as possible from the solute. Because the number of ions is fixed, the concentration of ions in the bulk could change if the solute undergoes a conformational change that affects the number of trapped ions. Ideally, this number should not be dependent on the specific conformation of the solute. In cases where this number depends on the solute conformation, one should use methods that keep a constant chemical potential of the ions instead of a constant number of ions. Simulating such ensembles is nontrivial, and has proven to be challenging both theoretically and in terms of practical implementation. A suitable approach to this problem has been proposed, but to date it has only been applied to urea, which is not charged.[588]

3.4.6. Problems with Experimental Positions of Ions. Another fundamental issue is uncertainty in experimental structural data. This is an absolutely critical problem when dealing with directly bound divalent ions because their incorrect placement at the start of an MD simulation may cause serious sampling problems. Indeed, due to the problems discussed above, incorrectly placed divalent ions may both stay trapped in the same place for the entire simulation and also trap the RNA structure in an incorrect conformation. NMR studies usually do not detect positions of ions around nucleic acids, or

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

only provide indirect information on their positions.[589−591] In principle, X-ray structures could provide reliable information on the binding patterns of Mg^{2+} ions. However, the relevance and accuracy of their structural information is sometimes compromised. First, some ions observed in X-ray structures may be present because they played some role in the formation of the crystal lattice, and may not interact in the same place in solution. Second, some centers identified as Mg^{2+} cations during refinement may actually be other particles.[513,592]

In an earlier MD study on the product[593] and inactivated precursor[594] of the HDV ribozyme,[503] a striking difference between the two experimental structures was noticed. The precursor X-ray structure featured only two bound Mg^{2+} ions, both neatly placed in dominant ESP minima identified in the simulations. Both ions behaved well when included in MD simulations. However, the product X-ray structure included nine bound Mg^{2+} cations, some of which were not close to any ESP minima. It is likely that some of these ions were not properly refined. A systematic analysis revealed that it is not uncommon for electron densities corresponding to water molecules or monovalent ions to be identified as Mg^{2+} ions during refinement.[592,595] Even more strikingly, some reported Mg^{2+} ions may correspond to misinterpreted anion densities.[592,595] Even though one might expect that these incorrectly placed ions would quickly dissociate in simulations using a well-parameterized force field, transferring such unnatural Mg^{2+} ions into simulations may cause substantial instability because the incorrectly bound divalent cation could immediately perturb the RNA structure.

A unique insight into the ambiguities of identifying cation binding in RNA X-ray studies was obtained in a benchmark study using X-ray structures of the duplex of the HIV-1 dimerization initiation site (DIS), which contains a known ion-binding internal loop.[596] The X-ray structures were solved in the presence of 13 different cations: K^+, Pb^{2+}, Mn^{2+}, Ba^{2+}, Ca^{2+}, Cd^{2+}, Sr^{2+}, Zn^{2+}, Co^{2+}, Au^{3+}, Pt^{4+}, $[Co(NH_3)_6]^{3+}$, and $[Ru(NH_3)_6]^{3+}$. This RNA molecule exhibits sequence symmetry but lacks true crystallographic symmetry, so each structure provided two crystallographically independent but equivalent ion-binding patterns. The study revealed very diverse ion-binding patterns that were often not very consistent with the commonly assumed basic rules governing the binding properties of different ions. More strikingly, in many instances the binding of a given cation to the two presumably equivalent halves of the duplex was very different. The authors suggested a complex and strong dependence of the observed cation binding on hidden parameters that could be related to biochemically irrelevant differences in crystal packing.[596]

Binding of Mg^{2+} to the pocket of HIV-1 DIS kissing-loop complexes provides another illustration of the difficulty of identifying cation binding to RNA by X-ray crystallography. Some X-ray structures clearly show the binding of an Mg^{2+} ion to the pocket, while other crystal structures do not identify bound ions in the same position.[597,598] It has been shown by atomistic MD simulations that, in the absence of divalent ions, the pocket of HIV-1 DIS kissing-loop complexes is occupied by a set of 2−3 long-residency but delocalized monovalent ions (see section 4.3.10).[489] The X-ray structures have also revealed that the Mg^{2+} ions in the kissing-loop complex may be displaced by a spermine molecule used for the crystallization;[598] note that positions of spermine molecules are rarely identified in X-ray structures. Simulations of many other RNA systems have similarly shown that monovalent ions can occupy the same ion binding sites as divalent ions. These observations confirm the complexity of ion-binding to RNA, which involves both competition among different types of bound ions and that among different ion-binding patterns. Consequently, one X-ray structure of a given molecule may suggest the presence of an important divalent ion binding site, while another structure of the same molecule may show no ion binding because the site is, for example, interacting with fluctuating monovalent ions that cannot be discerned from the densities. In the case of the HDV ribozyme, the MD prediction of a monovalent ion binding site at the catalytic pocket[503] was subsequently confirmed by an X-ray structure obtained in the presence of Tl^+.[599] The absence of bound ions in some X-ray structures of the HIV-1 DIS kissing-loop complexes is readily explained by the presence of fluctuating monovalent ions or spermine that are not experimentally resolvable because of averaging. Even dynamic fluctuations of divalent ions can prevent their detection in many refined X-ray structures. Thus, while some X-ray structures may feature incorrectly refined ions or ions whose presence is due to crystal packing, others may fail to resolve important bound ions. As concluded by Ennifar et al.,[596] it is inadvisable to draw firm conclusions on the mode of binding of a given cation on the basis of a single crystallographic example. This is obviously a substantial problem when preparing starting structures for MD simulations, where one must rely on single crystallographic structures.

With regards to incorrectly assigned Mg^{2+} ions, newer studies[514,600] suggest that at least for high-resolution crystallographic structures one can reliably determine that a resolved species that may be an ion is not an Mg^{2+} center by applying a set of distance constraints derived from the known chemical behavior of Mg^{2+}. However, a much more careful analysis would be needed to identify a directly bound Mg^{2+} without any uncertainty. This would require consideration of geometric constraints on the first coordination shell of magnesium and typical distances from its most common ligands (phosphate oxygens and guanine O6 oxygens).[513,514]

It should be noted that, at first sight, MD simulations would be ideal tools for studying the structural dynamics of the diffusive ion atmosphere around RNA molecules and the coupling of ion atmosphere dynamics with solute structural dynamics and transitions. There are currently no experimental methods that permit quantitative characterization of the ion atmosphere.[601] As noted above, the presence of an RNA polynucleotide causes compensatory ion redistribution, which can be characterized, for example, by measuring preferential ion interaction coefficients that reflect the cation accumulation and anion depletion around the RNA (resulting in a positive excess of cations and negative excess of anions).[496,502,601] However, this information alone is not sufficient to characterize the ion atmosphere because the overall preferential ion interaction coefficients are entirely insufficient to quantify the ion distributions reflected by radial distribution functions, never mind ion dynamics. Thus, in principle, MD simulations provide an ideal method for capturing the structural and dynamic details of the diffusive ion atmosphere. However, because of the limits discussed above, MD data must be interpreted with great care. Comparison of cation accumulation and radial ion distribution patterns obtained for B-DNA duplex using atomistic MD simulations and the 3D-RISM (3D reference interaction site model) formalism showed that while both methods gave similar preferential ion interaction coefficients,

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

they predicted different spatial distributions of ions around the solute molecule.[602]

3.4.7. How to Choose the Ionic Conditions for a Simulation? Given the above limitations, it is actually quite reasonable to perform MD simulations of RNA molecules without including divalent cations except in cases where there is unambiguous and exact structural information about a specifically bound Mg^{2+}. In the absence of such information, including divalents may lead to more losses than gains because incorrectly bound divalent cations can severely distort the simulated structure in a way that is not reparable on moderate simulation time scales. Divalent ions also complicate sampling. Regarding the issue of the sensitivity of the simulations to the different monovalent ion treatments (i.e., Na^+ vs K^+ and net-neutral vs excess-salt simulation conditions), several studies explicitly testing diverse ion conditions have been reported, and it was found that the simulations (with presently achievable time scales and box sizes) are fairly insensitive to the type and concentration of monovalent ions.[137,422,603−605] In fact, the results were often more sensitive to the use of specific water models.[137,604,605] At first sight, ions might seem to be more important components of simulation boxes than water molecules. However, the number of water molecules interacting with the RNA at any instant is much larger than the number of ions, explaining the simulations' sensitivity to the water model's parameterization. This point in no way excludes the possibility that simulations of some systems may be highly dependent on the type and concentration of monovalent cations.[128]

If one accounts for all of the limitations discussed above, Mg^{2+} ions can be usefully incorporated in MD simulations. Some newer Mg^{2+} models focus on describing specific types of interactions. For example, the model of Allnér et al.[504] was developed to better describe the interactions between Mg^{2+} and the nonbridging oxygens in the RNA backbone, together with the ion−water exchange rate. However, while properties like hydration free energies, ion−water coordination numbers, and oxygen distances are correlated, given the discussed limitations it might be not possible for a single set of LJ parameters to accurately depict all of these properties simultaneously.[110,542,574,578] Introducing extra LJ terms (12-6-4) or distributing the charge over dummy atoms (multisite ion models) does not always yield a better model for an arbitrary property because the model's quality is conditioned to the quantities it was designed to match. Furthermore, when a new model is proposed, a certain amount of time is needed for both proper testing and design improvements. Consequently, newer models may be less accurate than older ones when tested against large numbers of benchmarks or for specific goals (see section 3.4.3). York et al. compared the thermodynamic, kinetic, structural, and mass transport properties of 17 available Mg^{2+} MD models and found that, while many models accurately reproduced specific properties, only one 12-6-4 LJ model, coupled with a specific water model (SPC/E), managed to reproduce all of the studied properties reasonably well.[578]

3.4.8. Selected Simulation Studies Investigating the Ion-Binding to RNA. The paragraphs above discuss selected studies on the interactions of RNA with ions. However, there are numerous other studies in this area that warrant discussion (some of which are covered in other parts of this Review). For example, MD was used to investigate the structural roles of both mono- and divalent cations in the Varkud satellite ribozyme stem-loop V (SLV).[490,491,606] The SLV is an interesting RNA system because its structure and dynamics are Mg^{2+} dependent, and both Mg-free and Mg-bound structures are known.[607] The simulations revealed that the Mg-free SLV structure exhibited a partial transition into a structure resembling the Mg-bound state upon binding of K^+, but the correct Mg-bound structure was established only in the presence of Mg^{2+}. While one would not expect Mg^{2+} to reversibly sample every possible binding site on the studied time scale, it was interesting to see that adding Mg^{2+} in the simulation led to formation of the experimentally observed structure. The authors also reported a systematic analysis of different modern Mg^{2+} force-field models. While the relative binding patterns and occupancies were model-dependent, all of the models led to formation of the correct Mg-bound SLV structure. Surprisingly, the 12-6-4 model, which was expected to be the most accurate, was the poorest at capturing the occupancy of the directly bound Mg^{2+} ions.[490] This reemphasizes the warning that the supposedly best models are conditioned to specific properties, and that the problem of selecting the best model from the available options is nontrivial, and possibly not even well-defined. The inclusion of divalent ions in RNA simulations will likely remain a tricky problem with no flawless solution. MD simulations were also used to investigate the ion dependence of unfolding of GTPase associated center rRNA.[608,609]

Another work used atomistic MD to show that multivalent counterions increase bending rigidity (i.e., the persistence length) of double-stranded RNA by at least 30%, which contradicts the data for DNA.[610] This counterintuitive effect is observed for various tri- and tetravalent ions, and is robust to methodological details and changes in the RNA sequence. The simulations suggested that, in contrast to B-DNA, multivalent counterions bind inside the A-RNA major groove, causing significant contraction of the molecule along its helical axis. This makes further deformation due to bending more energetically expensive than it would be in the absence of the bound multivalent ions (for further discussion of differences between A-RNA and B-DNA helices, see section 4.3.9).

Finally, it should be noted that modeling of transition metal cations is even more complicated than modeling Mg^{2+}.[542] While ions are mostly described via nonbonded (intermolecular) force-field models in RNA simulations, modeling of transition metal centers may often benefit from the use of a bonded particle approach that is specifically parameterized for a certain coordination number and chemical environment. While the parameterization of multipurpose potentials is difficult for organic compounds, it is virtually insoluble for transition metal elements. Transition metals use higher-angular momentum orbitals (d and f) to establish complex chemical bonding patterns, may exist in different oxidation states, and can adopt a range of different coordination symmetries. In many cases, various spin states of the transition metal systems may be close in energy, complicating the calculations or even leading to problems due to real degeneracy of the electronic states. As a result, it is difficult to obtain quantitatively accurate descriptions of many transition metal elements in certain open-shell electronic states even when using the best available QM-based approaches.[611−613] However, this topic is beyond the scope of this Review.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

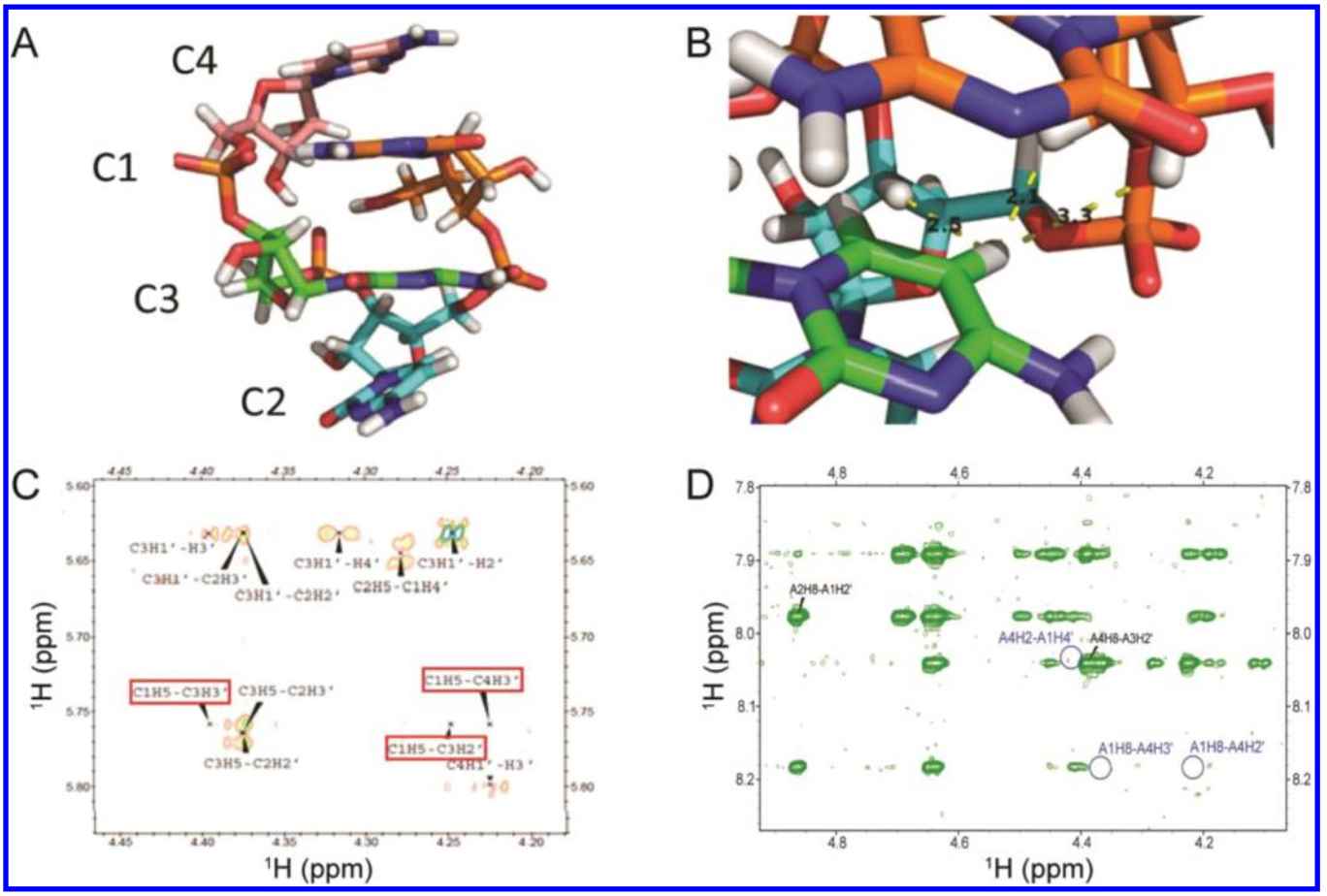


Figure 30. Sample MD structures and NMR data for tetranucleotides. (A) Intercalated structure for the r(CCCC) tetranucleotide obtained after ~800 ns MD simulation starting from A-form using AMBER99TOR force field. (B) Close-up of the same structure, showing the distances corresponding to C1H5−C4H3′ (2.5 Å), C1H5−C3H3′ (2.1 Å), and C1H5−C3H2′ (3.3 Å). (C) NOESY spectrum for the same system, showing the absence of the hypothetical H−H cross-peaks (red boxed labels). Adapted with permission from ref 615. Copyright 2013 American Chemical Society. (D) NOESY spectrum for r(AAAA) tetranucleotide. Regions corresponding to peaks that would be observed if intercalated states had a significant population are shown with blue circles. Adapted with permission from ref 222. Copyright 2017 American Chemical Society.

4. MD SIMULATIONS OF SPECIFIC RNA SYSTEMS

4.1. RNA Tetranucleotides as Key Benchmarks for Force Fields and Enhanced Sampling Methods

The smallest model systems studied widely are tetranucleotides. Although they do not have any particular biological function, the structural dynamics of single-stranded RNA is of relevance for understanding RNA folding processes and for description of unstructured RNAs. In addition, some folded RNAs contain single-stranded segments. Further, tetranucleotides have become key benchmarks for RNA force fields, because unique NMR data for several sequences are available.[614−616] Distances obtained from transient nuclear Overhauser enhancement (NOE) experiments are particularly important because they directly report on the proton−proton contacts present in solution, together with boundaries of their populations. In these experiments, all of the observed signals have similar line widths, and only nonexchangeable protons are analyzed.[222] Furthermore, these experiments reveal which contacts are absent in solution and should thus not be observed in MD simulations. In other words, they provide both positive and negative information that can be used to characterize the structural ensemble.[617] The NMR data indicate that RNA tetranucleotides mostly adopt an A-form-like conformation in solution,[616] with the exception of the disordered r(UUUU).[616] Importantly, so-called intercalated structures with out-of-order stacking are either absent or only populated at a level below the limit of detection. However, intercalated structures are highly populated in MD simulations and thus appear to be force-field artifacts (Figure 30).

According to the NMR data, tetranucleotides composed of purines (e.g., r(AAAA)) mostly sample the A-helix form.[616] As such, it makes sense to compare the MD-sampled structures to a single reference structure. In contrast, tetranucleotides composed of pyrimidines (e.g., r(UUUU)) are highly flexible, and even solution-phase experiments suggest the coexistence of several different structures.[616] As such, any comparison of simulated and experimental results should preferably use primary solution-phase NMR data. Every tetranucleotide that has been studied by NMR has exhibited *anti* glycosidic torsions and 3′-endo puckered sugars, with the exception of the 3′ terminal ribose in r(UUUU).

The structural dynamics of RNA tetranucleotides is highly sensitive to stacking interactions, backbone conformations, base−phosphate, and ribose−phosphate hydrogen bonds. The quality of force fields' descriptions of these energy contributions can thus be assessed by studying these systems. Conversely, although base pairing interactions play central roles in determining structures and dynamics of larger RNA molecules, they are not present in these short oligomers, so

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

studies on tetranucleotides cannot provide much information on force-field descriptions of base pairing. Tetranucleotides are therefore important model systems for selectively probing performance of the force field for a subset of factors determining the RNA structure. However, they should not be used as the sole benchmarks for force-field tuning because doing so could lead to overfitting.

The initial insights into the difficulties of simulating RNA tetranucleotides with the available force fields emerged from simulations that accompanied the original NMR studies and were already complemented by some attempts to improve the force fields. For the r(GACC),[614] a spurious population of the *syn* conformation of the glycosidic bonds was observed in MD simulations using the at the time standard ff99 force field (see section 3.1.2). A previously reported refinement of the glycosidic torsion potential χ_{YIL}[136] yielded better agreement with the NMR data (see section 3.1.2.2). A later work presented NMR data for the r(CCCC)[615] and compared them to the results of simulations using a different force-field reparameterization, AMBER99TOR (see section 3.1.2.2).[138] A subsequent comprehensive work[616] compared reference NMR data for the r(AAAA), r(CAAU), r(GACC), and r(UUUU) to the results of MD simulations. While the reported simulations were too short to permit ergodic sampling, these three seminal papers[614−616] provided very important reference experimental data and introduced the idea of benchmarking MD ensembles against solution-phase experiments. Importantly, the MD results were compared to primary NMR observables rather than to a single structure deduced from a model. Ultimately, the authors found that all of the tested force fields predicted significant populations of structures that were incompatible with the NMR data, including the aforementioned spurious intercalated structures.[614−616]

Tetranucleotides are also challenging from the point of view of sampling. Despite their small size, it is difficult to achieve their ergodic, statistically converged sampling. This would require extremely long MD simulations with time scales of at least hundreds of microseconds. Therefore, RNA tetranucleotides are also good benchmarks for assessment of the performance of enhanced sampling techniques (see section 3.2). In particular, the r(GACC) has been used as a test system in several methodological papers examining techniques including T-REMD, reservoir T-REMD,[258] multidimensional REMD,[271] accelerated MD coupled with T-REMD,[269] and replica exchange with collective-variable tempering (see sections 3.2.4 and 3.2.6).[259] These studies were primarily intended to benchmark the chosen enhanced sampling methods. However, they all predicted large populations of the intercalated structures and other spurious conformations, fully confirming the Turner group's results from medium-length plain MD simulations.[616]

A subsequent study presented probably the most extensive enhanced sampling simulations of RNA tetranucleotides reported to date,[139] likely achieving for the first time statistically converged results. The study clearly revealed that none of the presently available force fields could produce conformational ensembles for r(GACC) and r(CCCC) that are consistent with the available solution-phase experimental data. Another work from the same group[126] showed that an improved OPC water model[127] and modified phosphate vdW parameters[150] somewhat reduced the population of the intercalated structures. In both studies, comparison between MD and NMR was done mainly by monitoring the population of A-helix-like structures rather than using primary experimental data (Figure 31).

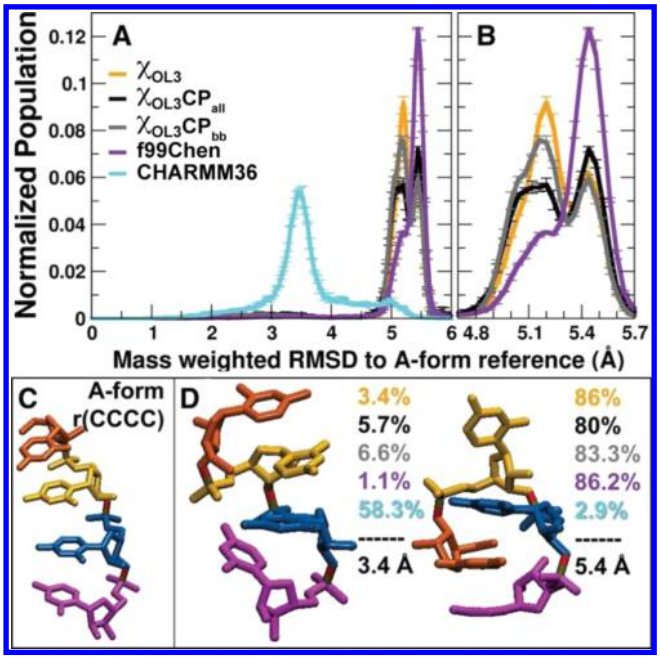


Figure 31. (A) Histogram of RMSD from A-form for r(CCCC) tetranucleotide as obtained from M-REMD simulations (see section 3.2.4).[139] Different colors represent different force fields, as noted. CP_{all} stands for a force-field variant that utilizes modified vdW parameters for all of the RNA oxygens while CP_{bb} is only for the phosphate oxygens in the backbone and the O4′, in both cases using parameters taken from ref 150; ff99Chen is the Chen and Garcia force field (see section 3.1.2).[120] Note that the structures at RMSD ≈ 3.4 Å from A-form obtained with the CHARMM36 force field are not intercalated but are still reported not too agree with experimental NOEs.[139] (B) Zoom on the region with RMSD 4.8−5.7 Å, corresponding to the intercalated structures. (C) Reference structure (A-form). (D) Two most populated clusters. Populations are written in colors compatible with those in panels (A) and (B). RMSD from A-form is also shown. It is evident that none of the force fields is capable of reproducing the A-form structure. Adapted from ref 139; http://creativecommons.org/licenses/by/4.0. Copyright 2015 Cold Spring Harbor Laboratory Press.

Subsequently, Bottaro et al. reported T-REMD simulations of five tetranucleotides for which experimental data were available.[242] They confirmed disagreement between MD and NMR by comparing the simulations' results to primary NMR data, including both NOEs to probe intercalated structures and scalar couplings to probe backbone conformations. Although these simulations were shorter than those reported by Cheatham et al.,[139] they achieved significantly more robust sampling than those reported in the original works by the Turner group,[616] with relatively low statistical errors. Interestingly, the study demonstrated that ensembles of tetranucleotide fragments with a given sequence extracted from the NDB/PDB database of RNA crystallographic structures agreed more closely with primary NMR data than did ensembles generated using MD. Similar observations have been made for protein systems.[618] This prompted an attempt to test empirical corrections that enforce dihedral distributions obtained from the PDB database.[143] The idea of using rotamer distributions from the PDB database to refine a force field has been adopted earlier by the CHARMM community to improve

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

modeling of protein backbones.[96] Unfortunately, the resulting force field modifications ultimately did not improve the agreement of MD with primary NMR data for RNA tetranucleotides.[143] Nevertheless, the computations identified some major structural determinants of the spurious intercalated structures, that is, an overpopulation of *gauche+* states for the α and ζ backbone dihedral angles. This led to the suggestion that ad hoc corrections penalizing these rotamers could improve agreement between MD and NMR ensembles. Such ad hoc corrections did indeed prove useful for several tetranucleotides and tetraloops, although they remained short of fully resolving the force-field problems.[163]

Another systematic study of RNA tetranucleotides simulated 48 different tetranucleotide sequences together with their DNA counterparts.[619] The RNA tetranucleotides were simulated using the χ_{OL3} force field (see section 3.1.2.1) for 1 μs, and the authors used the TIP4P-Ew model[620] for water instead of the more common TIP3P.[576] While simulations on this time scale could not be converged, they confirmed that intercalated structures with out-of-order stacking are a hallmark of all RNA tetranucleotide simulations.

Kinetics of sampling of different parts of the conformational space of tetranucleotides has been investigated with Markov state models (MSM) (see section 3.2.3) in conjunction with simulations using the χ_{OL3} force field.[222] Initially, the simulations consistently predicted significant populations of the spurious intercalated structures. To overcome this problem, the authors constructed a reduced MSM in which the spurious intercalated structures were manually removed from the data. After this cleaning of the data, the authors were able to extract kinetics for the formation of stacking interactions that were qualitatively consistent with experimental results.[224] For r(AAAA), the process occurring on the slowest time scale after filtering out the intercalated structures was associated with conversions between different stacked structures rather than a helix−coil transition. Although the comparison with experiment in this case was only indirect, calculated kinetic properties could be useful for validating estimates of barriers associated with conformational transitions.

In conclusion, RNA tetranucleotides represent valuable systems for testing both the accuracy of molecular mechanics force fields and the validity and efficiency of enhanced sampling methods. They are the first (and so far only) RNA systems for which a full sampling of the folding landscape has been achieved by using enhanced sampling methods and pushing the limits of the capabilities of contemporary computers.[139] Existing force fields do not offer satisfactory accuracy for these systems.[120,139,242,259,271,614−616] Exact reasons for these force-field imbalances are currently unknown, and may include a combination of diverse factors such as inaccurate descriptions of base stacking, overstabilization of H-bonds involving the nonbridging phosphate oxygens, and inaccurate description of RNA backbone rotamer families. In addition to effects arising from the deficiencies of the solute force field, inconsistencies between the RNA force field and the simple water models used in the simulations may also contribute to the observed simulation behavior.[126] This suggestion is supported by the above-noted sensitivity of the results to the choice of water model, and by the inability to identify salient and systematic force-field imbalances using large-scale QM calculations on RNA tetranucleotides.[113] The currently available results have confirmed that tuning of dihedral parameters is not sufficient to fully correct the tetranucleotide simulations, although some improvement can be achieved in this way. Undoubtedly, NMR studies on RNA tetranucleotides[614−616] constitute a vital source of experimental data against which all future force-field modifications must be benchmarked, although achieving good agreement with experiment for RNA tetranucleotides will not necessarily guarantee improved global force-field behavior due to the high risk of overfitting to a specific class of systems (section 3.1.1.3).

4.2. RNA Hairpin Tetraloops: Pushing Predictive Simulations to Their Limits

The RNA stem-loop (hairpin) motif is the most common secondary structure element found in naturally occurring RNAs (Figure 4).[621−623] Hairpin motifs consist of a loop containing 3−7 nucleotides capping an A-form helix stem.[624] Hairpin loops are usually highly organized RNA motifs with conserved structures that dictate their consensus sequences, and they allow the direction of the RNA backbone to be reversed. Most common motifs in this class are the U-turn and tetraloop (TL). U-turns are commonly found in the anticodon loop of tRNAs[625,626] but also occur in many other RNA structures.[627−629] TLs are the most common class of hairpin loops in rRNA and other large RNAs.[60] Hairpin loops guide folding of RNA structures,[67,68] protect mRNA from degradation,[623] and provide interaction sites for RNA tertiary contacts and protein binding.[622,630−632]

Because of their biological relevance, small size, and complexity of their internal interactions,[448,622,633−635] RNA hairpins have served as model systems for many experimental studies on RNA structure, dynamics, and folding.[172,636−647] They have also been the subject of numerous theoretical studies utilizing a wide portfolio of methods ranging from Rosetta de novo modeling[648] through Gō-like models,[446,447] kinetic models,[649] up to standard and enhanced sampling atomistic MD. Small hairpin motifs, and TLs in particular, have been used since the 1990s to test the accuracy of force fields used in MD simulations.[129,133,139,163,650] The most ambitious applications have been simulations of folding/unfolding pathways[163,228,237,240,241,313,651,652] conducted to predict the conformation and stability of a sequence's native structure and key features or properties of its folding landscape such as its melting temperature, folding transition states, and kinetics.

TLs have been classified into different families on the basis of phylogenetic analyses, backbone geometry, and interactions between the loop nucleotides.[60,622] Two families that are consistently recovered using different approaches are GNRA[621,653] and UNCG[654] TLs, where N is any base and R is a purine. These two groups account for over 70% of all TLs found in rRNA.[621,653] The other suggested families are CUYG,[621] GANC,[655] (A/U)GNN,[656,657] and UUUM[658,659] TLs (where Y is a pyrimidine, and M is either adenine or cytidine).

The UNCG family consists of particularly stable TLs characterized by a peculiar *trans*-Watson−Crick/sugar-edge base pair between G4 in the *syn* orientation and U1[40] (see Figure 32), complemented by a C3−G4 stacking interaction and a U2−C3 type 7 base−phosphate[50] (7BPh) interaction.[633,660] GNRA TLs are characterized by a *trans*-sugar/Hoogsteen (tSH) G1/A4 base pair, an N2−R3−A4 stacking interaction, and a G1(2′-OH)···R3(N7) sugar−base hydrogen bond.[661] They adopt structures featuring a sharp turn between G1 and N2 that is reminiscent of the U-turn motif. GNRA and UNCG sequences generally adopt distinct turn structures, and

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

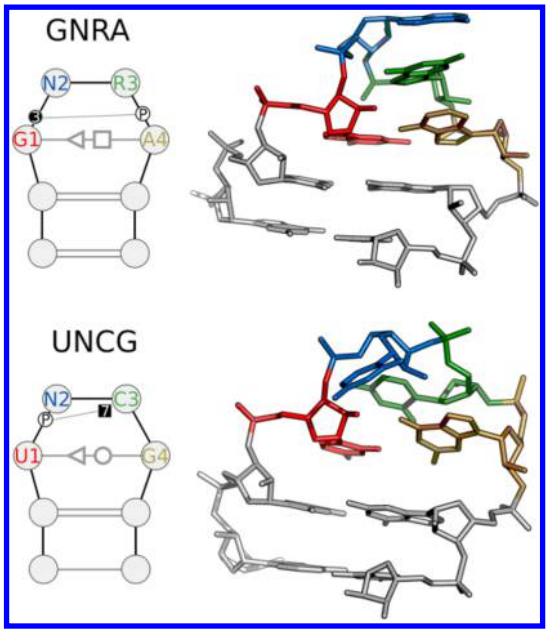


Figure 32. Annotated 2D (left) and 3D structures (right) of two representative RNA hairpins capped with GNRA (top) and UNCG (bottom) TLs. Their molecular interactions are annotated according to the Leontis/Westhof/Zirbel nomenclature.[39,40,50]

the turn in the latter family was named the Z-turn because of its similarity to Z-DNA.[662] GNRA TLs are generally considered more flexible than UNCG TLs, and often participate in tertiary contacts with other RNA elements.[663]

The feature that makes GNRA and UNCG TLs attractive targets for computational study is their exceptional thermodynamic stability,[645] together with the fact that they are well-structured and capable of autonomous (context-independent) folding.[67,624] However, like all RNAs, they are difficult to simulate because they have complex and rough free-energy landscapes that make simulations very susceptible to becoming kinetically trapped in local minima. TLs thus represent the smallest nontrivial model systems that exhibit every type of interaction and structural feature that is generally important in the structural dynamics of RNA, including stacking, canonical and noncanonical base pairing, base−phosphate interactions, canonical and noncanonical sugar−phosphate backbone conformations, and *syn*/*anti* nucleobase orientations. This makes these systems rather challenging to describe well with force fields. Below we present a comprehensive review of MD simulation studies on RNA TLs published between 1995 (when reasonable methods for accurately describing long-range electrostatic interactions were introduced) and 2017, including the latest applications of enhanced sampling methods and microsecond- to millisecond-long simulations. The history of TL simulations provides insight into the progress of MD methods over the last 20 years.

4.2.1. Structure and Dynamics of RNA Tetraloops. MD simulations of TLs have been surveyed with varying degrees of comprehensiveness in earlier reviews.[93,154,176,664−666] Simulations performed toward the end of the previous century primarily focused on testing the accuracy of force fields on picosecond−nanosecond time scales by evaluating structural stability of TLs in the vicinity of their native folds. Other subtle effects, such as the impact of the orientation of ribose O2′H groups on the TL structure, specific ion interactions, ionic strength, and hydration, were also explored.

TLs are highly stable in solution, and short MD simulations are therefore expected not to disrupt their experimentally derived native structures. However, TLs are not rigid:[667] they may sample alternative substates[668,669] (called minor conformations or excited conformational states) with low populations (<10%), and their dynamics may involve conformational transitions occurring on the microsecond−millisecond time scales (e.g., disruption of base pairs, base flipping, unstacking, and sugar repuckering).[668,669] The identification of these minor conformations can be essential for understanding how the internal dynamics of an RNA molecule relate to its function. However, unambiguous experimental characterization of sparsely populated RNA conformational states is very difficult, and their occurrence even for small systems such as TLs remains somewhat controversial. Given the long time scales associated with intersubstate transitions, long MD simulations would be needed to provide a complete structural description of the conformational states of TLs and their relative probabilities. It is always advisable to use the longest simulation times that are affordable with the available computer hardware, although even pushing the limits in this way will generally not be sufficient to achieve convergence in simulations of TLs. Moreover, force-field inaccuracies can wildly distort the folding landscape.

Another way to identify putative minor conformations of RNA TLs is to consider the entire collection of available experimental structures. For example, the PDB database contains several examples of GNRA and UNCG sequences in conformations that differ substantially from those associated with their canonical fold.[242,662,670] These alternative conformations are usually stabilized by contacts with other RNAs as well as proteins and ions, but they may provide useful information on the intrinsic internal dynamics (flexibility) of TLs. It is reasonable to assume that while the local context may remodel some RNA sequences, this remodeling will proceed via conformations that are energetically easy to access (see section 4.3.5).

It should be emphasized that a prerequisite for any computational method to properly describe the rich ensemble of minor conformations is the ability to correctly capture the native ground-state fold and all of the signature interactions that dictate the consensus sequence associated with that native fold. Without this ability, any theoretically observed minor conformations or folding events may be spurious.[154] TLs clearly display structural conservation across a broad set of structures determined by X-ray crystallography and NMR spectroscopy, so despite the potential (and likely) occurrence of minor conformations, they have clearly defined dominant native folds.

Fluorescence spectroscopy studies have suggested that these TLs may have some inherent flexibility even in isolation. However, most of the suggested alternative conformations, such as hairpins exhibiting terminal base-pair fraying or GNRA TLs with unstacked second nucleotide of the loop, retain the motif's signature interactions.[640,642,671] The exception is a "5′-stacked" conformation in which the first three bases on the 5′-end of the stem are stacked instead of the last three bases on the 3′-end as in the native pattern.[640,642] However, the 5′-stacked conformation appears to be absent or rare in the structural database. Its rarity is unlikely to be due to the effects of crystal packing because the native GNRA conformation has been observed in many different X-ray crystal structures and a

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Table 3. Key Simulation Studies of RNA TLs, Either in Isolation or as Part of a Larger System

year	ref	loop	force field[a]	solvent	initial native	sampling
1995	Zichi[650]	GCAA	OPLS	explicit	Y	MD
1995	Cheatham[675]	UUCG	ff94	explicit	Y	MD
1997	Miller[677]	UUCG	ff94	explicit	Y	MD
1999−2000	Willams[678−680]	UUCG	ff94	implicit	Y	MD
2002	Sorin[228]	GCAA	ff94	implicit	Y	MD
2003	Sorin[651]	GCAA	ff94	implicit	Y	folding@home
2003	Sarzynska[366]	GCAA	CHARMM27	explicit	Y	MD
2004	Cheng[275]	UUCG	ff94	implicit	Y	LES
2005	Koplin[277]	UUUU	ff98	explicit	Y	MD, LES
2005	Sorin[652]	GCAA	ff94	explicit	Y/N	folding@home
2006	Spackova[530]	GAGA	ff94	explicit	Y	MD
2006	Villa[688]	UUCG	ff98	explicit	Y	MD
2007	Zhuang[234]	UGGAAC (T-loop)	ff94	explicit	Y	T-REMD
2007	Deng[235]	UUUU	CHARMM27	both	Y	MD, T-REMD
2008	Ferner[689]	UUCG, CACG	ff98	explicit	Y	MD
2008	Villa[236]	UUCG, CACG	ff98	explicit	Y	T-REMD
2008	Bowman[690]	GCAA	ff94	explicit	Y	SREMD
2008	Garcia[237]	UUCG	ff99	explicit	N	T-REMD
2009	Zhang[243]	GCAA	ff98	explicit	Y	T-REMD
2010	Deng[159]	UUCG, UUUU	ff99; CHARMM27	explicit	Y	umbrella sampling
2010	Zuo[244]	UUCG	ff99	explicit	Y	T-REMD
2010	Banas[133]	UUCG, GAAA, GAGA	AMBER several variants including new χ_{OL3}, CHARMM27	explicit	Y	MD
2010	De Paul[221]	GCAA	ff94	explicit	Y	folding@home
2013	Chen[120]	UUCG, GCAA, CUUG	Chen−Garcia	explicit	N	T-REMD
2013	Krepl[507]	UCCG, GCCA	χ_{OL3}	explicit	Y	MD, bioinformatics
2013	Kuhrova[240]	UUCG, GAGA	χ_{OL3}	explicit	N	T-REMD
2014	Chakraborty[691]	UUCG, GCAA	bsc0	implicit	N	discrete path sampling
2015	Giambasu[692]	UUCG	ff99; χ_{OL3}	explicit	Y	MD
2015	Haldar[313]	UUCG, GAGA	χ_{OL3}	explicit	Y	MD, WT-metadynamics
2015	Bergonzo[139]	UUCG	AMBER several variants	explicit	Y	M-REMD
2016	Kuhrova[129]	GAGA	AMBER several variants	explicit	Y/N	T-REMD, WT-metadynamics, solute tempering
2016	Miner[241]	GAAA	Chen−Garcia	explicit	N	T-REMD
2016	Bottaro[163]	GAGA, UUCG	χ_{OL3}	explicit	N	WT-metadynamics, T-REMD

[a] χ_{OL3} is always used as ff99bsc0χ_{OL3}.

wide range of structural contexts; the effects of crystal packing in at least some of these structures are certainly negligible. Consequently, the significance (population) of the 5′-stacked GNRA TL (if existing) is difficult to estimate.

In summary, despite the existence of alternative structures, the dominant native TL conformations revealed by structural data are vital references and benchmarks for simulations of isolated TLs. Systematic analyses of X-ray structures have clearly shown that TLs (and especially GNRA TLs) quite frequently adopt some other alternative conformations.[670] However, the formation of these alternative structures may be related to specific contexts, interactions with other RNA molecules, proteins, or crystal contacts. This suggests that these states should be less intrinsically stable than the native state and so would not be adopted by isolated TLs.

4.2.2. MD Simulations of RNA Tetraloops. To evaluate a force-field performance and/or the extent to which a simulated ensemble is representative of the full conformational space, simulations have been compared to experimental data obtained by X-ray crystallography, solution-phase NMR, and optical spectroscopy. As will be shown in the following subsections, agreement between MD simulations and experiments depends strongly on the length of the simulations and the choice of force field. For example, NOE distances and scalar couplings, which are primarily used for solution structure determination, typically agree well with nanosecond-long MD simulations initialized from experimental structures. Transitions to potentially artificial parts of the ensemble may only occur after hundreds of nanoseconds or a few microseconds. However, once simulations reach these minima (alternative structures, often spuriously stabilized by the force field), the agreement with experiment can deteriorate considerably. One should also keep in mind a fundamental difference between computer simulations and experiments: most experiments provide information on ensemble averages, but simulations usually track one molecule over time. Consequently, simulations are more like single-molecule experiments than conventional ensemble experiments.[176] Table 3 summarizes the most important features of the discussed TL MD simulation studies.

4.2.2.1. MD Simulations of Tetraloops Conducted between 1995 and 2000. Most reported MD simulations of

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

TLs have used an experimentally observed native structure determined by NMR or X-ray crystallography as their starting point. The very first MD simulation of an RNA TL was reported in 1995 by Zichi et al.[650] Simulations of a GCAA hairpin were initialized from the NMR structure[661] in explicit water[575] and ions.[572] The authors performed a ∼200 ps-long simulation that required over a month of wall-clock time, corresponding to approximately 35 s per MD step. Electrostatics was modeled using Ewald summations,[672] which was crucial for maintaining the stability of simulations of highly charged polymers such as RNA systems. Simulations of RNA hairpins that used truncation methods (the standard way of treating long-range electrostatics at the time) were much less stable.[673,674] The reported trajectory showed that the system was structurally stable, and the NOE distances computed for the simulated structure were consistent with the available NMR data.[661] Note that the starting NMR structures[661] were determined using a simpler force field combined with NOE distance restraints, whereas the chemical model used by Zichi et al. was much more detailed and included explicit water. This more sophisticated model made the NMR-derived structure stable in simulations even without restraints, at least over the short simulated time scale. In retrospect, the results obtained by Zichi et al. can be considered trivial and entirely expected given the short simulation time scale. However, this paper provided early evidence that even the potential energy functions available in the 1990s could preserve the correct native structure of a TL on time scales of hundreds of picoseconds, indicating that the native structure corresponded to at least a local minimum on the force-field free-energy surface. At the time, the main outstanding problem with classical MD simulations was believed to be the treatment of long-range electrostatic forces.

In the same year, Cheatham et al. reported one of the first nanosecond-time scale nucleic acid simulations using the particle-mesh Ewald (PME) method,[579,675] and obtained stable trajectories for several systems (including a UUCG TL) using the ff94 force field (see section 3.1.2) in explicit water.[80,576] PME revolutionized the field of RNA simulations by speeding up simulations while providing accuracy equivalent to that achieved by the Ewald summation method. It has since become a standard technique that is routinely applied in simulations of biomolecular systems in explicit solvent and using periodic simulation boxes. Therefore, 1995 can be regarded as the year in which the modern approach to MD simulation of nucleic acids was born.

At the time, two alternative structures for the UUCG TL had been suggested on the basis of NMR experiments,[634,676] which differed in the details of the U1/G4 base pairing (see Figure 33). An early structure[634] was subsequently replaced by a more accurate model.[676] Attempts to simulate a spontaneous transition from the incorrect conformation to one with the correct hydrogen-bond pattern were made to test the reliability of the available force fields and to evaluate different sampling methods.

In 1997, Miller and Kollman[677] performed explicit solvent MD simulations of a TL but were unable to explore the transition from the incorrect TL conformation to the correct one within 2.5 ns. The authors had to replace the native riboses with deoxyriboses to facilitate a spontaneous transition between the two structures. These results highlighted the sampling limitations of MD and suggested that the optimism expressed in the earliest publications utilizing PME may have been somewhat misplaced.

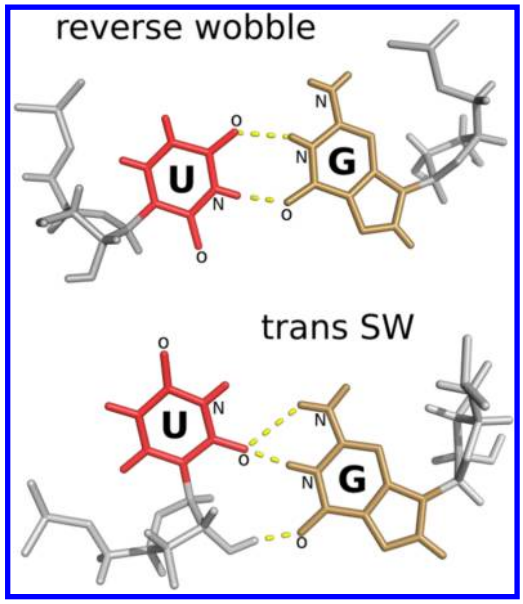


Figure 33. Correct and incorrect structures of the U/G base pair in the UNCG TLs. The incorrect structure has a reverse-wobble base pair, whereas the correct loop conformation has a bifurcated hydrogen bond in which one of the uracil carbonyl oxygens interacts with the imino and amino groups of the guanine.

One year later, Kollman et al.[274] used the locally enhanced sampling (LES, see section 3.2.4) method to explore the same transition in UUCG. Using five copies of the entire loop, an incorrect-to-correct structure transition was observed after ∼200 ps. On the other hand, when the simulation was started from the correct structure, no significant change was observed for ∼750 ps. These results demonstrated the usefulness of enhanced sampling techniques and the accuracy of the available force field on short time scales in the vicinity of the native fold.

Another way to substantially accelerate sampling is to replace explicit water molecules with a continuum model (see, however, section 3.2.8 for a cautionary comment). The group of Kathleen Hall published several joint computational and experimental papers that used stochastic dynamics with implicit solvent to describe the UUCG TL.[678−680] The authors compared various implicit solvation models and found that the best description of RNA molecules was obtained with the generalized Born/surface area (GBSA)[681] model. Their simulations were shorter (∼1.2 ns) than those of Kollman,[677] but were approximately 5 times faster. Transitions from the incorrect to correct UUCG structures were observed. Interestingly, LES and implicit solvation can also be combined in a cooperative manner.[275]

Observing the transition from the incorrect to the correct conformation is a useful test, but at that time the simulations were too short to quantitatively estimate the relative stabilities of these different states. One way to estimate free energies for very different structures is to perform MD simulations in explicit water and then average the molecular mechanical energies and continuum-calculated solvation free energies of representative solute configurations sampled in the simulation.[682−684] This popular technique is known as MM-PBSA (see section 3.2.8).[371] Srinivasan et al.[685] used MM-PBSA to compare the free energies of the two alternative UUCG conformations as well as two different conformations of the UGAA hairpin. In both cases, the correct structures had lower free energies than the incorrect ones. Specifically, for the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

UUCG TL, the difference in free energy between the correct and incorrect conformations was estimated to be −2.7 kcal/mol. The same work also correctly predicted the relative stability of the r(GGAC-UUCG-GUCC) dimer and hairpin as a function of the salt concentration, which was a more ambitious application.

During the 1995−2000 period, alchemical free-energy simulations using the free-energy perturbation (FEP) method[359] were relatively new (see section 3.2.7). These techniques were used to characterize free-energy changes caused by replacing standard nucleobases with modified ones in RNA hairpins. Alchemical free-energy methods should in principle be more accurate than continuum solvent free-energy methods, although their accuracy is typically limited to only very small structural or chemical changes.[686] For instance, one group studied the C3U substitution in the UUCG TL,[687] and another used FEP to quantify the effect of G1 to inosine substitutions in the GCAA TL,[366] showing that this mutation significantly destabilizes the loop. Because of the studied systems' construction, the initial and final states of all of these calculations[89,687] were assumed to be very similar and easy to sample over a short time scale. This was probably a reasonable assumption, supported by the good agreement of the results with experimental data.

4.2.2.2. MD Simulations of Tetraloops from 2000 to the Present. In the early 2000s, researchers started to explore the rugged free-energy landscape of TLs, expanding simulations beyond local subnanosecond motions around the native minimum. Pande et al.[228] investigated dynamics of the GCAA TL by performing multiple 10 ns MD simulations with the GBSA implicit solvent model.[681] These calculations provided insight into conformational changes within the loop associated with unstacking and flipping of residues C2 and A4, which correlated with sugar pucker dynamics. The unstacked C2 conformation was consistent with NMR measurements,[667] and unstacking of C2 was also later detected in analysis of X-ray structures.[242] However, the loop distortion created by flipping out of the A4 residue, although very infrequent, deviated from the native NMR ensemble and was not compatible with the experimental data.[640] Observation of this relatively large conformational transition on such a short time scale could have been an artifact of the chosen force field and solvation model.[80,681] Interestingly, this distortion is also observed in the crystal structures of some GNRA tetraloops bound to toxins,[622] suggesting that it is an accessible higher free-energy substate. Ribonuclease toxins may exploit the loop's partial unfolding upon binding to facilitate RNA cleavage.

In 2005, the groups of Stock and Schwalbe reported a very thorough comparison between computational and NMR studies of the UUUU TL using the ff94 AMBER force field.[277] Several 10 ns simulations in explicit solvent were performed, as well as high-temperature MD and LES MD. None of the simulations yielded converged data, and some of the standard MD results had to be discarded due to base pair disruptions that occurred after a few nanoseconds. Notably, the analysis of the backbone dihedral angle distributions with respect to experimental ^3J couplings and dipole−dipole cross-correlated relaxation rates showed that the conformational space sampled by the ε and ζ angles of some loop residues was inconsistent with the experimental averages. This discrepancy was not alleviated by the increased sampling of high-temperature and LES simulations. Interestingly, the enhanced sampling and standard MD simulations differed markedly in terms of their phosphodiester angle distributions; the fluctuations of the corresponding dihedrals were considerably more pronounced in the enhanced sampling simulations. Indeed, the flawed behavior of the ε and ζ dihedrals of the original AMBER force field[80] was later identified as one of the potential sources for the stabilization of spurious intercalated structures in simulations of single-stranded RNA tetranucleotides (see section 4.1).[143,614] On the other hand, the MD ensemble was in good agreement with experimental NOE distances. Finally, the authors[277] found no correlation between the motions of the sugar rings and the stacking dynamics of the loop bases, which was partly inconsistent with the results of Pande et al.[228]

Later, Villa and Stock[688] reported internal dynamics of a 14-mer UUCG hairpin in explicit solvent on a tens-of-nanoseconds time scale, and compared the results to ^{13}C NMR spin relaxation experiments.[693] They observed a good agreement between calculated and experimental relaxation rates, steady-state NOEs, and S^2 order parameters, suggesting that AMBER ff94[80] force field correctly reproduced the studied system's dynamics on the subnanosecond time scale. Interestingly, the TIP3P[576] water model's high diffusion constant strongly affected the estimated rotational correlation time τ_C, which is proportional to the inverse of the RNA hairpin rotation diffusion constant. This forced the authors to use the experimental τ_C value. However, the internal correlation times of the bond dipoles were calculated using results from the simulation. It is to be noted that these experiments were not sensitive to motions occurring at time scales much slower than the overall molecular tumbling time (∼2 ns). Consequently, any internal motions occurring on longer time scales, such as large amplitude collective motions, would not be reflected in the experimental data.[668]

In a later work,[689] the Stock and Schwalbe groups reported a combined ^{13}C NMR spin relaxation and MD study on the CACG and UUCG TLs at various submelting temperatures. These TLs have similar native conformations but differ in their thermal stability, having T_m values of ∼300 and ∼350 K, respectively. The only structural element that exhibited significant dynamics in both hairpins was the flipped out A2/U2 nucleobase. It should be noted that the simulations were not performed at the same temperature as the experiments. The justification for this choice was presented in an earlier work conducted by the same group, in which T-REMD simulations of the same systems were used to calculate their melting temperatures.[236] The force field appeared to overestimate the melting temperatures for both TLs. Despite this, a comparison of the measured and calculated S^2 order parameters suggested a convincing overall agreement between experiment and theory. Because the MD simulations were run for about 50 ns, they revealed some unexpected conformational transitions that occurred on a nanosecond time scale, such as *syn*/*anti* transitions and losses of stacking interactions, which could not be detected by the NMR data. This was an early indication that longer simulations might reveal new problems.

A related work compared NMR and MD, with 100 ns MD simulation using the CHARMM36 force field[98] characterizing the internal motions of the apical RNA stem of the human hepatitis B virus.[694] Again, the NMR relaxation rates calculated directly from the trajectory agreed well with experiments.

Although the works cited above achieved good agreement with experiment, their results should be considered in light of

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

later efforts that have highlighted the force-field problems exposed by long TL simulations (see section 4.2.2.3).

In 2007, Deng and Cieplak[235] estimated the relative stability of single-stranded, duplex, and hairpin conformations of the UUUU TL monomer and dimer as a function of the temperature. They used the MM-PBSA method to process the MD trajectories.[371] The resulting free-energy profiles supported a shift in the conformational equilibrium from duplex to hairpin as the temperature was increased, in agreement with NMR results.[695] ΔH values calculated at different temperatures for the hairpin to single-stranded chain transition were qualitatively consistent with the experimental values. However, it is not clear whether the reported ~100 ns simulations were sufficiently long to correctly sample the single-stranded RNA's conformational space.

The same authors also reported an enhanced sampling simulation of UUCG and UUUU TLs[159] in which the potential of mean force (PMF) as a function of the end-to-end distance was determined by umbrella sampling (US; see section 3.2.5) in explicit TIP3P[576] water. Results obtained with both ff94[80] and CHARMM27[89] were reported. The authors compared relative stabilities of the native structure (which has a low end-to-end distance) and the unfolded structure (which has a high end-to-end distance). The AMBER force-field simulations predicted a native state stability consistent with experiment,[641] whereas CHARMM27 simulations clearly predicted one of the loops to be unstable. Note that the results for both loops may have been biased by the fact that the simulations were initialized by steering in one direction only and starting from the native structure (see section 3.2.5). This procedure is expected to overestimate the stability of the folded state. Hysteresis effects are common in US but not always taken into consideration.[283]

The first comprehensive comparison of the most common force fields for TL simulations was presented in 2010 by Banas et al.[133] The authors performed MD simulations of UUCG, GAGA, and GAAA TLs on the submicrosecond time scale using several AMBER force-field variants with different dihedral corrections[94,95,136,696] as well as CHARMM27.[89] None of the studied TLs was described entirely satisfactorily, and at least some signature interactions were lost in all cases. Note that, unlike some other studies, this work monitored all of the signature features of the studied TLs. Importantly, the TLs were found to be stable on submicrosecond time scales with the χ_{OL3} RNA force field,[94] which was used for the first time in this work (see section 3.1.2.1). Without this critical correction of the RNA dihedral potential, the GNRA TL degraded into ladder-like structures (see Figure 7). In addition to eliminating ladder-like artifacts, the χ_{OL3} parameters greatly improved the description of the UUCG TL because the modification of the *syn* region stabilized the signature U/G(*syn*) base pair. The CHARMM27 simulations predicted an unrealistic progressive fraying of the stem due to the disruption of canonical base pairing. A-RNA melting is a common problem in CHARMM RNA simulations, and may still be problematic with the CHARMM36 reparameterization (section 3.1.3).[134,137]

Despite the growing amount of evidence highlighting the deficiencies of RNA force fields (see section 3.1), two publications reported computations that successfully recapitulated experimental results for TLs. Chakraborty et al.[691] studied folding kinetics of UUCG and GCAA hairpins in implicit solvent and characterized their kinetics by analyzing preliminary 200 ns-long plain MD simulation using the discrete path sampling method, which is a general non-MD method for computing transition rates between metastable states.[697] The MD simulations were only used to seed the subsequent discrete path sampling mapping of the conformational space. It should be noted that because of the implicit solvent approximation, the local minima identified in this work can be regarded as free-energy minima for which the solvent degrees of freedom have been averaged out (see section 3.2.8). The authors reported results in agreement with experimental kinetic measurements, suggesting that the process of folding from relatively compact states was characterized by short paths and occurs on microsecond time scales. Conversely, extended states (which were only sparsely populated at 298 K) reached the native state via much longer routes, and were associated with lower folding rate constants. Obviously, systematic conformational searches in implicit solvent have their own limitations, and there is currently limited experience with their application to nucleic acids.

Giambasu et al.[692] reported an extensive comparison to NMR data for the UUCG RNA hairpin based on a high-resolution solution structure.[698] Multiple 300 ns simulations on a UUCG TL construct initialized in the native state were analyzed using both ff99[90] and χ_{OL3}[94] force fields. Production runs were performed in the constant energy ensemble to minimize the thermostat's effects on the system's dynamics. The work found good agreement with NMR data, including NOEs, order parameters, and residual dipolar couplings (RDCs). RDCs report on the relative orientation of chemical bonds at distant positions, thus providing long distance information. The experimentally observed NMR structure and fluctuations were accurately reproduced by the simulations, suggesting that the experimental ensemble was well represented by the free-energy basin sampled in the simulations. The χ_{OL3} force field offered better accuracy than the ff99 variant. However, exploration of the conformational space was still limited to the region close to the native structure, so these simulations cannot be regarded as a test of global force-field accuracy.

Besides modeling of true RNA hairpins, MD simulations of TL-like single-strand overhangs of A-RNA stems were used to explain a series of experiments in which slow spontaneous ligation and cleavage reactions were detected in complementary oligoC/oligoG RNA constructs.[699] These results may be essential for understanding the transition between organic molecules and biopolymers leading to the emergence of life on the Earth (see section 4.8.8).

4.2.2.3. Folding Simulations of RNA Tetraloops. All works discussed above presented trajectories that were initialized in the native state. This induces a strong initialization bias (see section 3.2.1) toward the known experimental structure. Ideally, MD simulations should be able to predict the native structure from an unfolded one. One possible approach to structure prediction is to use fast modeling tools[700] to produce putative structures and then use a force field for structural refinement, as was done by Maier et al.[701] These authors generated structures for the GGGCGNRAGCCU sequence (R = A/G, N = any) using MC-SYM,[467] a structural bioinformatics tool, and then minimized these structures using an implicit solvent model. However, energy minimization alone cannot generally be expected to find structures that differ substantially from the starting structure, so the accuracy of predictions based on this method is very sensitive to the accuracy of the modeling tool used in the first step, and we will not address such procedures further. The only way to search for the global free-energy

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

minimum of an RNA TL using MD is to perform extremely long MD simulations or to use enhanced sampling methods (see section 3.2).

In 2003, the Pande group[651] reported a study involving a large number of independent simulations of a GCAA TL performed with a distributed computer framework (folding@home)[74,75] and multiplexed replica-exchange MD method, a variant of the T-REMD technique (see section 3.2.4).[702] They collected over 500 μs of total simulation time using multiple simulation methodologies (with the individual simulations starting from both native and extended states), representing a significant advance in the field. For comparative purposes, the year before, the same group published multiple MD simulations of the same system with a total time of 0.1 μs.[228] Distributed computing using ~40 000 processors via folding@home made it possible for the first time to compare experimental data to an MD ensemble beyond the native minima and thus to test the force field's predictive power by starting simulations from the denatured state at room temperature. This was also the first application of a methodology based on the parallel tempering (T-REMD) strategy to RNA folding (see section 3.2.4). Simulations that began close to the native structure achieved good agreement with experimental NOE distances. However, despite the impressive overall simulation length, the ensemble was not converged, and it was thus not clear whether the ff94 energy model was capable of correctly predicting the experimental structure. Relaxation times estimated for the collapse of the hairpin into a native-like conformation from the unfolded states were consistent with the kinetics observed experimentally for DNA hairpin folding (1−10 μs). Two competing folding/unfolding mechanisms were observed, but these results should be treated with caution because of the known inaccuracies of the force field (see section 3.1.2.1) and mainly the use of an implicit GB solvent model used in the study.

Two years later, the same group studied the role of explicit water and ions on the folding mechanism of GCAA TL.[652] The use of explicit solvation yielded a more varied and complex folding mechanism than was previously reported.[651] This work suggested that explicit water plays an important structural role in RNA folding that cannot be easily captured by implicit solvent models. For this reason, explicit solvation has been preferred to implicit solvent models in most computational studies.

The folding@home infrastructure[74] was also used by Bowman et al.,[690] who reported a structural analysis of folding intermediates for the GCAA TL. Here, a total of 54 μs of MD was performed using a replica-exchange-like scheme in explicit water. Statistical convergence was not achieved, but the authors did observe some folding and unfolding events. However, the TL structure was not discussed in detail, and the detected folding events may actually have formed non-native TL structures. Yet another study using the folding@home massively parallel MD infrastructure[74] examined the equilibrium conformational dynamics of the GCAA TL,[221] with total simulated time of more than 100 μs. Despite this, the individual trajectories were probably too short to reveal the problems of the available ff94 force field. In fact, Figure 3 from ref 221 clearly shows that the results were not converged and that the population of the native cluster decreased over the full simulation time scale, suggesting the native state would vanish if the simulations were extended for long enough.

Several other works have reported nonconverged replica exchange MD simulations starting from the native TL structure. For instance, various groups have performed T-REMD simulations starting from the native structures of several sequences.[234,236,243,244] While these simulations could provide insights into the large-scale dynamics around the native state and the thermostability of the studied molecules, they cannot be used to evaluate a force field's ability to predict known experimental structures. In fact, the published data (where available) typically indicate that the force fields tend to lose at least some of the studied system's native interactions.

Perhaps the first report of predictive folding of an RNA TL was presented by Garcia and Paschek.[237] These authors used the AMBER ff99 force field[90] with explicit water[576] in conjunction with T-REMD based on 52 replicas at temperatures between 270 and 600 K. The simulations were initialized in the unfolded state, with no bias toward the native structure. The authors reported the folded state to be more stable than the unfolded state by 2.4 kJ/mol. However, some of the structures regarded as "folded" deviated markedly from the native structure, having RMSD values of up to 5−6 Å. While it is possible that the reported simulations were statistically converged, they did not reproduce the stability observed experimentally, again probably due to problems of the ff99 force field that was later shown to be inadequate for TLs.[94] Figures included in the paper clearly show that the TL itself was not structured, and the stem appeared to adopt a spurious ladder-like conformation.[133] A later publication by the same group discussed the limitations of this attempt in detail.[120]

A subsequent attempt at folding from the unfolded state was made by Kuhrova et al. using extensive T-REMD simulations of UUCG and GAGA TLs.[240] The χ_{OL3} force field[94,95] was used, and 48 replica simulations of 400 ns each were initiated from the unfolded state. Essentially only non-native states were populated. Although transient native-like structures were found, the results suggested that even the latest force field could not reproduce the correct experimental structures. Crucially, the authors used very strict criteria for identifying the folded structure, requiring all of the signature interactions to be present at the same time for the TL to be considered folded. This definition is consistent with that preferred by the RNA structure prediction community,[52] and must be used to avoid including false positive folding events. Inadequacy of the RMSD metric for identifying native-like structures, especially when using loose thresholds of the order of 4 Å or more, was clearly discussed.[240] For the GAGA TL, one complete folding event was observed. The UUCG TL was even more difficult to fold due to the G in the loop, which should flip from *anti* (favored in single-stranded RNA) to *syn* (favored in the TL). The UUCG structure sampled during the MD that was closest to the experimental one had the G in the *anti* state. The authors suggested, based on a thermodynamic integration calculation, that the force field would have correctly predicted the *syn* conformation to be more stable for the TL, but the transition was not observed, probably because of the still short time scale of the T-REMD run. The same group subsequently reported unbiased 10-μs-long MD simulations of the same two TLs using the χ_{OL3} force field and WT-metadynamics.[313] That work again confirmed that the native structure was a local free-energy minimum of the χ_{OL3} force field, achieving stable dynamics on time scales of a few microseconds. However, the metadynamics simulations confirmed that the native structure does not

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

correspond to the global minimum in terms of the force-field free energy.

In an attempt to resolve the TL folding problem, Chen and Garcia modified the vdW term of the force field (see section 3.1.2.4).[120] Because base stacking propensity was reported to be overestimated by the original force field,[111,124] they revised Lennard-Jones nonbonded parameters. They attempted to fold three TLs using the revised Lennard-Jones parameters for the nucleobases together with a new reparameterization of the glycosidic torsion, providing total dihedral energy profiles similar to χ_{OL3} (Figure 9). The bsc0 correction for the backbone was omitted in this work. Perhaps most importantly, to allow separate tuning of the RNA−water and RNA−RNA interactions, the suggested modifications included a breakdown of the standard combination rules for vdW parameters adopted in the AMBER force field. This approach is known as the NBfix method and effectively increases the number of tunable parameters (see section 3.1.2.4). Without this, the new vdW parameters might perform worse than the original ones. Promising results were obtained: by simulating 64 replicas from the unfolded state for 200−400 ns each, the authors reported folding events for two of the three investigated TLs. However, the agreement between the reported ensembles and the experimental structures was still not completely satisfactory (Figure 34). The fraction of folded structures was defined using a quite generous threshold for agreement with experiment (4 Å RMSD), which is not sufficient to unambiguously identify folding, as shown subsequently,[223] and could even include ladder-like artifacts.[240] From cluster analysis, the cluster with the highest relative population was within 1 Å RMSD of the native structure. However, given the reported populations of structures with low RMSD separations from the native conformation, it seems that the absolute population of this cluster was small. Nevertheless, these results suggested that correcting nonbonded force-field parameters (most likely including NBfix, see section 3.1.2.4) may be essential for prediction of RNA structures by MD.

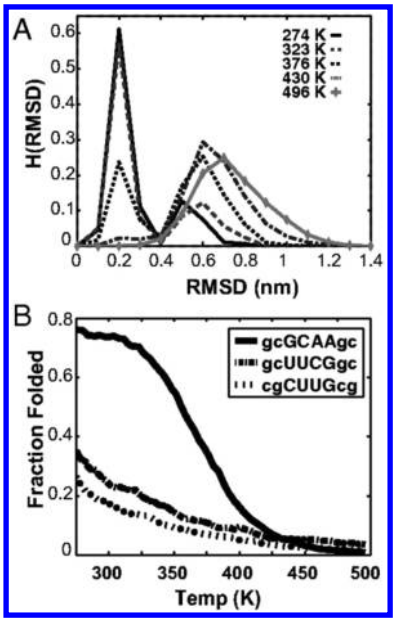


Figure 34. (A) Histogram of RMSD for GCAA TL with respect to the native structure at different temperatures as obtained from T-REMD simulations from unfolded state with the Chen and Garcia force field. A clear bimodal distribution distinguishes low RMSD (putatively folded) and high RMSD (putatively unfolded) structures. (B) The fraction of folded structures, defined as those with RMSD less than 4 Å from the native structure, for three TLs as a function of the temperature. The GCAA TL shows a larger fraction of the lower-RMSD structures as compared to the other two TLs. Adapted from ref 120. Copyright 2013 National Academy of Sciences.

Cheatham's group subsequently reported an exhaustive benchmark for tetranucleotides and TLs using the most relevant RNA force fields,[139] including that developed by Chen and Garcia[120] and χ_{OL3}.[94,240] The alternative vdW parameters for the phosphate group suggested by Case and co-workers (CP)[150] were also tested. The stability of the UUCG TL was evaluated by performing multidimensional replica exchange MD simulations (see section 3.2.4) with hundreds of replicas lasting for multiple μs each.[139] The authors claimed convergence for the sampling of the TL structure when full pairing was enforced in the base pair stem below the TL. Because of this enforcement of stem base pairing, the simulation still cannot be regarded as an example of predictive folding. In addition, the convergence was reported to be convincing for the results obtained at 277 K but somewhat less convincing for those at 300 K.[139] This highlights the unprecedented scale of the sampling problem even for the folding of very short RNA sequences. The key finding of these impressive calculations was that none of the tested force fields could reproduce the correct experimental structure. Of the tested force fields, that of Chen and Garcia provided the highest populations of properly folded TL structures at ∼277 K: ∼10% using the signature interactions[133,240] criteria, and ∼35% using a 4 Å RMSD[120] threshold. However, at room temperature, the population of properly folded TL dropped to ∼1%. Perhaps, the native-like cluster is enthalpically stabilized, and other conformations with greater conformational entropy might occur at higher temperatures. When discussing these results, it is also important to remember that the simulations were started from the folded state. This, together with the restraints imposed to preserve pairing in the stem, may artificially facilitate the persistence of the population of folded TL structures at low temperatures. Further, different force fields might need different time scales to achieve convergence.

These studies were extended by Kuhrova et al.,[129] who performed extensive T-REMD, REST2, and WT-metadynamics (see sections 3.2.4 and 3.2.5) simulations of the GAGA TL, which is believed to be an easier folding target than the UNCG TL, using every relevant RNA force field available at the time. The results clearly showed that no force field could maintain the correct folded TL structures, as confirmed by different enhanced sampling simulations starting from unfolded state and folded state ensembles (Figure 35). The criterion used to detect correct formation of the native structure was based on the presence of all of the TL's signature interactions. While some folding events satisfying this criterion were observed in 2-μs T-REMD simulations, bootstrapping analysis revealed that none of the force fields achieved a folding of the TL with a statistically significant population at any temperature. That is, all of the error bars corresponding to confidence intervals constructed at the 5% significance level contained zero population. The study also analyzed the key force-field deficiencies that prevent successful folding of RNA TLs. Underestimation of the free energy of base pairing was suggested to be a fundamental problem. Subsequently, the authors proposed a gentle tunable local stabilizing potential (HBfix, see section 3.1.2.5) to support the native H-bonds.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

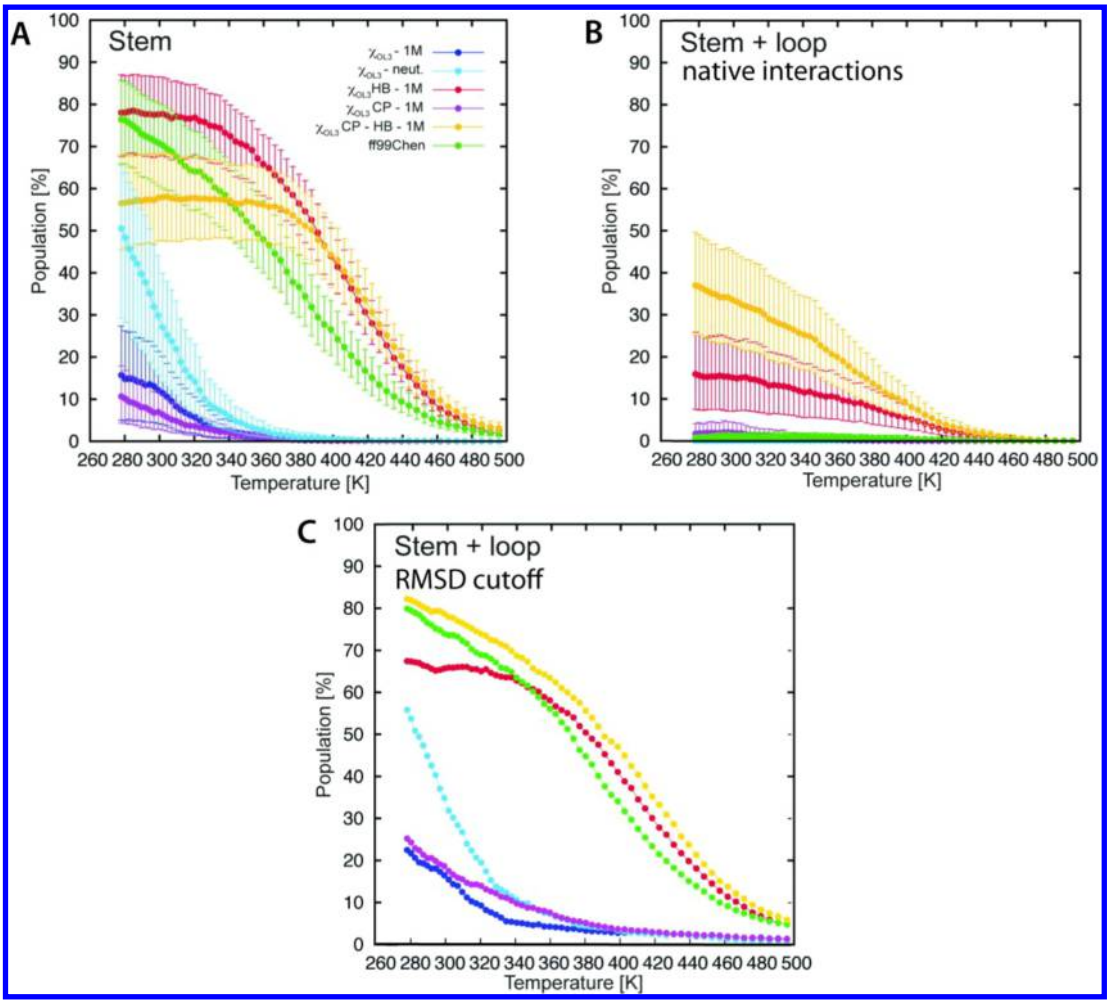


Figure 35. Folding of gcGAGAgc TL using extended T-REMD simulations from the unfolded state. The folding in this figure is assessed on the basis of strict structural criteria, thus avoiding the false-positive results that commonly occur when folding is judged on the basis of some RMSD metrics. Results are shown for a number of force fields: χ_{OL3} in net-neutralizing cation atmosphere (χ_{OL3} - neut.), χ_{OL3} in 1 M salt (χ_{OL3} - 1M), χ_{OL3} with hydrogen-bond fix (HBfix) in 1 M salt (χ_{OL3}HB - 1M), χ_{OL3} with modified phosphate parameters in 1 M salt (χ_{OL3}CP - 1M), χ_{OL3} combining HBfix with the modified phosphate parameters in 1 M salt (χ_{OL3}CP - HB - 1M), and the Chen and Garcia force field (ff99Chen). (A) Population of structures where the stem is correctly folded (base paired), but the TL is not folded. (B) Population of native structures where both the stem and the TL are correctly folded. This figure clearly shows that while the force fields can fold the stem, only the simulations with the structure-specific HBfix potential energy bias achieve statistically significant folding of the TL. (C) Population of folded state as suggested by RMSD metrics (using 4 Å cutoff to RMSD calculated over heavy atoms of stem and backbone heavy atoms of the loop); note that the RMSD metrics leads to a high portion of false-positive results (cf., parts B and C) dominating the populations of putatively folded states in (C), which is especially evident for the Chen and Garcia force field (the green curves). The error bars in (A) and (B) were calculated by bootstraping (using resampling of both time blocks and coordinate following replica; see ref 129). (A) and (B) adapted with permission from ref 129. Copyright 2016 American Chemical Society.

When using this bias potential for all base pairs and TL signature interactions, the T-REMD simulations finally achieved successful complete and statistically significant folding of the GAGA TL, with all signature interactions present. We note, however, that adding a native-centric potential significantly simplifies the problem and does not correspond to a predictive simulation. Despite its structure-specificity, the HBfix potential can be clearly physically justified (see section 3.1.2.5). The authors suggested that the use of such force-field corrections may be necessary in future folding simulations of RNA molecules due to the fundamental limitations arising from the simplicity of general force fields. Other problems with current force fields were also discussed, such as over-stabilization of base−phosphate and sugar−phosphate interactions.

Free-energy landscape of gcGCAAgc TL was characterized by performing massive T-REMD calculations (64 replicas, 3 μs per replica) using the Chen and Garcia force field.[241] The simulations revealed reversible folding/unfolding transitions of the TL into the canonical A-RNA conformation and the presence of two alternative configurations, including a left-handed Z-RNA conformation (the same cluster was reported in an independent work[129]) and a compact purine triplet. These structures were not directly compared to experimental data. The authors also reported that the stability of the folded structure at 310 K was approximately 10 kJ/mol. However, the folded structure was identified by dihedral principal component analysis, so it was not easy to determine whether the corresponding structures actually exhibited the signature interactions present in the native structures.

If the global free-energy minimum predicted by the force field corresponds to a structure that differs from the correct native structure, any sufficiently long simulation will depart from the native structure without any significant probability of

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

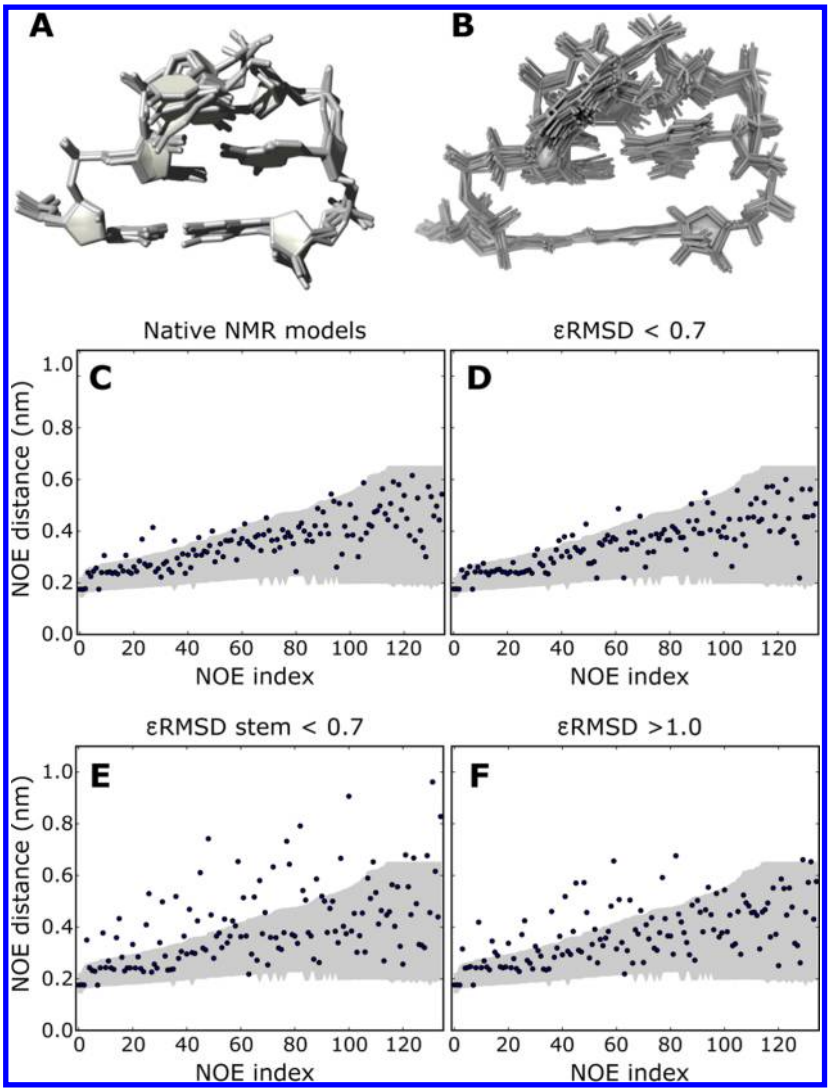


Figure 36. Structural analysis of UUCG tetraloop, panels (A) and (B). Only the tetraloop and one closing base pair are shown. Superposition of four X-ray structures (PDB codes: 1F7Y (tetraloop 1 and 2), 1I6U, 1FJG) (A). Superimposition of 20 solution structures from ref 698 (PDB code: 2KOC) (B). Comparison between experimental and calculated NOEs as reported in ref 163, panels (C), (D), (E), and (F). Blue points indicate the calculated average NOE distance, while the gray area indicates the experimental range. Four different data sets are considered: native NMR models (PDB code: 2KOC) (C); and selected snapshots from MD simulations where the whole structure is correctly folded (D), only the stem structure is correctly folded (E), or the whole structure is disrupted (F). Selection criteria are discussed in detail in ref 163 and are based on εRMSD. Whereas native NMR models and structures selected to be similar to the native one satisfy almost all of the NOEs, the other two data sets display a large number of violations. Remarkably, if all of the snapshots where the stem is correctly folded are included (E), the number of violations is larger than in the case where only incorrect structures are included (F). This indicates that the employed χ_{OL3} force field alone is not capable of reproducing a loop structure compatible with solution NMR data even if the stem is constrained. Panel (A) is adapted with permission from ref 698. Copyright 2009 Oxford University Press. Panels (C), (D), (E), and (F) are adapted with permission from ref 163. Copyright 2016 American Chemical Society.

visiting it again. In such cases, most of the simulation time will be wasted only to confirm the eventual loss of the structure. This problem can be alleviated by combining T-REMD with WT-metadynamics, especially when using a collective variable (CV) that properly distinguishes the native structure from the competing ones. This approach has been applied to the UUCG and GAGA TLs,[163] using εRMSD[223] as a biased CV. εRMSD measures the discrepancy between two structures by comparing the relative positions and orientations of nucleobases that are within a given distance cutoff. It can thus provide a continuous measure of the "discrepancy in the annotation" made using the Leontis–Westhof definition.[40] εRMSD-based biasing allows near-native structures to be visited even if they do not correspond to the force-field free-energy minimum, enabling estimation of their (in)stability (see section 3.2.5). This work confirmed the instability of both TLs under the current AMBER force field. However, the work also suggested that small empirically based changes in the dihedral potential parameters for α and ζ might reduce the problem, as these modifications improved the stability of GAGA TL. However, their general applicability to common RNA molecules remains to be determined. Furthermore, for the UUCG TL, the dihedral corrections did not significantly improve the results. In addition to analyzing the energetics of the folding landscape, the authors also compared their simulations to previously published NOE data.[667,698] This analysis confirmed that, at least for the UUCG TL, primary experimental data can only be reproduced using conformations with well-structured TLs (see Figure 36). Thus,

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

to be compatible with experimental data, MD simulations must predict dominant populations for structures with low εRMSD deviations from the native structure; that is, all of the TL signature interactions must be correctly formed.

4.2.2.4. Take-Home Message from Tetraloop Simulations. As described above, many published works have attempted to elucidate the structure and dynamics of GNRA, UNCG, and selected other TLs using MD simulations. Most of them used experimental structures as their starting points. These simulations were very informative in that they predicted the local dynamics of the TLs, sometimes achieving good agreement with experimental data. However, this was only true for simulations that were short enough to avoid leaving the vicinity of the initial native structure.

The availability of more powerful computers combined with massively parallel simulation methods and enhanced sampling techniques revealed several problems with modern force fields. In particular, enhanced sampling simulations can easily access structures that would only be observed in plain MD simulations that were several orders of magnitude longer. This is very useful for exposing spurious structures that are unrealistically stabilized by the force field. The reviewed works clearly show that current force fields cannot blindly predict the structure of RNA TLs. Because TLs are relatively small systems, which nevertheless contain all key interactions and energetic contributions that control RNA folding, they are important test systems for evaluating the accuracy of RNA force fields. Note that some claims in the literature about successful folding of TLs must be carefully interpreted; cf., discussion in refs 129 and 139 and section 4.2.2.3.

In terms of sampling, the studies discussed above show that TLs are either beyond or at the very limits of the degree of complexity that can be explored ergodically by MD, even when using enhanced sampling methods. This makes the necessary simulations very expensive and difficult to reproduce. We suggest that it could be useful if researchers in the field cooperated more closely, notably by freely sharing simulation data that are useful for force-field testing and parameterization. Obviously, sharing trajectories that include all solvent molecules would not be straightforward. However, removing water greatly reduces the amount of data to share. For example, TL trajectories encoded in the gromacs compressed format result in ∼20 MB per 20 000 frames (https://github.com/srnas/tetraloops-trajectories), which corresponds to 1 μs with a 50 ps stride. Such sharing of data could enable small changes to variables such as dihedral parameters to be tested using reweighting techniques (see section 3.1.4).[143,163]

It is always important to consider how MD data are compared to experiments. TLs tend to be highly structured RNAs, and one can safely assume that to a first approximation the native state will be represented by a single conformation. This conformation should be dominantly populated in an accurate MD simulation. This also implies that one must be careful when comparing predicted structures to experimental ones. In particular, commonly used trivial metrics such as RMSD may not reliably identify correctly folded RNA structures.[223] It is worth noting that the RNA structure prediction community[465] uses several different criteria to score the agreement of predicted structures with experiment. Measures such as εRMSD[223] or interaction network fidelity[332] directly report on the differences in the base-pairing and base-stacking with respect to a reference structure,[52] and should be preferred.

Agreement based on these metrics correlates strongly with the formation of all of the correct signature interactions. However, the ideal way of verifying the correctness of MD simulations of TLs and other RNA systems is still comparison to primary data from solution-phase experiments (i.e., various NMR parameters) if available.

Seen in their wider context and with an appropriate degree of realism, the results obtained to date in MD simulations of TLs are not especially promising, in terms of either force-field performance or sampling. The take-home message from benchmark RNA simulations that have been reported (see also section 4.1) is that modern force fields struggle to achieve a balanced description of stacking, hydrogen bonding, backbone substates, and solvation. In addition, the difficulty of achieving convergence for TLs raises major questions about the likelihood of achieving convergence (and thus obtaining a comprehensive description of the full folding landscape) for more complicated RNA systems. It should be noted that this pessimistic outlook does not preclude the possibility that MD can be useful in studies on various RNAs. The goal of simulations discussed in section 4.2.2.3 was to achieve full folding of TLs, that is, to make structure predictions without using any starting knowledge about the sequence's native structure. This is definitely the most challenging task in the field, and is currently beyond the capabilities of atomistic MD simulations.[175] However, when MD simulations are applied to larger systems with the goal of sampling local fluctuations around the native structure, they are much more reliable. In addition, due to compensation of errors, the simulation quality does not need to deteriorate as the studied system's size and complexity increase (see, e.g., section 4.7.1). In fact, although TLs are physically small systems, their folding is governed by a very intricate balance of various energy terms, making them naturally exceptionally difficult to describe accurately with force fields. Moreover, conformational transitions to spurious structures usually involve cooperative rearrangements associated with large free-energy barriers. Therefore, they are likely to be rare events that happen on relatively long time scales, especially for large RNA structures. Consequently, large RNA molecules can be often simulated on time scales of hundreds of nanoseconds or even microseconds without incurring significant problems. Such simulations can provide very useful information that would not be accessible using experimental methods alone. Nevertheless, MD simulations of RNA must be done with great care and validated against experimental data, with in-depth analysis of their structural details. In addition, to facilitate the improvement of force fields, sampling techniques, and simulation methods, the occurrence of spurious structures stabilized by force fields should be reported openly. This has not always been the case in the MD simulation literature, which has slowed progress in the force-field development. Further, simulations that are not based on reliable experimental starting structures should be done only exceptionally, and their outcomes will always be speculative.

4.3. MD Simulations of Internal Loops and Other Small RNA Molecules

As noted in section 2.2, canonical helices are represented as formally base-paired segments in RNA 2D diagrams, and sequence alignments usually reveal a free covariation of G═C, C═G, A═U, and U═A base pairs in canonical RNA segments. The G/U wobble base pair (Figure 2G) is the third RNA canonical base pair.[47] Thermodynamically, canonical

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

A-RNA regions represent the dominant source of stability of RNA molecules, and kinetic studies indicate that they usually form before tertiary structural features.[45,443] Canonical helices alternate with formally unpaired 2D loop regions, generally the most functionally important parts of RNA molecules.[41,52−55,59,61,703,704] Many of them form highly specific 3D structures that serve as molecular building blocks and have characteristic sequence patterns.[41,49,52,53,59,705] They can be formally classified as hairpin,[621−629] internal,[51,52,54,55,61,521,525,527−529] and multihelix junction loops.[431,706−713]

4.3.1. Basic Introduction to the Structural Organization of RNA Internal Loops. So-called RNA motifs are small ordered arrays of noncanonical base pairs[39] and interactions such as base-phosphate H-bonds[50] that are subject to sequence constraints, as characterized by seminal works by Leontis, Westhof, and co-workers.[41,52−55,59,61] This means that a change in one nucleotide of the motif must be accompanied by changes in other coupled positions to preserve the motif's function (which is given by its 3D structure). In other words, each RNA motif has some consensus sequence that is determined by its signature interactions. Canonical base pairs are present in some RNA motifs but not all. Roughly speaking, RNA motifs constitute a second level in the structural hierarchy of RNA, above the level of base pairs.[39,40] They have the following characteristic properties: (1) They are modular and can thus be seen as undividable, discrete units. This precludes experimental dissection of the thermodynamic effects of their individual H-bonding and stacking interactions because disrupting one native interaction may cause the entire motif to collapse.[54] (2) They are usually autonomous (but not inevitably; see below for examples); that is, they may occur in different molecular contexts but will generally fold into their sequence-dictated characteristic geometry independently of the context. (3) They are usually recurrent.[54,55] This means that they occur in diverse evolutionarily unrelated RNA molecules or even multiple times in the same molecule. The same motifs thus appear to have evolved convergently in different molecules. (4) They are multipurpose. For example, depending on the context, a given motif may participate in protein binding or RNA−RNA interactions, or may undergo dynamic motions. Over 100 distinct RNA motifs have been recorded in a database of RNA 3D structures (http://rna.bgsu.edu/rna3dhub/pdb).[49,61] Some motifs have only been observed in one particular RNA, while others are ubiquitous. Most known motifs were identified by analyzing ribosomal X-ray structures. (5) RNA motifs adhere strongly to the isostericity principle, as demonstrated by their signature interactions.[42,49,61] In keeping with these properties, sequence alignments of corresponding RNAs from diverse organisms have shown that if a motif is present at a given position in the RNA of one organism, other organisms will typically have a sequence that complies with the consensus of that motif in the same position. At specific locations harboring an RNA motif that allows sequence variability, evolutionary processes have often produced a context- and function-dependent variant of the motif that preserves the motif's core signature interactions while modulating its variable features. This enables development of new signature-like interactions and sequence constraints that characterize the corresponding region of the RNA in question.[55,61] Canonical A-RNA could be regarded as a motif with the $(N{=}N)_n$ consensus, that is, an array of n consecutive canonical (N═N) base pairs with any sequence.

If a sequence pattern associated with a recurrent RNA motif is detected in an RNA sequence of unknown structure, one can quite confidently predict that the corresponding section of the RNA will adopt the motif's 3D structure.[49,61] Such predictions are especially reliable when a sequence pattern consistent with the motif consensus sequence is identified by sequence alignment in multiple equivalent RNAs from different species. Conversely, the sequence patterns of smaller submotifs (see section 2.2) are too ambiguous to be a reliable basis for structural predictions. A notable omnipresent submotif is the UA_handle, which beautifully illustrates the extreme versatility of RNA structural organization.[65] Another well-known example is the RNA dinucleotide platform.[714−718]

Because of their small sizes, internal structural stability, and biological relevance, RNA motifs represent obvious targets for studies using explicit solvent MD simulations. MD simulations can provide information about the impact of base substitutions in RNA motifs as well as their structured hydration, ion-binding properties, intrinsic flexibility, interactions with ligands and proteins, and other features of interest. Several following sections provide an overview of the MD simulation literature relating to mainly internal loop RNA motifs and some other small RNA building blocks.

4.3.2. 5S rRNA Loop E and Other Well-Paired Symmetrical Internal Loops. The bacterial 5S rRNA Loop E (LE) motif is an internal 14-nucleotide loop forming seven consecutive noncanonical base pairs.[51,484−487,525−527,719] It is localized roughly in the middle of the bacterial 5S rRNA (its loop E region), the smaller rRNA of the bacterial large ribosomal subunits. It is one of the longest known fully paired double helical regions with consecutive noncanonical base pairs. In bacteria, the first submotif of the LE interacts with ribosomal protein L25,[51,527] while its second submotif participates in RNA−RNA interactions. Although the precise function of the 5S rRNA is unknown, it is widely assumed to play a role in dynamic intersubunit communication during elongation.

The original nomenclature for LE motifs was somewhat ambiguous and was confused with that of another RNA motif, the Sarcin-Ricin loop[528,529] (SRL, see section 4.3.6). This is because the equivalent regions of archaeal and eukaryal 5S rRNAs are different from the bacterial region, and form SRL motifs (for more details, see refs 56 and 51; the latter paper is separating the bacterial LE from SRL) (Figure 37). The LE motif consists of two submotifs separated by a water-mediated A/G base pair and related by dyadic rotational symmetry. The LE and SRL were among the first RNA motifs to be discovered, and were structurally characterized before the solution of any ribosomal X-ray structures.[525−529] Their investigations were central to the formulation of the rules of RNA sequence conservation (i.e., the isostericity principle).[51]

The bacterial 5S rRNA LE has been subject of several MD simulation studies that helped to establish MD as a valid complement to experimental methods.[484−487] Although these studies were conducted with time scales ranging from 1 to 25 ns and using the ff94-ff99 AMBER force fields (see section 3.1.2), their results are considered to be reliable. For example, MD was used to clarify the relationship between the bacterial and spinach chloroplast 5S rRNA LE motifs.[484] Chloroplasts are evolutionarily related to bacteria, suggesting that their ribosomal structures should closely resemble those in bacteria. However, the sequences of the *E. coli* 5S rRNA LE (for which X-ray structures are available) and the presumably equivalent spinach chloroplast rRNA region differ substantially, with only

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

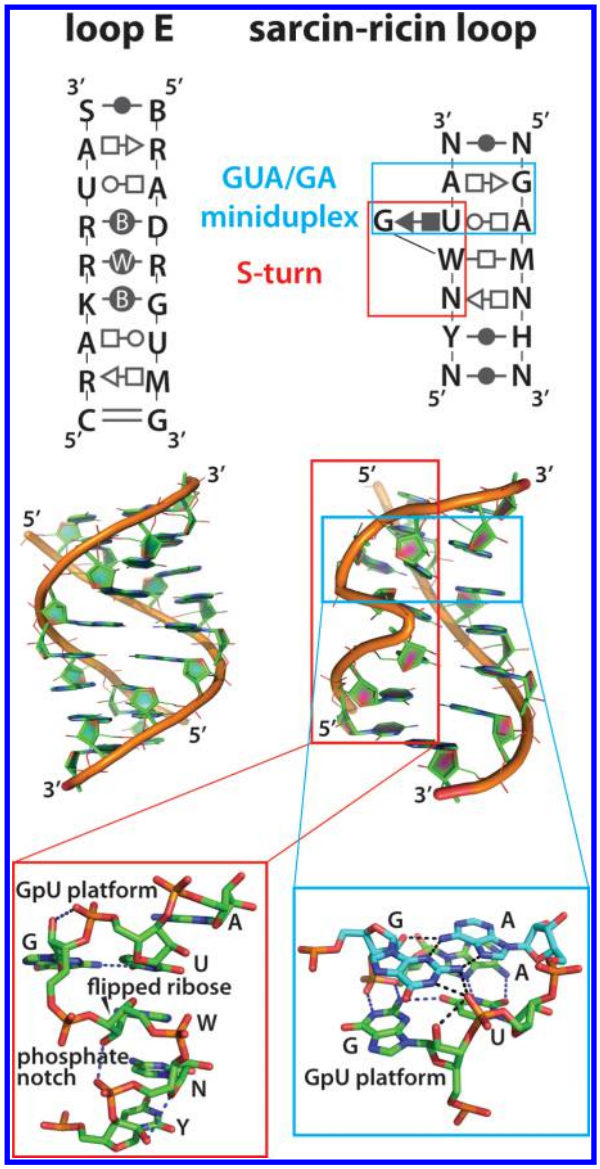


Figure 37. Annotated 2D and 3D structures of LE and SRL RNA motifs. Red and blue boxes highlight S-turn and GUA/GA miniduplex submotifs of SRL. The GpU platform and phosphate notch formed by three adjacent phosphates in the S-turn motif are highlighted in the bottom left panel. The base pair family symbols are taken from ref 39; B and W in base-pair symbols stand for bifurcated and water-mediated base pairs. The 2D structure is annotated according to Leontis and Westhof nomenclature (the sequence is annotated as follows: S, strong (G or C); W, weak (A or U); R, purines (A or G); Y, pyrimidines (C or U); K, keto (G or U); M, amino (A or C); B, not adenine; D, not cytosine; H, not guanine; and N for any base).[720] The annotation reflects the isostericity.

64% sequence identity (9 nucleotides out of 14). Nevertheless, the early works of Leontis and Westhof on base pair classification and the isostericity principle indicated that these two RNA regions may in fact form identical (isosteric) structures.[51] Leontis and Sponer then probed this prediction using MD simulations.[484] A hypothetical starting structure of the spinach chloroplast LE motif was prepared by homology modeling, using the bacterial LE X-ray structure as a template. Subsequent simulations neatly relaxed the chloroplast sequence into a structure isosteric with the bacterial LE. It should be noted that by modern standards, these simulations were rather short, so the results obtained may have been significantly affected by the chosen starting structure. In addition, they were performed with an outdated force field, so it might be interesting to reinvestigate these structures using modern tools and methods. However, the simulations' results were independently confirmed by NMR experiments on the spinach chloroplast LE, which strongly supported its isostericity with the bacterial variant.[719] In addition, the spinach chloroplast LE variant binds the bacterial L25 protein.[719]

Good correspondence between MD simulations, experiments, and bioinformatics data for the LE motif shows that when an appropriate question is asked, MD can provide correct answers even with modest computational efforts. In fact, studies on the impact of base substitutions in known RNA structures represent one of the most appropriate applications of MD simulations.

The MD method is especially efficient for well base-paired regions, that is, canonical helices and fully paired symmetrical *N*×*N* internal 2D loops. The *N*×*N* internal loop is an RNA segment in which each strand contains *N* nucleobases that do not form canonical base pairs with their partners. Thus, the LE motif is a 7×7 internal loop. Such long internal loops are quite rare, and shorter 1×1, 2×2, and 3×3 symmetrical internal loops are more common.

The reason why MD simulations of well base-paired double helical RNA segments using the AMBER variants of the force field are usually successful stems from the fundamentals of the force field's derivation (see section 3.1.1). This force field provides a rather good description of both base stacking and base pairing because of the original parameterization strategy of the Cornell et al. force field, whose partial atomic charges (which are still used even in the latest variants of the force field) were derived by fitting to the molecular electrostatic potential (ESP).[80] QM studies have demonstrated that such ESP-fitted charge distributions are very suitable for base stacking and base pairing calculations, including for RNA base pairs that interact via the 2′-OH hydroxyl group.[104,721,722] The alternative CHARMM force field provides less structurally stable trajectories for paired RNA duplexes; see also section 3.1.3. For example, simulations using the CHARMM27 force field[155,156] predicted nanosecond-scale base pair opening rates for A═U base pairs in canonical A-RNA duplexes.[723] This result was initially[723] considered to agree with the available experimental base pair opening data,[724] but this suggestion was later corrected.[98] In addition, large-scale fraying of RNA duplexes has been observed in longer CHARMM27 simulations,[160,235] resulting in an effort to stabilize the CHARMM RNA duplex simulation.[98] The CHARMM36 RNA force field[98] is improved in this respect, but some problems with structural stability may remain.[134] Our tests show that, as compared to all of the available alternatives, the AMBER force fields (especially the newer versions that are not prone to the formation of ladder-like artifacts[94,154]) yield more structurally stable simulations of RNA double helical regions over longer time scales, and do not require the imposition of restraints on the base pairs. These properties of the CHARMM and AMBER force-field families should be borne in mind when interpreting literature reports discussing MD simulations that predict large fluctuations in symmetrical internal RNA loops.[725−727] A degree of common sense is needed when evaluating any MD data, and we strongly caution against interpreting every single fluctuation or change in base pairing as being biochemically relevant. We in fact suggest that the basic approximations of

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

pair-additive force fields (specifically, the lack of polarization) cause them to generally underestimate the structural stability (i.e., base pairing) of RNA double helical regions (see sections 3.1.1.3 and 3.1.2.5).

Note that not all symmetrical internal loops are fully base paired, as there may be slippage of the interacting partners in the loops. The presence of unpaired nucleotides complicates computations because (i) it drastically increases the sampling requirements, and (ii) it increases the simulation's sensitivity to inaccuracies in the description of solvation energies and other force-field issues. Both problems arise when simulating not fully paired RNA internal loops because it is necessary to sample nucleotides that fluctuate between being stacked inside the helix, bulging out into the solvent, and interacting with the double helix grooves. Any imbalance in the description of these interactions might stabilize a conformation not compatible with the experimental structure. The same problems are encountered in studies of nonsymmetrical *N*×*M* internal loops.

4.3.3. Can MD Simulations Help Explain Thermodynamic Rules? As discussed above, symmetrical internal loops are ubiquitous building blocks of RNA structures and profoundly affect the thermodynamics of RNA folding. It would therefore be exceptionally useful if MD simulations could be used to characterize their thermodynamic properties as well as their structures. The sequence-dependent thermodynamics of canonical A-RNA is very well established thanks to the seminal measurements of Turner and co-workers.[44,728,729] These benchmark thermodynamic measurements constitute the core of the nearest-neighbor thermodynamic stability model for RNA, and are essential for predicting 2D structures of RNA sequences.[46,439,730] In contrast, predicting thermodynamic properties of internal 2D loops remains challenging.[731] One reason for this is that it is not possible to easily predict the structures of internal 2D loops from their sequences; in addition, conformations of some loops can be context-dependent and thus ambiguous (see section 4.3.5).[732−739] The lack of reliable information on the thermodynamics of short internal RNA loops was actually identified by experimentalists a decade ago as an important issue that could potentially be addressed by modern computational chemistry.[740] However, it is still much more difficult to accurately compute the free-energy profiles of RNA duplexes than to simulate their structural dynamics.[365,741,742] This is illustrated by a study that used thermodynamics integration (TI) free-energy calculations (see section 3.2.7) to characterize the context- and sequence-dependence of the structures of tandem A/G base pairs in RNA duplexes.[365] Tandem arrangements of A/G base pairs are common in RNA, and tandem A/G base pairs play important roles in determining the structure, dynamics, and stability of larger RNA sequences. The TI simulations evaluated free-energy differences between r(5′-GCGGACGC-3′)$_2$, r(5′-GCiGGAiCGC-3′)$_2$, r(5′-GGCGAGCC-3′)$_2$, and r(5′-GGiCGAiGCC-3′)$_2$ self-complementary duplexes (the 2×2 "GA/GA" internal loop in each duplex is underlined). Here, iG and iC respectively denote isoguanosine and isocytidine, which are derivatives of guanosine and cytidine with transposed amino and carbonyl groups. The NMR structures of these short sequences suggest that the A/G base pairs adopt either imino (cis Watson−Crick/Watson−Crick A/G) or sheared (trans Hoogsteen/Sugar edge A/G) conformations, depending on the identity and orientation of the adjacent (closing) canonical base pairs. Such coupling of the A/G base pair geometry to the sequence and orientation of the closing base pair is known as the context-dependence of the structure of the tandem A/G base pairs in internal loops.[733,734] This context-dependence has been partly attributed to potential formation of an out-of-plane H-bond involving unpaired guanine amino group in the imino A/G base pair.[106] Thus, a reparameterization of the amino group valence and dihedral angle parameters was proposed[365] in an attempt to mimic partial sp^3 pyramidalization of the free amino group,[105] but this did not improve the results obtained. It is likely that the force-field form did not allow one to make the parameterization of the amino group pyramidalization sufficiently robust, due to its inability to include the electronic structure changes that are coupled with the nonplanarity. The computations were directly compared to benchmark NMR structural data and thermodynamic measurements,[734] but the agreement with the experimental results was poor despite the use of advanced computational methods.[365] It is possible that better results would have been obtained using improved force fields, which had become available after the study was conducted (see sections 3.1.2.1 and 3.1.2.2). However, newer works suggest that the difficulties encountered also reflected general limitations of free-energy computations when applied to nucleobase substitutions (section 3.2.7).

Tandem A/G base pairs were further investigated in a later study, using the χ_{OL3} force field (marked as ff10).[743] The authors considered the r(5′-GCGGACGC-3′)$_2$ duplex with experimentally observed tandem imino A/G base pairs and r(5′-GCGGAUGC-3′)$_2$ structure, which has G/U closing pairs sandwiching two sheared A/G base pairs. They have found that modification of the nonplanar guanine aminogroup[365] improved the simulations of native structures. Alternative (unobserved) A/G base pairing patterns then were constructed, and free-energy differences between imino and sheared conformations were estimated in both sequence contexts using two-dimensional US method (see section 3.2.5); the reaction coordinate was suggested using the nudged elastic band approach.[741] US computations predicted the native A/G pairing patterns to be more stable, although the free-energy difference of ∼0.2 kcal/mol was evidently underestimated.

It should be noted that the thermodynamics of RNA folding is difficult to analyze experimentally as well as computationally, and some common assumptions made when analyzing experimental data are oversimplifications. The thermodynamic experiments do not provide an unambiguous decomposition of total free energies into individual energy contributions. There are two fundamental types of interactions between nucleobases whose energetics affect the structure and stability of RNA: base stacking and base pairing. Traditionally, it has been assumed that the sequence dependence of free energies measured in thermodynamic experiments primarily reflects the combined intrinsic strength of base pairing and stacking. If this view is correct, it would ultimately be possible to derive plausible free-energy models by analyzing the energetics of the direct base pairing and stacking interactions (i.e., dimerization energies between two nucleobases). Nowadays these intrinsic interaction energies can be estimated quite accurately using QM methods, although, to assess the nucleic acids thermodynamics, it would be also necessary to properly consider factors such as solvent screening of the electrostatic interactions.[118,119] However, the above assumption is probably a major oversimplification: it is more likely that the measured free energies result from a very complex mixture of contributions, with base pairing and stacking energies being just part of the overall

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

energy balance. In addition, the balance may differ from system to system, limiting the transferability of thermodynamic measurements. Measured free energies therefore also include some less salient and more elusive contributions such as those of ion-binding sites and particularly hydration; both stabilizing tightly bound water molecules and destabilizing unsaturated (unsatisfied) H-bond donors/acceptors can be important, as discussed below. These contributions may "overwrite" effects of base stacking and base pairing, making it impossible to identify any systematic correlation between the system's thermodynamics and its base pairing and stacking interaction energies.

It is possible to also obtain thermodynamic parameters from single-molecule pulling experiments, as shown for instance for DNA in ref 744. Although single-molecule experiments are in principle similar to MD simulations experiments, we should recall that they are performed on time scales that are several orders of magnitude slower than those of typical MD simulations. Simulations corresponding to unzipping dynamics using, for instance, steered MD (see section 3.2.5) would not allow realistic thermodynamic parameters to be recovered. However, pulling simulations can be used to qualitatively recover unfolding pathways, as discussed in section 4.3.10.

It is also important to take into consideration that free energy of the folded RNA is always measured with respect to the unfolded reference state. For example, in the case of duplex formation, the reference state is the single-stranded state (Figure 19). However, conformation of the unfolded single-stranded RNA is usually unknown; it may exist as a very complex ensemble of diverse structures, and will almost certainly contain some amount of stacking. However, QM interaction energy calculations typically derive stacking energies with respect to noninteracting nucleobases (zero stacking) used as the reference state.[118,119] The structure of the unfolded state can thus influence the kinetics and thermodynamics of folding/binding. If (for example) a substitution expected to stabilize the folded state stabilizes the unfolded ensemble to a greater degree, the experimental results may be counterintuitive, and the results obtained in typical free-energy simulations may contradict those observed experimentally.[119,745] Difficulties in sampling single-stranded RNAs are exemplified in section 4.1. The extent to which the unfolded state has been properly sampled is one of the largest sources of uncertainty in computational free-energy studies such as TI simulations (section 3.2.7); full sampling of unfolded states is essentially unachievable with current methods. In view of these considerations, it is not surprising that there have been no successful attempts to correlate experimental thermodynamic data with base stacking energies sampled in MD simulations (i.e., evaluations of base−base interaction energies derived from the force field's nonbonded terms).[742,746]

The complexity of molecular interactions is probably one of the factors responsible for so-called non-nearest-neighbor effects, which give rise to substantial coupling (nonadditivity) between diverse mutation sites and can be detected using double mutant cycle experiments.[747] The intricacy of this phenomenon can be visualized by experimental studies that have revealed surprisingly large differences between canonical A-RNA and B-DNA with respect to the free-energy changes that occur upon substituting guanine with inosine (I) in identical sequence contexts.[584,747] This clearly shows that experimental free-energy changes cannot be rationalized using (direct) base pairing and stacking energies, because the G═C → I═C substitution should have an identical effect on base pairing and a similar effect on stacking in A-RNA and B-DNA. G → I (and similar) substitution biochemical experiments are commonly used to estimate free energies associated with single H-bonds. However, it can never be guaranteed that the substitution affects only the studied interaction that is abolished. In an ideal case, the G═C → I═C substitution would cause loss of a single H-bond in both A-RNA and B-DNA, so, in the absence of any other factors, one should see near-identical free-energy effects for both helices. This is not the case in reality.[584,747] These experimental findings have been qualitatively reproduced by MD-based TI free-energy simulations with the bsc0 and χ_{OL3} AMBER force fields.[367]

Probably the technically most advanced MD thermodynamics study on duplex free energies was performed to evaluate the free-energy impact of base pair substitution in canonical A-RNA.[362] This benchmark study considered only the most conservative mutations from one canonical base pair to another canonical base pair (up to two mutated base pairs in duplexes consisting of six base pairs in total), and excluded purine to pyrimidine exchanges. A smart combination of the FEP thermodynamics cycle method with the Hamiltonian REMD made it possible to achieve improved convergence by performing simulations with different λ-values in parallel with exchanges between neighboring replicas using the Metropolis criterion (see sections 3.2.4, 3.2.6, and 3.2.7). The χ_{OL3} force field was used, together with impressively long simulation times. Comparisons with UV melting experiments[438] for nine substitutions revealed an amazingly good correlation of $\Delta\Delta G_{calc} = 1.16(\pm 0.04)\Delta\Delta G_{exp}$, with $R^2 = 0.97 \pm 0.01$ and a mean absolute deviation of 0.55 kcal/mol. However, for the reasons noted above, it is likely that the accuracy of even this modern procedure would decline sharply if applied to more challenging substitutions. Experimental data are available[438] to test this suggestion.

Earlier, ff99 force field has been used to predict relative free energies of RNA helix formation, using three hexaloop hairpins with identical loops and varying stems.[748] The potential of mean force of stretching the hairpins from the native state to an extended conformation was calculated with US (section 3.2.5). The simulations were able to correctly predict order of stabilities of the hairpins, although the magnitude of the free-energy change was larger than that determined by optical melting experiments. As pointed out by the authors, the unfolded state in the optical melting experiments is a random coil, while the end state in US simulations was an elongated chain, complicating a direct comparison between theory and experiment. Thus, the calculations were compared to reference data by the thermodynamic cycle (section 3.2.7) applied to transitions between hairpins estimated using simulations and nearest-neighbor thermodynamics data. Once more we recall that evaluation of the stability (i.e., full sampling) of the random coil is out of reach for current methods.

Some other quite successful free-energy calculations are discussed in sections 4.3.9 and 3.2.7, but it is obvious that, despite occasional optimism in the literature, current MD methods are not generally capable of providing consistently accurate free-energy calculations for RNA internal loops.

We would like to emphasize that, despite all limitations, the MD technique does not have any alternative in computational free-energy studies. This is because evaluation of thermodynamic stabilities always requires proper consideration of Boltzmann sampling. Despite common attempts to estimate stabilities of nucleic acids using for example accurate QM

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

calculations of conformational or interaction energies, such "single-structure" calculations do not reliably reflect free energies. Single-structure calculations just provide "snapshots" at the potential energy surface, while Boltzmann averaging over the potential energy surface is required to get free energies. For a more detailed explanation of how, for example, description of stacking differs in QM, MD, and experimental studies, see refs 118 and 119. To illustrate the overwhelming complexity of the RNA conformational space, the discrete path sampling method identified around 80 000 local minimal on the potential energy surface of the RNA TLs (section 4.2.1) when representing the solvent by an implicit model averaging out solvent degrees of freedom (section 3.2.8).[691]

4.3.4. Long-Residency Hydration and Ion-Binding Sites in RNA Motifs. Simulations of the LE motif have also been instrumental in studies of hydration.[484,487] It was suggested a long time ago that hydration makes water an integral part of RNA's structure.[749] However, while the term "integral part of nucleic acid structure" sounds very appealing, its exact meaning remained somewhat elusive, especially in quantitative terms. X-ray crystallography studies have clearly demonstrated structured hydration of nucleic acid molecules.[750−755] However, X-ray structures provide only static averaged pictures of hydration, and do not always clearly show which hydration sites are important and which are not. This is partly because the hydration patterns of crystallized nucleic acids may be affected by rigidification of structures in the crystal lattice, to which the inherently dynamic hydration may be sensitive. Additionally, determination of solvent molecules in biomolecular X-ray crystallography is more ambiguous than that of solute atoms[756−760] because the hydration network is affected by subtle biochemically irrelevant differences in crystal packing and the genuine thermal fluctuations of the often flexible hydration shell.[369,756−760] With some exceptions,[761] NMR structural experiments typically do not provide any information about hydration. By contrast, MD stands out as a natural and very suitable tool for studying the dynamic aspects of hydration, and has been used for this purpose almost since it was first developed.[369,484,487,503,522,762−775]

Early MD studies strongly supported the hypothesis that hydration of RNA molecules is very dynamic, and this has been confirmed by later studies. The residence (binding) times of individual water molecules (i.e., exchanges between the first hydration shell and the subsequent layers or bulk solvent) in common RNA hydration sites are on the order of 50−500 ps.[369,503,769,771] Thus, atomistic MD simulations do not support the idea that rigid spines of hydration and structured water clusters commonly form around nucleic acid molecules. However, a small subset of RNA hydration sites known as long-residency hydration sites may be structured, with longer binding.[369] Even the early LE motif simulations identified long-residency water molecules that remained bound to specific positions for up to 5 ns.[484] Some static long-residency water molecules remained bound for even longer (for the simulation's full 25 ns duration) in the LE−L25 protein complex.[487] Thus, specific buried waters may help to stabilize protein/RNA interfaces.[369] The repertoire of long-residency hydration sites contains a much greater diversity of structures than just water-mediated base pairs and H-bonds (water bridges). The current literature reveals an astonishingly wide variety of hydration sites. Water molecules can bind to RNA at sophisticated hydration sites with coupled water networks or in sites that are entirely buried, long-residence water molecules may "rattle around" inside a hydration pocket rather than being firmly H-bonded to one donor or acceptor, and water molecules can contribute substantially to the flexibility of an RNA molecule by mediating large-scale movements of adjacent segments via dynamical water insertion.[522] Hydration can also be intimately related to ion binding, and exchanges (competition) between water molecules and monovalent ions are common. MD is the only method able to probe detailed solvent dynamics and identify specific long-residency hydration sites. Unfortunately, despite the existence of highly refined force fields and the ability to perform simulations with time scales that are several orders of magnitude greater than was historically feasible, many modern MD studies neglect hydration analysis, partly because hydration analyses have always been tedious and time-consuming. It is a pity because descriptions of the structural dynamics of hydration are among the most reliable kinds of data that can be gleaned from MD simulations.

Although MD can clearly visualize structured hydration, it is still not clear to what extent tightly bound (long-residency) waters stabilize folded RNAs in terms of free energies. We are not aware of any computational method that can be used to derive the free-energy effects of specific hydration sites, and it is not straightforward to formulate a computational strategy that would subtract the free-energy cost of disrupting a structured hydration network from the overall free-energy balance. Although methods such as grid inhomogeneous solvation theory (GIST)[776,777] might allow the effect of water displacement to be quantified, in the case of RNA folding a full sampling of the unfolded ensemble would be required. It is also not clear how to design a generally applicable computation strategy that would quantify the free-energy contributions of tightly bound waters relative to those of "common" hydration sites. It is quite possible that long-residency water molecules affect only dynamics and not stability. If a given water molecule takes a long time to exchange with the bulk solvent, it is so because the kinetic barrier to exchange is high. Typically, barriers are visited very rarely and do not contribute to the overall partition function of the system. However, a combined biochemical, NMR, and MD simulation study revealed, using serine-to-alanine mutations, a visible free-energy effect of structured hydration sites in a protein/RNA complex.[369]

It should be noted that the absence of hydration (i.e., hydration holes) can be as important as visible hydration. Hydration holes occur when the overall disposition of donors and acceptors within a molecule is such that they cannot be efficiently hydrated simultaneously. Such hydration patterns are much harder to identify than conventional hydration sites in MD simulation trajectories because they are inconspicuous. However, they can significantly affect the thermodynamics of RNA motifs by penalizing those conformations in which they occur, due to the presence of unsatisfied H-bond donors and acceptors.[367,584] Their effects can be indirectly visualized (at least in theory) using accurate explicit-solvent MD free-energy methods such as thermodynamics integration (section 3.2.7).

Simulations of the LE have also been instrumental in revealing tightly bound and long-residency monovalent ion binding sites.[484−487] Although RNA is a polyanion, the surfaces of RNA molecules do not usually experience any tight binding of monovalent ions. Common ion binding sites including the anionic phosphate groups exhibit occupancies of around 10−20% for direct (inner-shell) monovalent ion binding, with fast exchange. These phosphate groups are thus primarily hydrated, and, in the presence of explicit water, bidendate binding of

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

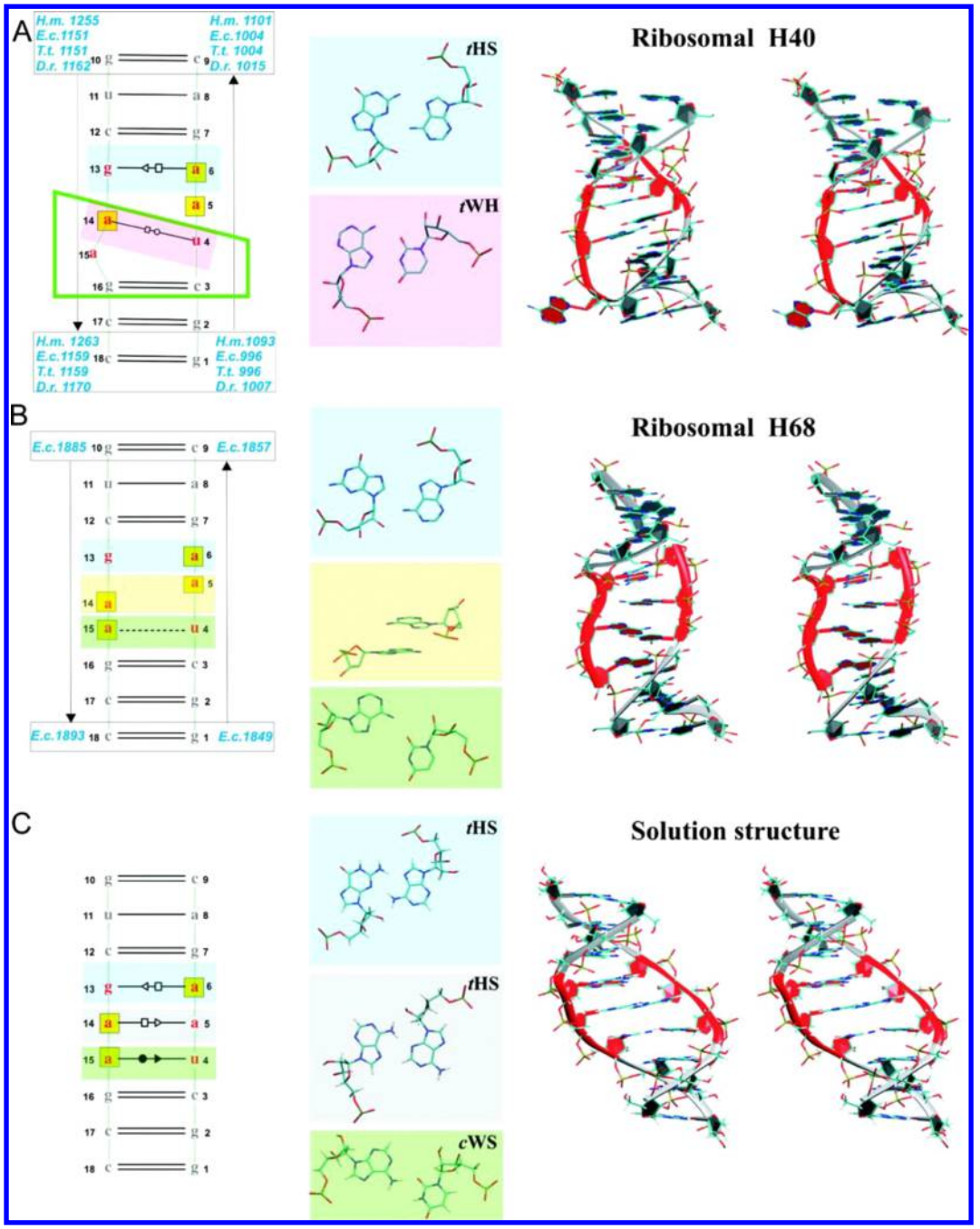


Figure 38. RNA molecular switch. 2D structures (left) and 3D stereo views (right) of different conformations of the internal UAA/GAA loops (middle).[737] (A, left) 2D annotation of the X-ray structure of typical ribosomal conformation, with unified canonical flanking sequence. (A, middle) Sheared (i.e., tHS)[39] A/G and reverse Hoogsteen (*t*WH)[39] U/A base pairs. (A, right) Stereo view of *E. coli* ribosomal helix 40 X-ray structure. (B, left) 2D X-ray structure of *E. coli* 23S rRNA helix 68 exhibiting an alternative conformation of the UAA/GAA motif with unified canonical flanking sequence. The black dashed line indicates a single H bond. (B, middle) Unpaired G and A bases, stacked middle adenines, and single-bonded A/U base pair. (B, right) Stereo view of this structure. (C, left) The 2D NMR structure.[736] (C, middle) Sheared A/G, sheared A/A, and incomplete *c*WS A/U base pairs. (C, right) Stereo view of the NMR structure. In all parts, bases of the UAA/GAA internal loop are in red, 3D structures are colored accordingly, hydrogens are not shown in the X-ray structures, bases in yellow boxes in the 2D structures are involved in stacking, and the marks between the bases indicate the base pairing family according to the Leontis and Westhof classification.[39] X-ray structure nucleotide numbers are in blue; NMR numbers are in black. The green rectangular trapezium for H40 structure marks nucleotides forming the UA_handle submotif.[65] Reprinted with permission from ref 737. Copyright 2010 American Chemical Society.

monovalent ions to the nonbridging phosphate oxygens does not occur.[503,769,771] However, folded RNAs can develop specific ion-binding pockets (the LE deep groove is a textbook example), which are always occupied by cations. Such pockets were later identified in many other RNAs and, like water binding sites, were shown to be highly variable in terms of their structure and ion residence times; for more details, see section 3.4.1.

Finally, simulations of the LE also provided some early insights (despite limitations noted in section 3.4) into the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

binding of Mg^{2+} ions to RNA, specifically to the deep groove.[485,486] These simulations captured long-lived and specific water-mediated interactions between hydrated ions and RNA atoms. It has been suggested that Mg^{2+} binding specificity is modulated by electrostatic complementarity and structural correspondence between the hydrated ion and its binding pocket. The latter can be estimated by measuring the degree of ion dehydration and the resulting number and lifetime of intervening water-mediated contacts.[486] Significant freezing of tumbling motions of bound Mg^{2+} ions has been observed. The monovalent and divalent ion binding sites of the LE motif overlap, so water molecules, monovalent ions, and divalent ions compete for the same interaction sites; for more details, see section 3.4.1. For some newer work, see ref 778.

Many subsequent MD simulation studies characterized binding of monovalent and divalent ions to RNA molecules. As is also true for hydration sites, there is currently no straightforward procedure for calculating the contribution to the folding free energy associated with individual ion binding sites whose structures are captured in simulations. Whereas in principle it would be possible to compare the computed affinity of the ions for different RNA conformations, one would have to face the challenge of properly sampling the unfolded ensemble to get the differences in folding free energies. Binding of ions to RNA, its effect on RNA structure, and the study of these factors using MD simulations are discussed further in section 3.4.

4.3.5. 5′-UAA/5′-GAN RNA Structural Switch. A striking example of an internal symmetrical 3×3 loop is the 5′-UAA/5′-GAN loop, a common and conserved motif occurring in seven ribosomal helices, RNase P, group I and group II introns, and other RNAs.[779] While individual instances of the UAA/GAN internal loop occur in highly variable contexts and are involved in diverse tertiary interactions, most of them adopt a distinctive structure with an unpaired stacked adenine and a bulged nucleotide (*N*). The three conserved adenines create a characteristic cross-strand AAA stack, although in some cases evolution has found an alternative secondary structure for this loop, as in helix 68 of the *E. coli* 23S rRNA (Figure 38). Thus, on the basis of the available structural and sequence data for folded RNAs, the 5′-UAA/5′-GAN 3×3 internal loop seems at first sight to have all of the hallmarks of a highly conserved autonomous RNA motif. However, this seeming autonomy was disproved by determination of the solution-phase NMR structure of an isolated UAA/GAA internal loop flanked by Watson−Crick base pairs,[736] which differs markedly from the recurring structure observed in folded RNAs by X-ray crystallography. In isolation, the UAA/GAA pairing is restructured into three consecutive noncanonical base pairs so that there are no unpaired bases. The solution structure, although never seen in crystallographic structures of folded RNAs, is much more consistent with the arrangement expected on the basis of thermodynamic predictions, except that the standard 2D prediction for 5′-UAA/5′-GAA would feature a 2×2 loop with a canonical A═U base pair instead of the observed noncanonical A/U base pair. To better understand this system's behavior, a series of ∼200 ns simulations were performed, supplemented by some enhanced sampling simulations.[737] Even though the simulations were performed with the pre-χ_{OL3} AMBER force fields, they quite convincingly demonstrated that the characteristic functional structures of 5′-UAA/5′-GAA loops extracted from folded RNAs are not structurally stable in isolation; that is, this loop does not seem to be separated from the rest of the ensemble by any significant free-energy barrier. This RNA element thus functions as a molecular switch that is remodeled by its interactions with the surrounding RNA sequences and proteins. The global minimum of this internal loop as captured by NMR is not known to be biochemically relevant, and the evolutionary pressures acting on it serve to conserve the functional 3D structure embedded in complete functional RNAs. The functional structure can be seen as a minor conformation and is very efficient at forming various tertiary interactions.

4.3.6. Sarcin−Ricin Loop as a Stiff RNA Motif. The SRL motif occurs in many different contexts and folds autonomously (Figure 37).[56,528−530,705] It can be found inside large RNAs and on their surfaces, and may exist as an internal loop within a helix or a component of a complex multihelix junction. It is usually formed at the level of 2D structure, that is, as a local internal loop formed by noncanonical base pairing of two strands. However, there are also composite SRL motifs involving more than two strands, which cannot be inferred from 2D structures. SRLs probably fulfill a variety of functions. The first SRL motif to be discovered is located in helix 95 in domain VI of the large ribosomal subunit. Together with an adjacent GAGA TL (section 4.2), it comprises one of the most conserved sequences of the ribosome, the sarcin−ricin domain (SRD).[780] The name derives from the toxins sarcin and ricin, which cleave the phosphodiester backbone at a specific position in the TL and depurinate the adenine stacked at the top of the TL.[781−785] The SRD forms part of the binding site for elongation factors EF-Tu and EF-G (in bacteria).[786] Disruption of the SRD by ribotoxins or deletion of SRD are lethal because they disable the binding of elongation factors to the ribosome and cause its inactivation.[787−789] There are multiple other SRLs in the ribosome, which are typically conserved across all three kingdoms of life. Multiple X-ray structures featuring SRL motifs are available, including examples located in ribosomes and riboswitches.[790−792] There are also high-resolution atomic structures of isolated SRL motifs.[528,716,793,794]

The SRL motif always includes a universally conserved 5′-GUA/5′-GA miniduplex that incorporates a GpU dinucleotide platform (Figure 37). Below the miniduplex, there is a more sequence-variable segment that usually consists of two noncanonical base pairs. The G residue of the platform is bulged, and the 5′-GUA/5′-GA miniduplex is stabilized by an unprecedented network of 13 H-bonds supplemented by extensive stacking.[717,718] A characteristic feature of SRL motifs is the S-turn conformation of the backbone around the bulged G,[795] which reverses the chain direction in the flexible region and then restores it in the G-bulge region; this feature is referred to as the "double flip over". The preceding adenine is said to be in a "3′-5′ like" arrangement (Figure 37). The SRL backbone features several unusual backbone dihedral angles.

Because of its importance and ubiquity, SRL has been the subject of several MD simulation and QM studies.[86,530,705] MD simulations have shown that it is one of the stiffest RNA building blocks and exhibits less pronounced thermal fluctuations than most RNA sequences.[530] It has been suggested that the stiffness of the SRL motif is important for its biochemical roles. Its so-called flexible region, which contains base pairs whose interactions would be expected to be quite weak, is stiffened by base−phosphate interactions[50] and S-turn backbone topology.[795] The SRL may be associated with simple water bridges, but it has no obvious ion binding site.[530] Superficially, the SRL motif is nicely stable in MD simulations. However, this first impression is misleading

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

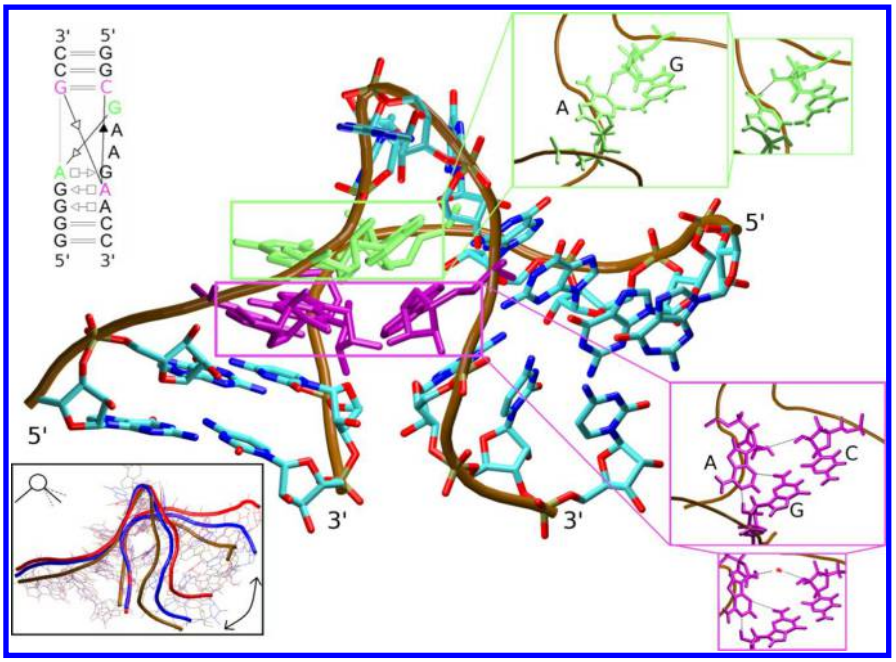


Figure 39. Annotated 2D[39] and 3D structures of a typical kink-turn.[521] Its essential components include a three-nucleotide bulge, an adenine-specific sugar/base interaction (green inset), and an A-minor interaction (purple inset).[810] MD simulations have revealed that kink-turns may act as flexible anisotropic and nonharmonic molecular hinges (black inset).[522] This motion is significantly aided by reversible insertions of water molecules between O2′···O2′ atoms in the A-minor interaction (smaller purple inset).[522] On the other hand, alternative conformation of the bulge signature interaction, commonly observed in MD simulations (smaller green inset), is most likely a force-field error.[507] For more details, see the text. Adapted with permission from ref 176. Copyright 2014 American Chemical Society.

because current force fields still do not reproduce important details of its unique backbone topology and the H-bond network around the G-bulge region.[705] The molecule appears very stable in MD simulations because it is firmly locked in its native conformation, while alternative or spurious structures are separated from the native basin by large free-energy barriers. It is unlikely that SRL would fold spontaneously from the unfolded state in MD simulations using current force fields. QM calculations[86,718] have shown that the different structural features of SRL give rise to a very complex energetic balance that is unlikely to be well-described by force fields.[132,705] SRL motif is thus an important benchmark for parameterization and verification of force fields, and especially for the reparameterization of dihedral potentials. It should be included in the standard portfolio of test targets for RNA force fields because of its uniqueness and the availability of unambiguous structural data. Despite the limitations of current force fields, MD simulations with a modest time scale of 200 ns have been successfully used to rank relative structural stabilities of different SRL motif sequence variants.[705] More specifically, MD predicted the impact of substitutions on SHAPE probing[796] data better than a bioinformatics scoring scheme based on isostericity.

4.3.7. Kink-Turns and Reverse Kink-Turns as Potential Molecular Elbows. Kink-turn is an asymmetric internal loop that adopts a unique V-shaped topology in its folded state (Figure 39).[55,282,507,521,522,797−808] Both local kink-turns and composites consisting of more than two strands have been observed.[55] Kink-turns consist of three distinct subsegments: a noncanonical (NC) stem composed of noncanonical base pairs, a canonical (C) stem with canonical base pairs, and a usually three-nucleotide bulge between the stems in the longer strand of the asymmetric internal loop. Each of these three segments contains characteristic base pairs and tertiary interactions that cause the kink-turn to adopt a specific functional fold with a sharp bend in the phosphodiester backbone in the bulge region, where the interstem angle is ∼120°. Kink-turns play a key role in protein-assisted RNA folding.[521] In addition, structural analyses, MD simulations, and cryo-EM data suggest that some kink-turns may contribute to the functional dynamics of RNA.[522,802,805,806,809] Kink-turns are stabilized by a set of five signature interactions: a canonical (usually G═C) base pair at the first position in the C-stem, tandem tHS[39] "sheared" AG base pairs with opposing polarities at the first and second positions of the NC-stem, and two tertiary tSS interactions. The first of these tSS interactions occurs between the 5′-end nucleotide of the bulge and the adenine of the first tHS pair of the NC-stem, and includes a 2′-OH−A(N1) H-bond that is the most important interaction for folding of a kink-turn.[55,801] The second tertiary contact connects the NC and C-stems, and is formed between adenine in the second NC-stem's tHS pair and the first C-stem canonical pair. The second tertiary contact is a common A-minor interaction.[810] Formation of the sharply bent (folded) kink-turn structure requires the presence of divalent cations or binding proteins.[523,524]

Kink-turns have been examined in several MD studies, which have investigated their flexibility, hydration, ion-binding, sequence properties, and protein binding.[282,507,522,603,802−808] These studies generally showed that kink-turns are well described by the AMBER force fields, which nicely capture their complex topology. The only known force-field issue is that it tends to predict that the key apical 2′-OH−A(N1) tertiary H-bond fluctuates between the native state and an alternative configuration in which the adenine amino group acts as the hydrogen-bond donor (Figure 39), which does not seem to be supported by any experimental data.[507]

This force field's good performance is somewhat surprising given the uniqueness of the kink-turn topology; it may be due

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

to the overall looseness of the kink-turn structure, in which the individual interactions do not act in opposition to each other. Therefore, modest inaccuracies in the force field's description of individual backbone elements' conformations, H-bonding, stacking, and solvation features do not amplify each other and are tolerated within the overall structure. For additional discussion of compensation of errors in networks of H-bonds, see the analysis of force-field performance for protein/RNA complexes (section 4.7.1).

Interestingly, MD simulations did not reveal any obvious ion binding sites around kink-turns.[603] This seems inconsistent with experimental findings, which clearly show that kink-turns are among the most divalent-ion-dependent RNA motifs. The contradiction probably stems from the intuitive assumption that tight structural ion binding and ion-dependent RNA behavior must always be directly related. However, there is no a priori physical requirement for prominent ion-dependence in RNA folding to always be due to tightly bound ions (or vice versa). Contemporary MD methods can reliably identify major ion-binding sites around RNA motifs, so if such a site existed in kink-turns, it would be detected. In fact, no experimental study has provided evidence of structural ion-binding in kink-turns. This shows that structural data alone are not sufficient to confidently predict the ion-dependency of RNA folding. In the case of kink-turns, nonspecific ion atmosphere screening is probably needed to stabilize tightly kinked structures in the kink-turn ensemble, which then allows formation of the motif's signature interactions. The remarkable ion-dependence of kink-turn folding clearly warrants further study using MD techniques.

MD simulations have revealed that kink-turns are exceptionally flexible elements that look like hinges and can act as passive molecular elbows that facilitate large-scale dynamical movements in larger RNAs.[522,603,802,805,806] By "passive elbows" we mean structures that permit fast barrier-free movements but do not drive them. Kink-turn flexibility is highly anisotropic and stems from the apex of the kink, while the attached arms are more rigid. The A-minor interaction between C- and NC-stems fluctuates between direct and water-mediated conformations (Figure 39).[522] In the water-mediated conformation, the distance between sugar−phosphate backbones entering the two stems of the V-shaped RNA is increased, causing an increase in the interstem angle. In this way, local fluctuations in the system's hydration contribute to the global flexibility of the RNA motif. The kink-turn is thus a unique hinge-like RNA element with high anisotropic flexibility. Because MD simulations indicate that its water insertion-expulsion dynamics occur on a time scale of nanoseconds, we suggest that the kink-turn elbow should be described as a floppy system with a single shallow (flat) nonharmonic free-energy minimum rather than a system that alternates between direct-binding and water-mediated substates.[522] Ribosomal X-ray structures show both water-mediated and direct H-bond variants of the A-minor interaction in kink-turns.[522] Importantly, solution-phase experiments have been unable to capture the kinked structure at atomic resolution because isolated kink-turn sequences tend to unfold.[811] These results show how MD simulations can complement experimental studies: we are unaware of any experimental method that could have revealed the unique flexibility of the folded kink-turn, or shed light on the relationship between this flexibility and the system's hydration dynamics.

It is important to note that MD simulations, solution structural experiments, and X-ray crystallography are mutually nonequivalent techniques that complement each other. Although MD studies examine isolated kink-turns in simulated solution-phase environments, they are not equivalent to solution experiments. MD primarily provides information on the physicochemical properties of folded kink-turns because the starting structures used in simulations are taken from functional conformations seen in X-ray structures of ribosomes, and they do not unfold on the simulations' time scale. In this case, one can take full advantage of the essentially unlimited temporal resolution of the MD technique, which allows one to directly visualize the large-scale thermal fluctuations. One can study intrinsic properties of the folded conformational subensemble for a sufficiently long time without being limited by the context of any surrounding molecules. Simulations of this sort help explain how the intrinsic flexibility of folded kink-turns is experienced and exploited by surrounding proteins and RNAs in the ribosome.[522,603,802,805,806]

A prominent and potentially flexible kink-turn 42 is located at the bottom of helix 42 in the helix 42−helix 44 (H42−H44) part of the large ribosomal subunit. This large RNA segment, which is also known as the GTPase-associated center (GAC) rRNA, moves during the elongation cycle. In the available ribosomal X-ray structures, the GAC region (also known as L11 or L7/L12 stalk or protuberance) adopts a set of "inward" and "outward" geometries with respect to the A-site of the large ribosomal subunit.[9] Its movement controls the position of the upper part ("head") of the H42−H44 rRNA relative to the body of large ribosomal subunit. Unfortunately, resolution of the available X-ray structures does not permit unambiguous determination of the exact source of observed mobility of the GAC RNA. In addition, X-ray structures may only capture part of the full system's flexibility, because the X-ray structures basically show it in random positions with different crystal packing arrangements. Consequently, different researchers have interpreted the available data differently. The observed flexibility of H42−H44 has been attributed to several factors, including: (i) hinge-like behavior and twisting flexibility of kink-turn 42, which was observed in MD simulations and coincides with the observed path of H42−H44 bending;[522] (ii) common A-RNA twisting flexibility of the lower part of H42, that is, the kink-turn C-stem;[10] and (iii) flexibility of noncanonical base pairs and RNA interactions at the base of H42.[9] However, the natural flexibility of kink-turn 42 and the other suggested movements can be suppressed by a tertiary interaction with a conserved guanine from helix 97, which is inserted into the NC-stem of the kink-turn between the "head" of the whole structure and the kink-turn. On the basis of the presence of this interaction, a newer MD study suggested that the observed flexibility might be best explained by directional flexibility of the type C three-way junction between helices 42, 43, and 44 above the kink-turn.[431] This junction's flexibility is anisotropic and spreads over multiple nucleotides (i.e., there is no easily identifiable hinge point), and coincides with the observed path.

It has also been suggested that kink-turn 38, which is located at the base of helix 38 (also known as the A-site finger), may enable dynamic behavior of this helix.[806] Helix 38 is the longest (110 Å) helix of the large subunit, and connects to the small subunit via dynamic intersubunit bridge B1a.[812−815] Cryo-EM studies revealed that the tip of H38 moves by more than 10 Å as the ribosome transitions from the ratcheted to nonratcheted state (Figure 40).[816,817] Mutational studies suggested that the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

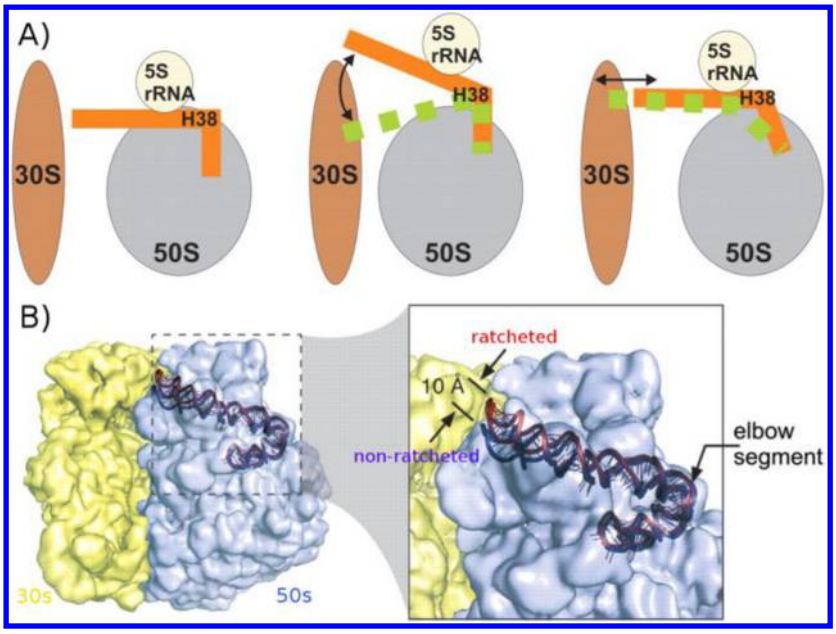


Figure 40. Elbow segments at the base of helix 38 (H38) can facilitate fast barrierless movements of the whole H38.[806] (A) Model of H38 substates as suggested by cryo-EM and MD data with H38 positioned with respect to the subunits and the 5S rRNA (left). Elbow elements at the base of H38 would support up-and-down (middle) and back-and-forth motions (right). (B) Cryo-EM density map of the 70S *E. coli* ribosome in a nonratcheted state with the H38 shown in nonratcheted (blue) and ratcheted (red) positions. A zoom-in view of the boxed region (right). Cryo-EM density maps of 30S and 50S subunits are outlined in yellow and light blue, respectively. Adapted with permission from ref 806. Copyright 2010 Oxford University Press.

flexible H38 plays a central role in attenuating translocation and dynamic signaling between ribosomal functional centers.[818−822] Intrinsic flexibility of kink-turn 38, which forms an elbow-like segment within helix 38, would nicely support the movements suggested by cryo-EM, although the experiments do not permit exact localization of the movement's origin. The elbow element could facilitate fast barrierless relocations of the entirety of helix 38 (Figure 40). However, the archaeal kink-turn 38 is not conserved in bacteria, which (as demonstrated by the examples of *E. coli*, *T. thermophilus*, and *D. radiodurans*) possess diverse and seemingly unrelated RNA elements in the positions corresponding to that of kink-turn 38.[806] There appears to be absolutely no conservation at the level of sequence or secondary structure within the bacteria kingdom, or between bacteria and archaea.[806] This is at first sight inconsistent with any putative functional significance of the elbow segment of helix 38. Strikingly, despite their complete lack of sequence similarity, all of these H38 segments ultimately fold into remarkably similar overall shapes (as demonstrated by structural data) and have similar levels of directional flexibility that closely mirror those of a kink-turn system (as demonstrated by MD simulation results).[806] Thus, important physical properties of RNA molecules (i.e., their overall topology and directional flexibility) can sometimes be conserved even in the absence of any conservation of not only in the sequence data but also of isostericity. One possible explanation for this in case of the elbow-like segment of helix 38 is that the ribosome merely requires the presence of an RNA sequence that is sufficiently amenable to bending into the required direction. Alternatively, it is possible that the sequence differences between the species reflect some fine-tuning of H38 elbow segment's properties. There are presently no data that could be used to say which of these scenarios is most plausible; this ambiguity illustrates the complexity of studying functional flexibility of rRNAs. The yeast ribosome has a similar topology around the H38 elbow segment,[11] suggesting that the topology of this segment is also conserved in eukaryotes.

Bioinformatics analysis of available structural data on flexible elements of the large ribosomal subunit suggested that pivoting positions for movements in the large subunit are typically associated with structurally weak motifs such as noncanonical pairs (primarily G/U wobble base pairs), bulge loops, and three-way junctions.[823] This indicates that RNA flexibility can be spread over multiple nucleotides rather than being associated with localized hinges. Such smooth distributions of flexibilities of (along) the molecules do not rely on highly specialized RNA motifs and may thus be more evolutionarily robust than those that do. This is consistent with several MD studies.[431,824] However, given the limited resolutions and inherently averaged nature of available experimental structures, further research into the distribution and properties of flexible elements in ribosome is needed. MD simulations are natural tools for studying flexible RNA elements. However, although simulations of isolated RNA elements such as those described above can identify the intrinsic flexibility of different RNA building blocks, they cannot be used to directly study coupling of those building blocks with their broader structural contexts inside the ribosome. Therefore, interpretations of such simulations may be ambiguous. More complete analysis of ribosome flexibility would require atomistic simulations of the whole ribosome, which, however, do not provide the accuracy required to understand its flexibility in detail, for a variety of reasons explained in section 4.6.

The crystal structure of the helix−loop−helix of the Azoarcus group I intron shows another bent RNA building block that superficially resembles the kink-turn consensus. However, this building block bends in the opposite direction; that is, it is kinked toward the major grooves of flanking helices. Additionally, there are no significant tertiary interactions between its stems. This structure has been named the reverse kink-turn,[825]

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

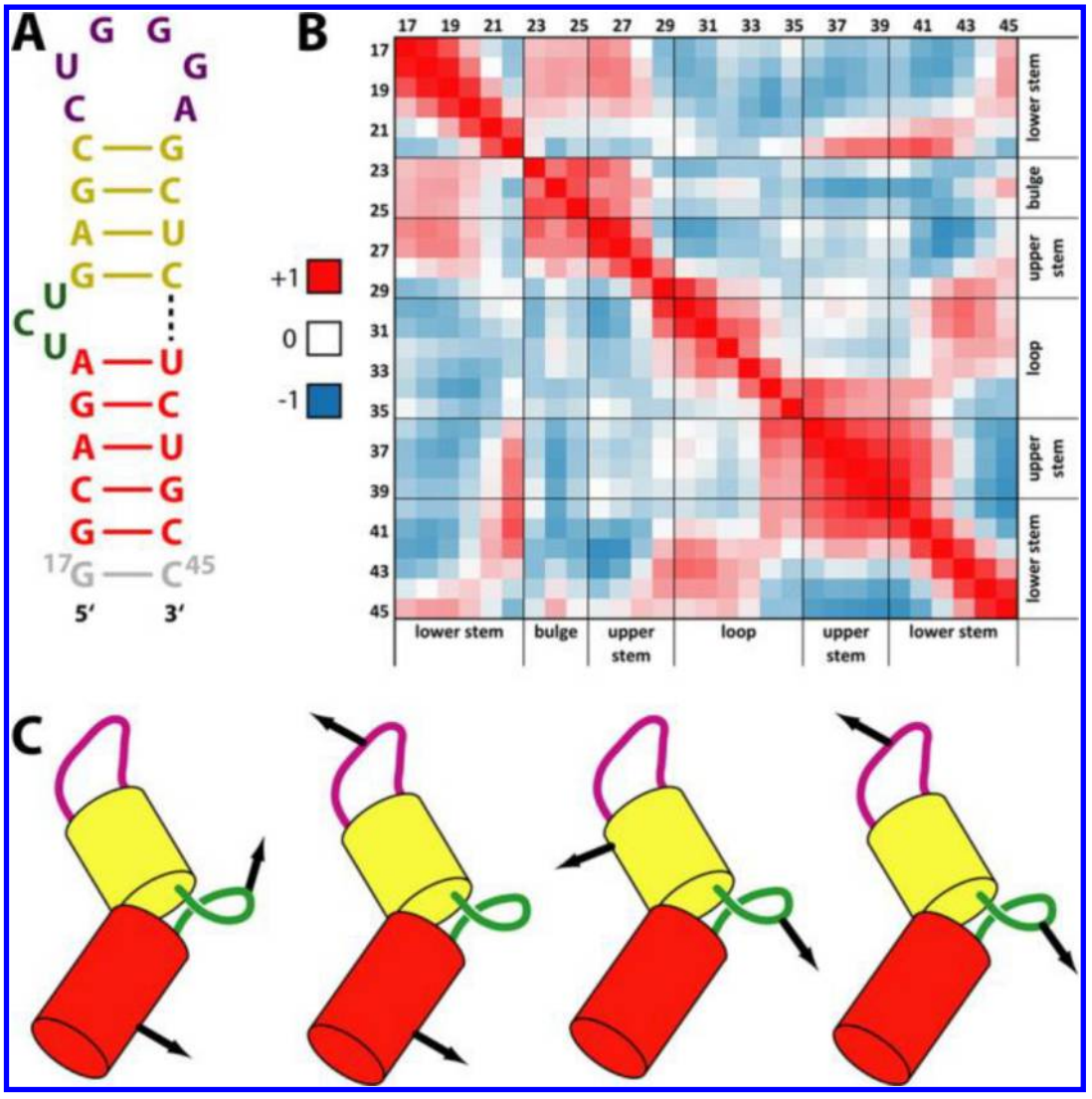


Figure 41. Quasi-rigid domains (QRDs) of TAR. (A) Calculated QRDs (shown by different colors) identified by MD simulations of a typical truncated TAR construct used in NMR experiments; the residue numbering corresponds to the full TAR RNA. (B) Residue-by-residue map of correlation matrixes for the C4′ atoms of TAR in MD simulations. Red and dark blue regions refer to correlated and anticorrelated movements, respectively. (C) Cartoon representation of correlated and anticorrelated motions of QRDs. Reprinted with permission from ref 831. Copyright 2014 American Chemical Society.

and has been characterized by extensive MD simulations.[604] The simulations supported the view that reverse kink-turn is, despite some sequence similarity, unrelated to true kink-turns. This study was one of the first microsecond-scale MD works on nucleic acids, and tested several different RNA force fields, ion conditions, and water models. It clearly demonstrated the necessity of using an RNA force field with a proper treatment of the χ dihedral potential, such as χ_{OL3}.[94] It also showed that simulations of some RNAs are sensitive to the choice of water model. The simulations were used to characterize the directional intrinsic flexibility of reverse kink-turn pertinent to its folded functional geometries, leading to the conclusion that it was the most flexible RNA motif studied to date using explicit solvent simulations. The simulated reverse kink-turns spontaneously and reversibly sampled many geometries, ranging from tightly kinked to flexible intermediates and extended unkinked structures. MD thus confirmed that, similarly to the 5′-UAA/5′-GAN internal loop (section 4.3.5), the native reverse kink-turn structure is primarily imposed by the surrounding context. Consequently, the reverse kink-turn is not an autonomous RNA motif. Conservation of the tHH A/A base pair that terminates the NC-stem with a flipped adenine is thus a consequence of the evolutionarily conserved outlying tertiary RNA−RNA interactions that determine the direction of the reverse kink-turn's bend. Reverse kink-turn topology is not compatible with the tHS A/G base pair characteristic of conventional kink-turns, suggesting these two RNAs are evolutionarily unrelated. Additionally, kink-turns are common but reverse kink-turns are quite rare.[826]

4.3.8. Depicting RNA Dynamics Using NMR and MD Data: The TAR RNA Element. The current understanding of RNA structures is primarily based on X-ray crystallography data. This means that our knowledge is inherently biased in favor of RNAs that fold into well-defined functional structures observed in X-ray experiments. However, many RNAs are inherently dynamic and in extreme cases may be unstructured like intrinsically disordered proteins. Therefore, to understand the behavior of many RNAs, it is necessary to move beyond static structures by describing their dynamic ensembles, preferably at atomic resolution. Unfortunately, the amount of

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

experimentally observable data is typically much smaller than the number of parameters that would be needed to capture the structures and populations of even just the most important conformers that define the free-energy landscape. Problems of this sort are best addressed by analyses that combine experiments with advanced MD simulations. One of the best methods for studying RNA dynamics in solutions is NMR, although it is fair to note that, for several reasons, the dynamics of RNA are much more difficult to study by NMR than are those of proteins. One of the reasons is low density of hydrogens along the nucleic acids backbone. In addition, most proteins are more "globular" than RNA molecules, which reduces robustness of structural determination methods based on NOEs.

An important model system for studying RNA dynamics and finding ways to link NMR and MD descriptions is the HIV trans-activation response (TAR) element.[62,303,346,630,827−841] TAR is a 59-nucleotide RNA hairpin essential for trans-activation of the viral promoter and viral replication (Figure 41). Although the TAR RNA has a defined secondary structure, it is a rather dynamic molecule. Its binding to the Tat protein is essential for activation of the long terminal repeat promoter.[62,833,838,840] The TAR RNA hairpin became one of the most studied RNAs due to its profound flexibility and potential as a target for antiviral therapy.[838] The available experimental data on TAR RNA are primarily derived from solution-phase experiments, but its size and flexibility have made it impossible to obtain a high-resolution structure of the complete TAR element. NMR experiments have provided some information on the properties of the apo form of the TAR,[62,837] but most studies have focused on its molecular complexes with cyclic oligopeptides.[346,630,833,835,837,839,840] These peptides mostly recognize the $U_{23}C_{24}U_{25}$ internal bulge and the sequences in its vicinity,[346,630,833,835,837,840] but they may also interact with the apical (N_{30}-N_{35}) hairpin (Figure 41).[835,839] In addition to structural insights obtained from various NMR experiments, analyses of primary NMR data on the TAR have yielded unique indirect information about its dynamics.[827,831,832,842] Space limits make it impossible to discuss these analyses comprehensively here. However, the data indicate that the TAR RNA is an archetype of a class of RNA molecules that cannot be satisfactorily represented by a single structure. While the apo structure has a large kink between its stems, the binding of the Tat protein is associated with a conformational change that produces a straight structure with coaxial helix stacking. Two potential mechanisms of action for antiviral compounds targeting the TAR element are direct inhibition of the TAR−Tat interaction and suppression of this conformational change.

In terms of its dynamics, the molecule can be divided into four segments: the upper and lower stem, the bulge between them, and the apical loop (Figure 41). The segment whose dynamics are believed to be most biochemically important is the bulge region, which functions as a molecular hinge permitting variation of the angle between the two stems. The interstem angles suggested for various TAR solution structures range approximately from 70° to 130°, although one must always treat quantitative structure determinations based on NMR experiments with some caution. Most studies consider the stems to be essentially rigid segments, although A-RNA helices are definitely not inflexible (section 4.3.9). To describe the dynamics of the stems and to interpret key residual dipolar coupling (RDC) NMR data, a model based on three interhelical Euler angles has been suggested.[836] In some studies, the helical stems were extended with additional base pairs (3−22) to obtain additional independent sets of RDCs and thereby improve the overall resolution of the NMR experiment.[173,668,836,842] RDC data are sensitive to submillisecond-scale dynamics, although they cannot be used to directly determine the time scale of a given dynamic process. Information on slow modes of conformational dynamics on the microsecond to millisecond time scale related to the presence of minor conformations (conformational "excited states"), which simultaneously remodel the structure of the bulge, stem, and apical loop, has been obtained using other NMR methods[843] including measurements resolving sugar pucker conformation and slow modes of sugar repuckering.[844]

It is to be noted that the apical hexaloop was in many above-described experiments replaced by the more stable UUCG tetraloop.

Natural flexibility of the TAR RNA limits the usefulness of X-ray crystallography; although a 1.3 Å resolution X-ray structure of a 27-nt model of the TAR bulge exists,[834] it is considered biochemically irrelevant because the flexible UCU bulge, which is important for molecular recognition, is immobilized by inner-shell binding of Ca^{2+} to all its phosphates. The structure is inconsistent with solution-phase data. However, even structures determined by conventional NMR experiments are unlikely to be fully representative because such experiments necessarily yield static, averaged models. The resulting PDB files indeed depict sets of structures compatible with the averaged NMR data rather than the true dynamic ensemble that is present in solution. To properly understand the structure and dynamics of TAR RNA, it will therefore be necessary to combine MD simulations with data from dynamic NMR experiments such as RDCs, as reviewed for proteins.[845]

The apo and holo forms of the TAR RNA have been studied extensively using MD.[303,346,827−832,841,842] In addition to conventional MD simulations, enhanced sampling[303,346,828] and Monte Carlo simulations of constrained (rigid) fragments[828] have been performed. Different groups have studied these systems using either the CHARMM or AMBER RNA force fields. Two studies reported extended simulations of TAR RNA and presented detailed comparisons of their simulated results to primary NMR data.[831,842] The first study[842] accumulated an 8.2 μs CHARMM36 trajectory using the Anton supercomputer.[73] Because of the huge amount of data generated in this trajectory, 10 000 snapshots were collected at 820 ps intervals. This interval between snapshots is noteworthy because just two decades ago, 820 ps would have been a respectable duration for a complete MD simulation trajectory. Such "coarse graining of the temporal resolution" of MD trajectories is becoming common or even unavoidable when processing long MD runs. However, it can lead to losses of useful information, for instance, when it is necessary to carefully follow a series of rearrangements to understand the origins of force-field imbalances[133] or to describe fast structural transitions. This time-resolution would definitely preclude analysis of dynamical hydration. The impressive 8.2 μs CHARMM36 simulation achieved a somewhat better agreement with the measured RDCs than an earlier set of 80 ns CHARMM27 simulations.[841] This may reflect improvements in both sampling and force-field parameters. However, discrepancy between the calculated and measured RDCs remained appreciably larger than the RDC measurement uncertainty.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

A considerably better agreement was achieved by processing the original MD ensemble using a Monte Carlo-based selection approach to extract a subset of MD-sampled structures that minimize the difference between measured and computed data. The resulting RDC-selected (sub)ensembles are assumed to better represent the atomistic dynamics of the real molecules because they integrate information from both MD and NMR.[842] Notably, an analysis of the sampled interhelical angles showed that many conformers in the RDC-selected ensemble occupied sparsely populated regions of the raw MD simulation trajectory. The RDC-selected ensemble exhibited, among other features, significantly larger bend angles between the helices than the raw MD simulation ensemble, and confirmed the occurrence of previously predicted large-amplitude bending and twisting motions.

Although the above-described sampling and selection procedure is a valid and potentially powerful way to correct the deficiencies of MD simulation trajectories, limited resolution of the experimental data may lead to "over-selection". This could cause population of unrealistic local structural features if a large force-field inaccuracy is the primary source of discrepancies between MD and primary NMR data. The RDC-selected ensemble suggested local dynamics in the bulge region (and its vicinity) involving rearrangements of stacking and H-bonding. These local dynamics, which are not detectable using traditional NMR techniques, may drive global changes in interhelical orientation, with transitions dominated by motions occurring on nanosecond-to-microsecond time scales. This relationship between local dynamics and global flexibility resembles the mechanism of flexibility suggested for kink-turns[522,805] (section 4.3.7). Such motions are detectable by analyzing RDCs but not spin relaxation data. The MD trajectory was also compared to NMR data using the maximum occurrence (MaxOcc) method, which analyzes averaged experimental RDCs by weighting contributions of a set of sterically allowed conformations.[832]

A series of 1 μs AMBER bsc0 simulations of the isolated TAR using starting structures taken from various NMR studies on apo and holo TAR variants were performed to re-evaluate the conformational selection hypothesis of TAR-ligand recognition.[831] The predicted unbiased structural dynamics were verified by comparison to the available RDCs,[846] order parameter (S^2) values,[830] and NMR structural data. In contrast to earlier simulation studies, including those with the CHARMM36 force field,[842] previously measured RDCs and order parameters were quantitatively reproduced. However, it should be noted that the reported direct comparison of the data from the two studies was based primarily on a visual assessment of a fraction of the RDC data set (cf., Figures S2 and S19 in refs 842 and 831, respectively). The bsc0 simulations confirmed that the TAR RNA undergoes a very rich set of dynamic processes occurring on time scales of nanoseconds to microseconds (the latter matching the duration of the simulations). Further analyses led to the suggestion that dynamics relevant to ligand binding may occur on a near-microsecond time scale. This suggestion appears to be consistent with the apparently central role of the stochastic hinge-bending motion of TAR, which occurs on a time scale of hundreds of nanoseconds. This bending motion is believed to induce conformations that facilitate Tat protein binding, as discussed by the authors in the context of the NMR data.[831] Because the relevant dynamics occur on time scales comparable to those of the simulations, it would be useful to perform much longer simulations with the χ_{OL3} force field (unpublished simulations using bsc0 conducted by some of us show that the collapse of the TAR RNA into a ladder-like structure might occur on submicrosecond time scales). The MD analysis suggested a visible coordination of the motions of the bulge and of the apical loop.[831] Because this very interesting observation is somewhat counterintuitive, it also merits further study by more extended simulations.

Despite evidence that the discrepancy between the raw MD CHARMM36 trajectory[842] and the experimental data for the TAR RNA arose primarily from differences in the description of global interhelical dynamics (good agreement was reported for backbone angles, pucker, and base-pair parameters), its physical origin could not be determined.[847] Thus, as part of their long-running effort to use NMR data and MD simulations to unravel RNA dynamics,[848] the group responsible for the CHARMM study shifted their focus to characterization of structural dynamics of the apical loop. In particular, they considered a UUCG-capped TAR sequence. They simulated this system using the χ_{OL3} AMBER force field instead of CHARMM.[849] The MD-generated ensemble qualitatively reproduced the measured RDCs, but the selection of a subensemble was required for quantitative reproduction. The largest discrepancies between the RDC-selected and MD-generated ensembles were observed for the most flexible TL residues and backbone angles connecting the TL to the helix, with the RDC-selected ensemble yielding more uniform predicted dynamics. A comparison of the RDC-selected ensemble to NMR spin relaxation data suggested that the dynamics occurred on time scales of picosecond−nanosecond, which was verified by analyzing relaxation dispersion data. The RDC-satisfying ensemble sampled some conformations adopted by the hairpin in crystal structures, indicating that intrinsic plasticity may play important roles in conformational adaptation. The approach used in this work would be very useful for development and back-verification of force fields, especially if similar data were obtained for other RNA motifs. However, when looking into the actual data in the paper, the reported geometrical variations of the TL appear rather subtle. For more discussion of the structural conservation of UNCG TLs, see section 4.2. In our opinion, the results obtained in this work are consistent with earlier analyses showing that the χ_{OL3} force field provides an improved description of UNCG TLs, but still has many flaws,[129,139,163] as discussed in sections 4.2.2.3 and 4.2.2.4.

MD simulations have also been used to study the binding of a short cyclic peptide to the TAR RNA.[303] The peptide was designed to mimic Tat, and an NMR structure[839] of the complex has been determined. The authors used steered MD simulations to bring the peptide close to TAR and vice versa. The "steered variable" was the net electrostatic interaction between the two molecules, which was computed using the Debye−Hückel approximation (section 3.2.5). This made it possible to drive the positively charged peptide toward the most negative region of TAR. Remarkably, these simulations predictively found the correct binding site. However, they also identified a second binding pose that was more stable than the one that matched the experimental data. This suggested that the protocol was not robust enough to distinguish between the two poses. Additionally, the authors combined forward and backward simulations to obtain absolute binding affinities, which proved to be significantly greater than those observed experimentally. Two possible reasons for this have been suggested: (a) the force field might be not accurate enough

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

to describe these strong electrostatic interactions, or (b) the simulations may have been too short and failed to properly sample the full free-state of TAR. The latter problem would lead to an underestimation of entropy in the unbound state and thus an overestimation of the complex's stability. This interpretation is also consistent with the typical time scales required to sample the free state discussed in the studies commented above.

Interesting insights into the principles governing TAR-Tat binding were obtained by analyzing the first high-resolution structure of an intermediate in the binding of TAR RNA to a cyclic peptide designed to mimic the binding of Tat while having a greater affinity for TAR.[346] The intermediate's structure was constructed by combining sophisticated NMR measurements with enhanced sampling MD simulations. The simulations were done with the χ_{OL3} force field, which was modified using measured RDC restraints, and utilized replica averaged metadynamics (RAM, section 3.2.6).[345] Results were verified by comparison to a number of independent NMR parameters, which were more accurately reproduced by the RAM ensemble than by the standard MD ensemble or the static experimental structure. The suggested binding intermediate was used to identify two mutants that were expected to affect interactions present in the intermediate complex but absent in the ground state. The predicted effects of these mutations were then confirmed by surface plasmon resonance kinetic experiments.

The authors subsequently used the RAM methodology with RDC data to study a 14 nt RNA hairpin containing the UUCG TL.[347] This study suggested that the UUCG TL exhibits rich dynamics and identified multiple substates that exist in equilibrium due to thermal fluctuations. However, in our opinion, the reported structural flexibility of the UUCG TL is somewhat counterintuitive (see section 4.2). We cannot exclude the possibility that the method may face some resolution limits, that could distort the balance between the RDC measurements and the force-field model, which is well-known to be far from being flawless for the UUCG TL.[129,139,163]

In summary, visualization of the TAR-Tat binding intermediate[346] is a textbook example of the emerging synergy between solution-phase experiments and modern MD simulations using powerful computers, improved force fields, and smart enhanced sampling methods. However, despite the potential of these synergies and the tremendous work done on the TAR RNA and the UNCG TL, the studies discussed above also show that obtaining unambiguous atomic-resolution insights into RNA dynamics using combined NMR and MD tools remains far from routine. The UNCG TL is in fact often assumed to be one of the stiffest RNA motifs, and to lack substantial dynamics! Therefore, it remains to be seen if it is possible to develop methods that would be capable of studying truly dynamical RNA molecules rearranging their 3D and 2D structures during their thermal dynamics.

4.3.9. Canonical A-RNA. The MD literature relating to DNA is dominated by studies on canonical B-DNA. This is not the case for the RNA simulation literature because of the different biochemical roles of canonical helices in folded RNAs, although A-RNA flexibility may be relevant in systems such as RNA-based nanostructures and double-stranded viral genomes.[850] Nevertheless, studies of canonical A-RNA are useful and are also important for benchmarking simulation methods. If one uses a stable simulation force field that does not cause substantial A-RNA degradation (either massive fraying or formation of ladder-like structures; see section 3.1.2.1 and Figures 7 and 8), canonical A-RNA simulations converge well within a few hundred nanoseconds.[137] This is because A-RNA simulations lack any alternative backbone substates that could retard convergence, in contrast to the B-DNA BI/BII substates.[115,146,149,851,852] We recall that this sort of convergence should be considered as a local one, and that there is no publication so far showing predictive hybridization of an A-RNA helix, which, with current force fields, may still not be the global free-energy minimum (see section 3.2.1).

Sequence-dependent local conformational variability is a key property of B-DNA. Because purely canonical helices in folded RNAs are short and canonical A-RNA variability is marginal as compared to variability of noncanonical RNA regions, it is not clear whether A-RNA sequence-dependent variability has any biochemical relevance. Nevertheless, MD simulation studies have been conducted to investigate this in A-RNA, revealing interesting sequence-dependent properties of A-RNA helices[137] that superficially resemble the variability of B-DNA.[851,853,854] The inclination of simulated A-RNA duplexes varied from 10° to 24°, and their major groove widths also exhibited substantial variability.[137] On the basis of simulations of sequences with modified bases (inosine and 2,6-diaminopurine), the calculations suggested that sequence-dependence of purely canonical A-RNA double helix is due to the steric shape of base pairs, that is, their vdW interactions.[137] The electrostatic component of stacking does not appear to affect the A-RNA shape. Especially visible is the role of minor groove amino group of purines. This is broadly consistent with the so-called Dickerson−Calladine mechanical rules suggested three decades ago for DNA double helices.[853,854] The simulations revealed that A-RNA helices are easily deformable in the direction of the inclination/groove width space. Another variable parameter is base pair roll, which is mathematically directly coupled to inclination through helical twist (i.e., the inclination and base pair roll are mutually corresponding helical parameters that are defined in the global and local coordination frames, respectively).[855,856] Thus, A-RNA can smoothly move between low and high inclination conformations on the basis of its environment.[137]

It is surprisingly difficult to compare these simulation results to experimental data.[137,857] The available A-RNA oligonucleotide X-ray structures are usually very short, affected by crystal packing, and often contain at least one noncanonical base pair. A-RNA helices in the ribosome are also very short, and their major groove widths show high variability with no apparent sequence-dependence. In other words, experimentally observed A-RNA groove widths may be affected by their environments, including the presence of partner molecules in the ribosome and crystal packing in the case of oligonucleotide X-ray structures. The only sequence effect that unambiguously emerges from experiments is a pronounced alternation of base pair roll in alternating −AU− sequences.[858] This is caused by balance of intrastrand and interstrand stacking in these sequences,[855,856] and is well reproduced in MD simulations.[137] It was also pointed out that given the substantial thermal fluctuations of the groove width seen in the simulations, it is not clear that the groove width can be adequately represented by a single static distance.[137]

Another study attempted to understand the effect of crystal packing on observed structures of canonical A-RNA by performing simulations of the alternating $r(UU(AU)_6AA)_2$ A-RNA duplex (PDB code 1RNA)[858] in a crystal lattice consisting

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

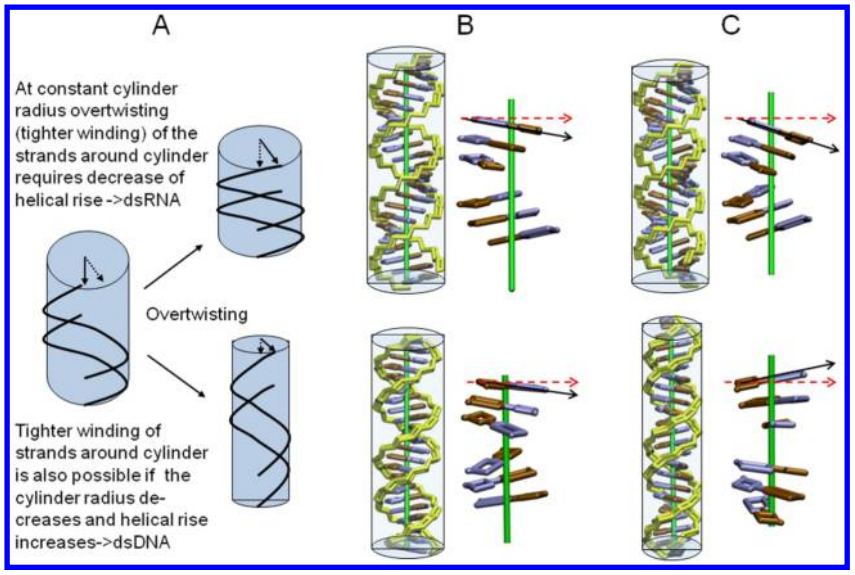


Figure 42. Model of stretch-twist coupling of A-RNA and B-DNA inferred from MD simulations. (A) Effect of tighter wrapping (overtwisting) of two simplified nucleic acid strands (black lines) around a cylinder (in light blue) at constant radius and with decreasing cylinder radius. The simplified RNA and DNA representations have been generated on the basis of the helical parameters sampled at low twist values (28°, B) and high twist values (34°, C). Backbone is shown in yellow, and bases of the two strands are in blue and brown, respectively. Enlarged blocks of three base pairs separated by approximately one-half a helical turn are shown without the backbone for clarity. In case of RNA (upper panels in B, C), the diameter of the double helix decreases only slightly and the length of the helix decreases with increasing twist. This is achieved with a minimum of steric strain by a more acute inclination (compare black and red arrows in B, C) and a slide motion (without disturbing the base pair geometry) resulting in a more narrow minor groove and compaction along the helical axis. For DNA, the increase of twist results in a significant narrowing of the helix (reflected by large change in *x* displacement). With a minimum of steric strain this is achieved by a change of inclination in the opposite direction as compared to RNA (compare black and red reference arrows in B, C), which extends the helix. Reprinted from ref 860; http://creativecommons.org/licenses/by/4.0. Copyright 2015 Oxford University Press.

of eight unit cells each containing four copies of the duplex, with an aggregate of 16 μs of data per duplex.[859] In principle, such simulations could be used to understand effects of crystal packing on A-RNA helices seen in different X-ray structures. The simulated A-RNA structure conformed more closely to the average structure seen in the crystals than to structures extracted from a solution simulation with the same force field. However, the integrity of the crystal lattice degraded slowly as the simulations progressed, causing growing heterogeneity at the interfaces between the chains.[859]

The flexibility of A-RNA has also been investigated,[860,861] using computational approaches similar to those developed to study B-DNA.[862−864] A-RNA and B-DNA actually have similar flexibilities, with comparable bending, twisting, and stretching deformabilities.[865] However, experiments with single-molecule magnetic tweezers revealed a striking qualitative difference between these canonical forms in that the twist-stretch coupling in A-RNA has the opposite sign to that of B-DNA.[866] dsRNA has a force-torque phase diagram similar to that of dsDNA, including plectoneme formation, melting of the double helix induced by torque, a highly overwound state, and a highly under-wound left-handed state. However, unlike dsDNA, dsRNA shortens upon overwinding, and its characteristic transition rate at the plectonemic buckling transition is 2 orders of magnitude slower than for dsDNA.[866] These observations were not explainable using existing models of nucleic acid double helix elasticities. The discrepancy was addressed in a subsequent MD simulation study:[860] unconstrained MD simulations on 16-base pair duplexes with modern force fields broadly reproduced the qualitatively different twist-stretch couplings for dsDNA and dsRNA, yielding semi-quantitative agreement with experiment. These results were confirmed by additional simulations in which an external torque was applied to the double helix to induce over- or unwinding of the double helix (Figure 42). The first two and last two base pairs of the duplexes were weakly restrained to limit their displacement in the plane perpendicular to the helical axis of the double helix. Such cylindrical restraints permit free rotation around the *z*-axis and therefore allow the duplex to display its full twist flexibility. This approach was designed to mimic the weak stretching force applied during torque tweezer experiments to align the duplex along a helical axis. An analysis of the observed helical deformations and their coupling to the twist suggested that the interplay of helical rise, base pair inclination, and displacement from the helix axis upon twist changes could explain the experimentally observed twist-stretch correlations. Overwinding of the RNA produced more compact conformations with a narrower major groove, and consequently reduced helical extension. Conversely, overwinding of DNA reduced the size of the minor groove, and the resulting positive base-pair inclination resulted in helix extension. An alternative MD-bases explanation of the twist-stretch correlation has been suggested more recently in a work that also investigated its dependence on the stretching force.[867]

Chemically modified bases can be used to tune thermodynamic stability and hybridization properties of DNA and RNA helices, and their effects can be rather straightforwardly assessed using MD simulations. Fluorinated apolar base analogues generally destabilize DNA and RNA helices and exhibit little binding sequence specificity.[868] They are considered to be isosteric substitutes for nucleobases that do not form H-bonds. An early nanosecond-scale MD study on incorporation of fluorobenzene (i.e., fluorophenyl) nucleobase analogues into A-RNA confirmed that these apolar universal bases can be neatly

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

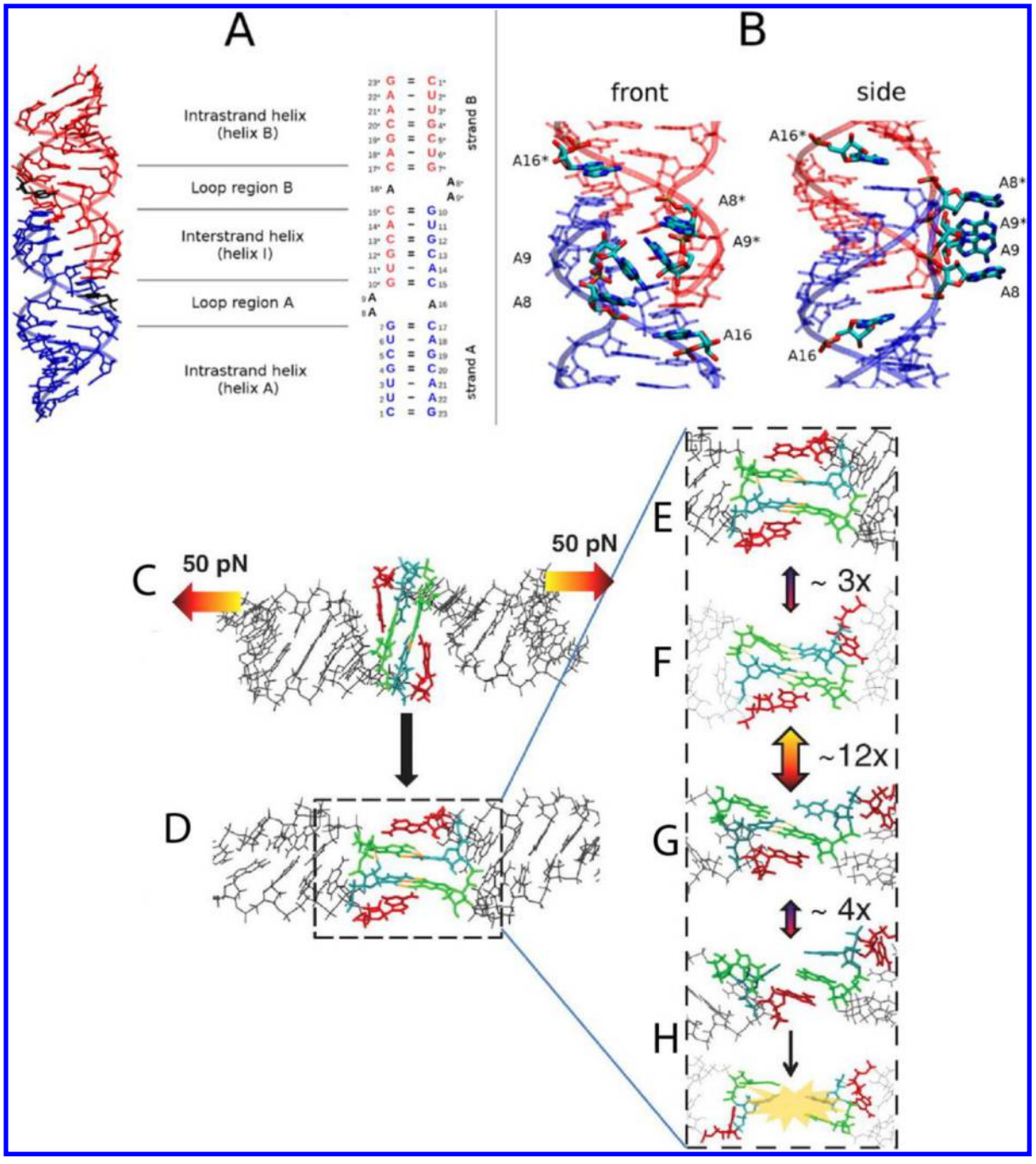


Figure 43. RNA kissing complexes. Top: (A) 2D and 3D representations of the HIV-1 subtype F kissing-loop complex in the X-ray structure (PDB file 1ZCI). The system is composed of two identical strands (blue and red; nucleotides of the red strand are marked by an asterisk). The two strands form stem-loop structures that interact by forming six-base-pair long interstrand helix I. Bulged adenines A8, A9, A8*, and A9* are omitted for clarity from the 3D structure, while adenines A16 and A16* are in black. (B) Detail view of position and numbering of the unpaired adenines, colored according to atom types. Reprinted with permission from ref 140. Copyright 2015 American Chemical Society. Bottom: Kinetic scheme for force-induced dissociation of the two-base-pair MMLV kissing-loop complex at 100 pN based on nonequilibrium MD. (C) The initial conformation of the kissing-loop based on the NMR structure; equal and opposite forces are applied to each strand. (D) Conformational rearrangement to the "bridging" form with H-bonds parallel to the applied force and stacked; the flanking adenine residues are in red. (E–H) The top adenine "unstacks" and "restacks" an average of three times. The top C═G base pair breaks and reforms an average of 12 times. The bottom C═G base pair breaks and reforms an average of four times, rescued by the bottom adenine stack. The final irreversible dissociation. Adapted from ref 888. Copyright 2012 National Academy of Sciences.

substituted into the A-form RNA helix.[869] The following pairs were studied: difluorobenzene:adenine (F:A), benzene:adenine (P:A), and F:F. The F:A-, F:F-, and P:A-containing RNAs all retained A-RNA structures. However, the duplexes with modified base pairs showed significantly larger local conformational fluctuations, as expected.

Free-energy computations using the TI method (see section 3.2.7) were later performed to systematically study free-energy effects of the presence of benzene, 4-fluorobenzene, 2,4-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

difluorobenzene, 2,4,5-trifluorobenzene, and 2,3,4,5-tetrafluorobenzene in the A-RNA duplex.[870] Thermodynamic stabilities of the modified RNA duplexes were also determined by thermal denaturation experiments based on UV absorbance measurements, assuming a two-state model for the transition from duplex to single strand. In these experiments, the nucleoside analogues were introduced into the tested sequences so as to form self-pairs. These measurements demonstrated that the pairing stability of fluorinated bases is higher for self-pairs than for pairs with natural bases. The stability of the pairs generally increased with the number of fluorine substituents in the base analogue. However, the greatest gain in stability was observed for the first two fluorinations. Additionally, the stabilities of RNA duplexes containing di-, tri-, and tetra-fluorobenzene self-pairs were similar to or greater than those for duplexes containing the canonical AU base pair at the same position.[870] It has been noted that DNA duplexes exhibit the opposite behavior, being heavily destabilized by the incorporation of a tetrafluorobenzene self-pair. All of these results indicate that the final thermodynamic stability of nucleic acid duplexes depends on an intricate balance of diverse contributions,[870] which is consistent with the results for the thermodynamics of RNA internal loops discussed in section 4.3.3. In other words, thermodynamic stabilities cannot be inferred from simple calculations of base pairing and stacking energies. Experimentally measured RNA stabilities were very well reproduced by the TI computations, which revealed that the at first sight counterintuitive nonadditive effect of increasing the number of fluorinated sites can be correlated with solvation energies and the solvent-accessible area.[870] This work nicely illustrates how theoretical computations can be used to analyze the complex balance of molecular forces in nucleic acids. Later, the same group performed extensive MD simulations with umbrella sampling to calculate pairing free energies between natural nucleobases and a series of fluorinated base analogues in aqueous solution. Although the studied systems lacked the context of a wider nucleic acid structure, a Watson−Crick-like orientation of the pairs was imposed by restraints.[871] These calculations provided further insights into the principles governing interactions of fluorinated bases in RNA and DNA duplexes, suggesting a pronounced enthalpy−entropy compensation during base pair formation.

4.3.10. Kissing-Loop Complexes. The genetic information in retrovirus virions is stored in two identical RNA molecules. The human immunodeficiency virus type 1 (HIV-1) packages its genomic RNAs as RNA dimers during the late stages of the viral replication cycle. This genomic RNA contains a specific sequence necessary for dimer formation known as the dimerization initiation site (DIS), and the first step of dimerization involves formation of a base paired dimer of two hairpins that is known as a kissing-loop.[597,598] Various kissing-loop complexes are present in other RNA systems and may be useful components of RNA-based nanostructures.[710,872−874]

The HIV-1 DIS (Figure 43) and some other kissing-loop complexes have been examined in several MD studies.[140,276,489,875−882] The history of MD simulations of RNA kissing-loop complexes illustrates the sampling limitations that arise when simulating even simple RNAs. RNA kissing complexes primarily feature canonical intermonomer base pairing, which is quite well described by the force fields used in most simulations provided that formation of ladder-like artifacts is prevented by using an appropriate χ parameterization (see section 3.1.2.1). However, their unpaired (bulged) nucleotides (Figure 43) represent a significant sampling problem. In fact, conformation of the unpaired bases has been a major focus of the simulation studies reported to date because experimental investigations have provided little clarity about this issue.

The NMR structures of HIV-1 DIS kissing complexes suggest that the unpaired purines are stacked inside the helices,[883−885] whereas available crystal structures suggest that they bulge out of the helices and are solvent-exposed.[597,598] However, bulging of the purines in the X-ray structures could be due to crystal packing effects, and it is not clear whether positions of the unpaired bases in the NMR structures are derived from the primary NMR data or are artifacts arising from the refinement protocol. While NMR studies agree that the unpaired purines are somewhere inside the kissing-loop pocket, they are mutually inconsistent regarding the bases' exact positions. NMR structures also generally exhibit structural deformations such as unrealistic values for the glycosidic χ angles of the unpaired purines.[140] Disagreement between the crystallographic and NMR structures complicates biological interpretations of the role of the unpaired purines.

One of the first MD studies of kissing complexes simulated the six-base-pair HIV-1 complex on a ∼5 ns time scale as well as a two-base-pair kissing complex on time scales of up to 16 ns with ff99 force field.[489] This study suggested that the kissing-loop inner pocket is a major ion binding site, a result that was confirmed by all subsequent studies. The results appeared to support the bulged-out position of the unpaired bases, and predicted a novel arrangement of four stacked bulged-out purines in the HIV-1 kissing-loop complexes that clearly differed from the two separate stacks seen in the X-ray structure[597] available at the time (which was used as the starting structure in the simulations). Several years later, this MD prediction was vindicated by a new X-ray structure showing a quantitatively identical conformation of the unpaired nucleotides.[598] The older X-ray structure appears to have been affected by crystal packing.

The picture painted by simulation results was complicated by a subsequent study on the HIV-1 kissing-loop complex that accumulated almost 1 μs of aggregated simulation data with ff99 force field, with individual runs extending to as much as 50 ns, and used the then-popular locally enhanced sampling (LES) method (see section 3.2.4).[276] This study still supported the bulged-out configuration of the unpaired bases, but because of a tendency to also sample bulged-in arrangements, it suggested that bulged-out geometries are likely to be only slightly preferred in terms of free energies.

The true scale of the sampling nightmare was revealed in the subsequent work on this system, which was based on ∼25 μs of accumulated simulation time primarily with χ_{OL3} force field.[140] Although this series of 1 μs-scale simulations was still very far from reaching convergence, it strongly suggested that all four unpaired bases of the kissing-loop complex were moving into the interior of the kissing-loop complex over time. This conclusion was based on a comparison of the short life times of the initial bulged-out conformations and the speed of the transitions to the bulged-in states as the simulation trajectories progressed, combined with the very limited fluctuations in the opposite direction. In other words, even if the bulged nucleobases could temporarily return to the bulged-out geometry in longer simulations, the subsequent short lifetime of the bulged-out conformations would keep their populations

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

low. When summarizing all of the studies, the simulation time scale had to be extended by almost 3 orders of magnitude (over a period of 12 years) to show the transition of the bulged bases of the HIV-1 DIS kissing-loop complex from the bulged-out to the bulged-in conformation.[140,276,489]

The ability to perform simulations over microsecond time scales represents a qualitative step forward in the application of MD to RNA, although even these time scales are short as compared to most dynamic processes occurring in nucleic acids. It is still notable that even the early short simulations provided several useful insights (including observation of the four purine stack and ion-binding pocket) that were confirmed in subsequent MD studies and experiments. For many other systems with well-defined experimental starting structures, short simulations can sometimes provide better results than more extended runs. This suggests that short simulations may often profit from a compensation of errors. Obviously, we do not advocate the use of short simulations in any situation; ideally, all trajectories should be extended as much as possible given the limits of the available hardware. However, quite often interesting insights can be obtained by studying the initial parts of trajectories, before the true complexity of the conformational space and the limitations of the chosen force field are fully expressed. Regarding the HIV-1 DIS kissing-loop complexes, we actually do not think that additional prolongation of simulations would provide any further decisive insights. For nucleic acids systems with different ensembles that are so close in free energy that they coexist or compete, it is unlikely that force-field approximations would allow a confident evaluation of the exact populations even when reaching fully converged sampling. Note that even in the field of the highest-quality gas-phase QM calculations, the principal accuracy limit with which one can determine relative energies of different conformations of systems having ∼20 atoms (such as parallel-displaced vs T-shape conformations of isolated benzene dimer or formation of an isolated base pair) is a few tenths of kcal/mol.[886,887] This is known as the chemical accuracy limit. After reaching this limit, attempts to further increase the accuracy are becoming ill-defined, oscillatory, and unproductive even upon extreme increases of computational costs. In other words, even for a single van der Waals or H-bonding interaction, we cannot expect that computations can provide entirely "exact" error-free numbers. Consequently, computations of large biomolecular systems where vast numbers of various diverse molecular interactions and energy contributions need to be simultaneously described are necessarily inherently less accurate than calculations of just a single interaction in complete isolation, irrespective of the quality of the used method.

The currently available experimental and MD data suggest that both bulged-in and bulged-out conformations of the unpaired bases in kissing complexes are biochemically relevant. The kissing complex may act as a molecular switch like the ribosomal helix 44 decoding center (section 4.5.2) or the 5′-UAA/5′-GAN internal loop (section 4.3.5). The kissing-loop complex actually shows sequence and structural similarities to the ribosomal helix 44 decoding center (and interacts with the same antibiotics), implying that its bases should be dynamic and sample coexisting substates that are very close in free energy. Thus, while crystallographic experiments may trap one of the possible conformations available to these systems, NMR ensemble experiments may provide information on mixtures of structures. The abundance of possible microarrangements sampled in the simulations suggests that the bulged-in conformation is a complex dynamic ensemble of structures, which could be the primary reason for the lack of clear NMR restraints on the bulged adenines.

Nonequilibrium pulling simulations (see steered MD discussed in section 3.2.5) with the ff99 force field were performed to follow up on the results of optical tweezer experiments on the two-base-pair kissing-loop complex of the Moloney murine leukemia virus (Figure 43).[888] To see structural transitions on the simulation time scale, the simulations were performed using applied forces of 100−400 pN, while forces used in experiment were 13.5−30 pN.[889] The resulting disruptions were observed over time scales of up to 400 ns. The simulations confirmed the suggestion that the external force causes the kissing complex to undergo a barrierless 90° reorientation of its core such that the intermolecular base pairs become parallel to the applied force.[888] This leads to an equal redistribution of the external force over both base pairs, which explains the system's initially surprising mechanical stability. The force-unfolding mechanism of a kissing complex is thus very different from that of hairpins, which are disrupted by an unzipping mechanism in which the H-bonds are broken one at a time. The simulations also showed that unpaired flanking bases increase the stability of the kissing complex by stacking on the intermolecular base pairs.[888] The stacked adenines are indispensable for the system's tertiary interactions: by shielding the tertiary base pairs from solvent and reducing their fraying, the stacked flanking adenines make terminal kissing base pairs act more like interior base pairs. The role of different flanking bases was investigated in more detail in a later study that combined MD simulations with new experiments using optical tweezers and electrospray ionization mass spectrometry.[890]

4.4. Riboswitches

4.4.1. Introduction to Riboswitches. Riboswitches are regulatory elements located in noncoding regions of mRNAs, mostly in their 5′-UTRs (untranslated regions), which control the expression of the associated genes.[891−895] Although they are found in all domains of life,[896] they are most abundant in bacteria, establishing them as potential targets for novel antibiotics.[33−37] Riboswitches primarily regulate gene expression by terminating transcription or inhibiting translation in response to the binding of a specific small-molecule ligand.[891−895] A notable exception is the *glmS* ribozyme, a riboswitch that regulates gene expression by catalyzing site-specific RNA self-cleavage facilitated by a glucosamine-6-phosphate cofactor (see section 4.8.3.4).

Transcriptional and translational riboswitches consist of two parts: an aptamer domain responsible for sensing and specific binding of the cognate ligand, and an expression platform that regulates gene expression. Binding of the cognate ligand leads to a change in the secondary and tertiary structures of the aptamer domain that is propagated into the expression platform. Transcriptionally acting riboswitches form a terminator loop that initiates transcription termination, while translationally acting riboswitches sequester the Shine-Dalgarno sequence, resulting in translation inhibition. Ligand binding can either suppress or activate transcription/translation.

Several riboswitch aptamer domains with bound ligands (i.e., the "holo" forms of the domains) have been extensively structurally characterized by X-ray crystallography and/or NMR spectroscopy. In contrast, only a few structures of the corresponding ligand-free apo-forms have been determined to

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

date. Therefore, most computational studies have focused on predicting structures of the apo forms and/or providing atomistic descriptions of the structural rearrangement of the aptamer domain induced by ligand binding or dissociation. At first sight, MD simulations can appear straightforward, and they can in principle yield unique mechanistic insights into the relationships between a riboswitch's structure, dynamics, and function. However, reliably describing the structural rearrangements associated with ligand binding is not trivial due to the long time scales over which these processes typically occur and the complexity of RNA conformations involved.

Several studies have used experimentally determined holo-form structures as starting points and attempted to predict the structural changes that the aptamer undergoes upon removal of its ligand. Careful comparisons between such simulations and experimental results for the ligand-free state can help to rationalize experimental findings at the atomic level. However, any native structural changes in the aptamer induced by ligand binding or dissociation will typically occur on time scales that are significantly longer (subsecond to second) than those accessible with contemporary atomistic simulations (microseconds). The structural changes observed in simulations must therefore be interpreted very carefully because they may be due to simulation artifacts such as force-field imbalances that are unmasked by long simulations based on a nonequilibrium starting structure obtained by removing the ligand from the structure of the holo complex. That is, the observed structural changes may be unrealistically fast or even wholly unrelated to those that occur in the real system if they result from force-field errors. In addition, as noted by several authors, the activity of at least some transcriptionally acting riboswitches is believed to be controlled by their folding kinetics.[71,72,167,897] Conformational switching typically occurs cotranscriptionally in the context of the RNA polymerase transcription complex, and the structure of an isolated aptamer domain or full-length riboswitch used in experiments and computations may be relatively insensitive to the ligand's presence or absence. In such cases, simulations of the structure obtained by removing the ligand may be irrelevant for studying the mechanism of riboswitch action. Later, evidence has emerged that assigns a similarly high importance to the cotranscriptional folding of some translational riboswitches, suggesting the riboswitch simultaneously regulates the accessibility of the Shine-Dalgarno sequence for ribosome binding and the rho utilization site (rut), binding site for rho-dependent transcription termination.[898]

Over 20 riboswitch classes have been characterized experimentally. However, computational studies have primarily focused on three groups of riboswitches: (i) purine riboswitches, including adenine-sensing and guanine-sensing riboswitches; (ii) $preQ_1$ riboswitches that sense 7-aminomethyl-7-deazaguanine ($preQ_1$); and (iii) SAM riboswitches that sense *S*-adenosylmethionine (SAM).

4.4.2. Purine-Sensing Riboswitches. Multiple crystal structures of purine riboswitches have been determined, revealing that all riboswitches of this class share a similar 2D and 3D architecture featuring three A-form helices (P1−P3) connected by a three-way junction.[899−907] The ligand-binding pocket is buried in this junction region (Figures 44 and 45). Loops L2 and L3, which cap helices P2 and P3, form a tertiary contact classified as a kissing-loop motif (section 4.3.10) with a Y- or wishbone shape; similar arrangement is also found in the full-length hammerhead ribozymes (section 4.8.3.1.2). The P1 stem partially overlaps with the expression platform and thus

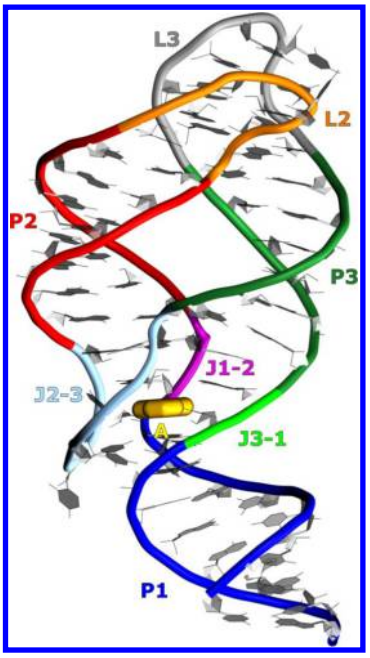


Figure 44. Structure of the A-riboswitch-adenine complex (PDB code: 1Y26). Different colors highlight the stems (P1, P2, and P3), loops (L1 and L2), and junctions (J1−2, J2−3, and J3−1) of the RNA structure. The adenine ligand is colored in yellow.

helps to transmit information about ligand binding. Although the overall structures of the aptamer domains are very similar, different purine riboswitches use different strategies for regulating gene expression. Both the *xpt* guanine-sensing and the *pbuE* adenine-sensing riboswitches act transcriptionally. However, whereas ligand binding to the *xpt* guanine riboswitch induces the formation of a terminator stem-loop that initiates transcription termination (i.e., switching expression OFF), the *pbuE* adenine-sensing riboswitch operates by switching gene expression ON.[908,909] Conversely, the *add* adenine-sensing riboswitch is a translationally acting riboswitch that exposes the Shine-Dalgarno sequence and thus activates translation upon ligand binding.[900,910]

Several computational studies have attempted to characterize structural dynamics of the apo-riboswitch upon ligand removal by performing simulations that start from an available holo-form structure with the ligand removed.[912−915] As noted above, such simulations must be interpreted carefully because their predictive power is limited by the time scale over which the rearrangement occurs and the difficulty of predicting the intricate apo-structures given the limitations of modern force fields. Despite these challenges, some studies have yielded interesting structural insights that have helped to rationalize experimental observations and suggested new experiments. For example, Priyakumar and MacKerell reported that removing the adenine ligand from the *add* adenine-sensing riboswitch increased the flexibility of the P1 stem and the J23 junction.[912] They also observed instability of the canonical base pairing between the adenine ligand and uracil U74 in the holo-form. The authors therefore suggested that this interaction is not essential for ligand binding but may somehow be crucial for ligand recognition.[912] However, this behavior could also be due to the use of CHARMM27 force field, which underestimates the stability of RNA base pairs (section 3.1.3 and Figure 8). The latter hypothesis is supported by the fact that this interaction was fully stable in similar simulations performed by

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

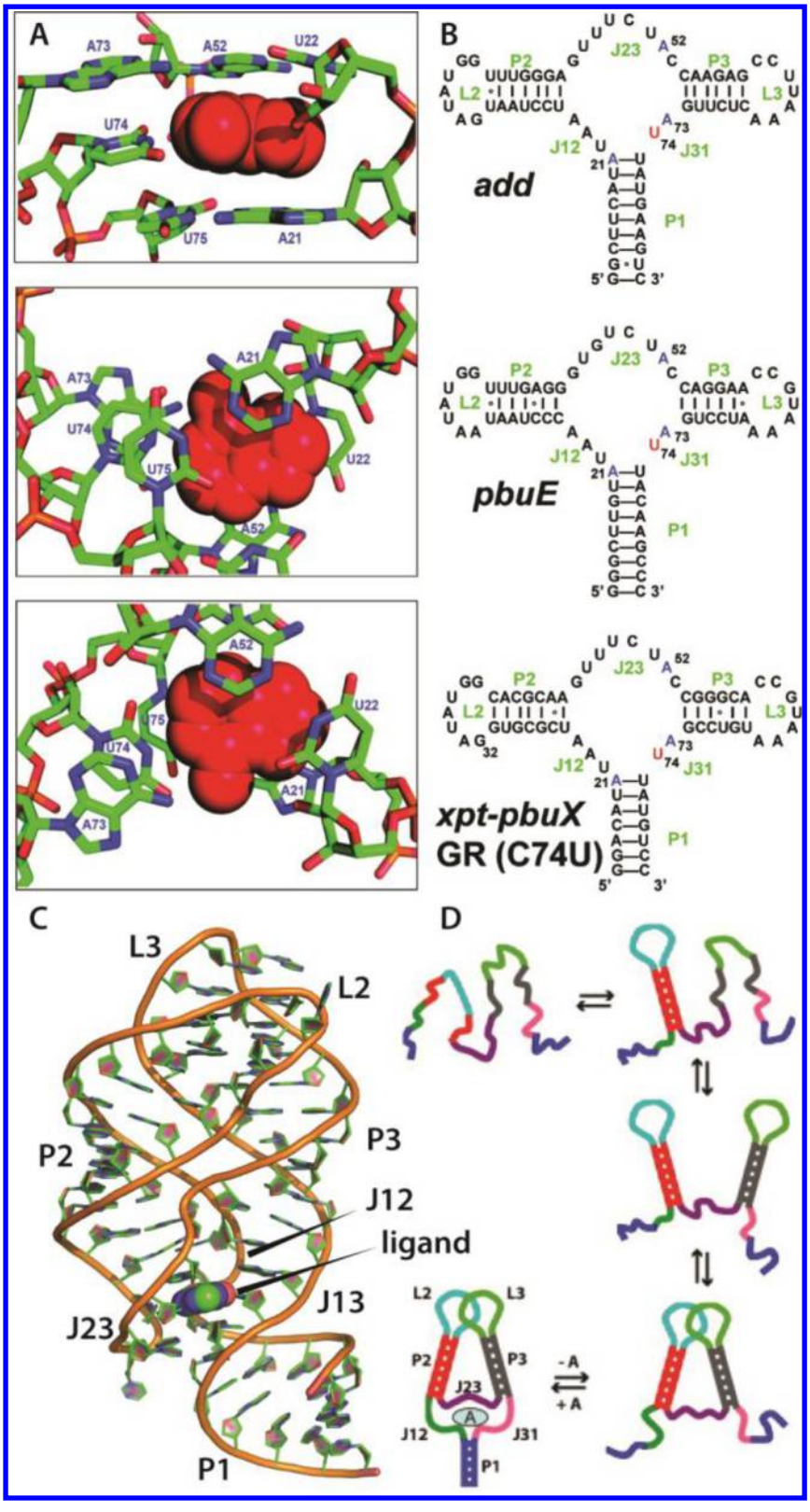


Figure 45. (A) Detail of the ligand-binding site, and (B) secondary structures of *add* and *pbuE* adenine-sensing riboswitches and *xpt-pbuX* guanine sensing riboswitch. (C) Typical 3D structure of a purine riboswitch (PDB ID 4TZY) involving three-way junction region binding its cognate ligand. (D) Suggested mechanism of folding of purine riboswitches. Panels (A) and (B) are reprinted with permission from ref 911. Copyright 2010 American Chemical Society. Panel (D) was adapted with permission from ref 912. Copyright 2010 Elsevier Ltd.

Gong et al. using the AMBER ff98 force field.[915] Villa et al. reported analogous simulations of the *xpt* guanine-sensing riboswitch after the cognate guanine ligand was removed, complemented by simulations placing the near-cognate adenine in the ligand-binding site.[913] Their simulations supported the hypothesis that U51 acts as a general docking platform for

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

binding purine ligands through hydrogen bonding with their N3 and N9 centers, while C74 is responsible for ligand specificity originating from WC interactions.[913]

To explain the mechanism of action of purine riboswitches more completely, their ligand-dependent folding pathways must be modeled and understood. To this end, several studies utilizing either classical MD simulations or enhanced sampling and coarse-graining techniques have been performed. Jain et al. combined classical MD simulations of three purine-sensing riboswitches with fluorescence spectroscopy experiments, using 2-aminopurine as the ligand.[911] They observed that, while the conformations seen in the crystal structures corresponded to global minima, their total populations in the simulated solution-phase conformational ensembles were only 50−60% because the ligand sampled multiple conformations in the ligand-binding site. Lin and Thirumalai used Langevin dynamics with a coarse-grained polymer model (see section 3.3.2) to probe folding mechanism of the *add* adenine-sensing riboswitch,[449] and predicted the folding of the P1 stem to be the final and rate-limiting step of the overall folding pathway (cf., also Figure 45D). In addition, they observed the P3 stem to be more stable than the P2 stem. This seemed to contradict single-molecule force spectroscopy experiments using optical tweezers, which suggested a higher stability of the P2 stem in the related *pbuE* adenine-sensing riboswitch.[449,916] This discrepancy was finally resolved when Lin et al. showed that the *add* and *pbuE* adenine-sensing riboswitches do indeed differ with respect to the relative stabilities of the P2 and P3 stems, demonstrating the fine-tuned stability of their secondary structures.[917]

In a later study, steered MD (section 3.2.5) was used to assess the stability of the terminal base pair of the P1 stem adjacent to the three-way junction in both apo- and holo-forms of a purine riboswitch.[298] The simulations suggested that direct interaction between the adenine ligand and P1 stem accounts only for a few kJ/mol and thus plays a smaller role than ligand-induced aptamer reorganization in conformational switching.[298] To put the experimental and computed data into the right context, one must recognize that the junction is expected to be significantly different in the apo- form,[906] and that MD simulations on the reported time scales cannot capture the large-scale rearrangement from the apo- to the holo-form, which affects the computed free-energy estimate.

Nonequilibrium MD simulation designed to capture the temperature-induced energy flow through the molecule suggested that ligand binding to the J23 junction of the *xpt* guanine-sensing riboswitch is a long-range effect that is coupled with formation of the kissing-loop interaction between the L2 and L3 loops,[918] like the coupling observed between the junction-based catalytic core and loops 2 and 3 of a full-length hammerhead ribozyme.[919] Similar coupling between ligand binding and global folding was suggested by Allner et al. for the *add* adenine-sensing riboswitch on the basis of umbrella sampling (US) simulations performed with the CHARMM36 force field, with the kissing-loop interaction being forced to open in both the apo- and the holo-forms.[284] These authors suggested that kissing-loop formation is anticooperative with ligand binding, leading to a $\Delta\Delta G$ of ∼10 kcal/mol.[284] However, these predictions are inconsistent with the experimentally observed cooperativity between kissing-loop formation and ligand binding.[920] A separate computational study using the AMBER χ_{OL3} force field did observe this cooperativity between the kissing-loop and ligand.[283] The authors of the latter study also discussed the limitations imposed by the decision to use the distance between the L2 and L3 loops as the collective variable (CV) in the US MD simulation of kissing-loop opening and closing. One such limitation is that CVs not explicitly related to the extent of opening may also play important roles in the process, effectively acting as slow orthogonal degrees of freedom.[283] The latter work thus provided an important discussion of the scope and limitations of the MD technique for riboswitch simulations (section 3.2.5). Together, these two works[283,284] can be seen as a textbook example of how results of MD studies may depend on various aspects of the simulation protocol and sensitivity of enhanced sampling methods to the completeness of the CV description that is chosen.

Finally, substrate specificity of the purine riboswitches was probed using molecular docking[921,922] and free-energy simulations.[923] Sund et al. applied the free-energy perturbation (FEP, section 3.2.7) approach to 14 different ligands in both adenine-sensing and guanine-sensing riboswitches, and obtained results suggesting that purine-sensing riboswitches exhibit a high degree of electrostatic preorganization of the ligand-binding site that favors their cognate ligands.[923] This was also supported by MM-PBSA free-energy calculations.[924]

4.4.3. PreQ$_1$ Riboswitches. The preQ$_1$ riboswitches regulate expression of several genes encoding enzymes participating in the queuosine biosynthetic pathway in bacteria. Their activity varies in response to the cellular concentration of 7-aminomethyl-7-deazaguanine (preQ$_1$), which is essential for tRNA maturation. Three different classes of preQ$_1$ riboswitches have been identified (Figure 46). Structural characterization of these classes by X-ray crystallography[925−929] and NMR spectroscopy[930,931] showed that preQ$_1$ riboswitches generally form a H-type or the closely related HL$_{out}$-type pseudoknot structure. The aptamers of class I and class II preQ$_1$ riboswitches partially overlap with their expression platforms such that the closure of the second helix of the pseudoknot upon ligand binding directly sequesters the sequence essential for gene expression. Unlike their class I counterparts, class II preQ$_1$ riboswitches have a small stem-loop inserted between the two pseudoknotted helices. Two different subtypes of class I preQ$_1$ riboswitches were identified: subtype I regulates gene expression via inhibition of translation, whereas ligand binding to subtype II induces transcription termination. In contrast, class III preQ$_1$ riboswitches have more separated expression and aptamer platforms. They utilize a unique mechanism in which ligand binding to the aptamer domain's three-way junction rearranges the orientation of the three flanking helices such that a docking platform comes within reach of the distal Shine-Dalgarno sequence and sequesters it to inhibit translation of the associated gene (Figure 46).[926]

The first insights into the folding mechanism of the class I subtype II preQ$_1$ riboswitch were obtained by Feng et al. using atomistic Gō-model (structure-based-potential) simulations.[933] These authors suggested that the riboswitch folds sequentially in the 5′ to 3′ direction; that is, the P1−L1 stem-loop folds first, followed by the formation of tertiary contacts with the L2 loop and then by P2 stem formation.[933] Similar folding pathways were suggested by all-atom simulations.[932,934−937] Petrone et al. used a combination of MD simulations and fluorescence experiments based on site-specific incorporation of 2-amino-purine to study structural dynamics of the class I subtype I preQ$_1$ riboswitch in its apo-form.[935] They proposed that the apo-form retains the pseudoknot structure of the holo-form, but has its ligand binding site occupied by nucleotides from the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

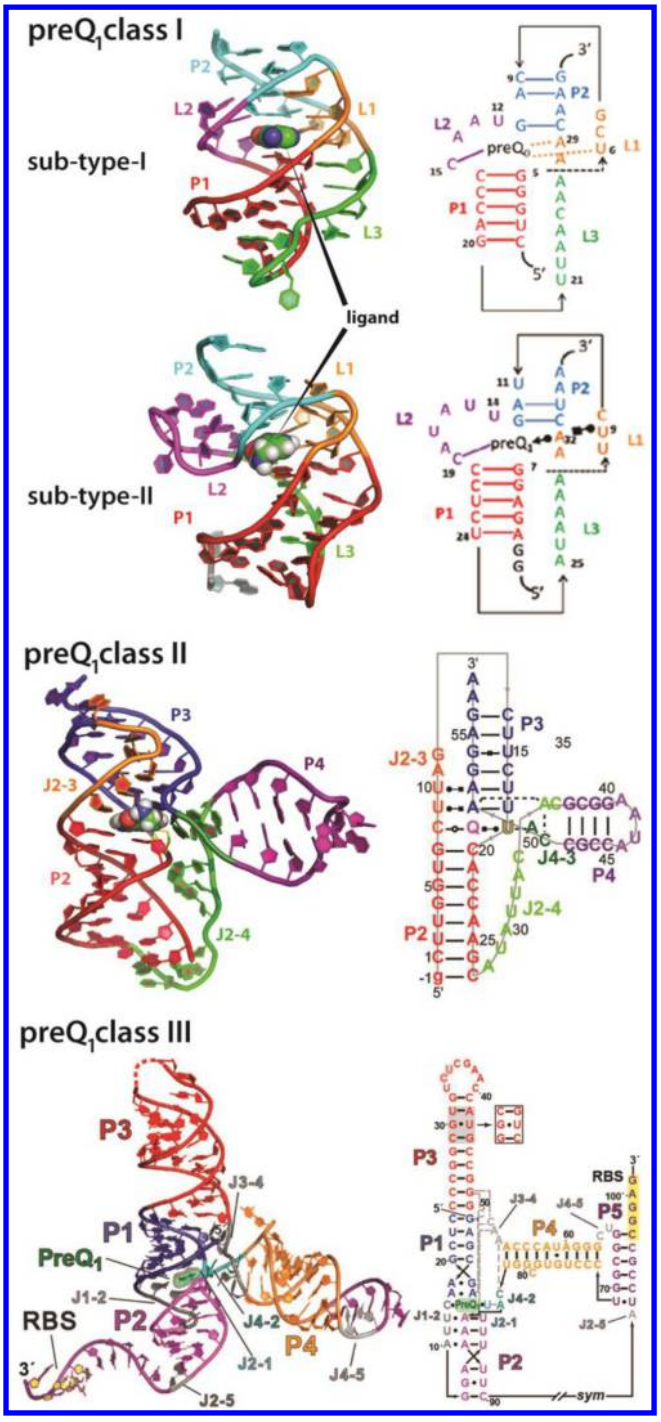


Figure 46. 3D (left) and annotated 2D (right) structures of currently known preQ$_1$ riboswitches. 2D structures of class I preQ$_1$ riboswitches are reprinted with permission from ref 932. Copyright 2012 Public Library of Science. 2D structure of class II preQ$_1$ riboswitch is reprinted from ref 930; http://creativecommons.org/licenses/by/4.0. Copyright 2013 National Academy of Sciences. Both 2D and 3D structures of class III preQ$_1$ are adapted with permission from ref 926. Copyright 2015 National Academy of Sciences.

L2 loop.[935] Conversely, on the basis of NMR experiments and MD simulations, Eichhorn et al. suggested that the A-rich L2 loop is rather flexible in the apo-form, adopting a single-stranded but highly stacked pseudohelical structure that more efficiently explores conformational space to facilitate ligand binding.[937]

The crystal structure of the class I subtype-II preQ$_1$ riboswitch apo-form was determined by Jenkins et al.,[929] showing that it retains the overall H-type pseudoknot fold, with adenine A14 of the L2 loop occupying the ligand-binding site. Interestingly, the P2 stem containing part of the Shine-Dalgarno sequence remained intact, so the riboswitch formally remained in the OFF state despite the absence of the ligand from its binding pocket.[929] This discrepancy was rationalized by MD simulations, which showed that the formation of a complete P2 stem was most likely due to crystal packing:[134] the P2 stem was indeed coaxially stacked with the same stem of a neighboring molecule in the crystal lattice, and removing this stabilizing crystal contact in the simulations disrupted the P2 stem, exposing the Shine-Dalgarno sequence to solvent as would be expected in the ON state.[134]

Structural dynamics of both apo- and holo-forms of a class II preQ$_1$ riboswitch were studied by Aytenfisu et al., who reported a destabilization of the H-type pseudoknot upon removing the ligand from its binding pocket, causing the Shine-Dalgarno sequence to become exposed to the solvent.[938] Coupling between ligand binding and formation of the P3 helix (which contains the Shine-Dalgarno sequence) was also suggested by an analysis of correlated fluctuations in the MD simulations.[939]

Finally, a unique mechanism of action was reported for the class III preQ$_1$ riboswitch on the basis of a combination of X-ray crystallography, MD simulations, single-molecule fluorescence resonance energy transfer (FRET) experiments, and ITC measurements.[926] As discussed above, this riboswitch is unique because it does not incorporate the Shine-Dalgarno sequence of the expression platform into the aptamer structure. Instead, central binding of preQ$_1$ causes the P4 helix to be positioned such that its capping J4−5 loop sequesters the Shine-Dalgarno sequence to form a new helix termed P5 (Figure 46).[926]

4.4.4. SAM Riboswitches. SAM riboswitches regulate expression of several genes involved in biosynthesis of sulfur-containing metabolites such as cysteine and methionine in response to the binding of *S*-adenosylmethionine (SAM).[940−944] Seven different families of SAM-sensing riboswitches have been identified (Figure 47). SAM-I riboswitch folds into a structure with a four-way junction surrounded by two sets of coaxially stacked helices, and regulates gene expression by transcription termination.[940,944−947] SAM-IV[948] and SAM-I/IV[949] riboswitches have similar binding pockets. In contrast, SAM-II riboswitch forms an H-type pseudoknot that sequesters the Shine-Dalgarno sequence upon ligand binding, regulating gene expression by translation inhibition.[950−952] Consensus sequences, structures, and ligand-binding pocket architectures of this family are all very similar to those of the SAM-V riboswitch.[953] SAM-III riboswitches adopt folds with a three-way junction, with the Shine-Dalgarno sequence being sequestered into the junction region upon ligand binding.[954−956] The last family, SAM/SAH riboswitches, regulates gene expression at the translational level upon binding to either SAM or *S*-adenosylhomocysteine (SAH).[896] Most computational studies on SAM riboswitches have focused on the SAM-I, SAM-II, and SAM-III families.

Several MD studies on the SAM-I riboswitch have compared the structural dynamics of its ligand-bound and ligand-free states.[957−960] While the SAM-bound state fluctuated stably around a conformation corresponding to the experimental crystal structure,[960] ligand removal was found to strongly affect dynamics of the J1/2 loop and the P1 helix, which overlaps with

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

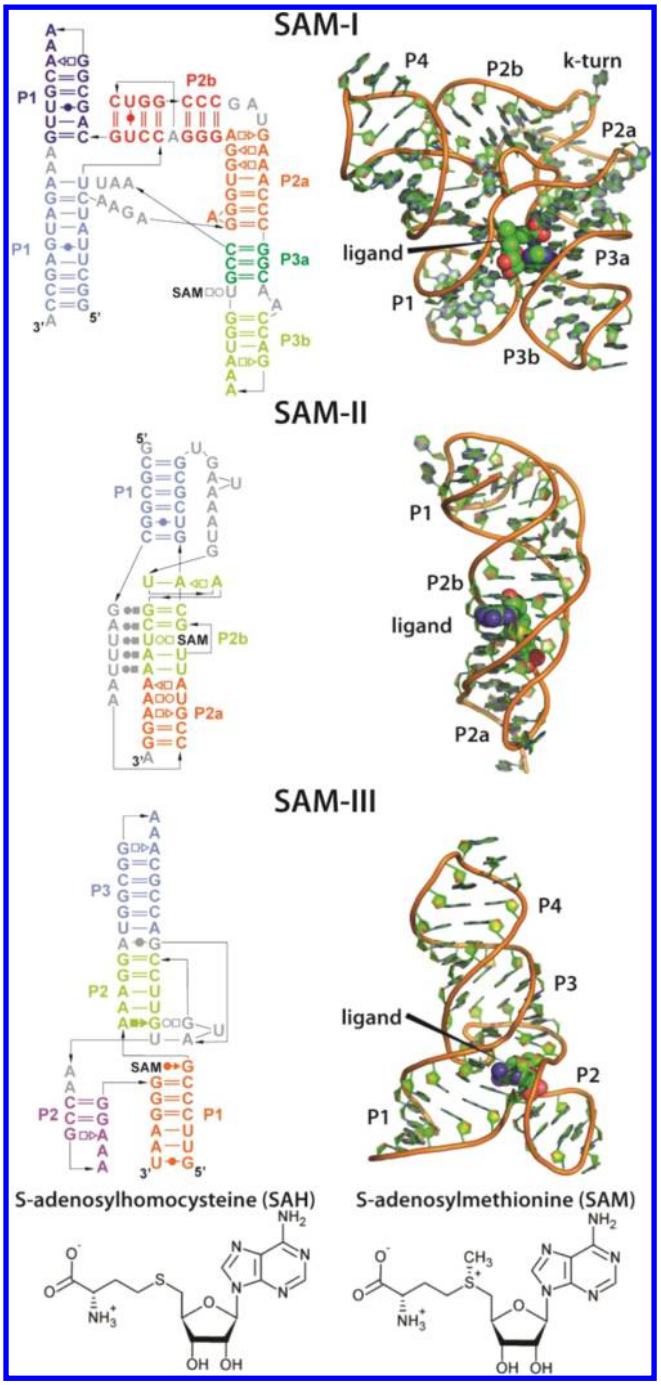


Figure 47. 3D (right) and 2D (left) structures of the most relevant families of SAM-sensing riboswitches: SAM-I, SAM-II, and SAM-III (S_{MK}-box) riboswitches. Chemical structures of cognate *S*-adenosylmethionine (SAM) and *S*-adenosylhomocysteine (SAH) are compared in the bottom panel.

the expression platform.[959] Stoddard et al. used a combination of chemical probing, SAXS, and T-REMD simulations (see section 3.2.4) to structurally characterize the apo-form of the SAM-I riboswitch.[957] Their work suggested that the riboswitch is prefolded in the apo-form, sharing certain tertiary interactions with the crystal structure of the holo-form but sampling an ensemble of conformational states.[957] Huang et al. constructed a model of the SAM-I riboswitch extended by the antiterminator helix.[958] They reported a spontaneous shift of three base pairs from the antiterminator helix to the P1 stem in the presence of SAM.[958]

Structural dynamics of the ligand-free SAM-II riboswitch was studied by Kelley and Hamelberg, who reported that after ligand removal the aptamer samples an ensemble of loose pseudoknot conformations including structures corresponding to the ligand-bound state.[961] Doshi et al. performed further simulations of the SAM-II riboswitch with cognate SAM and near-cognate SAH ligands in the ligand-binding site to rationalize its strong ligand preference for SAM.[962] They found that the positively charged sulphonium group on SAM acts as an anchor to the negatively charged ligand-binding site. Conversely, SAH binding was enthalpically disfavored and increased the pseudoknot's flexibility, exposing the Shine-Dalgarno sequence to the solvent.[962]

A similar comparison between the binding of the cognate SAM and near-cognate SAH ligands was presented by Priyakumar for the SAM-III riboswitch. This author also reported that the presence of SAH leads to partial liberation of the otherwise occluded Shine-Dalgarno sequence and its exposure to solvent.[963] However, in contrast to the SAM-II study, the selectivity of the SAM-III riboswitch was attributed to nonspecific interactions in the ligand binding pocket that favor SAM over SAH.[963] Finally, Suresh et al. performed ligand-free simulations of the SAM-III riboswitch that suggested that removing the ligand results in partial unfolding and weakening of the interactions between the Shine-Dalgarno and corresponding anti-Shine-Dalgarno sequences.[964]

In summary, atomistic MD simulations used in conjunction with other computational and experimental techniques have provided interesting insights into the structural dynamics of riboswitches. The main obstacle to characterizing these systems by MD is that their most interesting structural changes occur over longer time scales than are currently accessible in conventional MD simulations. Some of the dynamics observed in reported MD studies may thus be artifacts arising from force-field weaknesses, such as underestimations of base pairing strength (see sections 3.1.2.5, 4.2.2.3, and 4.2.2.4). Nevertheless, some of the observed structural changes may still closely resemble those occurring in the real systems, assuming that force fields correctly identify the weakest parts of the structures. However, the larger are the changes, the more carefully they should be interpreted. Typically, CHARMM simulations tend to predict more structural lability than AMBER simulations.

4.5. Simulations of the Functional Centers of the Ribosome

4.5.1. Introduction to the Ribosomal Decoding Center. The decoding center located on the small ribosomal subunit is one of the most sophisticated dynamic RNA/protein structures (Figure 48). It works with high speed and accuracy for an amazingly broad range of mRNAs and tRNA substrates,[3,965−968] performing two vital tasks: it accurately verifies the formation of canonical base pairing in the first two positions of the minihelix formed by the mRNA and tRNA at the ribosomal A-site, and it then transmits the resulting information to the other parts of the ribosome.[4,8,12,965−968] At the center of this decrypting device is the A-site segment of helix 44 of the small ribosomal subunit, which contains two universally conserved adenines (1492 and 1493) complemented by a single nucleotide in the other strand (the A1492−3 loop).

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

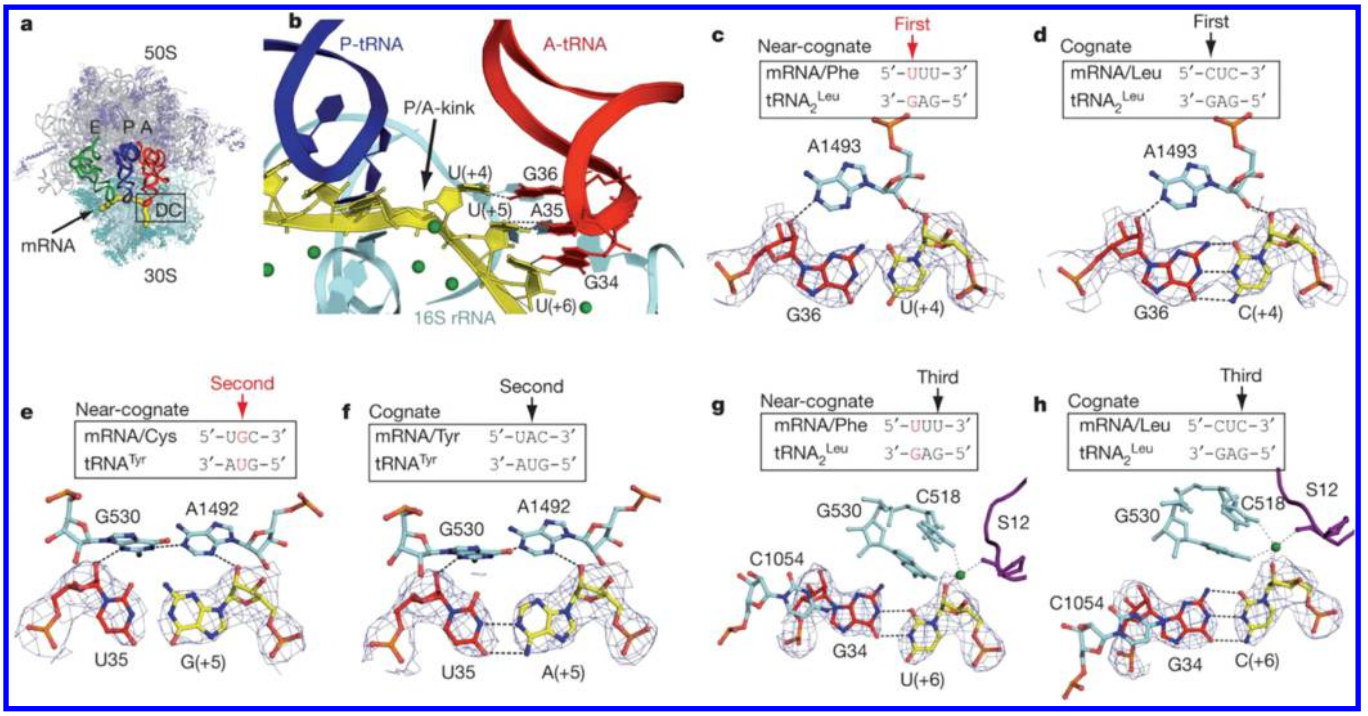


Figure 48. (a) The mRNA path on the 70S ribosome with the decoding area labeled as DC. (b) Close-up view of the mRNA P/A kink with near-cognate $tRNA_2^{Leu}$. Magnesium ions are in green. (c,d) The first base pairs of the near-cognate (c) and cognate (d) codon−anticodon duplexes in later X-ray structure[12] of a complete ribosome, and their recognition by A1493 of 16S rRNA via A-minor interaction. (e,f) The second base pairs of the near-cognate (e) and cognate (f) codon−anticodon duplexes and their interactions with G530 and A1492 of 16S rRNA. (g,h) Classical wobble G/U pair (g) versus canonical G=C pair (h) in the third position; a magnesium ion interacting with the base pair is coordinated by protein S12 and part of 16S rRNA. Note that the G/U base pairs in (c) and (e) adopt canonical-like geometry while the G/U base pair in (g) has the common wobble base pairing (see section 4.5.3 for a detailed explanation). Probable hydrogen bonds within 3 Å distance are indicated by dashed lines; $2F_o-F_c$ electron density maps are contoured at 1.2σ. Formation of the canonical-like G/U base pairs in the first two positions (c,e) interacting with the decoding center bases contrasts earlier observations of wobble G/U base pairs (the geometry shown in (g) or in Figure 2G) with decoding center in off-state.[3,4,969] This provoked a heated discussion in the literature (section 4.5.3). Adapted with permission from ref 12. Copyright 2012 Macmillan Publishers Ltd.

A traditional view of the role of the A1492−3 loop is that it performs dynamic proofreading of the codon-anticodon helix formed between the mRNA and tRNA; this view has been taken as an implicit initial assumption in most MD simulations of this system. In this scenario, the two adenines are predominantly stacked inside helix 44 in the ground state, although they also thermally sample other conformations. These adenines are activated by binding of the correct tRNA, and act in conjunction with the universally conserved G530 to monitor the exact shape of the codon−anticodon double helix, leading to on-state of the decoding center (a local configuration of the decoding center corresponding to binding of the cognate tRNA).[3,4,8] The decoding center has evolved to tightly discriminate between canonical base pairs and G/U wobble pairs (Figure 2G) at the first and second positions of the codon−anticodon double helix. Its capacity for discrimination stems from the energy penalty caused by the inability of the decoding nucleotides to efficiently interact with G/U wobble pairs leading to unsatisfied H-bond donors and acceptors.[3,4,969] With G/U wobble the A1492−93 loop is unable to stabilize the on-state and remains in the off-state. However, as discussed in section 4.5.3, this traditional dynamic model of the A1492−3 loop's function has been later challenged;[12] it was proposed that the decoding center of a complete ribosome might be so rigid that it can host only base pairs whose shapes are identical to the canonical ones (Figure 48). We will call these two mechanisms the statistical and static models of the A1492−3 loop's role.

The decoding center permits a variety of noncanonical base pairs to be accepted at the third (wobble) position, which allows a single tRNA to read multiple codons. The complementary wobble position 34 of the tRNA often exhibits sophisticated post-transcriptional modifications that tune the decoding process at the third position (see section 4.5.4). The complete decoding center also incorporates other RNA molecules and proteins; a comprehensive description of its structure is beyond the scope of this Review. The A1492−3 internal loop of helix 44 is its most heavily studied component because it is a small RNA internal loop that is responsible for direct discrimination between cognate and near-cognate tRNAs. It is also a target of the paromomycin class of antibiotics, which are assumed to stabilize the flipped-out conformations of the A1492−3 adenines by destabilizing their stacking inside helix 44. In addition to being visualized in numerous ribosomal X-ray structures,[3,12,965−968] the helix 44 part of the decoding center has been studied in isolation by X-ray crystallography, by looking at wild type, resistant mutant, human cytoplasmic, and mitochondrial sequences, and with and without bound antibiotics (Figure 49).[970−975]

4.5.2. Ribosomal Decoding Center: Small Models. The helix 44 part of the decoding center was investigated in a series of MD simulation studies designed to help understand the conformational space of the A1492−3 loop (i.e., the A-site

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

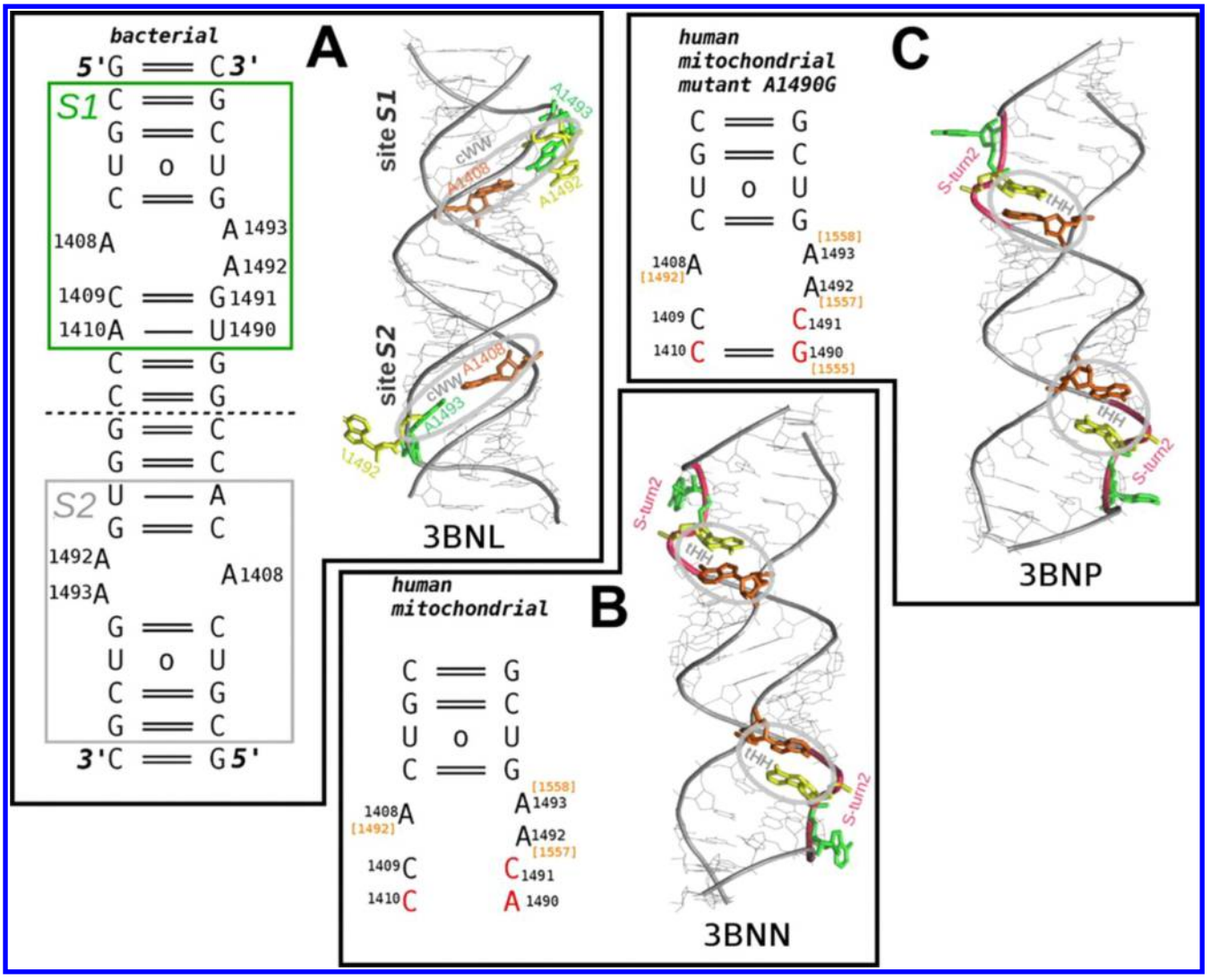


Figure 49. Crystallographic structures and 2D plots of selected A-site loop variants in oligonucleotide models. (A) The bacterial sequence (PDB: 3BNL) with two symmetrical A-sites S1 (green) and S2 (gray). (B) Human mitochondrial native sequence (PDB: 3BNN). (C) Human mitochondrial A1490G mutant (PDB: 3BNP). Sequence differences between the bacterial and human mitochondrial A-site are highlighted in red. The original nucleotide numbering as in human mitochondrial ribosomes is in orange. In the structure images, the gray ellipses and labels indicate the type of base pairing. The S-turn2 backbone is colored magenta, where relevant. A1492, yellow; A1493, green; and A1408, orange. Only one site is shown in 2D plots in (B) and (C). Adapted with permission from ref 980. Copyright 2014 American Chemical Society.

loop), its interaction with antibiotics, and the differences between bacterial, human cytoplasmic, and mitochondrial sequences.[131,239,773,976−983] Even from their earliest stages, all of the MD simulations suggested that the bulge adenines exhibit substantial flexibility and tend to sample a continuum of conformations rather than a few well-defined substates. The calculations supported the view that the bulged-out (i.e., flipped out) conformations of the A1492−3 bases are rarely populated in the absence of binding to a cognate tRNA or interaction with some ligand. Despite increases in simulation times, the application of enhanced sampling methods, and the use of more refined force fields, simulations have consistently shown that the kinetic barriers to the transition between the flipped-in and flipped-out states are low. Importantly, the results obtained appear largely independent of the chosen force field or enhanced sampling method. MD simulations of paromomycin bound to the helix 44 A-site indicated that the neamine part of the antibiotic is the main anchor for binding, whereas the sugar rings provide rather fragile contacts. Long-resident water molecules were also suggested to contribute to the binding event.[773]

The A-site loop of helix 44 was one of the first RNA systems on which extended T-REMD (see section 3.2.4) simulations were performed,[976] including a 15 μs run (320 ns per replica) for the gentamicin/A-site complex.[239] The simulations suggested that the decoding bases flipped on a time scale faster than that of gentamicin binding, supporting a conformation capture mechanism for the antibiotic's binding rather than an induced-fit model where the bases only flip in the presence of a ligand. The results suggested that the binding should not be viewed as a two-state process, because the ligand's dissociation involved shuttling between several metastable local minima on the free-energy landscape.

MD simulations were also performed to investigate the intrinsic properties of selected helix 44 A-site variants, including bacterial, eukaryotic, human cytoplasmic, and mitochondrial (native and A1490G mutant) sequences.[978−981,983] Simulations suggested sequence-dependent differences in the internal

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

mobility of the A-site, backbone conformations, and ion and water density distributions inside the binding clefts. Bioinformatics and MD analysis identified a specific RNA backbone conformation known as S-turn2, which occurs commonly in the structures of the human mitochondrial A-site but not in the bacterial variant (Figure 49).[980] When present within the helix 44 A-site loop, it spans from nucleotide 1491 to nucleotide 1494. S-turn2 was initially considered to be a variant of the S-turn backbone conformation,[795] a hallmark structural feature of the RNA backbone in the sarcin−ricin loop (see section 4.3.6).[528,529] However, the original bioinformatics analysis was incomplete because the S-turn was defined on the basis of the backbone conformation of just one strand. Later, a classification based on the conformations of five consecutive nucleotides in one strand and three in the other made it possible to clearly differentiate between the S-turn1 and S-turn2 (sub)motifs.[984] MD simulations in ref 980 were therefore complemented by a comprehensive bioinformatics analysis that identified numerous other S-turn2 instances in the large ribosomal subunit.[980] Analysis of their sequences did not reveal any consensus sequence pattern. The most weakly conserved nucleotide in the S-turn2 is the one that is flipped out (i.e., the third one, corresponding to A1493 in the A-site loop). Degree of sequence conservation in individual S-turn2 instances depends on their position in the ribosome. No obvious sequence preference for any nucleotide within the S-turn2 submotif emerges when the data for all known instances are grouped. The flipped out S-turn2 nucleotides form tertiary interactions with other parts of the rRNA, and are usually located in buried rRNA regions. Therefore, the S-turn2 was suggested to be a context-dependent structural element like the 5′-UAA/5′-GAN RNA switch (see section 4.3.5) and reverse kink-turns (section 4.3.7).[980] In contrast to the highly rigid "true" S-turn1 backbone conformation, S-turn2 is an easily accessible substate of a "common" RNA strand, that is, a minor conformation accessible to any single-stranded RNA. MD simulations clearly showed the lack of any relationship between the S-turn1 and S-turn2 conformations,[980] suggesting that the S-turn2 unit identified in the human mitochondrial A-site structures acts as a dynamic molecular switch that tunes the proper functioning of the A-site via tertiary interactions. The MD simulation data indicated that mitochondrial and bacterial A-sites show different propensities to adopt S-turn2 conformations due to the different base-pairing patterns of their flanking nucleotides.[980] This different propensity may contribute to the greater accuracy and lower speed of mRNA decoding in mitochondria relative to bacteria. Additionally, collective fluctuations of stacked A1408−C1409−C1410 nucleotides in the mitochondrial variant may modulate aminoglycoside antibiotic binding affinities.[980]

MD simulations were used to study the binding of antibiotics to the aminoglycoside modifying enzymes (AME) that enable bacterial resistance to aminoglycoside antibiotics.[985] A comparison of the aminoglycoside binding sites for several AME receptors revealed that the enzymes efficiently mimic the rRNA binding cleft. Although the internal dynamics of AMEs and their interaction patterns with aminoglycosides differ from those of the RNA binding site, the energy analyses suggested that key interactions in the enzymes may closely resemble those in the RNA.

Brownian dynamics (BD) simulations have been used to study binding of diverse antibiotics to the helix 44 A-site.[977] Although this alternative methodology has severe limitations (among other things, the RNA and encounter complexes are represented by a single conformation each, the internal mobility of ligands is not explicitly taken into account, and the continuum solvent approximation is used), it allows broader sampling of ligand association and dissociation pathways than is possible with MD methods. The BD simulations were performed to estimate diffusion-limited rates of association but also yielded visualizations of the way the antibiotic molecules "glide" around the RNA and predicted that antibiotic binding is only weakly dependent on the ionic strength. BD simulations were also used to study the kinetics and association of paromomycin to the entire 30S ribosomal subunit.[986] The resulting model suggested that specific binding at the ribosomal A-site starts with nonspecific antibiotic diffusion toward the 30S subunit and is followed by exploration of the subunit's surface.[977] The calculations did not suggest any specific long-range electrostatic navigation through the landscape to the A-site, but identified two entrances to the A-site around which the mobility of paromomycin is high. These entrances may help to funnel the binding process. The antibiotic also visited known binding sites for other drugs targeting the 30S subunit, and interacted with diverse positions along helix 44. This promiscuity of its binding may explain the overall tendency of paromomycin to stabilize the 70S complex. An alternative binding cleft in the 30S subunit may be related to the antibiotic's inhibitory effect on translocation. All of these results are consistent with the hypothesis that aminoglycoside antibiotics have multiple mechanisms of action, which was suggested on the basis of experimental findings.[987] Thus, while structural data strongly suggest that paromomycin affects the free-energy landscape of the A1492−3 loop, it also seems to interfere with translocation and other processes essential for proper elongation.[987] BD has also been used to study association of elongation factor G with the ribosome.[986] These BD calculations demonstrate that diverse computational approaches can complement each other in studies on various aspects of the ribosome's function.

MD simulations were also used to study the intrinsic flexibility of the complete helix 44 structure.[131] In ribosomal X-ray structures, helix 44 is bent into an arc. Further, cryo-EM experiments indicated that during EF-G dependent tRNA translocation, the A-site region of helix 44 moves toward the P-site by approximately 8 Å.[988] MD simulations revealed that the isolated helix 44 is intrinsically straight but undergoes pronounced and instantaneous bending processes due to isotropic thermal fluctuations. Simulation snapshots suggested that this helix undergoes spontaneous motions with even greater degrees of bending than are seen in the ribosomal X-ray structure. Helix 44 thus appears to have more than enough flexibility to tolerate the bending exhibited by its upper part in the cryo-EM studies within the genuine harmonic regime of elasticity. Thermal fluctuations of the simulated RNA were used to develop an elastic model of helix 44 in which the helix is treated as an extensible, twistable, and bendable anisotropic elastic rod. Analysis of local base-pair step deformabilities revealed a patch of more flexible base-pair steps in the upper part of the helix and the area proximal to the A1492−3 nucleotides, suggesting that this region has intrinsically enhanced flexibility.[131] Such coarse-grained elasticity models have been instrumental in studies of sequence-dependent DNA flexibility.[824,862−864] Although they are primarily applicable to canonical helices, they can also be used to describe other RNA segments that are assumed to exhibit functional flexibility.[824,989]

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Despite that the MD simulations discussed above provided important insights, we must caution that, although the helix 44 A-site loop appears to be a simple structure at first glance, it is very difficult to meaningfully simulate by MD even when treated in complete isolation. Unlike many other RNAs, no systematic analyses of the possible limitations and errors of simulations for the A-site loop of helix 44 are available. No specific problems were reported in the early publications, even though they relate to simulations performed with older force fields that should quite quickly develop simulation artifacts. Conversely, simulations of the complete helix 44 sequence performed by Reblova et al. using the ff99 force field predicted significant (~20%) populations of γ-trans states (which essentially correspond to a "crankshaft" conformation of the A-RNA backbone) along the whole helix.[131] Although formation of these states was reversible, they were long-lived and caused substantial local reductions of the helical twist. This behavior has been identified as undesirable, and demonstrated the usefulness of the bsc0[95] α/γ reparameterization (which was originally developed for DNA) in RNA simulations (section 3.1.2.1). Such γ-trans states are a common minor conformation in A-RNA X-ray structures, where they are associated with local A-RNA kinks likely stabilized by crystal packing,[858] or accompany some adjacent noncanonical features.[57] Therefore, the complete elimination of γ-trans backbone states from A-RNA in simulations using bsc0 is likely excessive; see section 3.1.2.1.[132,857] However, no better treatment of the γ-trans substates has been suggested so far. Nevertheless, the main uncertainty in helix 44 simulations relates to the description of bulged bases. Although this initially appears straightforward, it is actually very difficult to do accurately in terms of sampling, especially given force-field uncertainties (see sections 4.3.2 and 3.1.1.3). Because one needs to capture balance between substates that are roughly isoenergetic, even a small force-field bias could have a large impact on the calculated populations. Unfortunately, unambiguous benchmark experimental data that could be used to rigorously evaluate the descriptions of flexible bulge RNA regions[990] provided by various force fields are missing (see also the discussion on the difficulties of simulating kissing complexes in section 4.3.10). Precise sampling of bulges would require balanced descriptions of many difficult terms such as base stacking, solvation−desolvation, diverse intramolecular H-bonds, and noncanonical backbone conformations.

Inclusion of antibiotics adds further uncertainty to the modeling of the decoding center. Force fields for ligands are usually parameterized using "automated" procedures such as the general AMBER force field (GAFF).[991] While this is very convenient, it is a double-edge sword in terms of the accuracy and reliability of the resulting force fields. A more rigorous approach would be to prepare ligand force fields on a case-by-case basis using careful QM-based parameterizations. However, such procedures are vastly more time-consuming and require substantial skill.[372] In addition, the molecular and electronic structures of antibiotics are usually very complex, and it is not clear that they are fully describable within the basic approximations of typical force fields. It is particularly challenging to develop satisfactory force field parameterizations for the antibiotics' carbohydrate components, whose flexibility may have profound effects on binding.[992,993] In addition, the total net charge of aminoglycoside antibiotics (+5 in the case of paromomycin) presents a major problem. We consider the reliability of existing antibiotic force fields to be completely unaddressed in the literature.

An earlier comparison of the AMBER and CHARMM force fields for helix 44 simulations[978] illustrated the lower structural stability of A-RNA simulations performed with the latter force field, which is consistent with many other studies (see section 3.1.3). On the other hand, MD and free-energy simulations of the A1492−3 A-site loop in different studies suggested a broad qualitative consistency between the available AMBER and CHARMM RNA simulations.

4.5.3. Ribosomal Decoding Center: Larger Models. Even more challenging than simulations on isolated A1492−3 bulges (i.e., oligonucleotide models of the A-site, Figure 49) are MD studies involving larger segments of the ribosomal decoding center. Such studies are ultimately necessary for understanding the overall context of decoding, which cannot be obtained by studying isolated A1492−3 oligonucleotide constructs. Experimental studies have posed many fascinating questions about the decoding center that warrant computational analysis. For example, a series of newer X-ray structures on the complete 70S *T. thermophilus* ribosome including mRNA and near-cognate tRNAs revealed a striking incorporation of G/U base pairs at the first and second positions of the codon−anticodon double helix (Figure 48).[12,966−968] Amazingly, the G/U base pairs did not adopt the common wobble geometry, but appeared to be forced into a near-canonical (cognate-like, Watson−Crick-like) geometry. This can be explained either by a rare tautomerization (of either G or U) or by nucleobase ionization.[12,966−968] In line with suggestions of others,[994] we assume that formation of the guanine enol tautomer is the most likely explanation for the observed geometries because deprotonation of U or G would not resolve the close contact between the U(O4) and G(O6) oxygens, and the rare enol tautomers of U are less likely to form than that of G.[995−1000] The newer crystal structures are inconsistent with the earlier visualization of incorporation of the G/U base pair into the ribosome A-site, which showed the expected wobble G/U conformation and an inactive decoding center.[3,4,969] The earlier studies were done on the 30S subunit alone, and lacked a covalently linked mRNA in the P-site. Complete 70S ribosome structures show a much more tightly defined environment around the decoding center, including a sharp kink in the mRNA between the A and P sites (Figure 48b).[12,966−968] In fact, configuration of the decoding center around the "tautomeric" G/U base pairs is virtually indistinguishable from the arrangement of the decoding center when reading the canonical base pairs of the cognate tRNA. Importantly, the structures even featured domain closure of the small ribosomal subunit in the presence of the Watson−Crick-like G/U base pairs. Domain closure is a hallmark rearrangement of the small subunit involving head rotation and shoulder movement.[3,4] Domain closure movement in the complete ribosome appears to be similar but less pronounced than that in the isolated subunit.[12,966−968] Domain closure may be important for signaling to the large ribosomal subunit that a correct tRNA has been bound. These results collectively suggest that in the presence of Watson−Crick-like G/U base pairs, the small subunit in a complete ribosome (incorrectly) signals the binding of a cognate tRNA and the ribosome is fully competent to proceed with elongation.[12,966−968]

Similar Watson−Crick-like base pairing has previously been reported for the incorporation of GTP opposite to T in a crystal structure of human DNA polymerase λ. In this case, the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

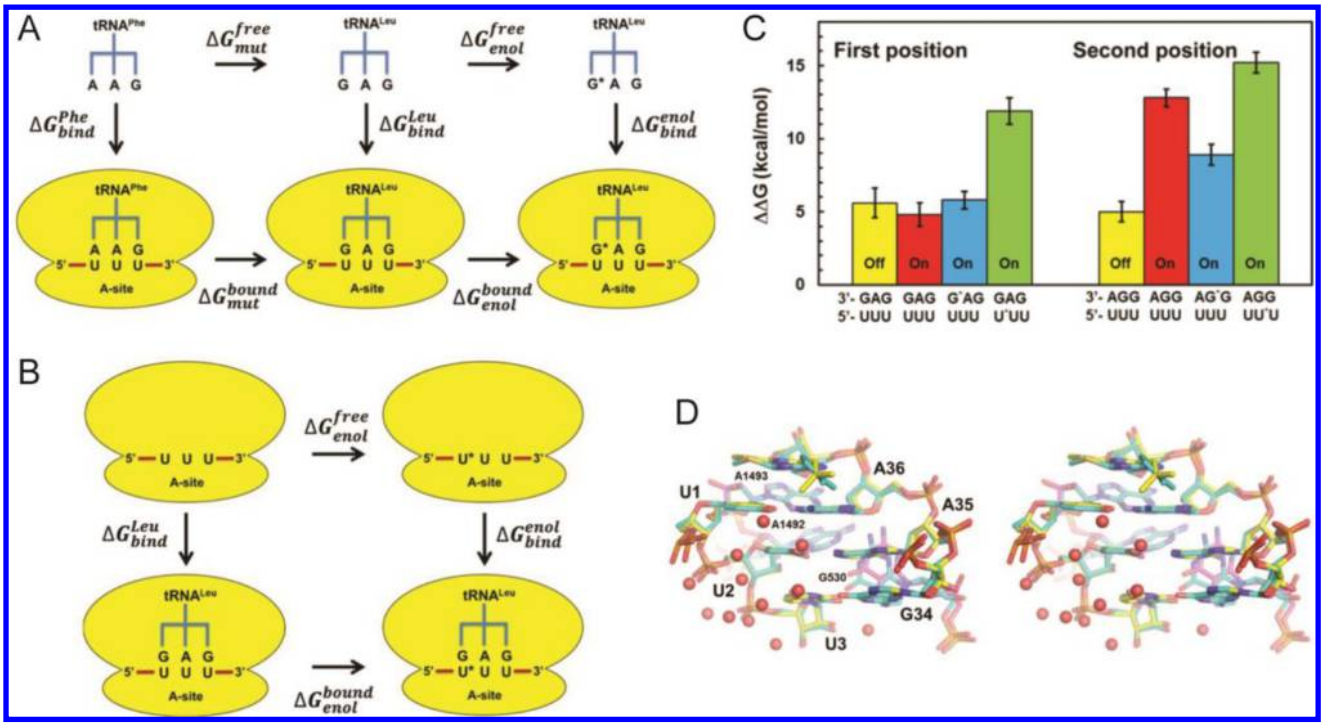


Figure 50. Thermodynamic cycles for evaluation of tRNA binding to the ribosome (see section 3.2.7). (A) The cycle used to calculate relative binding free energies between different tRNAs to the same codon (UUU), where G* denotes the enol form of G. (B) Thermodynamic cycle used for evaluating the effect of tautomerization of the mRNA codon for a given tRNA, where U* denotes the enol form of U. (C) Calculated energetics of mismatches in the first and second codon positions. Calculated binding free energies of $tRNA^{Leu}_{GAG}$ and $tRNA^{Ser}_{GGA}$ (in kcal/mol) are given relative to the cognate $tRNA^{Phe}_{GAA}$ complex with the UUU codon. MD free-energy calculations are carried out along the horizontal legs of the thermodynamics cycle in all cases, and the final $\Delta\Delta G_{bind}$ (C) is obtained from either $\Delta\Delta G_{mut}$ or $\Delta\Delta G_{enol}$ (i.e., the differences between the bound and free legs) together with the penalty for the enol tautomeric forms. The enol forms of G and U are again denoted G* and U*, respectively. Yellow bars denote calculations with the monitoring rRNA bases in their off-state. Error bars represent standard error of the mean. As the binding free energies in (C) are given relative to the cognate $tRNA^{Phe}_{GAA}$ binding, they are significantly higher than for the cognate complex; for a detailed discussion of the interpretation of data, see ref 994. (D) Comparison of the average MD and crystal structure for the cognate complex. Stereo view of the average MD structure (yellow carbons with the monitoring bases depicted with magenta carbons) overlaid on the crystal structure of the cognate $tRNA^{Phe}_{GAA}$ complex (cyan carbons). Water molecules in the vicinity of the codon−anticodon pair are shown as red spheres. Adapted from ref 994; http://creativecommons.org/licenses/by/4.0. Copyright 2014 Oxford University Press.

authors suggested that pH dependence of the misinsertion is consistent with ionization.[1001] Tautomer-like base pairing was also observed for an A/C base pair in a crystal structure of a high-fidelity DNA polymerase.[1002] Further, NMR relaxation dispersion experiments indicated that the common G/T and G/U wobble mispairs in DNA and RNA duplexes coexist in dynamic equilibrium with short-lived low-populated Watson−Crick-like mispairs that are stabilized by either rare enolic tautomers or anionic bases. Populations of the minor species were found to be from 10^{-3} to 10^{-5}.[1003] Such populations are consistent with numerous QM calculations showing that, although none of the natural nucleic acid bases forms stable rare tautomers in water, minor transient populations of these tautomers should be present.[998,1004] These independent observations of trapped Watson−Crick-like G/U, G/T, and A/C base pairs refreshed decades-old suggestions that rare tautomers may contribute to (or even determine) error rates in key biochemical processes responsible for coding and reading.[1005−1007] The main objection to the proposed biochemical relevance of tautomers stems from the substantial energetic penalty associated with their formation. These penalties were quite rigorously determined in earlier QM studies.[1004] The free-energy penalties of tautomerization are definitely larger than those associated with nucleobase ionization at neutral pH.[1008] U, T, and A should not tautomerize to any appreciable extent; G and C readily form rare tautomers in the gas phase,[995−1000,1004] but polar solvents strongly stabilize their major tautomeric forms.[998,1004] One might thus ask whether the observed rare tautomers are genuinely biochemically important or are just incidentally trapped in low-temperature X-ray crystallography experiments.

The difference between these two interpretations is as follows. Both of them agree that minor tautomers can be stabilized in the decoding center and thus observed. However, the original one assumes that the energy penalty of locally stabilizing the rare tautomer state is so large as to be incompatible with the overall free-energy balance of decoding or replication. Decoding also involves the transfer of information to the large subunit via different substates of the head of the small subunit, that is, the domain closure. According to the original interpretation, the error rate is primarily attributable to some pathways in which the bases are in their canonical tautomeric states and thus form a G/U wobble during the elongation process. Erroneous translation could proceed as a result of a rare event, with the decoding center occasionally reaching the on-state[3,4,8] even when the G/U wobble is present, or even when the decoding center is locally in the off-state if the whole subunit undergoes sufficiently large thermal fluctuations.[994] The newer interpretation assumes that the ribosomal decoding center is so

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

discriminative that it is absolutely incapable of binding the G/U base pair in its wobble form, leading to a zero error rate via all pathways involving the G/U wobble conformation.

Two comprehensive simulations and free-energy studies were performed to clarify the potential contribution of tautomers to ribosomal error rates using sophisticated models of larger fragments of the decoding site.[982,994] The first study analyzed ribosomal complexes with cognate and near-cognate tRNAs, with the aim of capturing the structures and energetics of G/U mismatches in the first two codon positions.[994] The simulations were done with the CHARMM22 force field using spherical simulation systems. The model system was derived from the crystal structure of the ribosome by defining a sphere with a radius of 25 Å centered on the N1 atom of the first codon position, and excising the structure enclosed by the sphere. The excised structure was then placed in a spherical water droplet with a radius of 37 Å, yielding the model to be studied. Water molecules at the surface of the droplet were subjected to radial and polarization restraints. Solute atoms 22−25 Å from the sphere's center were tightly restrained, while atoms inside an inner sphere of 22 Å radius that contained all of the important segments of the studied systems were unrestrained. MD simulations of this sort are less common than standard simulations using periodic boundary conditions. However, when studying a segment of the ribosome that cannot be meaningfully defined as isolated entity, conventional periodic box simulations are unsuitable. Free energies associated with different scenarios were evaluated using the free-energy perturbation (FEP) method (see section 3.2.7), revealing that tautomeric base pairs can indeed be trapped with the ribosomal decoding center in the on-state. The G enol form thus explains the crystallographic observations. However, the calculated overall free energies after adding the free-energy penalty of tautomerization suggest that the tautomer should not cause high codon reading error frequencies, as the calculated tRNA binding is too heavily destabilized relative to the cognate complex to cause significant error rates (Figure 50).

A similar picture was suggested by a newer study that combined fluorescence and NMR experiments with conceptually similar MD simulations using the CHARMM36 force-field variant.[982] The authors first performed two-dimensional umbrella sampling (US) free-energy simulations to obtain a reference free-energy landscape for A1492−3 flipping in an isolated helix 44 oligonucleotide model (see section 3.2.5). Pseudodihedral angles for flipping were used as the collective variables for the US calculations.[1009,1010] The calculations depicted the free-energy cost of extrahelical flipping (conformational excitation) of 1492−3 and showed that A1493 exhibits enhanced mobility. The free-energy cost of fully flipping A1492−3 into the active proofreading-like conformation was estimated to be around 7 kcal/mol in the isolated A1492−3 internal loop. Subsequent US simulations including paromomycin suggested that its binding reduces this cost to ∼3 kcal/mol, quantifying the free-energy change that leads to the loss of discrimination upon antibiotic binding. The authors then extended the US computations using a 25 Å spherical model of the decoding center with bound tRNA and canonical G/C, G/U wobble, or G/U tautomer base pairs in the first position of the mRNA/tRNA minihelix. Somewhat surprisingly, however, the authors chose to examine the enolic U form of the base pair, instead of the more likely G enol case. Full flipping of the A1492−3 adenines into a conformation capable of reading the codon−anticodon minihelix was associated with free-energy changes of ca. −1 and +2 kcal/mol for cognate and near-cognate tRNAs. Interestingly, upon calculating and adding the keto-to-enol tautomerization free-energy penalties, the obtained free energies again suggested that the "rare-tautomer" state was less favorable than either the cognate complex or the near-cognate complex with the G/U wobble. Thus, these calculations also suggest that the X-ray structures with Watson−Crick-like G/U pairs trapped a less stable state that is unlikely to give a major contribution to the overall error rate. The authors also proposed an overall free-energy model of tRNA selection and accommodation.

Note that, although both studies utilized seemingly similar spherical models, they employed different methods to calculate the free energies (see sections 3.2.5 and 3.2.7). The first study evaluated free energies via alchemical mutations and thermodynamic cycles.[358] The second study was based on calculation of changes in the free-energy landscape (reaction pathways) of the system with A1492−3 flipped out under different codon−anticodon minihelix scenarios. The qualitative consistence between the two studies is thus encouraging.

Although we assess the experiments and computations that have been performed to determine the role of tautomerization in the ribosome decoding and error rates, we take no position about which of the two views on its role is more likely to be correct. It is obvious that tautomerization must always be a rare event. However, once it happens in the decoding site, it is likely to lead to GTP hydrolysis on the elongation factor, accommodation, and translocation.[1011] An open question is "when does tautomerization occur, does it then permanently lock the tautomers in stable base pairing, and is it associated with the highest probability path to tRNA mis-accommodation?"

The X-ray crystallographic data are convincing and suggest that the ribosomal decoding center has a previously unanticipated capability to strictly enforce a WC-like geometry in the first and second positions of the codon−anticodon helix.[12,966−968] Structural evidence suggests that the decoding center has strong discriminatory power against both the wobble geometry (which is near-isosteric[40,47] with the canonical base pairing) and other undesired base pairs. It eliminates bulkier base pairs such as A/A, and penalizes the narrowing of the codon−anticodon helix that occurs in the presence of smaller base pairs such as U/U. Surprisingly, base pairing may also be closely monitored at the P-site, because even in this position the X-ray structures captured the tautomeric form of the G/U base pair.[966] This led to the suggestion that the tautomeric base pair may travel through the ribosome while locked in the tautomeric state from the early binding event.

It should be noted that the above-described computations that rather do not support the dominant contribution of the tautomers to the error rates may be affected by some uncertainties ranging from the used CHARMM force field (which in addition was used in its older version in the first study discussed) to the inability to describe the true coupling of the decoding center's conformations to the overall ribosome dynamics. Because it was only possible to perform calculations on medium-sized spherical models of the A-site, the simulations could not directly capture the coupling of the local action of the decoding center to domain closure. Additionally, the calculations may be sensitive to common uncertainties in various types of free-energy computations. For example, even when studying a simple isosteric guanine-to-inosine substitution in a short canonical A-RNA,[367] it was not possible to

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

achieve fully quantitative accuracy using a conceptually similar thermodynamics integration (TI) method (see sections 3.2.7 and 4.3.3). It is also not fully clear whether the equilibrium free-energy penalty of tautomerization of isolated nucleobases in water is fully transferable to the desolvated interior of the A-site of the elongating ribosome; we would like to reiterate that G readily tautomerizes in the gas phase.[996,1004] Finally, because the models were derived by excision of substructures from medium-resolution ribosomal structures, it is difficult to ensure their full equilibration when attempting to perform quantitative calculations (see discussion of the difficulties of obtaining good starting structures for protein/RNA complexes in section 4.7.1).

However, also the significance of the X-ray crystallography data is far from being "crystal-clear". One of the theoretical studies listed several reasons why tautomers might be trapped in crystal structures but not make a major contribution to the error rate.[994] For example, the low temperatures (100 K) used in crystallography may facilitate tautomer trapping. High Mg^{2+} concentrations (as used in crystallographic experiments) are known to increase error rates in protein synthesis and may support flipping out of the A1492−A1493 bases. The experiments were done in high excess of the noncognate tRNAs. These points do not yet seem to have been convincingly rebutted. As shown for systems such as small ribozymes, great uncertainties are common when attempting to use static ground-state structures or structural snapshots obtained by X-ray crystallography to analyze transient chemical processes that involve dynamic structural reorganization (see sections 4.8.3.2.2 and 4.8.3.3). Further, complete freezing of rare-tautomer base pairs since the early stages of the tRNA binding (i.e., the model assuming the preformed and rigid decoding center) is not in agreement with kinetic experiments on initial codon selection of tRNA in the ternary complex with the elongation factor.[1012] In addition, it would likely knock out subsequent contribution of ribosome proofreading mechanism to error reduction, which also contradicts biochemical experiments.[1012] This rigid-center model also does not explain the role of aminoglycoside antibiotics. The biochemical experiments thus suggest that if the tautomers contribute to the error rates, they would do so in a dynamical system allowing fast tautomeric form equilibrations and not in a process with frozen canonical-like tautomeric base pairing preventing relocations of the hydrogen atoms around the nucleobase rings. Further, cryo-EM studies support conformational mobility of the A1492−93 loop, which is consistent with the flexible model of the decoding center and dynamical error control.[1013,1014]

We speculate that errors caused by misreading of the G/U base pairs may occur via multiple pathways, and that all of the above-mentioned scenarios may coexist to some degree. This discussion resembles the never-ending debates about the relative importance of conformational capture and induced fit recognition mechanisms in biomolecular recognition: in reality, biochemical systems seem to often use a combination of both mechanisms, and such combined mechanisms may in fact be more evolutionarily robust. While the new X-ray structures featuring complete ribosomes[12,966−968] demonstrate the potential inadequacy of the earlier X-ray structures featuring only the small subunit,[3,4,8] they still do not (and cannot) show the ribosome when it is inserting the tRNA into the decoding site. This is probably a very dynamic process that occurs at temperatures much higher than those used in crystallography; moreover, the presence of the bound elongation factor that delivers the tRNA (which is not included in any currently available X-ray structure with the tautomers) may substantially affect the whole ribosome's sampling of its conformational space. We also feel that studies on ribosomal ambiguity mutations indirectly support some dynamic behavior of the small subunit.[1015] Thus, there are two limiting scenarios regarding the operation of the decoding center: the statistical scenario, in which the decoding center dynamically samples its conformational space; and the static scenario in which the decoding center is strictly structured. Both views are equally consistent with isosteric recognition of the codon−anticodon base pairs. It may be that the ribosome uses both scenarios in a multipathway process, and the balance between the two scenarios may vary during the tRNA binding process and depending on the external conditions. In conclusion, the contribution of rare tautomers to error rates in ribosomal protein synthesis remains one of the most challenging problems for future MD computations. The ultimate description of this process would require methods that reflect the ribosome's large-scale molecular dynamics. Without wishing to understate the power and importance of X-ray crystallography, inherently dynamic processes cannot be fully judged from static structures, and there is sometimes a tendency in the crystallographic literature to underestimate the need for Boltzmann sampling in the description of chemical processes.

Regarding further computational studies, it will be very important to see if calculations using other approaches, other force fields, and differently defined model systems provide the same quantitative results as the studies reported to date.[982,994] Studies on the potential involvement of tautomers in the mechanism of the DNA polymerase could be particularly interesting, because the polymerase is less complex than the decoding center of the ribosome.[1001,1002] Indeed, first free-energy perturbation study of DNA polymerase λ investigated incorporation of deprotonated and tautomer dGTP·dT base pairs into the active site.[1016] Spherical model of the polymerase has been used, with the center located at the initial position of the C4′ atom of dGTP. The simulations were performed with the AMBER $\chi_{OL4}\varepsilon\zeta_{OL1}$[100,146] force field and suggested that the canonical-like GT base pair with the G enol tautomer is equally stable as the wobble GT base pair in the context of the polymerase. The base pair variants with either thymine rare tautomer or deprotonated G^- were found to be less stable. The authors nevertheless cautioned, consistently with suggestions of our Review, that contemporary computational methodologies do not allow one to draw any quantitative conclusions from such computations due to the numerous approximations used. It is also important to point out that roles of rare nucleobase tautomers in the polymerase and the ribosomal decoding center may differ. In fact, different classes of polymerases may use different mechanisms to identify even canonical base pairs. While most of them definitely utilize the stericity of the base pairs as the primary discriminator, there are indications that some polymerases also monitor base pairing energies, that is, the direct strengths of the H-bonds.[1017,1018]

In summary, recent X-ray structures provide considerable indirect evidence suggesting a previously unrecognized role of the decoding center.[12,966−968] The suggested complex cascade of events involving the whole ribosome makes it even more difficult to reliably simulate ribosomal behavior than was previously thought. Instead of having flexible A1492−3 and G530 nucleotides that dynamically discriminate between canonical and wobble geometries of the tRNA-mRNA

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

interactions, the new structural data suggest that the decoding center (and the whole ribosome) has a fixed a priori structuration that forces the G/U base pairs into a "canonical" like geometry. The data indicate that A1493 is flipped out even in the absence of the A-site tRNA due to the steric effect of the sharp kink of the mRNA between the P- and A-sites, and the binding of the tRNA at the P-site. The A-site may then become fully structured after the nonspecific binding of the ternary complex (i.e., the tRNA and its elongation factor) to the ribosome. This could affect the complex balance of many interactions in the ribosome, including the B2a intersubunit bridge between helix 44 and helix 69.[12,966−968] The aminoacyl tRNA (aa-tRNA) anticodon would then enter a preformed stiff decoding center instead of a dynamic "proofreading" device. Instead of the A1492−3 loop, the anticodon of the tRNA would then extensively sample its conformational space to establish the codon−anticodon double helix. The error rate could thus be linked to the ∼4 kcal/mol free-energy penalty of tautomerization, which would reduce the binding free energy (due to the low probability of the tautomer being present at the right moment) of the near-cognate tRNA to the preformed stiff decoding site. However, as underlined above, this rigid model of decoding center contradicts many other experimental data. We still suggest that the statistical and static models of decoding center may to some extent represent two sides of the same coin, and, in reality, all of the involved partners may undergo substantial thermal sampling to establish the final interaction in a multipathway process. Complex chemical systems often do not act in a strictly single-pathway manner because the evolution of such single-pathway systems is less probable in entropic terms than that of multipathway reactions. The uncertainties in description of the role of monitoring bases, shapes of base pairs, formation of tautomers, domain closure, and numerous other factors involved in ribosome decoding are reminiscent of uncertainties surrounding the description of catalytic mechanisms of small ribozymes (see section 4.8.3.2.2). It has been recently suggested that the catalytic action of hepatitis delta virus ribozyme, a very simple system as compared to the ribosome, cannot be explained on the basis of a single structure, as it takes places on a very rugged free energy surface.[1019] It should be noted that even the existing kinetic models of the ribosome decoding, assuming participation of a few states, are likely oversimplified. In reality, the decoding might involve an enormously complex network of mutually interrelated and competing kinetic steps, which may lead to complex multiple-pathway overall process. As underlined by other authors, further experiments with improved structural and kinetic resolution will be essential to obtain a more accurate picture of the contribution of tautomers to decoding errors.[1012] Numerous real participating processes and states may merge in the current kinetics models into only few effective steps due to the natural resolution limits of experimental techniques. This is reminiscent of the description of complex multiple-pathway folding landscapes, where simple few-state models were shown to be often unrealistic, despite being formally in agreement with the available experimental data.[170] Advanced computations will almost certainly help, as complement to the experiments, to fully understand the decoding process.

4.5.4. Calculations on Other Ribosomal Centers. Simulations, free-energy calculations, and QM studies of reaction pathways using spherical models of the different centers of the ribosome have also been used extensively to understand other key aspects of ribosome decoding.[1020−1024] Free-energy simulations were performed to characterize different cognate and noncognate termination complexes.[1022] Simulations characterized the basic principles of decoding in all three codon positions and tentatively identified key elements responsible for the specificity of the release factors. They also revealed potential roles for several novel interactions, and emphasized the versatility of codon reading by proteins, which goes far beyond simple tRNA mimicry. Other simulations were performed to analyze the origin of discrimination between isoleucine (AUA) and methionine (AUG) codons in the decoding site, which relies on posttranscriptional tRNA modifications. The modification of the "wobble" C34 position of the tRNA anticodon was scrutinized by simulating the 2-agmatinylcytidine (agm^2C34) derivative found in archaeal tRNAs,[1024] and FEP simulations (see sections 3.2.7 and 4.5.3) were performed for cognate and noncognate ribosome-tRNA complexes. The starting structure of the AUG Met codon was obtained by mutating the AUA Ile codon inside a 25 Å spherical model of the ribosomal A-site. Calculations were performed for the archaeal $tRNA_2^{Ile}$ with the CAU anticodon, which would, without the C34 modification, readily read the Met AUG codon. Simulations were performed with and without the agm^2C34 modification, and with the UAU anticodon as a control. Starting structure was based on the crystallographic structure of the cognate 70S *T. themophilus* ribosome complex with archeal $tRNA_2^{Ile}$ accommodated in the A-site. The FEP simulations suggested that the tight local hydrophobic environment created by the side chains of the modification and supported by the ribosomal G530 penalized unsatisfied hydrogen-bond donors and acceptors for the AUG codon, yielding a large energetic discrimination of about 7 kcal mol^{-1} (cf., also comments about unsatisfied hydration sites in section 4.3.3). The effect of hydrophobicity was found to be bolstered by steric effects, and the formation of the cytosine imino tautomer was found to be unlikely.

Simulations based on the crystal structure of elongation factor-Tu (EF-Tu) ternary complex bound to the ribosome[1025] analyzed the role of the universally conserved EF-Tu histidine 84, which has been proposed to act as the general base in guanosine triphosphate hydrolysis.[1026,1027] Free-energy computations of different protonation states that could be involved in the activation of this reaction indicated that the histidine is unlikely to act as the general base, but must be protonated in an active conformation to enable GTP hydrolysis. The simulations also suggested that the sarcin−ricin loop of the ribosomal helix 95 spontaneously places the histidine in the catalytically active conformation (Figure 51). Conformational dynamics of the peptide plane preceding the conserved histidine have been proposed to be critical for facilitation of the reaction. The unusually high catalytic rates of EF-Tu were explained by a very large positive activation entropy of more than 7 kcal/mol at room temperature. It has been argued that this is a unique feature of translational GTPases, which achieve hydrolysis rates exceeding 500 s^{-1}. This entropy-driven mechanism may have evolved to ensure that the speed of protein synthesis is not limited by the rate of GTPase activity.[1028] A comparison with nonribosomal GTPases involving analyzing Arrhenius plots of the temperature dependence of the calculated free-energy profiles suggested that different mechanistic pathways are associated with distinct differences in activation entropies and enthalpies. The computed activation parameters have been extensively compared to experimental data. It also seems that

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

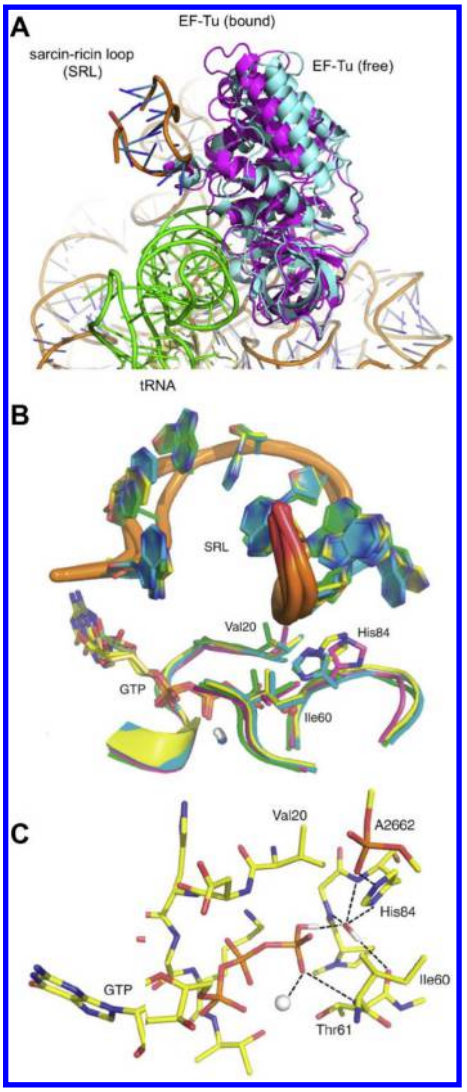


Figure 51. (A) Overall conformational change of the G-domain of EF-Tu when the ternary complex (aa-tRNA:EF-Tu:GTP) binds to the ribosome. This structural change can be seen by superimposing the tRNAs of the free ternary complex (PDB code 2C78) with that of the ribosome-bound complex (PDB codes: 2XQD and 2XQE). Adapted with permission from ref 1030. Copyright 2015 Elsevier Ltd. (B) Active site of EF-Tu during GTP hydrolysis as seen in the high-resolution crystal structure (green), and as computed by MD simulations during an initial prereaction state (magenta), an intermediate rotated state (yellow), and the final activated state (cyan). The sarcin−ricin loop (SRL, orange), His84, and Val20 and Ile60 (which form a hydrophobic gate) are highlighted. (C) MD structure of the EF-Tu catalytic center during GTP hydrolysis. Shown here is the resulting reactant complex after pre-equilibrium proton transfer from the catalytic water molecule to the c-phosphate. The resulting hydroxide ion is stabilized by the backbone NH groups of His84 and Gly83, the side chains of His84 and Thr61, and the protonated c-phosphate itself. The Mg^{2+} ion is depicted as a white sphere. Adapted with permission from ref 1026. Copyright 2013 Macmillan Publishers Ltd.

the details of the reaction mechanisms depend on the exact amino acid composition of the catalytic centers.[1029]

4.6. Atomistic Explicit-Solvent Simulations of a Complete Ribosome

Atomistic explicit solvent simulations of the whole ribosome represent a particularly attractive and challenging application of MD methods.[1031−1043] The first atomistic MD study of the whole ribosome (containing 2.46×10^6 atoms) simulated accommodation of the bound tRNA from the A/T state into the ribosome during decoding.[1038] The simulations required state-of-the art hardware with near-linear scaling to 768 processors, and were done on the nanosecond time scale, with a total sampling of 20 ns. One of the most delicate issues was the preparation of the starting structure, which was primarily based on a homology model of a previously reported complete *T. themophilus* ribosome structure model,[1044] with further structural refinements of the ribosomal A and P sites, modeling of the mRNA path, and some other necessary steps. The enormous size of the system and the limited time scale required the application of targeted MD simulation technique,[299] here with the ribosomal A/T state used as the starting structure and the A/A state as the target structure (see section 3.2.5). This work was the first to model the path of the tRNA during accommodation and identified ∼70 nucleotides of the 23S rRNA that may interact with the aminoacyl-tRNA (aa-tRNA) during the process, suggesting that the ribosomal A-loop and helix 89 play the dominant roles in guiding the tRNA.

Over the following decade, the sophistication of atomistic simulations of complete ribosomes increased substantially, and they were extensively combined with experimental studies. Series of 100−200 ns explicit-solvent simulations (3.2 million atoms, with an aggregate simulation time of 1.4 μs) combined with a Gō-type biased (structure-based) atomistic model and supplemented by single-molecule fluorescence resonance energy transfer (smFRET) experiments were performed to investigate the balance between configurational entropy and effective enthalpy during the accommodation process (Figure 52).[1035] This balance is suggested to lead to spontaneous reversible accommodation attempts. It has been proposed that the configurational entropy of the 3′-CCA end of the aa-tRNA opposes accommodation, leading to a multistep accommodation process involving multiple parallel pathways.

A 1.3 μs atomistic explicit-solvent simulation of the *E. coli* ribosome in a pretranslocation state was used to separate ribosomal movements according to the time scales on which they occur.[1032] To capture the collective rotational dynamics of the ribosomal subunits and separate them from fast thermal fluctuations of individual residues, the authors used an iterative protocol to identify a semirigid ribosome core that exhibits sub-angstrom internal displacements. The study identified several core regions (the 50S subunit, and the head and body of the 30S subunit) that moved relative to one another in a coordinated manner. Rotations and tRNA movements were measured using carefully selected reaction coordinates. 30S subunit body and head rotations during translocation over the course of the MD simulation were observed on scales of −2° and 5°, respectively. These movements are smaller than the functional angular displacements indicated by experiments (7° and 20°, respectively), suggesting that simulations cannot yet capture functional dynamics accurately. This may be related to the time scale of the simulation, uncertainties in the starting structures, or force-field deficiencies. Regardless, the data were successfully used to describe the 30S subunit body rotation, head movement, and tRNA displacement as diffusion processes along the reaction coordinates. This enabled separation of slow large-scale collective movements from fast local fluctuations. In contrast to the core regions, about one-half of the rRNA residues were reported to undergo uncoupled thermal fluctuations. This behavior was typical for the peripheral

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

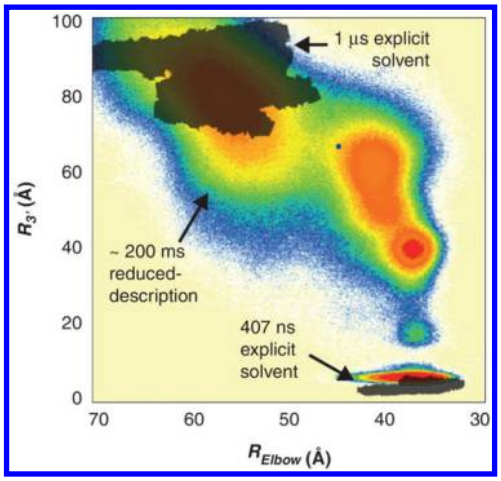


Figure 52. Simulations of accommodation of A-site aa-tRNA from the A/T state to the A/A state in a complete ribosome.[1035] The figure shows overlap of spontaneous fluctuations achieved in explicit solvent MD (gray, ~1.4 μs total sampling, ~3 200 000 atoms, most of them belonging to the almost 1 million of water molecules) with a structure-based-atomistic-potential simulation probability distribution plot (color, ~200 ms total sampling, ~140 000 atoms). The two-dimensional plot is built using two descriptors: R_{Elbow} is the distance between U47 of the incoming aa-tRNA and U8 of the P-site tRNA, while $R_{3'}$ is the distance between A-site and P-site amino acids. Note that while the structure-based-potential simulations sample the whole space, the explicit-solvent MD trajectories remain localized around the starting structures; that is, they sample the A/T basin (upper left) and the A/A basin (lower right). Reprinted with permission from ref 976. Copyright 2006 Elsevier Ltd.

regions of the ribosome. A more recent study using structure-based atomistic model has suggested that the aa-tRNA accommodation kinetics can be modulated by the L11 stalk mobility.[1045] It has also been suggested that the steric composition of EF-Tu can reduce the free-energy barrier associated with the initial stages of aa-tRNA accommodation.[1046] All-atom simulations of a complete ribosome model (3 000 000 atoms) were performed to study the dynamics and opening of the L1 stalk region.[1039] Full-ribosome simulations were combined with several simulations of reduced systems including the P/E state tRNA, the L1 stalk, and some proximal ribosomal elements (340 000 atoms). In a biased simulation of the entire ribosome with bound EF-Tu and three $tRNA^{Phe}$ units in the A/T, P/P, and E/E positions, the L1 stalk was pulled from a half-closed position to an open conformation within 3 ns, after extensive pre-equilibration. Pulling of the L1 stalk induced a movement of the E-site tRNA. Some of the atomistic models used in this work were constructed using molecular dynamics flexible fitting (MDFF) with primary cryo-EM data. This approach makes it possible to circumvent the lack of X-ray crystallography data for appropriate ribosomal states, and to model some mobile parts of the ribosome that are not resolved in the X-ray structures. The MDFF method was used for the pulling of the L1 stalk, targeting a partially resolved open structure known from X-ray crystallography data.[1039] It is useful to point out that resolution of cryo-EM studies has been recently substantially enhanced, due to new technical enhancements.[1047−1049]

Obviously, when evaluating simulations of the whole ribosome, we should be much more tolerant with respect to the various limitations of MD than when considering simulations of small RNA systems. We suggest that the largest uncertainty in such simulations does not relate to the choice of force field or the simulation time scale. Force-field artifacts are unlikely to fully develop on the simulations' time scale, and should not propagate through the whole system. In the peripheral regions, these errors should only propagate slowly throughout the structure, while errors in the rigid core region are likely to be attenuated by the structure's overall compactness. Thus, when studying the basic atomistic dynamics of such large systems, one can reasonably expect simulations to be pretty reliable. For large systems and short time scales, one can reasonably hope for a compensation of errors. Our general experience is that for larger systems, simulated for presently feasible time scales, force-field performance is often not as bad as for tetraloops (section 4.2) and tetranucleotides (section 4.1). In addition, whole-ribosome simulations do not attempt to capture folding, and remain within a given conformational basin of the studied structure. The simulation time scales achieved in the longest simulations should be sufficient to reveal many biologically relevant processes, especially when using well-chosen biases and enhanced sampling methods. The time scale of the ribosomal machine in operation (~22 amino acids per second for elongation) is not fundamentally different from the time scale of simple base pair opening in a duplex. We assume that the true Achilles heel of accurate whole ribosome simulations is the unavoidable uncertainties in the starting structures, which result from experimental resolution limits and ensemble averaging. In addition, there are countless sources of potential errors such as missing residues, missing internal waters, incorrectly placed or missing monovalent and divalent ions, incorrect backbone substates, *syn*/*anti* nucleotide bias, and so on. It was shown that the omission of a single monovalent ion from the starting structure can entirely derail microsecond-scale simulations of a fairly small protein/RNA complex (see sections 4.7.1 and 4.7.10).[508] This demonstrates the enormous sensitivity of MD simulations to the accuracy of the starting structure,[176,177] an issue often overlooked in the MD literature. Additionally, one should consider that a missing monovalent ion is a minor perturbation as compared to the omnipresent initial structural uncertainties that must be overcome when basing simulations on relatively low-resolution ribosomal X-ray structures.

4.7. Protein/RNA Complexes

4.7.1. Atomistic MD of Protein/RNA Complexes: General Considerations. MD simulations of protein/RNA complexes are inherently more difficult than simulations of the isolated monomers, and must be executed and evaluated with great care. Perhaps because of this difficulty, MD literature still includes only a rather small number of publications discussing studies on complexes of this sort even though RNA is almost always bound to proteins in vivo.[1050] Protein/RNA complexes are relatively large systems, and so are usually more computationally demanding to simulate than their constituent monomers in isolation.[508] However, the biggest barrier to atomistic simulations of protein/RNA complexes is often the unavailability of suitable experimental structures. For biologically interesting molecules, it is relatively common for the atomistic structures of one or both of the monomers (more often the protein) to be known, but not that of the complex.[1051] Some authors have sought to overcome this problem using various "docking" procedures, in which a knowledge-based approach is employed to derive a structure of the complex, which is then used as the starting structure for

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

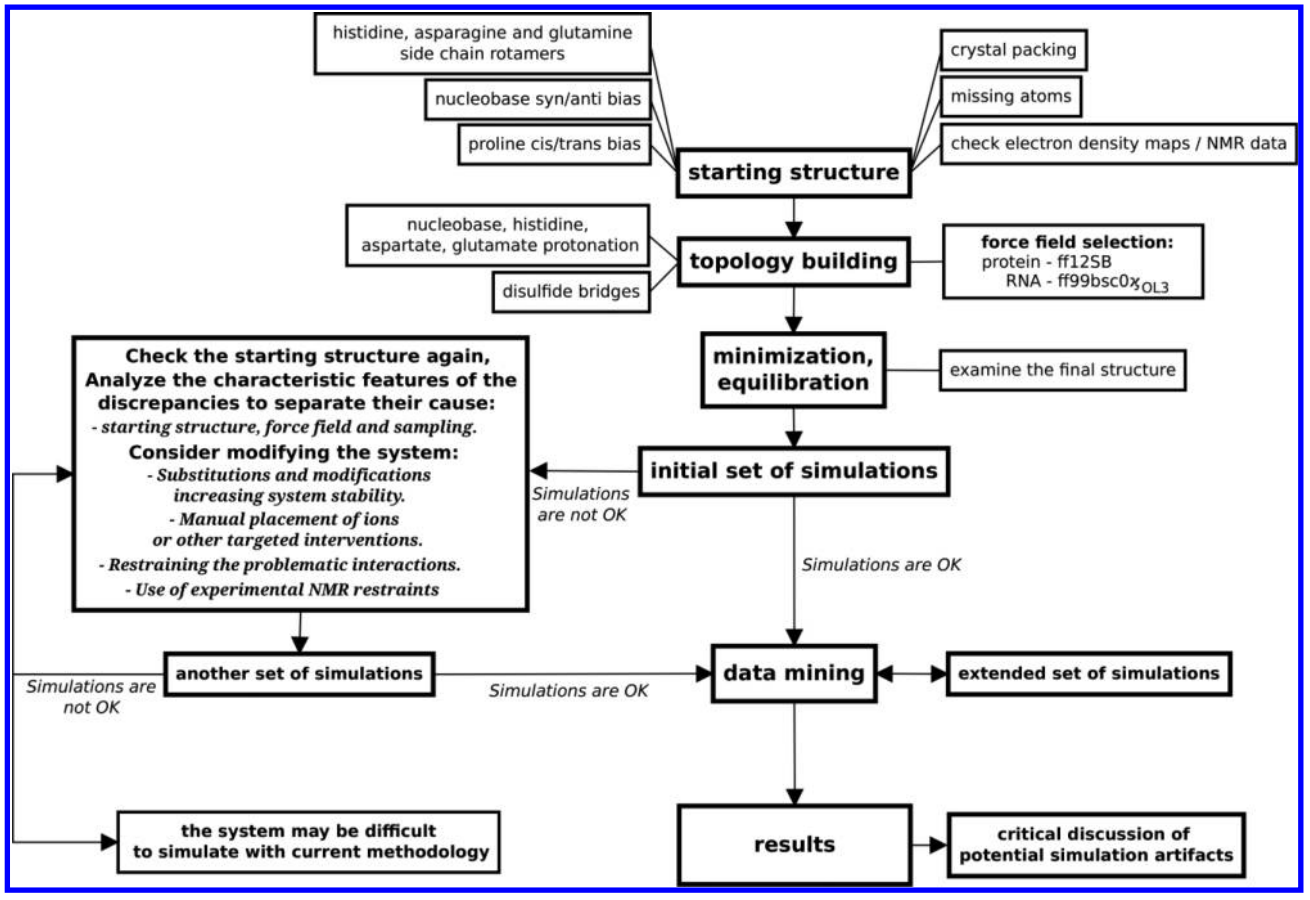


Figure 53. A basic workflow for performing MD simulations of protein/RNA complexes. Adapted with permission from ref 508. Copyright 2015 American Chemical Society. Some systems may require additional actions to be taken. Note that running MD simulations of protein/RNA complexes based on PDB files without a careful preparation and analysis should be strictly avoided.

MD simulations. While this may be the only way to simulate protein/RNA complexes in the absence of an experimental structure, any building-up approach such as a docking procedure is a source of potentially critical errors that may negatively affect the simulation results.[1052] Upon formation of protein/RNA complex, significant local and global remodeling of both structures may occur. Indeed, given the known sensitivity of simulations of the protein/RNA complexes to the properties of the starting structures, we argue that docking procedures are not acceptable substitutes for atomistic structures in such cases.[508] In other words, MD simulations of protein/RNA complexes that are not based on reliable experimental structures should only be performed under exceptional circumstances. Even if an experimental structure is available, there is no guarantee that it represents a suitable starting point for stable MD simulations, for a variety of reasons that are discussed in detail elsewhere.[154,508] Simulations of protein/RNA complexes are also more demanding in terms of the force-field balance: one may encounter force-field errors typical for the protein, the RNA, or both in the course of a single trajectory. In addition, the protein/RNA interface (the border between the interacting monomers) is necessarily subject to its own unique set of potential force-field imbalances. This is compounded by the fact that force-field parameters are developed separately for proteins and nucleic acids, so a reparameterization that improves simulations of isolated proteins or RNAs might not have the same beneficial impact at the protein/RNA interfaces. It is therefore not trivial to assess force-field performance for simulations of protein/RNA complexes. The literature on simulations of protein/RNA complexes is quite fragmented, with few systematic studies on issues such as the performance of force fields for complexes or the preparation of starting structures. As is common in the MD literature, most works in this area do not acknowledge or discuss the problems that the authors encountered while performing the simulations, even when they attempt to draw substantial biochemical conclusions from the results obtained in those simulations. It is therefore not clear how reliable some results presented in the literature are, or whether the force fields used in some studies are even capable of correctly describing protein/RNA complexes.

In 2015, Krepl et al. for the first time provided systematic benchmarking data for this field by simulating six structurally diverse protein/RNA complexes over multiple microsecond-scale MD runs and evaluating the simulations' stability, paying particular attention to the protein/RNA interface.[508] Figure 53 presents the basic workflow these authors used when performing protein/RNA simulations. They tested the ff99SB[1053] and ff12SB[142] force fields for the protein components, and χ_{OL3}[94] for RNA, that is, the current state-of-the-art AMBER force fields. The generally positive conclusion of the work was that these force fields are mature enough to handle microsecond simulations of protein/RNA complexes in many cases. However, there were specific issues that had to be considered, and the possible simulation outcomes were remarkably variable: some systems yielded

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

entirely stable trajectories in full agreement with the starting structures, while others yielded progressively degrading trajectories in every simulation that was attempted. For most systems, it was possible to achieve a good but imperfect agreement with the experimental structure. The differences between the simulations and experimental data were interpreted as resulting from an intricate mixture of force-field limitations and various factors related to the starting structures, which had to be analyzed on a case-by-case basis. Thus, there is no universal answer to the question: "How good or bad are MD simulations of protein/RNA complexes?" Every studied system must be judged individually, the simulations must be very carefully designed, and it is often necessary to regularly adapt the design of the study on the basis of preliminary results. The benchmark simulations suggested that with current force fields and simulation time scales the main limiting factors in simulations of protein/RNA complexes are the quality and properties of the experimental data used as initial structures. Many simulations were very sensitive to detailed aspects of the starting structures, which were often below the natural resolution/accuracy limits of the experimental data. It should be noted that these benchmark simulations were performed using a very carefully selected set of experimental structures that were expected to cause minimal uncertainties in the simulations. The results obtained by applying the same MD techniques to randomly chosen protein/RNA systems would probably be much less promising.

The simulations were generally unable to resolve many common errors in the experimental structures, such as extensive crystal packing contacts or incorrectly refined side-chain rotamers of amino acids. These issues often only became apparent after significant simulation time (hundreds of nanoseconds). Even a small incompleteness in the starting structure such as a missing internally bound monovalent ion can dramatically destabilize the subsequent simulations. In addition, the behavior of the studied systems often differed across series of otherwise equivalent simulations, for a variety of reasons. It was thus necessary to perform multiple independent microsecond-scale simulations of each protein/RNA complex because single simulations or series of shorter simulations typically provided misleading results. Performing multiple microsecond-scale simulations and verifying that they all yield similar results does not guarantee convergence, but seems to represent the minimum required to obtain results that start to make sense for a wide range of systems. This implies that in many earlier studies, certain force-field issues and errors in experimental structures may have been misinterpreted as biologically significant developments. While such flaws are fully understandable in older studies whose authors did not have access to sufficient computational resources, contemporary and future studies should be extended at least to the above-noted time scales. When basing simulations on NMR solution structures of protein/RNA complexes, some of these issues can be overcome through wise use of experimental data such as NOE restraints to bias the simulations.[346,368,1054,1055] The restraints can be applied during the initial phases of the simulations to stabilize them and then lifted, allowing the simulated complexes to settle and reducing the likelihood of abrupt changes during the early stages of the simulations.[368]

We would like to stress that a prerequisite for successfully studying protein/RNA complexes is to first obtain basic microsecond-scale simulations of the experimental structures that are stable with the exception of local fluctuations. This requirement is supported by the following argument. Healthy microsecond-scale simulations should not progressively corrupt the experimentally observed protein/RNA interface because this would be incompatible with the lifetime or k_{off} (often 10^{-2} s^{-1}) of the real systems.[1056] Therefore, large immediate distortions of the protein/RNA interfaces almost certainly indicate the presence of unacceptable errors in either the force-field balance or the starting structures. Obviously, it would often be useful to see more of the system's dynamics, but if this is the goal then one should complement the standard simulations with enhanced sampling simulation methods (section 3.2). Larger reconformations of the protein/RNA interfaces must be simulated in a controllable manner. On the basis of this argument, we consider the analyses of dynamics presented in many simulation-based papers on protein/RNA complexes to be suboptimal.

The outcomes of simulations of protein/RNA complexes depend on the mode of protein/RNA recognition: different forms of protein/RNA recognition appear to have different degrees of compensation of errors in the force-field description. Our experience is that MD usually reproduces (i.e., keeps stable) protein/RNA binding that relies on shape recognition without complicated H-bonding. However, simulations of some protein/RNA complexes are unable to simultaneously satisfy all native H-bonds, especially for systems with dense networks of protein/RNA H-bonds.[508] We suggest that this results from imperfect descriptions of the individual H-bonds by the force field. This leads to a tension between the individual H-bonds that strains the protein/RNA interface in the force-field description, as all of the H-bonds try to satisfy the overall structure. During the early stages of the simulation, the simulated system responds to this stress by losing a few H-bonds, which relieves the structural strain and allows the simulation to progress steadily thereafter. This is a common obstacle in MD simulations of protein/RNA complexes. This view is supported by the fact that some of the systems can be stabilized using the HBfix potential function (sections 3.1.2.5 and 4.2.2.3),[154] a local structure-specific bias that aims to compensate for the deficient description of H-bonds through the use of pair-additive force fields.[129] Perhaps, the structural stability of simulated protein/RNA complexes may also be sensitive to the exposure of the interface to the solvent.

When simulating protein/RNA complexes, it is also important to keep in mind that the standard equilibration protocol is less efficient than for small RNAs.[154,508] The experimental structures are, for a variety of reasons, always high in force-field potential energy. Equilibration relaxes the simplest high-energy features in these structures, such as nonoptimal bond lengths and angles. However, the equilibrated structures may still be biased by more complex unnatural structural features such as inaccurate RNA backbone conformations, wrong protein side-chain orientations, unfilled internal hydration sites, and missing ions. These strains are transferred into the production simulations. Some simulations proceed to relax smoothly, dissipating the excess energy through many degrees of freedom, and continue stably. In other cases, the simulation is derailed by abrupt structural transitions. In an ideal case, given a flawless force field and unlimited simulation time, the MD method would ultimately resolve all such problems. However, in reality this typically leads to a confusing set of equivalent simulations with at first sight randomly variable behavior.[154,508] That is why we emphasize the need to perform series of simulations, and the usefulness of applying

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

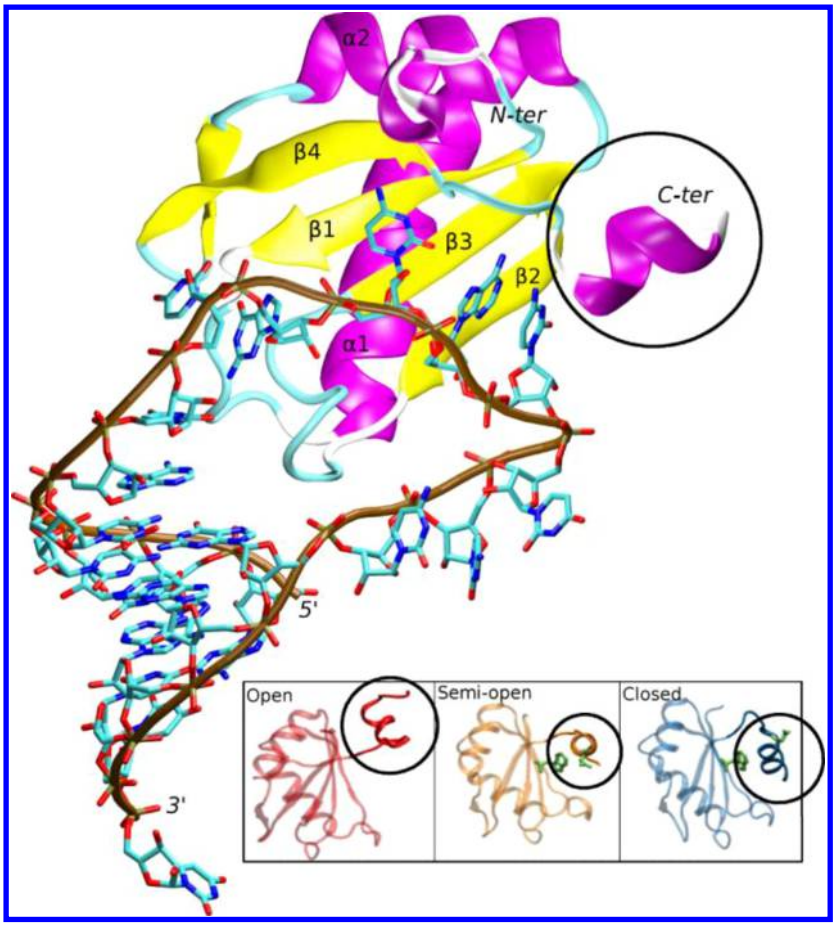


Figure 54. View of the U1A RRM protein/RNA complex, the most commonly simulated protein/RNA system. The requirement for conformational change of the C-terminal helix (black circle) for the RNA binding has long been suspected.[1068] However, only longer time scales of modern MD simulations finally allowed one to observe full spontaneous transition between open, semiopen, and closed conformations of the C-terminal helix.[1071] Figure partially adapted from ref 1071. Copyright 2015 American Chemical Society.

NMR restraints during the initial stages of the production phase[368] when performing simulations based on NMR structures, although it should be noted that common restraining potentials can sometimes also influence the simulated ensemble in an undesired manner.[1057] Manual modification of the starting structure is sometimes necessary, as demonstrated by the example of the Sam domain of the Vts1 complex. This experimental structure required removal of an overhanging nucleotide and stabilization of the end of the RNA stem by G═C base pairs, because excessive end-fraying of the RNA stem end destabilized the complex, despite that the stem end was not initially in a physical contact with the protein. This modification yielded neat simulations in which all of the important interfacial H-bonds were satisfied.[508]

The benchmark simulations[508] revealed that the ff12SB[142] protein force field is an improvement over the ff99SB. Additionally, the χ_{OL3}[94] parameters provided an adequate description for RNA, achieving accuracies compatible with previously established limits.[508] It was later observed that the ff14SB[142] protein force field may be (at least as implemented in the AMBER program package in the time of writing this Review) slightly inferior to ff12SB in simulations of protein/RNA complexes. This is due to a lower energy barrier for phenylalanine and tyrosine side-chain rotation, which allows for excessive fluctuations of these side-chains in ff14SB simulations and can perturb some protein/RNA complexes. An identical issue was observed in simulations using the older ff99SB parameters. The ff12SB parameters address this issue by doubling the size of the energy barrier, which significantly reduced the number of aromatic side-chain flips.[508] Somewhat surprisingly and for unknown reasons, this refinement was not carried over to the ff14SB parameters, leading to an observed regression in the ff14SB simulations.[154,369]

Below, we provide an overview of the MD simulation studies on protein/RNA systems that are available in the literature. Despite all of the limitations noted above, we believe that MD simulations of RNAs complexed with proteins will represent a major area of research involving RNA simulations over the next decade. Although not every protein/RNA system can be satisfactorily simulated, the vast and rich world of protein/RNA interactions guarantees that there will always be a lot of material for meaningful simulations. However, to avoid discrediting the MD method by populating the literature with poorly executed and interpreted simulations, it will be necessary for researchers working in this field to uphold certain standards when performing simulations, to provide full details of their methods, and to avoid overinterpreting their results.

4.7.2. RNA Recognition Motif (RRM) Protein/RNA Complexes. RRM domain is the most common RNA binding motif in eukaryotic proteins. RRMs are involved in most steps of RNA metabolism. Despite having a highly conserved β1-α1-β2-β3-α2-β4 fold, RRMs are known to specifically bind a great

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

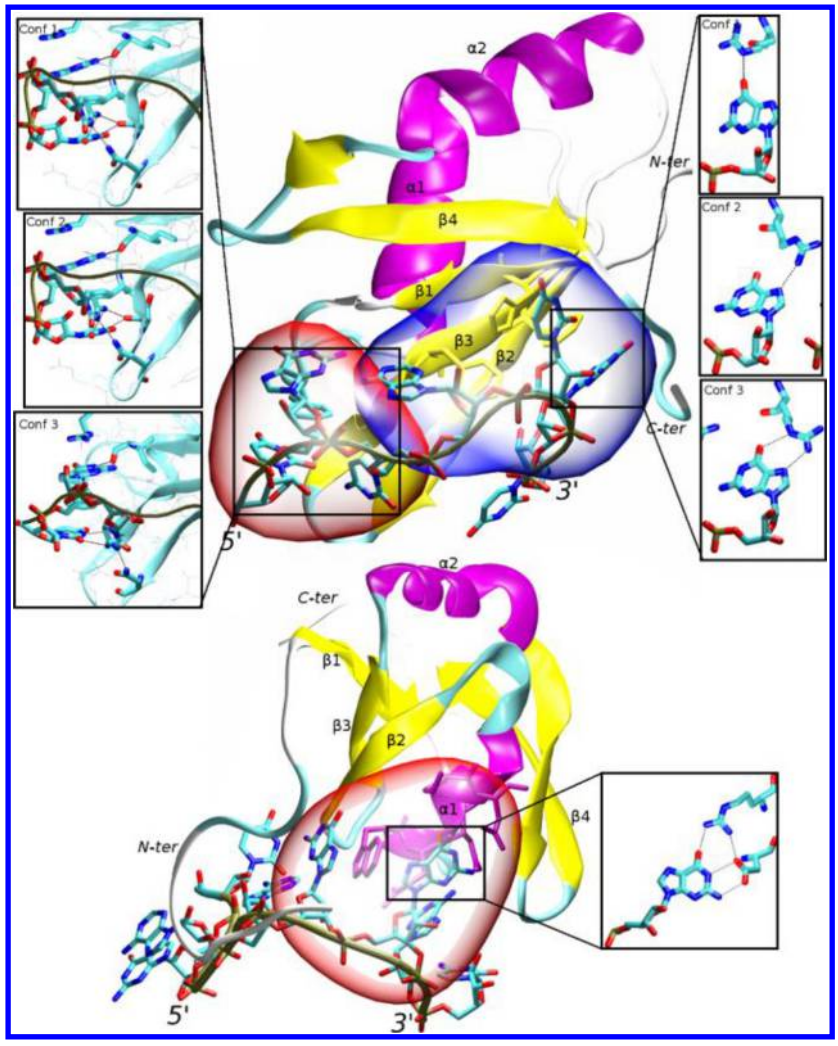


Figure 55. Views of the Fox-1 (top) and SRSF1 (bottom) RRM protein/RNA complexes. A different part of the protein domain is utilized for RNA recognition than in the U1A RRM; such recognition versatility is a common feature of the RRM domains.[1058] Specifically, in addition to canonical RNA recognition via the β-sheet surface (highlighted in blue), the Fox-1 RRM also uses noncanonical recognition via protein loops (highlighted in red). SRSF1 RRM recognizes the RNA solely in noncanonical mode via α-helix. MD simulations identified dynamical substates in the protein/RNA interaction with rapidly shifting conformations in the Fox-1 system, while a previously unknown protein/RNA interaction was observed in the SRSF1 RRM. The figure was adapted from ref 368; http://creativecommons.org/licenses/by/4.0. Copyright 2016 Oxford University Press.

variety of RNA sequences. This is due to subtle variations in their amino acid sequences and the coupled action of multiple RRM domains within a single protein. In addition, individual RRMs can bind to RNA via different parts of the protein domain. This combination of high versatility and high specificity makes the RRM a centerpiece of structural studies on protein/RNA complexes.[1058]

4.7.2.1. U1A Complex. By a wide margin, the RRM protein/RNA complex that has been most extensively studied using MD simulations is the U1A RRM complexed with its stem-loop RNA target (Figure 54). There are many reasons for this. The first is that it was one of the first RNA/protein complexes for which an atomistic experimental structure was determined. Specifically, a high-resolution X-ray structure was reported in 1994, and an NMR solution structure was presented in 1997.[1059,1060] Second, the complex has a very high (picomolar) binding affinity along with a well-defined network of H-bonding and stacking interactions at its protein/RNA interface. Last, the U1A protein is biologically significant and has been studied in detail by biochemists, allowing simulation results to be compared to experiments. For these reasons, studies on U1A RRM/RNA complexes constitute a prime example of MD simulations of protein/RNA complexes.

The U1A RRM was the subject of the very first study of protein/RNA complexes using explicit-solvent, unrestrained MD simulations.[1061] In this work, the authors first suggested that MD simulations may be suitable for study of protein/RNA complexes and discussed the dynamical properties of the protein/RNA interface on a 1 ns time scale. Later, they used MM-PBSA free-energy calculations (section 3.2.8) to estimate the absolute binding energy of the U1A complex, reporting a result within range of the experimental data.[1062]

In later years, other authors applied similar methods to elucidate different aspects of the U1A protein/RNA recognition, such as the role of the Phe56 residue, which stacks with one of the RNA bases in the complex.[1063] This work demonstrated the structural-dynamics and free-energy impact of the Phe56Ala mutation on the complex. The mutation alters the structure of both the complex and the free protein, necessitating inclusion of both in the calculations. It was found that the mutation's net effect was to impose a free-energy penalty on complex formation. Further information about the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

contribution of the Phe56 residue was provided by a second study that examined the Phe56Ala, Phe56Leu, and Phe56Trp mutants.[1064] On the basis of MM-GBSA calculations (see section 3.2.8) and the free-energy decomposition method, the authors argued that the loss of stacking caused by the Phe56Ala mutation is not the dominant contribution to the observed free-energy penalty to the complex formation. Finally, another study examined in detail the changes in the interaction network caused by the Phe56 mutations and proposed that the observed free-energy changes are due to the interplay of structural changes in both the complex and the free protein.[1065]

In unrelated studies, Law et al. used MD simulations in combination with surface plasmon resonance experiments to elucidate the role of electrostatic interactions between the U1A RRM protein and RNA by mutating the Lys20, Lys22, and Lys23 residues. These residues were identified to be important for the initial complex formation, and two of them also stabilized the formed complex.[1066] Additionally, Showalter et al. presented a study of the Gly53Ala and Gly53Val mutants. Using NMR spectroscopy, they demonstrated that both of these mutant proteins exhibit reduced affinity for the target RNA. They then performed MD simulations of the mutant complexes and suggested that the lower RNA affinity is due to changes in the dynamics of the free protein, which cannot undergo the same structural changes as the wild-type U1A protein upon binding of the RNA.[1067]

Many computational studies on the U1A protein/RNA complex have also described the structural and dynamical changes the monomers undergo upon complex formation. The first such study[1068] indicated that the protein structure capable of binding RNA is a minor substate observed in MD simulations of its free form, suggesting complex formation proceeds via conformational capture. The RNA molecule was predicted to undergo a larger structural rearrangement (induced fitting). However, the MD simulations of the free molecules were performed using the configurations of the monomers in the experimental structure of the complex as starting structures, and used rather short simulation time scales. Another study used fluorescence measurements and MD simulations to examine the behavior of the protein's C-terminal helix, which has been experimentally observed to adopt "closed" and "open" orientations in the free protein and the complex, respectively. While the C-terminal helix was found to be a flexible element in the free protein, no transition between closed and open conformations could be seen by the fluorescence measurements nor observed on the time scale of MD simulations.[1069] A full transition was finally observed in a later study using much longer simulation time scales,[1070,1071] confirming that the open conformation of the C-terminal helix is thermally accessible in the free U1A protein (Figure 54). The studies suggested that RNA binding occurs via conformational capture of the protein with C-terminal helix in the open conformation.

4.7.2.2. Other RRM Complexes. Despite the great diversity of RRM complexes (Figure 55) and their biological importance, there have been few reported studies on such systems other than those involving U1A.

Guo et al. studied a complex between the U2B protein and the U2 small nucleolar RNA (snRNA) hairpin IV.[1072] U2B protein is a component of the spliceosomal U2 RNP particle. It binds its target RNA in a similar but not identical manner to the U1A protein. The authors used MD simulations to study its dynamics and compared them to earlier simulations and structures of U1A. On the basis of alanine-scanning of the interacting regions, they suggested that electrostatic interactions play a stronger stabilizing role in U2B than in U1A. They also presented the first extensive simulation-based description of hydration of the protein/RNA interface. Notably, they observed a number of bridging water molecules and suggested that these molecules contribute to protein/RNA binding. This suggestion was corroborated by some ab initio QM calculations.[1072]

Schmid et al. studied the Polypyrimidine Tract Binding (PTB) protein/RNA complex.[1073] The PTB RRM protein contributes to RNA metabolism by recognizing pyrimidine-rich RNA sequences. The authors utilized an experimental structure of two joined RRM domains (RRM3 and RRM4) from PTB to study the dynamics of the entire complex and of the two RRM domains separately. The protein/RNA interface was analyzed by monitoring H-bonding, stacking, and salt-bridge interactions during simulations. In addition, they used experimental NMR data to verify the simulations' stability by monitoring NOE violations. Finally, they performed thermodynamics integration (TI) calculations to calculate free energies of binding for various RNA sequences.[1073] In a later study, the same authors simulated the Fox-1 RRM protein/RNA complex (among other systems) to benchmark a new version of the GROMOS force field.[1074] However, this force field has not yet been adopted by other research groups.

A study by Krepl et al. examined the Fox-1 and serine/arginine-rich splicing factor 1 (SRSF1) RRM protein/RNA complexes.[368] Both the Fox-1 and the SRSF1 proteins have roles in regulating splicing. On the basis of their experimental NMR structures, the authors conducted MD simulations and assessed the performance of selected force fields by comparing the resulting trajectories to NMR data. In total, this work presented 50 μs of simulation data and compared these results very carefully to experimental findings. It was shown that both complexes are quite well described by the MD technique. An initial use of experimental NMR restraints to increase the stability of MD simulations was suggested. The NOE-based restraints made it possible to stabilize the initial stages of the production simulations. The authors suggested that protein/RNA recognition may be intrinsically dynamical, which is not reflected in NMR structures based on ensemble averaged NOE data. Several substates contributing to the "dynamical recognition pattern" at the protein/RNA interface were suggested. Such interaction patterns cannot be identified in time-averaged NMR experiments but can be detected by MD simulations. Further, a new protein/RNA interface interaction was identified in simulations of the SRSF1 complex and subsequently confirmed by new NMR and ITC measurements (Figure 55). This interaction was also studied by MD thermodynamics integration (TI) free-energy calculations.[368]

Later, the authors revisited the Fox-1 RRM protein/RNA complex in a study focused on its hydration.[369] First, a new high-resolution X-ray structure of the free Fox-1 RRM was determined. The authors then used this structure, along with the earlier NMR structure of the protein/RNA complex, as a starting structure for MD simulations. A high degree of agreement between the experimental and simulation hydration sites was achieved, demonstrating that MD is a suitable tool for predicting the hydration patterns of RRM protein/RNA complexes. In fact, as discussed, MD simulations can significantly reduce some ambiguities in visualized hydration patterns that are inherent to X-ray crystallography data sets.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

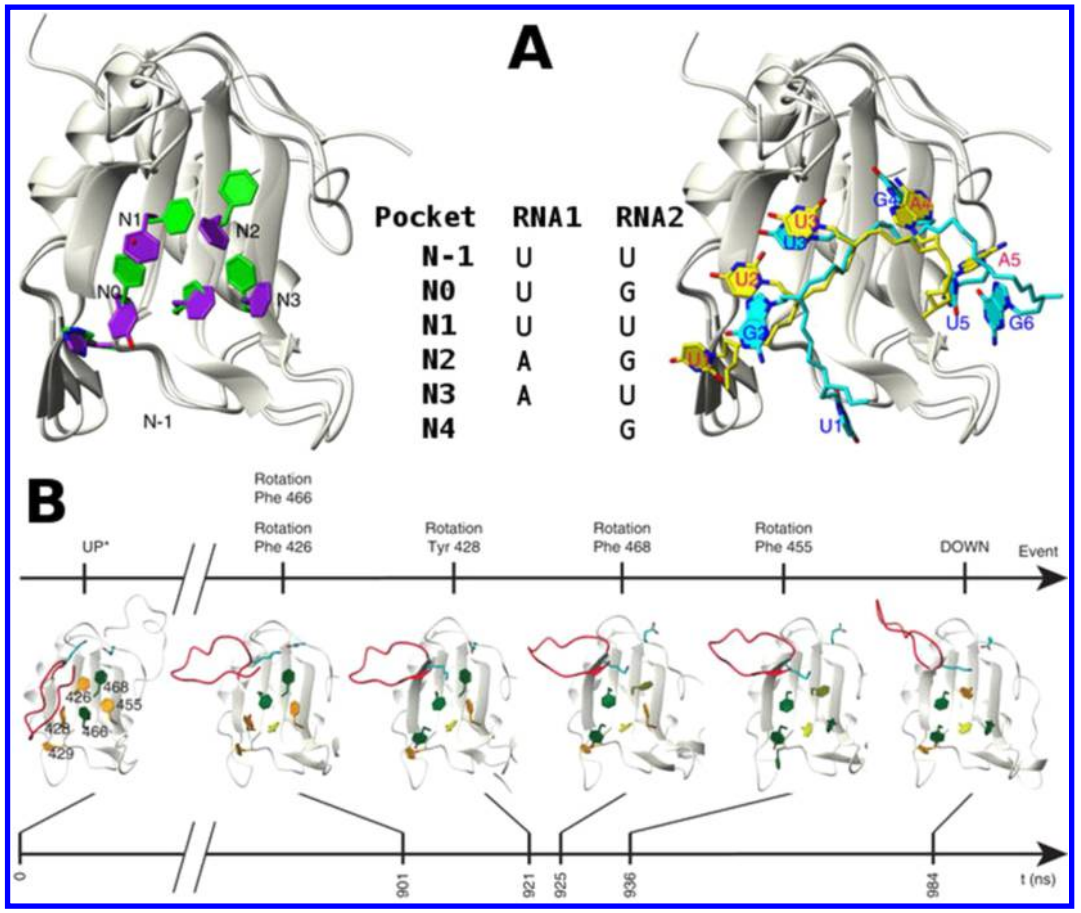


Figure 56. (A) Comparison of CUG-BP2 RRM3 bound to 5′-UUUAA-3′ and 5′-UGUGUG-3′ complexes, respectively.[1075] The CUG-BP2 RRM3/5′-UUUAA-3′ complex is represented with green protein side-chains and the RNA in yellow, and the CUG-BP1 RRM3/5′-UGUGUG-3′ complex with purple protein side-chains and the RNA in cyan. Schematic representation of the arrangement of the nucleotides of 5′-UUUAA-3′ and 5′-UGUGUG-3′ RNAs in the protein binding pockets (N1 to N4) is shown. (B) MD simulations of the free protein have shown spontaneous transition of the aromatic residues between respective conformations associated with binding of the two RNA sequences. The C-terminus is colored in red, and aromatic residues in the trans, gauche+, and gauche− conformations are colored in green, orange, and yellow, respectively. Time points and key events in the side-chain rearrangement are indicated. Reprinted from ref 1075; http://creativecommons.org/licenses/by/4.0. Copyright 2017 Nature Publishing Group.

Highly structured hydration sites with variable water residence times were predicted by MD. Last, two of the hydration sites predicted by simulations that could be abolished by single residue mutations were examined by preparing suitable mutant Fox-1 RRM proteins. NMR spectroscopy, switchSENSE experiments, and TI calculations confirmed the presence of the hydration sites and their contribution to the protein's RNA affinity. Moreover, full quantitative agreement between theory and experiment was achieved for one of the sites, which was shown to be conserved in other RRM domains. For the other site, the TI procedure did not achieve quantitative accuracy, illustrating the general limitations of free-energy computations (see section 3.2.7). The authors also tried to correlate the hydration patterns of the isolated protein with its RNA binding pattern, but found that knowledge of protein hydration pattern is not sufficient to predict RNA binding, in contrast to some more optimistic suggestions in the earlier literature.[369]

Subsequently, Konté et al. studied the CUG-BP1 RRM3 in complex with two different target RNAs (5′-UUUAA-3′ and 5′-UGUGUG-3′).[1075] In addition to NMR experiments, the authors used MD simulations to characterize the structural differences in the binding of the two sequences. The simulations indicated that multiple distinct conformations of aromatic residues (i.e., phenylalanine, tyrosine, and histidine) can exist on the RNA binding surface of the free protein on a microsecond time scale (Figure 56). This allows the protein to bind both RNA sequences even though their preferred aromatic residues conformations are very different. Thus, this protein's RNA binding proceeds via conformational capture of a specific arrangement of the aromatic residues. This conformational plasticity of the RNA binding site allows the RRM domain to show good affinity to diverse RNA sequences, and may help to guide the binding site to high-affinity targets.[1075]

The approach suggested in ref 368 (see above) was later utilized by Wang et al.[1076] who used MD simulations to explore the protein/RNA interactions of the quasi-RRM2 domain of hnRNP F protein with its target G-tract RNA. The authors described structure and dynamics of the wild-type protein/RNA complex and its response to several different point mutations in the G-tract sequence. Using PCA and MM-GBSA calculations and by analysis of protein/RNA interaction patterns, they provided reasoning for the experimentally observed reduction of binding affinity in the mutants.

4.7.3. dsRNA Binding Protein/RNA Complexes. The RRM domains primarily bind single-stranded or stem-loop RNAs. However, some proteins selectively bind to double-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

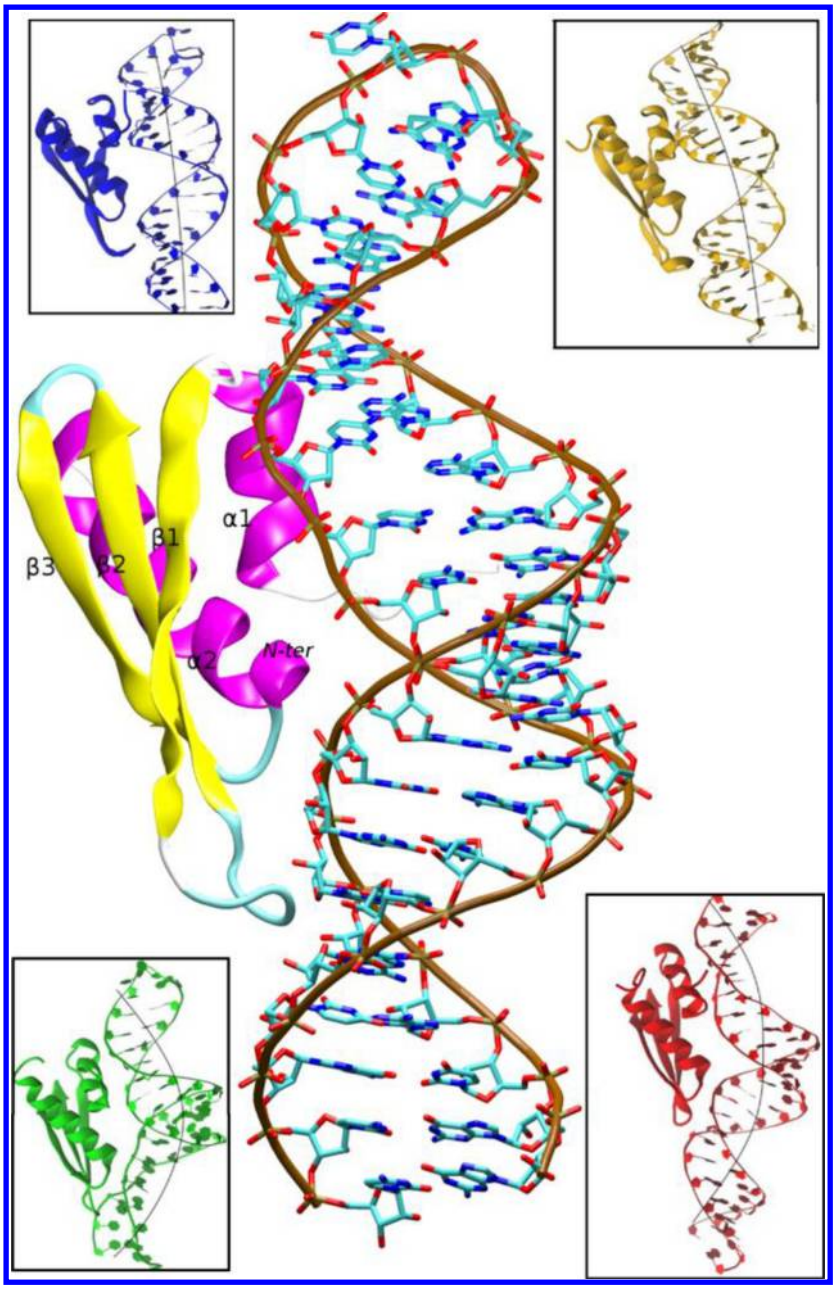


Figure 57. View of a dsRNA helix in complex with dsRBD domain of the ADAR2 protein.[1081] The figures in black boxes represent final structure from MD simulations of the crystal structure of ADAR2/dsRNA-A complex (blue), and three modeled complexes, ADAR2/dsRNA-D (yellow), DCL1/dsRNA-A (green), and DCL1/dsRNA-D (red). The simulations suggested that despite the DCL1 possessing shorter protein loop for RNA interaction, it achieves a similar level of recognition as ADAR2 by inducing a helical bend in the RNA substrates (the RNA helical axis bend is indicated by the black line).[1082] Note that the initial structures of the model systems were obtained by structural alignment over the crystal structure of the ADAR2/dsRNA-A complex. Whenever possible, this approach is generally recommended over de novo molecular docking. The figure was adapted with permission from ref 1082. Copyright 2016 Elsevier Ltd.

stranded RNAs, and such complexes have also been investigated by MD simulations. As with the RRM domain for ssRNA, there is a common protein motif for dsRNA binding in eukaryotes that is known as the double-stranded RNA binding domain (dsRBD) motif (Figure 57). The dsRBD motif has a conserved $\alpha1$-$\beta1$-$\beta2$-$\beta3$-$\alpha2$ secondary structure, and appears to participate in all cellular processes involving dsRNA molecules, including miRNA maturation, RNA interference, viral infection, and the antiviral response. Most dsRBD proteins derive their specificity from shape recognition of the A-RNA helix rather than a direct readout of the bases.[1077] This is because protein binding occurs exclusively via the A-RNA shallow minor-groove, which allows the protein to discriminate for RNA via O2′-hydroxyl recognition but is too narrow for the protein to directly access the bases. However, some dsRBD complexes reportedly exhibit specific distortions in the RNA helical structure (e.g., bulges and noncanonical base pairs) that widen the minor-groove and thus enable direct base pair readout.[1078]

The first MD simulation study of a dsRBD protein/RNA complex was done by Castrignano et al. and focused on the complex of the third dsRBD domain from the *Drosophila*

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Staufen protein. The study reported significant dynamic behavior of the amino acid side-chains involved in the protein/RNA interface. In particular, the authors described a "lysine walking" mechanism on the RNA phosphate backbone that involved a repeating movement of lysine side-chains between neighboring phosphate groups, forming alternating H-bond interactions with them. The authors observed that these dynamics of the protein/RNA interface were facilitated by the involvement of fast-exchanging water molecules.[1079]

Yang and co-workers used MD simulations to study the dsRNA in complex with adenosine deaminase acting on RNA 2 (ADAR2).[1080] ADAR2 is a deaminase that turns adenine into inosine as part of the RNA editing process. It acts on the dsRNA molecule, and experiments have suggested that its binding specificity is achieved by detecting a base pair mismatch in the dsRNA helix.[1081] The authors studied complexes between the ADAR2 dsRBD domain and dsRNA sequences with and without the mismatched base pair. Using MM-PBSA (see section 3.2.8), they reported a reduced binding affinity in the latter case and attributed this to a reduction in minor groove width and an altered helical bend in the RNA sequence lacking the mismatch.[1080]

Another study of ADAR2 in complex with a different dsRNA was done by Drusin et al.[1082] These authors used the ADAR2 structure to model a similar complex between dsRNA and the N-terminal domain of the dicer-like 1 (DCL1) protein. Their simulations suggested that even though the binding interface of the DCL1 dsRBD domain is smaller than that of most dsRBD domains, the protein can compensate this by inducing a bend in the dsRNA substrate (Figure 57). Further, the authors suggested that the Arg8 residue of DCL1 is responsible for mismatched base pair recognition.[1082]

Acevedo et al. used MD simulations and isothermal titration calorimetry (ITC) measurements to study the complex of dsRBD2 domain of the HIV-1 TAR RNA binding protein (TRBP) with dsRNA (TRBP-dsRBD2/dsRNA).[1083] An earlier X-ray structure of this complex suggested that the RNA helix becomes bent upon protein binding.[1084] However, on the basis of the simulations and ITC data, the authors suggested that the helix does not actually bend, that TRBP preferentially binds to an ideal A-RNA helix, and that the helical geometry is not appreciably altered upon protein binding.[1083] This stands in contrast to the behavior observed during DCL1/dsRNA binding,[1082] and suggests that the dsRBD proteins RNA recognition patterns are diverse.

Last, in addition to their single-molecule experiments, Wang et al. used MD simulations to examine the atomistic details and dynamics of the protein/RNA interface in several different dsRBD protein/RNA complexes.[1085] They analyzed size of contact areas and force-field interaction energies (i.e., not free energies) between dsRBD and dsRNA and proposed that, despite the high conservation of dsRBD RNA binding residues, individual dsRBD may have different affinities to various dsRNA. The results agreed with experimental data, suggesting that a binding affinity of dsRBD to dsRNA could be estimated from MD simulations.

4.7.4. Non-dsRBD Complexes. There also have been simulation studies on dsRNA complexes with proteins that lack a dsRBP domain. The first two[1086,1087] were viral proteins that evolved to protect the viral genome from the cellular RNA silencing mechanism during infection by coating the dsRNA helix. Specifically, Xue et al. studied a UP35 protein (from filoviruses) complexed with dsRNA, and performed MM-GBSA free-energy calculations (see section 3.2.8) to identify residues responsible for the protein's RNA affinity.[1086] Allen et al. studied a B2 protein/dsRNA complex from the Nodamura Virus.[1087] Their study described the dynamics of the wild-type protein/RNA complex and that of complexes with several mutant proteins. Steered MD was used to predict the free-energy impact of these mutations on the protein/RNA binding affinity. The authors then used experimental mutagenesis to successfully confirm some of the simulation results.[1087] However, it is difficult to assess the results of this study more generally because the computational methods chosen by the authors are quite risky, and the paper does not even properly specify which RNA force field was used (a presumably incorrect reference to a protein force field is given, cf., the comment in Table 1), rendering the work difficult to reproduce. We encountered similar problems in several other publications, some of which were so severe that we excluded the studies in question from this Review.

Xia et al. examined the binding of the p19 protein (Tombusvirus) to siRNA.[1088] Unlike UP35 or B2 proteins, which can nonspecifically bind the entire viral genome (see above), p19 binds short siRNAs with high affinity. It thus also blocks the RNA silencing mechanism, but at a different level. The authors used MD simulations of the wild-type p19 protein and two mutants to examine the atomistic details of its RNA recognition and the structural effects of the two mutations. The mutations' effects were quantified by performing steered MD and free-energy perturbation calculations (see sections 3.2.5 and 3.2.7). The authors suggested that the mutations reduce the protein's binding affinity by eliminating stacking interactions with the terminal base pair of the siRNA, which is vital for p19's specificity for short dsRNA molecules.[1088]

Last, Harikrishna et al. used MD simulations to examine the effects of ribose chemical modifications on siRNA binding to human Argonaute 2 protein (hAGO2).[1089] The authors first used molecular modeling to construct a model of double stranded siRNA complexed with hAGO2. They then separately introduced 2′-*O*-methyl, 4′-*C*-aminomethyl-2′-*O*-methyl, 2′-*O*-(2-methoxyethyl), or 2′-*O*-benzyl chemical modifications at selected riboses and performed microsecond MD simulations. The authors described structural and free-energy impact (using MM-PBSA and free-energy decomposition) of each modification as compared to the wild-type complex. Some modifications led to destabilization of the protein/RNA complex, while others had the opposite effect.[1089]

4.7.5. HIV-1 TAR Element. Another protein/RNA complex often studied by MD simulations is the HIV-1 TAR element RNA complexed with the Tat peptide or its analogues. The structure of the HIV-1 TAR element and the corresponding MD studies are described in section 4.3.8; here, we present some additional details of simulations of the protein/TAR interactions. The high biological relevance of this system together with its relatively simple protein/RNA interface have prompted the discovery of several potentially therapeutic peptide mimics of the Tat peptide whose interactions with the TAR element have been described by MD.

An early study by Reyes et al. used MM-PBSA (see section 3.2.8) to compute the binding affinity of the Tat peptide to RNA for the wild-type peptide and several mutants.[1090] Mu et al. studied an interaction between the TAR element and a heterochiral tripeptide[1091] using an initial structure obtained by rigid-body docking. Do et al. analyzed an interaction between the TAR element and a cyclic peptide analogue (termed "L22")

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

of the Tat protein. Their paper provides extensive descriptions of the changes in solvent and ion distribution that occurred during the MD simulations as a result of L22 binding to TAR.[830] The same complex was also the subject of an attempt for blind prediction of the bound structure (see section 4.3.8).[303] The TAR/L22 interaction was also studied by Li et al., along with another cyclic peptide (KP-Z-41).[1092] These authors discussed the specific binding interactions of the two peptides and reported MM-PB(GB)SA free-energy calculations of their binding affinity. They also proposed a consensual binding motif for both cyclic peptides. Finally, Borkar et al. used NMR residual dipolar couplings (RDC) in replica-averaged metadynamics simulations to identify intermediate states in the binding of the Tat peptide to TAR (see section 4.3.8).[346]

4.7.6. Ribosomal Protein/RNA Systems. In addition to simulations of the whole ribosome (section 4.6), some MD simulation studies have examined small portions of rRNA complexed with ribosomal proteins. Such structures can be obtained either from structural experiments involving the isolated molecules or by excising segments of interest from experimental structures of entire ribosomes. The latter approach is more commonly used.

Reblova et al. examined the behavior of the loop E/helix IV rRNA in complex with the L25 protein of the large ribosomal subunit (see sections 4.3.2 and 4.3.4)[487] and analyzed protein/RNA interactions, structural role of divalent Mg^{2+} and monovalent cations, and long-residency hydration sites. The role of divalent ions in protein/RNA interactions in the ribosome was further explored by Créty et al., who described the importance of Mg^{2+} ions in stabilizing the complex between protein S15 and the intersection of helices H20−22.[1093] Later, Chen et al. examined the behavior of the complex between helix H16 and the N-terminal fragment of protein S4. Their simulations predicted that the unbound protein chain is intrinsically disordered, which was in agreement with fluorescent spectroscopy and circular dichroism measurements.[1094] Krepl et al. examined the complex formed between helices H76−78 and the L1 protein (collectively known as the L1 stalk). MD simulations of this complex suggested a possible mechanism for the release of the deacylated tRNA molecule from the ribosome involving changes in an intricate system of RNA/RNA and protein/RNA interactions.[507] Wolf et al. studied a complex between helices H43−44 and the L11 protein. Using simulations, they examined the behavior of this complex with and without bound thiostrepton antibiotics, which inhibit bacterial protein synthesis. They suggested that the binding of the thiostrepton rigidifies the binding region and blocks contacts with the incoming tRNA molecule.[1095] MD simulations of ribosomal protein/RNA complexes can also be excellent complements to structural experiments, as demonstrated by Li et al., who used simulations of the H43−44/L11 complex to generate structural models for fitting to cryo-EM maps (see section 4.9).[1096]

4.7.7. Kink-Turns with Proteins. Kink-turns are tertiary structural elements of RNA that consist of an RNA helix with a bulge on one strand flanked by canonical base pairs and noncanonical A/G base pairs on each side, respectively (Figure 39). Structurally, kink-turns facilitate sharp bends in the RNA helix,[800] and they can act as flexible molecular hinges.[802] Although kink-turns are known to fold only in the presence of excess divalent ions or when bound to a protein,[800] most MD simulations have examined isolated kink-turns (see section 4.3.7). Here, we review only MD simulations of kink-turns in protein/RNA complexes.

Cojocaru et al. studied the protein/RNA complex between the human 15.5K protein and a segment of the spliceosomal U4 snRNA that contains a kink-turn.[804] Starting from a structure of a fully folded kink-turn, they showed that degradative structural changes (e.g., losses of A/G base pairs and the interhelical bend) readily occurred only in MD simulations without the protein, where the RNA was much more flexible. They also performed experiments using chemical probes, which demonstrated that the characteristic kink-turn interactions only formed upon protein binding, suggesting that the protein is essential for the folding and stability of this kink-turn.[804] Although the stability of the simulated kink-turn may have been adversely affected by the limitations of the available force fields, the relative trends should have been correctly captured. The protein/RNA interactions in kink-turns were later reexamined by Spackova et al., who studied the box C/D kink-turn (which is similar to that in the U4 snRNA) in complex with either the archaeal L7ae protein or the human 15.5K protein.[807] This study also showed that protein binding reduced the kink-turn's flexibility and stabilized its signature interactions. The complex of the box C/D kink-turn with the archaeal L7ae protein was also studied by Ye et al., who used standard and high-temperature MD simulations to study the dynamics and unfolding of this protein/RNA complex. Principal component analysis performed by the authors indicated that an induced fitting of the RNA occurs upon binding of the L7ae protein.[808] Finally, the system simulated in the study on the ribosomal L1 stalk discussed in section 4.7.6 included two different kink-turns (Kt-77 and Kt-78).[507] Both of these turns natively interact with the L1 protein but also participate in extensive RNA/RNA tertiary interactions within the stalk. Thus, they were fully stable even in simulations without the protein. However, when simulated in complete isolation, both kink-turns showed increased flexibility and occasional structural changes.[507]

4.7.8. tRNA Complexes with Aminoacyl tRNA Synthetase. Aminoacyl tRNA synthetase (AARS) is a protein that assigns amino acids to their cognate tRNA molecules, "charging" them for use in the ribosome during proteosynthesis. It accomplishes this in two separate steps. It first catalyzes the reaction between a free amino acid and ATP to produce an aminoacyl-AMP, which is then attached to the cognate tRNA molecule. An organism typically needs a full set of 20 unique AARS proteins (corresponding to the 20 biological amino acids) to survive, and many structures of AARS proteins complexed with their cognate tRNAs have been experimentally determined.[1097] MD simulations can be used to examine the dynamics of the AARS/tRNA interaction and to identify the properties responsible for the specificity of their recognition. Molecular modeling and subsequent simulations also make it possible to examine the complexes of AARS proteins with noncognate tRNA molecules. AARS/tRNA complexes are among the most studied protein/RNA systems in the simulation literature, second only to the U1A RRM complex.

In one of the first works on these systems, Yamasaki et al. explored the dynamics of the *E. coli* glutaminyl-RS complexed with its cognate tRNA and with a variant of this tRNA carrying a mutation in its variable loop.[1098] The authors wanted to understand why the mutant is bound with a higher affinity despite the variable loop being located far away from the protein/RNA interface. Their simulations revealed that the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

internal interactions of the free wild-type tRNA were dynamically rearranging, and that these dynamics were suppressed upon binding to the AARS. Conversely, the mutant exhibited no such dynamic behavior, either as a free species or when complexed with the protein. The authors thus proposed that the greater binding affinity of the mutated tRNA might be due to a greater entropic penalty associated with the binding of the wild-type tRNA to the AARS.[1098]

Mutual recognition between the AARS and tRNA molecules is primarily based on RNA sequence determinants occurring in the evolutionarily oldest tRNA acceptor stem, which probably reflects features that evolved in the early stages of the RNA world.[1097] Nevertheless, it is likely that modern AARS also have allosteric communication mechanisms involving distant parts of the protein, allowing some coupling between the anticodon and aminoacyl sites that lie on opposite sides of the tRNA molecule.[1099] Therefore, a goal of several theoretical studies has been to explore the structural basis for the coordination and communication between distal segments of the AARS/tRNA complex. Ghosh et al. used MD simulations to explore the cross-correlations of residues in the *E. coli* methionyl-RS/tRNA/methioninyl complex and its individual components.[1100] They also applied network analysis to elucidate the paths of communication within the complex. A strong communication path was observed within the AARS when both the tRNA and the activated methionine were bound. The authors suggested that this path may contribute to the fidelity of AARS/tRNA/amino acid recognition.[1100] Later, the same authors used the network analysis method in another study where they performed MD simulations of the human tryptophanyl-RS complex.[1101] Then, they used this method to analyze MD simulations of the cysteinyl-RS complex,[1102] focusing on a communication path between the aminoacyl and anticodon sites of the tRNA. Involvement of several key protein residues in this path was suggested and then explored by simulating mutant protein complexes.[1102] The dynamical network of interactions in the AARS/tRNA complexes was also studied by Sethi et al., who performed MD simulations of bacterial glutamyl-RS and archaeal leucyl-RS complexes.[1103]

Bushnell et al. explored the catalytic mechanism of the threonyl-RS aminoacylation reaction.[1104] Specifically, they wanted to clarify the structural role of the His309 residue, which has been experimentally shown to be important for catalysis.[1105] The authors used molecular modeling to obtain a catalytically relevant starting geometry for the AARS/tRNA/threonyl complex because no suitable experimental structures capturing this geometry exist. They then explored the system's stability in MD simulations involving different His309 protonation states and with or without the presence of a bridging water molecule. The system featuring a δ-protonated His309 without a bridging water proved to be the most structurally stable, and the authors suggested that His309 provides structural stabilization for the A76 nucleotide that receives the threonyl moiety in the aminoacylation reaction.[1104] Finally, among other studies, Grant et al. used MD simulations to further refine their experimental structural model of the eukaryotic glutaminyl-RS/tRNA complex, which was originally derived by combining X-ray and SAXS data.[1106]

4.7.9. Endonuclease Protein/RNA Complexes. The RNA endonucleases (endoribonucleases) are protein enzymes that cleave the RNA sugar−phosphate backbone at an internal register of the RNA chain. The endonucleases are involved at all levels of RNA metabolism. There have been several simulations of endonucleases complexed with their RNA substrates. MD simulations are useful when studying these complexes because they make it possible to reverse the structural effects of protein or RNA mutations that suppress the complex's catalytic activity but are often necessary to obtain usable data in structural experiments. This allows one, for example, to examine the dynamics and population of the relevant catalytic geometries and to conduct in silico mutagenesis studies. Simulations can also clarify the protonation states of the catalytic residues, which can be only indirectly guessed from the X-ray structures.

The study by Estarellas et al. on the clustered regularly interspaced short palindromic repeat (CRISPR)/Csy4 endoribonuclease in complex with RNA clearly highlights the strengths and shortcomings of the simulation approach.[1107] In an extended set of MD simulations, the authors tested all protonation states of the catalytically important His29 residue, using the Csy4 protein complexed with the RNA substrate as the starting structure.[1108] In a broader sense, the authors discussed the influence of starting structures and force-field limitations on simulations of protein/RNA complexes.[1107] The results are summarized in section 4.8.7.

The ribonuclease H (RNase H) complex with its DNA−RNA hybrid substrate was studied by Suresh et al.[1109] RNase H is an enzyme capable of cleaving the RNA strand in RNA/DNA mixed substrates without harming the DNA strand. Its catalytic site contains two magnesium ions, and the reaction occurs via a two metal ion mechanism.[518] The authors used MD simulations to examine the dynamics of the complex, the free protein, and the substrate. Their results suggested that the catalytic site is preorganized in the protein before the substrate binds, and that the protein induces geometrical changes in the RNA/DNA helix, particularly in the DNA strand.[1109]

The RNase H endonuclease was further explored by Figiel et al., who studied the HIV-1 reverse transcriptase (RT) in complex with an RNA/DNA substrate.[1110] In addition to its polymerase activity, the HIV-1 RT also has an RNase H domain that facilitates the breakdown of the RNA primer during the reverse transcription of the viral genome. It has been assumed that a substrate cannot simultaneously interact with both catalytic sites, implying that the enzyme periodically halts during the reverse transcription to switch between the polymerase and RNase binding modes.[1111] Figiel et al. used a combination of MD simulations and biochemical experiments to prove that the substrate can, in fact, interact with both sites simultaneously in a catalytically productive way (Figure 58).[1110]

In a follow-up study, Figiel et al. further explored the substrate specificity of the HIV-1 RT's RNase activity, and its interaction with the polypurine tract (PPT) primer sequence.[1113] The PPT primer RNA nucleotide sequence is refractory to cleavage by RNase H, and its protection from cleavage is critical for propagation of HIV-1 genome. The mechanism of this protection was, however, previously unknown.[1114] The MD simulations indicated that the enzyme can discriminate against certain base pair sequences through a combination of direct base/protein contacts and the intrinsic inability of some RNA/DNA substrates to assume a productive catalytic conformation. A set of simulation restraints was used to establish the productive conformation of the RNA/DNA substrate with the HIV-1 RT transcriptase, revealing that a substrate's ability to adapt to the RNase active site is sequence-dependent: some RNA/DNA duplexes are too rigid or too brittle to satisfy the structural requirements. This was

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

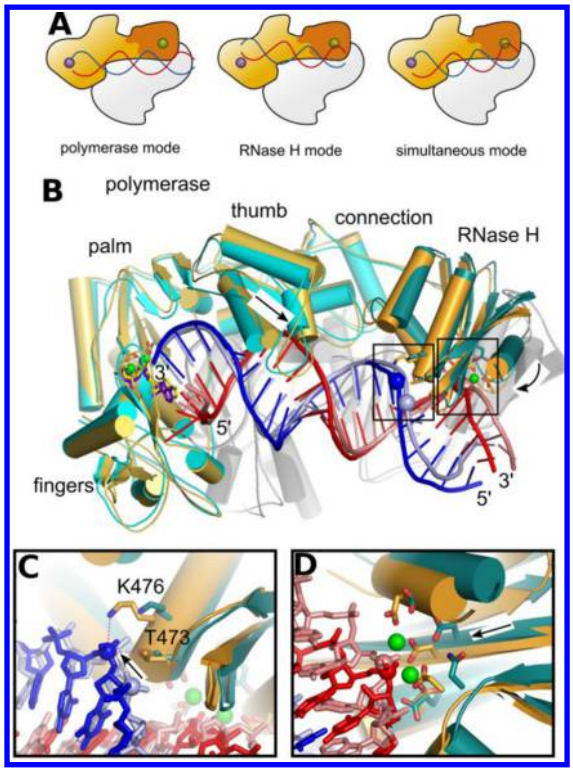


Figure 58. (A) Three possible modes of HIV-1 RT interaction with an RNA/DNA substrate. The RNase H domain is in dark orange. The polymerase and RNase H active sites are represented by purple and green circles, respectively. DNA and RNA strands of the hybrid substrate are shown in blue and red, respectively. (B) Superposition of starting structure (protein in cyan and light gray; RNA/DNA in light shades of red and blue, respectively) and final structure in MD simulations (protein in orange or darker gray; RNA/DNA in darker shades of red and blue, respectively).[1110] The starting structure was based on the crystal structure of HIV-1 RT bound to RNA/DNA substrate in polymerase mode (PDB ID: 4PQU).[1112] Residues that form the active sites are shown as sticks. Scissile phosphates and phosphates bound in the phosphate binding pocket are shown as spheres. Incoming nucleotides are shown in purple and yellow for the starting structure and the final MD structure, respectively. Mg^{2+} ions are shown as green spheres. Movements of the thumb and RNase H domains in the MD simulation are indicated with arrows. (C) Close-up of the phosphate-binding pockets of the starting and final MD structures. Residues that form the phosphate-binding pocket are shown as sticks and labeled. The positions of the phosphate group of nt-3 in the beginning of the simulation and at its end are indicated by spheres. The direction of movement of the DNA strand is indicated with an arrow. (D) Close-up of the RNase H active sites of the starting and final MD structures. Active site residues are shown as sticks. Scissile phosphates are shown as spheres. The direction of movement of the RNase H domain is indicated with an arrow. The figure was adapted with permission from ref 1110. Copyright 2017 Oxford University Press.

manifested in distortions of the helical geometry and base pairing when restraints were imposed to force the RNA substrate into the catalytic conformation. The greatest distortions were observed when the poly rA/dT tract of the PPT sequence was restrained at the catalytic site, providing a structural explanation of its ability to resist cleavage. The results of these simulations were subsequently verified by extensive biochemical experiments. Together, the simulations and experiments suggest the existence of a transiently populated conformation of the HIV-1 RT substrate complex that is important for modulating and coordinating the enzymatic activities of HIV-1 RT.[1113]

Last, Palermo et al. studied the CRISPR/Cas9 complex.[1115] CRISPR/Cas9 is a complex enzyme that binds a guide RNA molecule to cleave a complementary DNA strand.[1116] It is thus a DNA endonuclease. By applying cross-correlation and principal component analyses, the authors explored how the movement of the individual protein domains allows for the binding of the nucleic acid substrate. They also describe conformational changes that, in their opinion, lead to a structure capable of catalysis.[1115] Later, they further expanded their study by performing Gaussian-accelerated MD simulations[270] (section 3.2.4) in which they explored the intermediary conformational states of the Cas9 between its free and RNA-bound forms and described the associated dynamics of the individual protein domains.[1117] The CRISPR-Cas9 was also studied by Zuo et al., who explored the role of the catalytic Mg^{2+} ions and their influence on the substrate's conformation.[1118]

4.7.10. Some Other Protein/RNA Complexes. There have also been simulations of protein/RNA systems that do not fall into any of the above categories.

HIV Rev-Responsive Element (RRE) RNA was studied by Michael et al., who examined its binding to the arginine-rich Rev peptide and its analogue RSG 1.2.[1119] RRE is a specific sequence in the HIV genome whose binding to Rev affects the regulation of the viral life cycle.[1120] The authors used MD simulations to examine the dynamics of both complexes and to clarify the greater affinity of the RSG 1.2 peptide for the RRE RNA. They found that both peptides achieve specificity through a combination of direct base/arginine H-bonds, recognition of shape via the sugar−phosphate backbone, and cation/π interactions. Stronger binding of the RSG 1.2 as compared to the Rev peptide was explained by a greater number of base/arginine interactions, greater involvement of water molecules in stabilizing the protein/RNA interface, and an intricate network of salt-bridge interactions in the former system.[1119] The elucidation of the water molecules' role is a good illustration of the capabilities of MD simulations: such information cannot be obtained from NMR structures and can only be studied on the basis of static pictures when working with X-ray structures.

A structurally similar system is the arginine-rich P22 N-peptide/boxB complex. The function of the P22 protein is to inhibit transcription termination of the bacteriophage genome by binding to the boxB RNA.[1121] This protein/RNA complex was studied by Bahadur et al.,[1122] who used MD simulations and MM-PBSA calculations to study its dynamics and assess the contribution of individual residues to the protein/RNA binding energy, obtaining qualitative agreement with experimental results. As in the HIV-1 RRE/Rev complex, several stable hydration sites were identified at the protein/RNA interface. The authors also observed a conformational change associated with the protein/RNA binding event, and suggested that a coil to α-helix transition of the P22 N-peptide occurs immediately prior to binding under the influence of the RNA substrate's electrostatic field.[1122]

Mori et al. explored the HIV-1 nucleocapsid protein-7 (NCp7) complexed with both DNA and RNA stem-loop substrates. The authors described the interactions in both complexes and discussed their differences as well as the structural roles of individual protein residues in the binding.[1123]

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

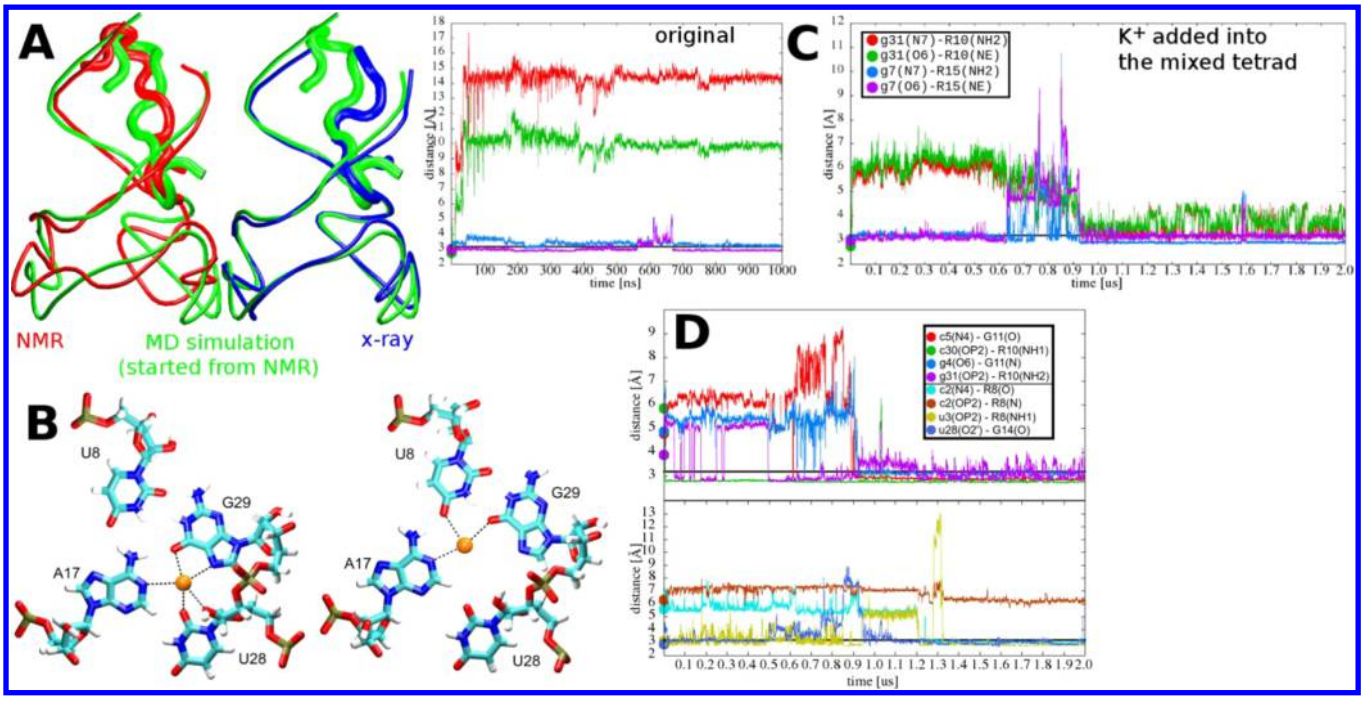


Figure 59. (A) Overlay of RNA and protein backbones of RNA duplex/quadruplex junction complexed with FMRP RGG peptide as observed in NMR (red),[1125] X-ray (blue),[1126] and an averaged MD structure (green); the latter corresponds to the last 100 ns of a 2 μs MD simulation.[508] (B) NMR structure of the mixed quartet upon manual insertion of a K^+ ion (left) and the final conformation after ~620 ns. (C) Simulation time development of H-bond heavy atom distances of the four protein/RNA H-bonds present in the NMR structure. The interactions were unstable in the initial simulations (left) but were eventually stabilized in simulation with K^+ ion manualy inserted into the mixed quartet (right). (D) Simulation time development of eight H-bond heavy atom distances of the interactions predicted by the MD and seen in the X-ray structure, which are not seen in the starting NMR structure. The MD simulation spontaneously moves from the NMR structure to the more accurate X-ray structure. Panels (A) and (D) have been adapted with permission from ref 154. Copyright 2016 John Wiley and Sons. Panels (B) and (C) have been adapted with permission from ref 508. Copyright 2015 American Chemical Society.

The complexity of protein/RNA simulations is well illustrated by the complex of the fragile X mental retardation protein (FMRP) arginine-glycine-rich RGG peptide with RNA duplex/quadruplex junction (Figure 59). FMRP protein is a human regulatory protein that binds with high affinity to guanine-rich RNA sequences capable of forming G-quadruplexes.[1124] MD simulations of this complex[508] were based on an NMR structure that was available at the time,[1125] and could not stabilize its seemingly simple recognition pattern, which is based on just four H-bonds. After many unsuccessful runs, the authors recognized that a monovalent ion was trying to enter a mixed RNA quartet, serving as an interface between the G-quadruplex and an A-form helical region. However, the ion was not able to enter because entry would require a temporary opening of the mixed quartet, which was not achievable on the simulation's time scale. The authors therefore manually placed an ion inside the mixed RNA quartet structure and resumed the simulation; after more than 0.5 μs, the ion found its final position, and within a further 0.2−0.3 μs, the structure neatly repaired its native protein/RNA contacts.[508] This is a striking example showing that seemingly marginal omissions in the starting structure can have dramatic effects on simulations (Figure 59). It also shows the need to adapt one's simulation strategy on the basis of interim results, given the unfeasibly long simulation times that would have been required for spontaneous relaxation of the initial structure with the ion outside the quartet. The same system was subsequently characterized by X-ray crystallography,[1126] which confirmed the mixed quartet ion-binding site predicted by the simulations. However, it also revealed a much more structured protein/RNA interface than was apparent in the NMR structure,[1125] because the peptide was much more structured in the crystallographic structure. The discrepancy was primarily attributed to under-determination of the solution structure, although some NOEs indicated the presence of modest real differences between the solution ensemble and the X-ray structure. Upon reinspecting the earlier trajectory after manual insertion of the monovalent ion,[508] the MD group found out that the simulation was actually converging from the NMR structure to the X-ray structure, and captured all but one of the protein/RNA H-bonds seen in the X-ray structure (Figure 59). This can be considered as an independent test of the reasonable performance of recent AMBER force fields (χ_{OL3} and ff12SB for the RNA and protein components, respectively).[154]

The NS3 helicase of the hepatitis C virus was studied by Pérez-Villa et al.[1127] NS3 is an ATP-dependent enzyme that translocates along an ssRNA substrate. The authors utilized two available experimental structures of the NS3/complex with and without bound ATP.[1128] In addition, they used molecular modeling to study the ADP-bound system to explore the structure of the complex after the hypothetical ATP hydrolysis event. For each of these systems, the authors described the available conformational space, the protein/RNA interface, and the ligand binding pocket. Their results suggested that the presence of ATP stabilizes a closed conformation of the NS3/RNA complex, whereas an open conformation is favored upon ATP hydrolysis, explaining the differences observed in the experimental structures.[1127]

Sharma et al. studied different isoforms of pre-Let-7 miRNA in complex with LIN28 protein.[1129] On the basis of the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

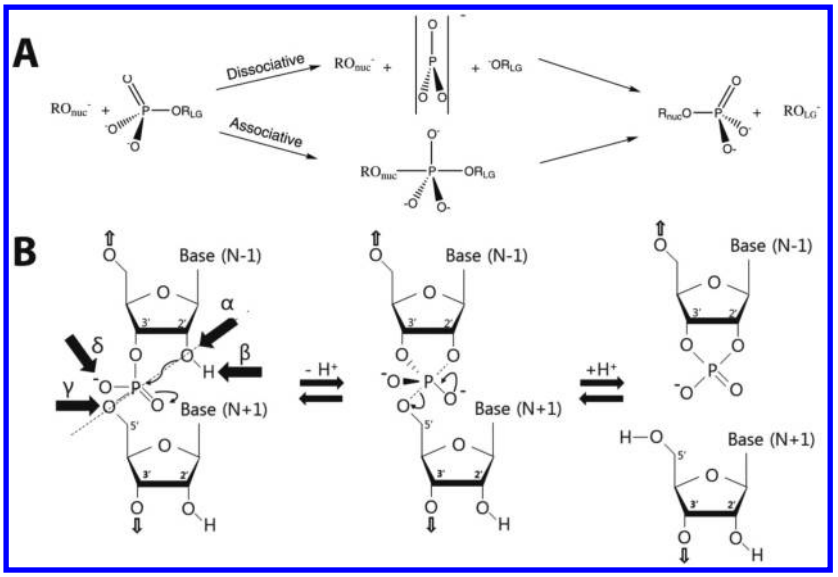


Figure 60. (A) Associative and dissociative mechanisms of the transesterification reaction. Adapted from ref 1143; http://creativecommons.org/licenses/by/4.0. Copyright 2017 MDPI AG. (B) A general mechanism of sugar−phosphate backbone cleavage with four highlighted catalytic strategies: (α) preorganization of in-line-attack conformation, (β) activation of the 2′-OH nucleophile, (γ) protonation of the leaving group, and (δ) electrostatic stabilization of the negative charge developed on the scissile phosphate. Reprinted with permission from ref 1147. Copyright 2007 American Chemical Society.

simulations, they suggested that the structural features of this miRNA are more critical for protein/RNA recognition than the sequence conservation. They also explored the influence of the loop length of the pre-Let-7 miRNA stem-loop on the protein/RNA binding.[1129]

A series of studies used Markov state models (MSMs, see section 3.2.3) to study kinetics of RNA polymerase activity while in complex with its RNA/DNA substrate.[1130] Specifically, they investigated processes of pyrophosphate ion release from RNA polymerase II (Pol II), comparing it with bacterial and viral RNA polymerase;[1131−1133] they showed that in the first case the process occurs in a four-state fashion, while in viruses and bacteria it consists of a single two-state transition. Silva et al. studied the translocation step of Pol II, building a model consistent with the previously proposed Brownian ratchet mechanism, in which the translocation is driven by thermal oscillations of the bridge helix.[1134] The MSM results of refs 1131 and 1134 have later been combined with rates estimated from single-molecule experiments, to build a kinetic model of the full nucleotide addition cycle in Pol II.[1135] Also, the process of backtracking in Pol II was studied,[1136] revealing an intermediate state in which the RNA 3′-end frays due to the mismatch with its DNA counterpart, helped by the bridge helix bending; after this, the DNA nucleotide stacks with the bridge helix, and the backtracking of the RNA/DNA complex occurs in the last step. These studies show how MSMs can be used to combine multiple smartly initiated simulations, each no longer than tens to hundreds of nanoseconds, to estimate the thermodynamic and kinetic properties of processes with time scales on the order of hundreds of microseconds or even longer.

Last, Huang and co-workers also used a combination of MSM and docking to study the miRNA−Argonaute interaction.[1137]

4.8. RNA Catalysis

4.8.1. Introduction to RNA Catalysis. Studies of RNA catalysis are among the most interesting but also most challenging RNA computations. In addition to using extended explicit solvent MD simulations, a number of studies have employed electronic structure methods, mainly in the form of various hybrid QM/MM calculations.[1138] While MD simulations aim to complement the available structural data by exploring structural excursions around them, QM/MM calculations are used to predict the most likely chemical reaction pathways.[1139] Mutational analyses and chemical modifications combined with classical enzymology are often able to identify specific chemical groups that are crucial for catalysis; however, a detailed description of their particular atomic-level chemical roles in the catalytic mechanism is significantly more challenging due to the underlying notorious ambiguity in distinguishing functionally related roles.[1140,1141]

QM, QM/MM, and MD calculations can complement the experimental data by evaluating various mechanistic hypotheses, yielding important insights into the catalytic mechanism with unrivaled spatial and temporal resolution. In particular, theoretical calculations can estimate the chemical reaction barriers and corresponding kinetic rate constants for comparison with experimental observations, provided appropriate starting geometries are available. The naturally occurring currently known nine classes (see below) of small ribozymes catalyze self-cleavage and ligation of their sugar−phosphate backbones.[1142,1143] The majority of them have been crystallized and extensively studied by computational approaches. To suppress self-cleavage for structural characterization of their precursor forms, chemical modifications are typically introduced into their active sites.[1144,1145] These modifications often significantly perturb the active site architecture and thus complicate interpretation of the structural data. MD simulations can deduce the effects of the introduced modification by computationally removing it. In addition, simulations can probe the effects of crystal packing, giving them significant potential for complementing structural data.

The S_N2 type self-cleavage reaction of the small ribozymes is initiated by a nucleophilic attack of the 2′-OH group on its

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

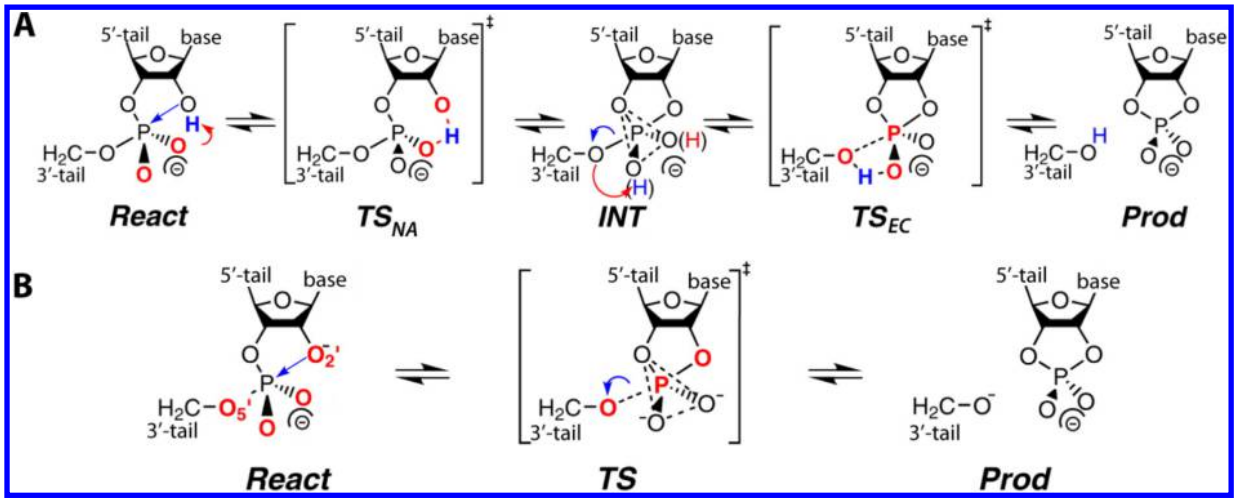


Figure 61. General reaction mechanisms of (A) monoanionic and (B) dianionic self-cleavage of the sugar−phosphate backbone. In the monoanionic mechanism, the 2′-OH nucleophile is activated by nonbridging oxygen of scissile phosphate and the proton is shuttled toward the leaving O5′ group. In the dianionic mechanism, the nucleophile might be activated either by a general base or by a specific base, while the leaving group is to be protonated by a general acid. Adapted with permission from ref 1148. Copyright 2008 American Chemical Society.

adjacent scissile phosphate, proceeds through a pentacoordinated phosphorane transition state (and eventually a phosphorane intermediate state), and generates 2′,3′-cyclic phosphate and 5′-OH termini as products.[1146] The 2′-OH nucleophile needs to be activated (deprotonated) prior to or during the initial phase of the reaction, while the O5′ leaving group needs to be protonated upon the cleavage reaction. It is therefore commonly assumed that both of these proton transfers can be facilitated by general acid−base catalysis. QM/MM calculations have been used extensively to probe the underlying mechanism (Figures 60 and 61).

What follows is a comprehensive survey of the main results of MD simulations and QM calculations on the small ribozymes and some other systems, with analysis of the significance and limitations of the results.

4.8.2. Uncatalyzed Self-Cleavage Reaction as an Important Reference. The catalytic effect of enzymes and ribozymes is typically defined as a rate enhancement relative to the corresponding uncatalyzed reaction in water.[1149] Therefore, an unavoidable part of investigating RNA catalysis is a detailed characterization of the uncatalyzed O2′-transphosphorylation reference reaction, including the description of all possible reaction pathways and evaluation of the effects of introducing site-specific modifications such as thio-substitutions or external chemical groups directly or indirectly affecting the reaction. In pioneering works, Warshel et al. focused on the description of mono- and diphosphoester hydrolysis.[1150−1152] In particular, the authors compared two general mechanisms: the associative path and the dissociative path proceeding through a phosphorane transition state (or intermediate) with pentavalent phosphorus and a PO_3^- anion, respectively (see Figure 60A). They concluded that both paths have similar reaction barriers, so either of them can be favored within a ribozyme's active site, depending on the particular electrostatic environment[1152] and/or pK_a of the leaving group.[1150,1151]

Breaker et al. defined four independent catalytic strategies that each can contribute to acceleration of the O2′-transphosphorylation reaction:[7,1153] (i) preorganization of the active site in an in-line attack conformation of the 2′-OH nucleophile with respect to the scissile phosphate (with the O2′···P−O5′ angle approaching 180°), (ii) activation of the 2′-OH nucleophile via deprotonation by a general base, (iii) activation of the O5′ leaving group via protonation by a general acid, and (iv) stabilization of the negative charge developing on the scissile phosphate during the reaction (see Figure 60B). Yang et al. used free-energy MD calculations with CHARMM27 force field to probe the first strategy, the effect of preorganization into an in-line attack conformation.[1154] From their calculations (using a splicing endonuclease complex for comparison with the uncatalyzed reaction), they deduced that the contribution of an in-line attack conformational effect to catalytic activity is rather small. The authors predicted an ∼12-fold enhancement of the reaction rate,[1154] while ribozymes typically display much larger rate accelerations of $\sim 10^9$.[1153] These specific calculations would thus predict that the remaining three strategies play more dominant roles in RNA catalysis.

Depending on the architecture of the active site and mode of 2′-OH nucleophile activation, the cleavage reaction might proceed through either a deprotonated dianionic phosphorane or a single-protonated monoanionic phosphorane, leading to distinct dianionic and monoanionic mechanisms, respectively (see Figure 61). The dianionic mechanism, expected to be the preferred reaction pathway under alkaline conditions, was studied by Karplus and co-workers using DFT calculations to show that solvent stabilization has a crucial effect on the reaction barrier.[1155] York, Harris, and co-workers proposed, on the basis of QM/MM calculations with semiempirical Hamiltonians, that the exocyclic cleavage step leading to disruption of the P−O5′ bond represents the rate-determining step of the uncatalyzed reaction.[1156] In a subsequent review, the same team acknowledged that the functional context of the ribozyme active site might alter the rate-determining step.[1157] Leclerc and co-workers studied the mechanism of 2′-OH nucleophile deprotonation and compared the activation by hydrated magnesium ion and a guanine nucleotide.[1158] Their calculations indicated that, in addition to its possible role as the general base, a magnesium ion can act as the Lewis acid to lower particularly the pK_a of the 2′-OH group, facilitating the activation step of the reaction.[1158]

York, Harris, and co-workers combined QM calculations with measurements of kinetic isotope effects (KIEs) to probe several plausible pathways of phosphoryl transfer reactions and built a

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

model for the underlying linear free-energy relationships that can be applied to more complex systems such as ribozymes.[1159−1163] Huang and York found that the barrier connected with the exocyclic cleavage step depends on the pK_a of the leaving group.[1159] Reactions with a poor, unprotonated leaving group are expected to proceed via a stepwise mechanism characterized by a late, rate-determining transition state barrier. By contrast, a more facile leaving group results in a concerted mechanism characterized by a single, early transition state barrier.

The effect of thio-substitutions was also extensively studied by York and co-workers using QM and QM/MM calculations in combination with KIE and equilibrium isotope effect measurements.[1164−1169] They found that thio-substitution of either the *pro*-R_P or *pro*-S_P nonbridging oxygens, or of both, has no effect on the reaction barrier of the uncatalyzed transphosphorylation reaction,[1164−1166] but may stabilize the otherwise kinetically transient phosphorane intermediate; such a longer-lived thio-substituted phosphorane may then undergo a pseudorotation, yielding a 3′,5′-phosphodiester side product of the reaction.[1170,1171] By contrast, thio-substitution of the O2′ or O5′ is predicted to have a significant impact on the reaction, leading to a profile with a single transition state corresponding to exocyclic cleavage (O2′ thio-substitution) or nucleophile attack (O5′ thio-substitution).[1164−1166] Finally, thio-substitution of the O3′ oxygen is expected to result in a decrease of the rate-determining barrier associated with the exocyclic cleavage.[1164−1166]

Extensive QM calculations were performed to fully map the chemical space of the RNA sugar−phosphate backbone self-cleavage reaction, including the effects of various groups acting as general bases and/or general acids, or to electrostatically stabilize the transition state.[1172] These calculations investigated 115 different pathways and revealed that involvement of a general acid seems to have a dominant catalytic effect, while nucleophile activation by a general base and an electrostatic stabilization of the transition state are predicted to be less important. Significantly, the presence of both general acid and base leads to a catalytic enhancement comparable with that found in naturally occurring ribozymes as discussed below. This observation suggests that ribozymes do not need to exploit the long-range electrostatic effects common in protein catalysis. Rather, they likely achieve catalysis primarily via preorganization of general acid and base functional groups within their active sites.[1172] In other words, in contrast to most protein enzymes, a large part of ribozyme catalytic effects seems to be attributable to a change of the reaction mechanism as compared to the uncatalyzed reaction. This suspected difference between protein and RNA catalysis might be caused by the polyanionic nature of RNA, which makes it more difficult for evolution to tune a strong specific electrostatic field at the active site to steer catalysis. This finding may rationalize why proteins were ultimately favored over RNA enzymes for most biological processes not involving direct RNA:RNA contacts.

Note that uncatalyzed cleavage and its dependence on RNA structure is at the root of the in-line probing method.[1173] An MD approach to model in-line probing reactivities has been presented, although results were not completely satisfactory due to the poor performance of the employed semiempirical potentials.[323]

4.8.3. Small Self-Cleaving Ribozymes. The small self-cleaving (or endonucleolytic) ribozymes are RNA enzymes less than 200 nucleotides in length that catalyze an internal transphosphorylation reaction to cleave their own sugar−phosphate backbone (in *cis*).[1174] Because of their small size, they are popular targets for theoretical studies. So far, nine naturally occurring classes of small self-cleaving ribozymes have been discovered: the hammerhead,[1175] hepatitis delta virus (HDV),[1176] hairpin,[1177] Varkud satellite,[1178] *glmS*,[1179] and later twister,[1180] twister sister,[1142] pistol,[1142,1181] and hatchet ribozymes.[1142,1182] In addition, several artificial small self-cleaving ribozymes were evolved in vitro to probe the catalytic potential of RNA enzymes. Small self-cleaving ribozymes usually utilize their active site nucleobases as general acid and base catalysts. Only some of them (such as the HDV, hammerhead, and twister sister ribozymes)[1183] may also use active site divalent ions, although they are generally at least somewhat active in high concentrations of monovalent ions,[510,515] so they cannot be considered obligatory metalloenzymes (section 4.8.3.2). By contrast, the transphosphorylation reactions of the self-splicing (intron) ribozymes typically rely on direct participation of active site metal ions. Although the self-splicing ribozymes are significantly larger systems than the small self-cleaving ribozymes, they are also becoming computationally accessible as shown in the work by Magistrato and co-workers,[1184] and rapid progress in this field is expected in the near future due to progressively improving computational power and optimized computational algorithms.

4.8.3.1. Hammerhead Ribozyme. Hammerhead ribozyme, only the third ribozyme to be discovered, first in plant virus satellites and later a number of eukaryotic genomes,[1185] was the first small self-cleaving ribozyme to be studied by an interdependent combination of MD simulations and QM calculations (Figure 62).

4.8.3.1.1. Minimal Hammerhead Crystal Structures. The first crystallographic studies revealed the structures of minimal hammerhead ribozymes, consisting of the minimal length required for catalysis, with stems I, II, and III flanking the catalytic core (Figure 62A).[627,1188−1194] Two Mg^{2+} ions were found close to the scissile phosphate.[1189] However, none of these crystal structures reported the expected in-line attack conformation of the C17(2′-OH) nucleophile with respect to the scissile phosphate. Early MD simulations aimed to probe the conformational dynamics of the active site in a catalytically potent state.[482,1195−1198] Predating any hammerhead ribozyme structures, a pioneering MD study by Bruice and co-workers was based on a structural model generated on the basis of the T-arm observed in a tRNA crystal structure.[1196] The authors predicted that formation of a reactive in-line attack conformation would require a C3′-endo to C2′-endo reconformation of the C17 nucleotide.[1196] Such a repuckering was in fact observed in subsequent MD simulations starting from the minimal hammerhead ribozyme crystal structures.[482,1198] Westhof and co-workers showed by MD simulation that two active site Mg^{2+} ions may provide a binding site for a bridging hydroxyl ion, which in turn may form a direct interaction with the C17(2′-OH) nucleophile upon the C17 C3′- to C2′-endo conformational flip.[482] Their simulations thus supported the potential formation of a reactive in-line attack conformation, in which a μ-bridging hydroxyl ion (a hydroxyl bridging the two above-mentioned active-site Mg^{2+} ions) activates the C17(2′-OH) nucleophile and thus plays the role of general base.[482] However, Karplus and co-workers argued that such a mechanism should be deemed unlikely on the basis of the fact that the pK_a value of the C17(2′-OH) nucleophile, calculated by QM to be 14.9, is too high, rendering

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

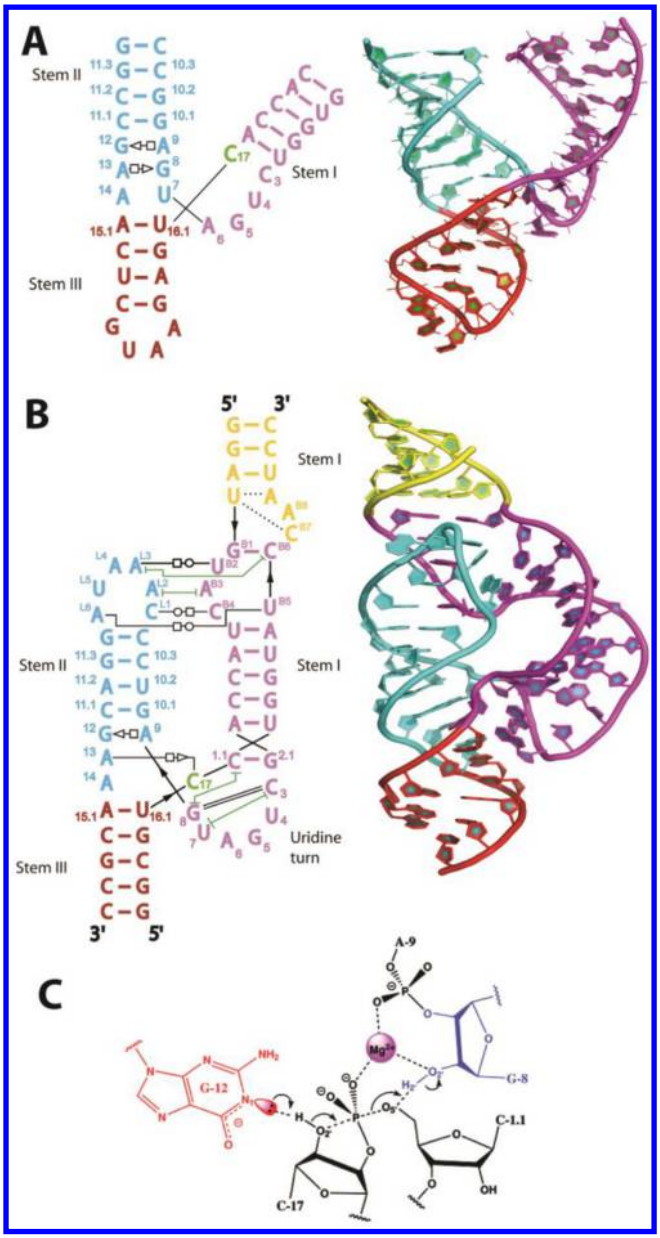


Figure 62. 2D and 3D structures of (A) minimal and (B) full-length hammerhead ribozyme. The figure was partially adapted with permission from ref 1186. Copyright 2016 Elsevier Ltd. (C) A self-cleavage mechanism suggested on the basis of crystal structures of full-length hammerhead ribozyme. Adapted with permission from ref 1187. Copyright 2010 American Chemical Society.

it unlikely that it can be deprotonated by a metal-coordinated hydroxyl ion with its comparatively lowered basicity.[1199] Instead, they suggested that an active site Mg^{2+} ion may play a role in transition state stabilization.[1199] Another mechanism was suggested by MD simulations performed by Bruice and co-workers,[1197] reporting an alternative path for the necessary sugar−phosphate backbone reconformation to yield an in-line attack conformation. The latter simulations provided support for a mechanism wherein only one of the active site magnesium ions is directly involved in the self-cleavage reaction, while the other instead plays a structural role.[1197] Furthermore, the authors suggested that the rearrangements observed in MD simulations may be caused by the structure relaxing from the pseudocontinuous helix formed between stems I and II of neighboring molecules in the crystal lattice.[1198] These observations motivated subsequent QM studies to probe specific plausible reaction mechanisms.[1200,1201] In particular, Lovell and co-workers reported reaction profile of the pathway involving direct participation of a single active site Mg^{2+} ion in catalysis. In this mechanism, a hydroxyl ion coordinated to the catalytically important active site Mg^{2+} ion is acting as the general base activating the C17(2′-OH) nucleophile, while a water molecule coordinated to the same Mg^{2+} ion acts as the general acid protonating the O5′ leaving group.[1201] This mechanism further assumes inner-shell coordination of the catalytic Mg^{2+} ion to the *pro*-R_P nonbridging oxygen of the scissile phosphate so that it can act as a Lewis acid by stabilizing the negative charge developed on the scissile phosphate during cleavage.[1201] In contrast, Leclerc and Karplus modeled the reaction profile involving the direct participation of two active site Mg^{2+} ions.[1200] In the latter mechanism, a hydroxyl group coordinated to one Mg^{2+} ion acts as the general base to activate the C17(2′-OH) nucleophile, while a water molecule coordinated to the second ion serves as the general acid to protonate the O5′ leaving group. In addition, both ions act as Lewis acids to stabilize the negative charge developed on O2′ and the nonbridging oxygens of the scissile phosphate during self-cleavage.[1200] The overall self-cleavage mechanism was further described by Radhakrishnan using a combination of classical MD simulations, umbrella sampling (see section 3.2.5), and QM/MM calculations.[1202] The suggested reaction path started from the minimal hammerhead crystal structure, proceeded via conformational rearrangement up to the chemical cleavage catalyzed by two Mg^{2+} ions, and indicated a coupling between these conformational and chemical steps.[1202] Taken together, all of these MD simulations and QM calculations based on the early crystallographic data supported mechanisms requiring an initial conformational switch followed by self-cleavage catalyzed by either one or two active site Mg^{2+} ions.

4.8.3.1.2. Full Length Hammerhead Crystal Structures. The above data were found to be inconsistent with biochemical experiments, in particular with measured thio substitution effects[1203,1204] and mutational data.[1205] The discrepancy between mechanistic and structural data was finally explained by a new crystal structure of a full-length hammerhead ribozyme presented by Martick and Scott (see Figure 62B).[1186] In particular, this crystal structure revealed a significantly different arrangement of the active site with close contact between the A9 phosphate and the scissile phosphate that were proposed to be bridged by an active site Mg^{2+} ion that itself was not resolved.[1204] This motivated York, Scott, and co-workers to probe the structural dynamics of the new active site architecture of the full-length hammerhead ribozyme by MD simulations.[161] This work supported a catalytically active conformation with an Mg^{2+} ion bridging the A9 and scissile phosphates, G12(N1) in a position to act as the general base after itself becoming deprotonated by an exogenous base, and the 2′-OH group of G8 in position to act as the general acid (see Figure 62C).[161] In addition to its structural role, the authors suggested that the active site Mg^{2+} ion may act as a Lewis acid via direct coordination to the C1.1(O5′) oxygen and may shift the pK_a of the G8(2′-OH) general acid.[161] The same team further supported the general base and general acid roles of G12(N1) and G8(2′-OH), respectively, by additional crystal structures and MD simulations.[162] Their work also suggested that the G12(N1) general base may become deprotonated by a

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

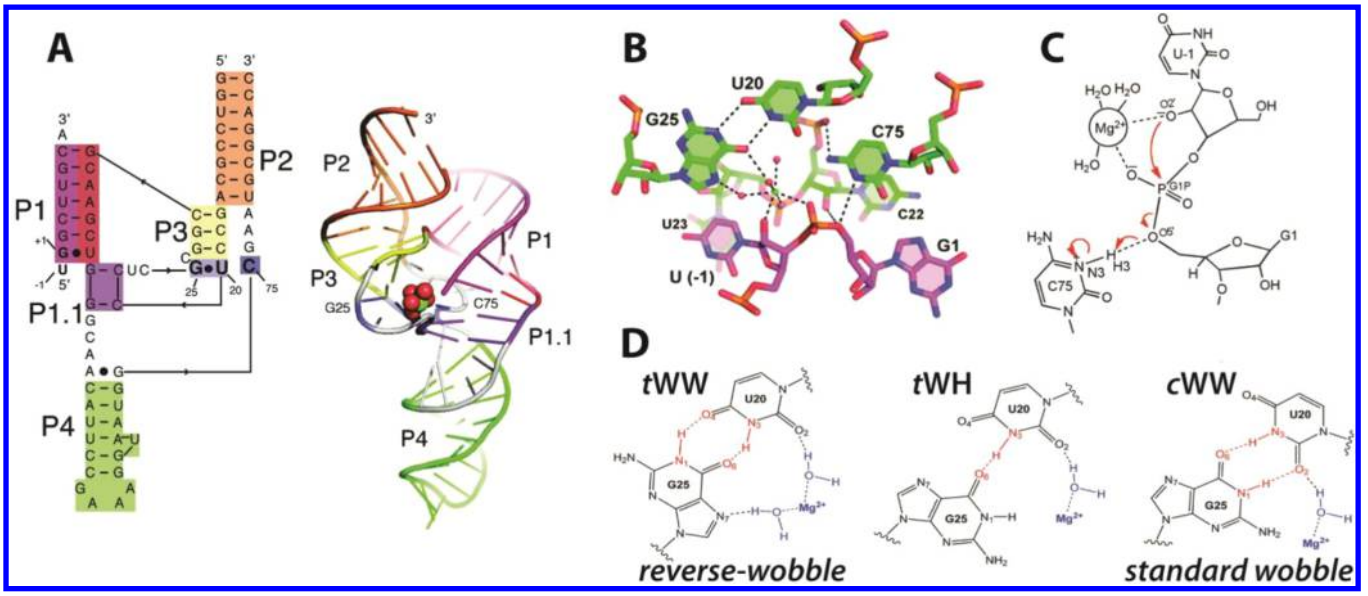


Figure 63. (A) 2D and 3D structures of the HDV ribozyme, and (B) a catalytically potent structure of the active site including hydrated Mg^{2+} ion, scissile phosphate, and U-1 nucleotide that were modeled by molecular modeling methods and MD based on the structure of hammerhead ribozyme and the newer X-ray structure of HDV ribozyme. (C) Suggested reaction mechanism involving hydrated Mg^{2+} ion and protonated cytosine $C75H^{+}$ acting as the general base and general acid, respectively, and (D) three different conformations of G25/U20 base pair of the L3 loop observed in X-ray structures and MD simulations; note that the *t*WW conformations have G in *syn* conformation. Panels (A) and (B) are reprinted with permission from ref 1221. Copyright 2010 American Chemical Society. Panel (C) is reprinted with permission from ref 1222. Copyright 2014 American Chemical Society. Panel (D) is reprinted with permission from ref 1223. Copyright 2016 American Chemical Society.

hydroxide ion coordinated to the active site Mg^{2+} ion,[162] and its pK_a is shifted toward neutrality by binding of another Mg^{2+} ion into G-site, that is, near the Hoogsteen edge of G12.[1206] The team then studied the structural role of the active site Mg^{2+} ion in more detail to conclude that the native fold of the hammerhead ribozyme creates a pocket with high negative electrostatic potential near the active site that recruits the catalytic divalent ion, which in turn helps to adopt the reactive in-line attack conformation.[1207] Further MD simulations by Lee and York[1187,1208] showed that the active site arrangement observed in the full-length hammerhead ribozyme is consistent with mutagenesis experiments.[1205] Specifically, the authors showed that the C3═G8 base pair, the hydrogen-bonding interactions between C17 and G5, and stacking of G8 on C1.1 all play crucial roles in buttressing the reactive architecture such that nonisosteric mutations of these crucial interactions are expected to lead to loss of catalytic activity.[1187,1208] Subsequent QM/MM-based free-energy calculations[1209,1210] of the suggested mechanism supported the notion that the active site Mg^{2+} ion indeed acts as a Lewis acid to stabilize the negative electrostatic potential developing on G8(2′-O^{-}) as it acts as the general acid during self-cleavage; in addition, the Mg^{2+} stabilizes both the in-line attack conformation and the interaction between the C17(2′-OH) nucleophile and the G12(N1) general base.[1209,1210] The authors predicted the exocyclic cleavage to be the rate-determining step of catalysis.[1210]

A slightly different mechanism was suggested by Golden and co-workers based on their later crystal structures.[1211,1212] Here, a water molecule coordinated to the active site Mg^{2+} ion acts as the general acid instead of the Mg^{2+}-coordinated G8(2′-OH). Still, consistent between the two mechanisms of hammerhead ribozyme self-cleavage, the C17(2′-OH) nucleophile is activated by the N1-deprotonated G12, and the leaving group is protonated by either G8(2′-OH) or a water molecule coordinated to an active site Mg^{2+} ion. It should be noted that the striking difference between these newer mechanistic proposals and the earlier ones based on a more direct involvement of one or two active site Mg^{2+} ions does not necessarily make them mutually exclusive; the latter class of mechanisms may be relevant for the minimal hammerhead ribozyme at high Mg^{2+} concentration where a significant conformational rearrangement needs to precede catalysis. After all, the hammerhead ribozyme can self-cleave in the presence of both millimolar concentrations of divalent metal ions and molar concentrations of monovalents,[515,1213,1214] and other ribozymes such as the hepatitis delta virus ribozyme have been suggested to also follow a multichannel reaction mechanism.[510] Indeed, MD simulations suggested that the distal loop−loop interactions of a full-length hammerhead ribozyme promote the necessary rearrangement to activate catalysis at much lower Mg^{2+} concentrations through long-range coupled molecular motions,[919] whereas a newer crystal structure suggests that monovalents can functionally replace active site divalents.[1214]

4.8.3.2. Hepatitis Delta Virus Ribozyme. Hepatitis delta virus (HDV) ribozyme is another extensively studied small self-cleaving ribozyme, and the only known ribozyme embedded in a human pathogen, HDV, as well as the human genome and genomes from many other branches of life (Figure 63).[1215] The first attempt to theoretically describe its mechanism was performed by Westhof and co-workers using modeling to predict its 3D structure.[1216] On the basis of the model and extensive mutational analysis, the authors correctly predicted that the ribozyme folds into a pseudoknot structure (see Figure 63A) with crucial C75, U20, and C21 nucleotides localized close to the catalytic core.[1216] This prediction was confirmed by the first crystallographic studies reporting structures of a cleaved product state,[593] and a precursor state inhibited by C75U mutation.[594] Biochemical data strongly suggested direct participation of nucleobase C75 (equivalent to C76 of the antigenomic sequence) in the self-cleavage reac-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

tion.[1144,1217−1219] Two different mechanisms of HDV ribozyme self-cleavage were initially suggested: (i) A mechanism with a neutral C75 acting as the general base activating the U-1(2′-OH) nucleophile for nucleophilic attack, and (ii) a mechanism in which C75 instead is assumed to be N3-protonated to be able to pass this proton to the G1(O5′) leaving group and thus act as the general acid. In both mechanisms, a hydrated magnesium ion was suggested to play the complementary role of general acid and general base, respectively. While the crystal structure of the cleaved product state involved a C75(N3)···G1(O5′) hydrogen bond, suggesting that C75 acts as the general acid via protonation of the leaving G1(O5′) group, C75U-inhibited precursor structures were instead consistent with the general base role of C75. However, the mutated precursor C75U crystal structures revealed neither the U75(N3−H)···U-1(2′-O) hydrogen bond corresponding to the C75(N3)···U-1(2′-OH) H-bond crucial for activation of the U-1(2′-OH) nucleophile nor the in-line-attack angle of U-1(2′-OH) needed for nucleophilic attack on the scissile phosphate. In addition, the very low-resolution C75 precursor crystal structure completely lacked resolution for the U-1 nucleotide and thus also did not directly support crucial structural features for the general base role of C75. Nonetheless, the authors suggested that structural adaptations toward the catalytically potent geometry would be accessible from the crystal structure by a simple rigid-body rotation around the U-1(O3′-P) bond.[594] In addition, the base pairing of G25/U20 differed in the crystal structures of the precursor and product states. More specifically, the available electron densities were interpreted as a G25/U20 *t*WH[39] base pair in the inactivated precursor, and an unusual *t*WW base pair with the G flipped into a *syn* conformation in the postcleavage product (see Figure 63D). Such a *syn*/*anti* flip of a base pair would require large-scale structural dynamics in the flanking loop region upon cleavage, supporting the transition through a conformational switch as indicated by fluorescence resonance energy transfer (FRET) experiments.[5,1220]

4.8.3.2.1. Computations Based on the Earlier Crystal Structures. The availability of atomistic crystal structures inspired a series of computational studies, first aiming to explore the active site conformational landscape in the reactive precleavage state.[503,1224,1225] Simulations starting from the C75U-inhibited precursor structure revealed, after replacing U75 with C75, that the U-1(2′-OH) nucleophile indeed can spontaneously flip into a reactive in-line-attack conformation. The simulations observed a minor population of the C75-(N3)···U-1(2′-OH) hydrogen bond required for the activation of U-1(2′-OH) when a neutral C75 acts as the general base.[1224] Other researchers showed that the simulations of the activated U-1(2′-O$^-$) precursor resulted in spontaneous formation of the in-line attack conformation required for self-cleavage.[1226] In contrast, the C75H$^+$(N3−H)···G1(O5′) hydrogen bond required for C75 to act as the general acid was never observed in the first set of precursor MD simulations, suggesting that the available precursor structures were rather incompatible with this mechanism.[1224] However, the catalytically unfavorable position of C(U)75 with respect to the scissile phosphate observed in the original precursor crystal structure could have been affected by crystal packing. The adjacent G76 nucleotide is restricted by stacking with G76 from a neighboring molecule in the crystal lattice. Indeed, MD simulation lacking these crystallographic constraints showed that the G76 tended to interact with the P1 stem instead,[1225] which released the strain on the sugar−phosphate backbone around G76, encompassing C75. A close contact between C75 and the scissile phosphate was in turn predicted to be supported by a U-turn motif conformation of the U-1 nucleotide.[1227] In addition, the simulations revealed dynamic behavior of the L3 loop, which may be catalytically important via dynamics of the embedded G25/U20 base pair, in turn controlling the electrostatic environment of the active site.[503,1224] The particular conformation of G25/U20 may therefore affect binding of the Mg^{2+} ion into the active site pocket. While in simulations of the precursor structure the Mg^{2+} ion remained stably bound at the active site near the G25/U20 *t*WH wobble pair, it was not similarly stable in the product form with the G25/U20 *t*WW reverse-wobble G-*syn* base pair.[503] In addition to the conformation of the L3 loop, more distal interactions may also influence the active site architecture over a long distance. For example, Hammes-Schiffer and co-workers suggested that conformational dynamics of a base triple located at the top of the P4 stem containing cytosine C41 may also affect the active site.[1228] The conformational dynamics of this base triple is sensitive to the pH through the C41 protonation state. While the protonated form of C41H$^+$ is required to properly fold the base triple in the wild-type ribozyme,[1224] a double mutant replacing the G73═C44 base pair with A73═U44 stabilizes the triple in the presence of a neutral C41.[1228]

Several QM studies have employed small model systems to mimic the active site of the HDV ribozyme. For example, to probe the chemical feasibility of the C75 general acid and C75 general base mechanisms, Gauld and co-workers compared the basicities of a hydrated magnesium ion and a neutral cytosine using a QM method (DFT) within a minimized cleavage site.[1229] While they observed that neither of these potential catalysts was able to directly deprotonate the 2′-OH nucleophile, the basicity of the hydrated magnesium ion was found to be higher. Consequently, they modeled a C75 general acid mechanism onto their truncated active site model, compared it with the uncatalyzed reaction, and concluded that such a mechanism was chemically feasible.[1229] Subsequently, Guo and co-workers aimed to compare the C75 general acid and general base mechanisms using DFT calculations with a slightly expanded minimal active site model and concluded that the C75 general acid mechanism would be energetically favored.[1230] However, as discussed elsewhere,[1231] the relevance of this particular comparison may be compromised by a steric clash of the participating functional groups with those left out through truncation. In addition, the reaction barrier assigned to the C75 general base mechanism in the truncated active site model was significantly higher than the barrier obtained from subsequent hybrid QM/MM calculations of the entire ribozyme,[1231] suggesting that the reaction paths based on the truncated models were not optimal. In contrast, using QM/MM calculations of the complete HDV ribozyme, it was predicted that the neutral form of C75 acting as the general base during cleavage in conjunction with a hydrated magnesium ion acting as the general acid could lead to a reaction barrier in agreement with the experimental kinetics.[1231] Therefore, the structural, simulation, and QM/MM data available at that time, relying on the C75U-mutation inhibited crystal structure, largely suggested that the C75 general base mechanism is both structurally and chemically feasible. The only contradicting computational results supporting a C75 general acid mechanism relied on less sophisticated calculations using truncated active site models. Still, the C75 general acid mechanism was

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

supported by some biochemical data, in particular the observation that the hyperactivation of the protonated G1(O5′) leaving group through thio-substitution muted the effects of C75 mutations and pH changes.[1232]

4.8.3.2.2. New Crystal Structures. The long-lasting discrepancy between biochemical and available structural data was finally resolved by a new crystal structure of *trans*-acting precursor ribozyme by Golden, Bevilacqua, and co-workers.[1221] In contrast to the earlier crystal structure inhibited by a C75U mutation, the new ribozyme precursor was inhibited by chemical 2′-deoxy modification.[1221] The crystal structure retained the *t*WW G25/U20 reverse-wobble pair with G in *syn* conformation from the earlier postcleavage product structure after which it was modeled. The U-1 nucleotide and scissile phosphate, however, were not resolved due to crystallographic disorder. Instead, the authors modeled this part of the active site using the kinked conformation of the scissile phosphate from the hammerhead ribozyme.[1221] Stability of this conformation was subsequently supported by MD simulations showing that protonated C75H$^+$ is able to donate the C75H$^+$(N3−H)···G1(O5′) and C75H$^+$(N4−H)···G1(*pro*-R$_P$) hydrogen bonds required for C75 to act as the general acid. Furthermore, the active site Mg^{2+} ion was positioned to coordinate the modeled U-1(2′-OH), supporting its role as a Lewis acid (see Figure 63B).[1233] The *t*WW G25/U20 reverse-wobble pair was found to create a pocket of negative electrostatic potential that stabilized binding of the catalytic Mg^{2+} ion into the active site and helped favorably shift the pK_a of the catalytically important C75 toward neutrality.[1233] Whereas the G25/U20 reverse-wobble pair helped anchor the catalytic Mg^{2+} ion, another Mg^{2+} occupying a second ion binding site formed by the standard G1/U37 wobble pair appeared rather diffuse.[1234] Overall, the two G/U base pairs helped preorganize the active site architecture to facilitate self-cleavage. While specifically G25/U20 was found to promote an in-line-attack conformation and the C75H$^+$(N3−H)···G1(O5′) interaction, G1/U37 was suggested to act more locally.[1235] The crucial structural role of particularly the G25/U20 base pair was also supported by MD simulations and kinetic measurements of an A25/C20 double mutant, found to retain the overall fold of the wild-type ribozyme including the L3 loop and reverse-wobble pairing of A25/C20. However, kinetic measurements revealed an inverted pH-rate profile, possibly explained by a loss of the catalytic Mg^{2+} ion from the active site pocket.[1236] This active site Mg^{2+} ion thus appears to be catalytically particularly important and may play a role as Lewis or Brønsted acid (or both).[1234,1236]

The C75 general acid mechanism based on the active site conformation modeled into the 2′-deoxy-modified crystal structure[1221] was probed in several subsequent QM/MM studies. Hammes-Schiffer, Bevilacqua, and co-workers showed that indeed a protonated C75H$^+$ could act as the general acid to protonate the leaving G1(O5′) group (see Figure 63C).[1222,1237] These QM/MM calculations started from an activated U-1(2′-O$^-$) precursor and thus did not include its activation by a general base. However, the authors suggested different mechanisms of self-cleavage depending on the nature of the active site ion to enable the observed multichannel reaction.[510] While in the case of a divalent ion (Mg^{2+} or Ca^{2+}) in the active site the reaction was proposed to proceed through a phosphorane transition state with an (uncorrected) free-energy barrier of ∼13 kcal/mol, the reaction with an active site Na^+ monovalent ion was suggested to proceed sequentially through the phosphorane intermediate with an (uncorrected) free-energy barrier of ∼3.5 kcal/mol.[1222,1237] "Uncorrected" refers here to the fact that no thermodynamic corrections were included for residues that react starting from rare (transient) protonation states (see below).[1238−1240]

The QM/MM analysis suggested that the active site Mg^{2+} ion could act as a Lewis acid via direct coordination to the increasing negative charge of G1(*pro*-R$_P$) and the U-1(O2′) developed during cleavage of the scissile phosphate and during deprotonation of the U-1(2′-OH) nucleophile, respectively. Such coordination is consistent with sensitivity of the U-1(2′-OH) pK_a to the presence or absence of the divalent ion in the active site.[1222] The interaction between active site divalent metal ion and the G1(*pro*-R$_P$) nonbridging oxygen was further supported by thio-effect and ion rescue experiments in conjunction with QM/MM calculations.[1241] In turn, significant inverse thio-effect for the G1(*pro*-S$_P$) nonbridging oxygen was observed in the absence of divalent ions, which might be related to an inhibitory U-1(2′O-H)···G1(*pro*-S$_P$) hydrogen bond that is weakened by thio-substitution of the latter.[1242]

Activation of the U-1(2′-OH) nucleophile as part of the C75 general acid mechanism was subsequently studied by a series of QM/MM calculations, which revealed that the nucleophile can be deprotonated by a hydroxide ion coordinated to the active site Mg^{2+} ion.[1240] These calculations showed that the activation of the U-1(2′-OH) nucleophile and the nucleophilic attack are two steps separated by a shallow and likely kinetically insignificant minimum corresponding to the activated precursor intermediate.[1240] The full path starting from activation of the U-1(2′-OH) nucleophile exhibited an (uncorrected) reaction barrier ranging from ∼6 to 13 kcal/mol, depending on the specific positioning and coordination of the Mg^{2+} ion.[1240] However, the C75 general acid mechanism relies on the presence of the rare N3-protonated form of C75 in conjunction with a hydrated $Mg^{2+}OH^-$ bound at the active site. Formation of these rare protonation species is not included in the QM/MM-calculated pathway. Consequently, the calculated reaction barrier needs to be a posteriori corrected for the penalty of adopting the two rare protonation states (C75H$^+$ and $Mg^{2+}OH^-$).[1239,1240,1243] After this correction, the barrier ends up in the range of 15−21 kcal/mol,[1222,1240] corresponding well with the experimentally observed kinetic rate constant of HDV ribozyme self-cleavage.[1241,1244] By comparison, as pointed out by York and collaborators,[1243] the reaction barrier calculated for the C75 general acid mechanism involving an active site monovalent ion[1222] is unexpectedly low even when all thermodynamics corrections are taken into account. Such a low barrier cannot even be explained by the absence of a pK_a shift of the U-1(2′-OH) nucleophile in the absence of the active site divalent metal ion,[1222] suggesting that the proposed pathway is not fully correct.[1243] Alternatively, the unexpectedly low reaction barrier may be due to the omission of the U-1(2′-OH) nucleophile activation. While in case of the Mg^{2+} ion path the activation of the U-1(2′-OH) nucleophile is not expected to contribute significantly to the overall reaction barrier,[1240] in the case of the reaction involving a monovalent ion at the active site this assumption may not necessarily be valid so that the description of the entire reaction path including U-1(2′-OH) deprotonation would instead be required.

On the basis of all available experimental and computational data, as well as additional extensive MD simulations, York and collaborators suggested an alternative complex (multistep) mechanism involving C75 in the role of general acid.[1223] This

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

mechanism assumes participation of two Mg^{2+} ions in the two experimentally observed binding sites. The authors suggested that, in an initial step, the ribozyme binds Mg^{2+} ion in the active site close to the G25/U20 wobble pair, which in turn would induce a flip of this base pair into the *t*WW reverse-wobble conformation. This flip then helps shift the pK_a of the active site C75 to protonate it. Subsequently, the presence of the charged $C75H^+$ species leads to unbinding of the active site Mg^{2+} ion, back-flipping of the *t*WW G25/U20 reverse-wobble pair into the *t*WH/*c*WW standard wobble-like base pair, and binding of another Mg^{2+} ion into the second ion-binding pocket close to the G1/U37 wobble base pair. Next, the authors suggested for this ion to deprotonate the U-1(2′-OH) nucleophile, which then would flip into the in-line-attack conformation and attack the scissile phosphate, while the protonated $C75H^+$ would act as the general acid protonating the leaving G1(O5′) group.[1223] This mechanism assumes that several structural changes accompany self-cleavage. Some significant conformational changes were indeed suggested by FRET[5,1220] and footprinting experiments.[1245] In addition, the proposed complexity of conformational changes is qualitatively consistent with a preceding study by Walter and collaborators based on extended MD simulations that suggested that the different active site architectures observed in precursor crystal structures based on different strategies of inhibition may represent the intermediates on a rugged folding free-energy landscape.[1019]

The mechanism suggested by York and collaborators also has some problematic parts that require further validation. For example, the mechanism assumes that protonation of C75 requires prior binding of the active site Mg^{2+} accompanied by flipping of the G25/U20 base pair, immediately followed by Mg^{2+} dissociation and back-flipping of G25/U20.[1223] Yet, the transient binding of an Mg^{2+} ion does not affect the thermodynamics of C75 protonation in the final (Mg^{2+}-unbound) state. The only explanation for the requirement of transient Mg^{2+} ion binding would therefore be that the protonation of C75 in the Mg^{2+}-unbound state is impaired kinetically. However, this is typically not the case for fast, reversible proton transfer reactions. Furthermore, in contrast to the C75 general acid pathway studied by Hammes-Schiffer and collaborators,[1222,1237] and Mlynsky et al.,[1240] in which the U-1(2′-OH) nucleophile is activated by an active site Mg^{2+} ion bound close to the G25/U20 reverse-wobble pair, the mechanism suggested by York and collaborators[1223] assumes activation by an Mg^{2+} ion bound to the G1/U37 standard wobble pair. However, when the U-1(2′-OH) interacts with the solvation sphere of this Mg^{2+} ion, it is not in the required in-line attack conformation. Although MD simulations support that an in-line attack conformation may be formed spontaneously after proton transfer, that is, after formation of the U-1($2'-O^-$) nucleophile, such a conformational change would occur on a much longer time scale than the proton transfer. Thus, it is likely that the nucleophile would become reprotonated in route to the in-line attack conformation.

Finally, the proposed mechanism aims to rationalize the different base-pairing patterns of the G25/U20 base pair in the crystal structures of the precursor and product forms. However, the original 1CX0 and 1VC7 structures used as starting geometries in the early computational studies have since been removed from the PDB database as obsolete (i.e., inaccurate) in 2014, and replaced by newly refined structures 4PR6 and 4PRF.[1246] Importantly, the *t*WH G25/U20 base pair (with G25 in *anti* conformation) of the precursor 1VC7 crystal structure was corrected to the *t*WW reverse-wobble base pair (with G25 in *syn* conformation) also found in the product and newer precursor structures. Retrospectively, the initially available X-ray structures were thus not sufficiently accurate for the presently available computational methods, and it seems likely that the G25/U20 base pair consistently forms a *t*WW reverse-wobble base pair with G25 in the *syn* conformation in all functional states of the HDV ribozyme. G25/U20 flipping and back-flipping thus appears as a questionable structural dynamics intermezzo that complicates the reaction.

In summary, the HDV ribozyme is perhaps the perfect example for highlighting the scope and limitations of computational approaches to understanding RNA enzymatic mechanisms. Numerous theoretical studies of this ribozyme have led to a multitude of distinct and partially inconsistent mechanistic proposals. Early studies based on a high-resolution C75U mutant crystal structure and modeling of the complete ribozyme supported what now appears to be an incorrect or minor mechanism, which underscores the dependence of computations on the availability of reliable structural data. Retrospectively, the "correct" prediction of the C75 general acid mechanism based on simple truncated structural models may be considered a "false-positive" result, where the correct mechanism was supported incidentally due to severe limitations of the calculations. Even using the later more refined structural information, different computations have arrived at distinct plausible reaction mechanisms and pathways, which cannot be further parsed within the error margins of computation. A resolution for this dilemma is offered by the latest two MD studies that support the idea of a rugged native conformational space of a ribozyme's active site, leading to multichannel and multistep reactivity that is not exhaustively represented by a single geometry,[1019,1223] as first formulated by Walter and collaborators.[1019] This would explain the insufficiency of the available structural data to pinpoint the exact atomistic mechanism of the reaction and the ambiguity of any subsequent theoretical treatment. None of the available computational methods has the capacity to sample such a rugged conformational space exhaustively in unbiased fashion.

4.8.3.3. Hairpin Ribozyme. Another small endonucleolytic RNA is the hairpin ribozyme, which seems largely confined to plant viral satellites.[1247,1248] Like the hammerhead[515,1213] and HDV ribozymes,[1219] the hairpin ribozyme does not absolutely require divalent metal ions for its cleavage activity.[515,516,1174,1249] Crystal structures of a precleavage state inhibited by a 2′-methoxy modification, a product state, and a transition state analogue have consistently revealed the catalytically important nucleobases G8 and A38 as in close contact with the A-1(2′-OH) nucleophile and the G+1(O5′) leaving group, respectively (Figure 64).[1250−1252]

Hairpin ribozyme has been the subject of many computations, yielding a wide range of mechanistic proposals. On the basis of initial classical MD simulations of the vanadate transition state analog, Park and Lee proposed that the active site nucleobases G8, A9, and A38 electrostatically stabilize the negative charge developed on the scissile phosphate upon cleavage via a network of hydrogen bonding interactions.[1254] In addition, they suggested that active site water molecules may be involved in such electrostatic transition state stabilization.[1254] Such long-residency active site water molecules were observed also in MD simulations of the precleavage form of the hairpin ribozyme.[1255] These water molecules are trapped in a metal-

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

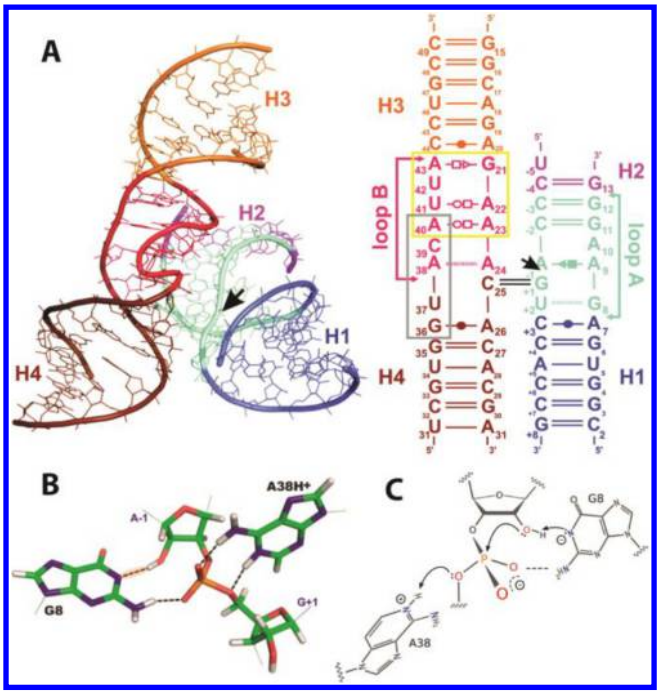


Figure 64. (A) 2D and 3D structure of the hairpin ribozyme. Reprinted with permission from ref 99. Copyright 2010 American Chemical Society. (B) Detail of the catalytically potent active site arrangement involving deprotonated guanine $G8^-$ and protonated adenine $A38H^+$. Adapted with permission from ref 147. Copyright 2015 American Chemical Society. (C) Suggested dianionic mechanism of its self-cleavage. Adapted with permission from ref 1253. Copyright 2014 American Chemical Society.

ion-free active site cavity with very deep negative electrostatic potential so that they are only occasionally exchanged with bulk solvent. In fact, they were found to be lined up like a proton-wire to connect the A-1(2′-OH) nucleophile with bulk solution. Accordingly, it was suggested that trapped water molecules may participate in specific base catalysis by mediating deprotonation of the A-1(2′-OH) by an exogenous base via a proton-wire.[1255] Conversely, these observations support the notion that the catalytic pocket of the hairpin ribozyme may be uniquely structured to create a very deep electrostatic potential minimum while also sterically disfavoring binding of cations that would compensate this negative potential. In addition, the structured water molecules were found to support long-range coupled molecular motions that communicate dynamic rearrangements throughout the catalytic core in response to site-specific chemical modifications.[1255] Because the simulations were initiated using an earlier crystal structure that did not resolve water molecules, but subsequent structures confirmed the computationally predicted hydration network in the catalytic pocket,[1252] the study demonstrated the capacity of MD simulations to spontaneously hydrate an initially empty catalytic pocket and correctly predict structured hydration on an accessible simulation time scale.[1255]

An alternative reaction mechanism was proposed on the basis of conformational behavior observed in MD simulations of the hairpin ribozyme precleavage form for the catalytically most critical A38.[1256] In these simulations, the neutral A38 moved into close proximity of the A-1(2′-OH) nucleophile. In contrast, simulations with an N1-protonated $A38H^+$ populated a hydrogen bond between the $A38H^+$(N1) and the G+1(O5′) leaving group. Accordingly, the authors suggested a mechanism in which A38 acts first as the general base activating the A-1(2′-OH) nucleophile and then shuttles this proton toward the G+1(O5′) group, where it acts as the general acid to protonate the leaving group.[1256] However, subsequent analyses evaluated this mechanism as less likely because the contact between A38 and A-1(2′-OH) requires an unfavorable in-like attack conformation.[99]

York and co-workers used short QM/MM-based free-energy calculations with a semiempirical Hamiltonian to study the various possible catalytic roles of guanine G8 and adenine A38 in the active site.[1148,1257] They probed a general in-line monoanionic mechanism, in which the A-1(2′-OH) nucleophile is activated by transfer of its proton to one of the nonbridging oxygens of the scissile phosphate and the proton is subsequently shuttled to the G+1(O5′) leaving group.[1148,1257] They compared this general in-line monoanionic path with mechanisms directly involving a deprotonated $G8^-$ as the general base[1257] or a protonated $A38H^+$ acting as the general acid.[1148] These calculations indicated that the hairpin ribozyme may achieve effective catalysis even without direct involvement of a deprotonated $G8^-$ and/or protonated $A38H^+$ as general acid–base catalysts.[1148,1257] In contrast, Mlynsky et al. demonstrated by QM/MM potential energy scans based on a robust hybrid DFT functional that both the general in-line monoanionic mechanism as well as a mechanism involving $G8^-$ and $A38H^+$ as general base and acid, respectively, are in good agreement with the experimentally observed cleavage rate constant.[1239] In fact, the latter general acid–base mechanism was found to be slightly more favorable.[1239]

Importantly, a direct comparison of potential-energy scans and free-energy profiles of hairpin ribozyme self-cleavage as calculated by QM/MM schemes utilizing semiempirical methods, DFT functional, and benchmark high-level ab initio calculations revealed that the accuracy of the semiempirical methods may be comparably limited, even though they were specifically parameterized for the ribozyme reaction.[1258] This finding illustrates the trade-off between the accuracy of the potential-energy surface description and the need for sampling as much as possible of the free-energy surface. Thus, Mlynsky et al. concluded that an intrinsically accurate description of the potential energy surface (i.e., the quality of the QM method utilized for its description) in most cases is more important than the use of free-energy simulations. The advantages due to increased sampling in QM/MM simulations with simplified QM descriptions are thus on many occasions not sufficient to justify their use.[1258]

Although QM/MM calculations by Mlynsky et al. supported chemical feasibility of the $G8^-/A38H^+$ general acid–base mechanism (after including the intrinsic thermodynamic penalty associated with the rare protonation states of the two active-site nucleotides),[1239] the remaining question for such a chemical pathway was the mechanism of G8 deprotonation. MD simulations suggested that the canonical form of G8 and a protonated $A38H^+$ are stable in the ground-state structure observed by X-ray crystallography.[99] In contrast, a deprotonated $G8^-$ is quickly expelled from the active site in all simulations.[99] Mlynsky et al.[99] and York's simulations[1253] revealed that adoption of the in-line attack conformation required for cleavage is correlated with the formation of an A-1(2′-OH)···G+1(*pro*-R_P) hydrogen bond. In this conformation, a canonical G8 donates a hydrogen bond to A-1(2′-OH), which in turn donates a hydrogen bond to the nonbridging oxygen of the scissile phosphate;[99,1253] G8 thus cannot easily be

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX–XXX

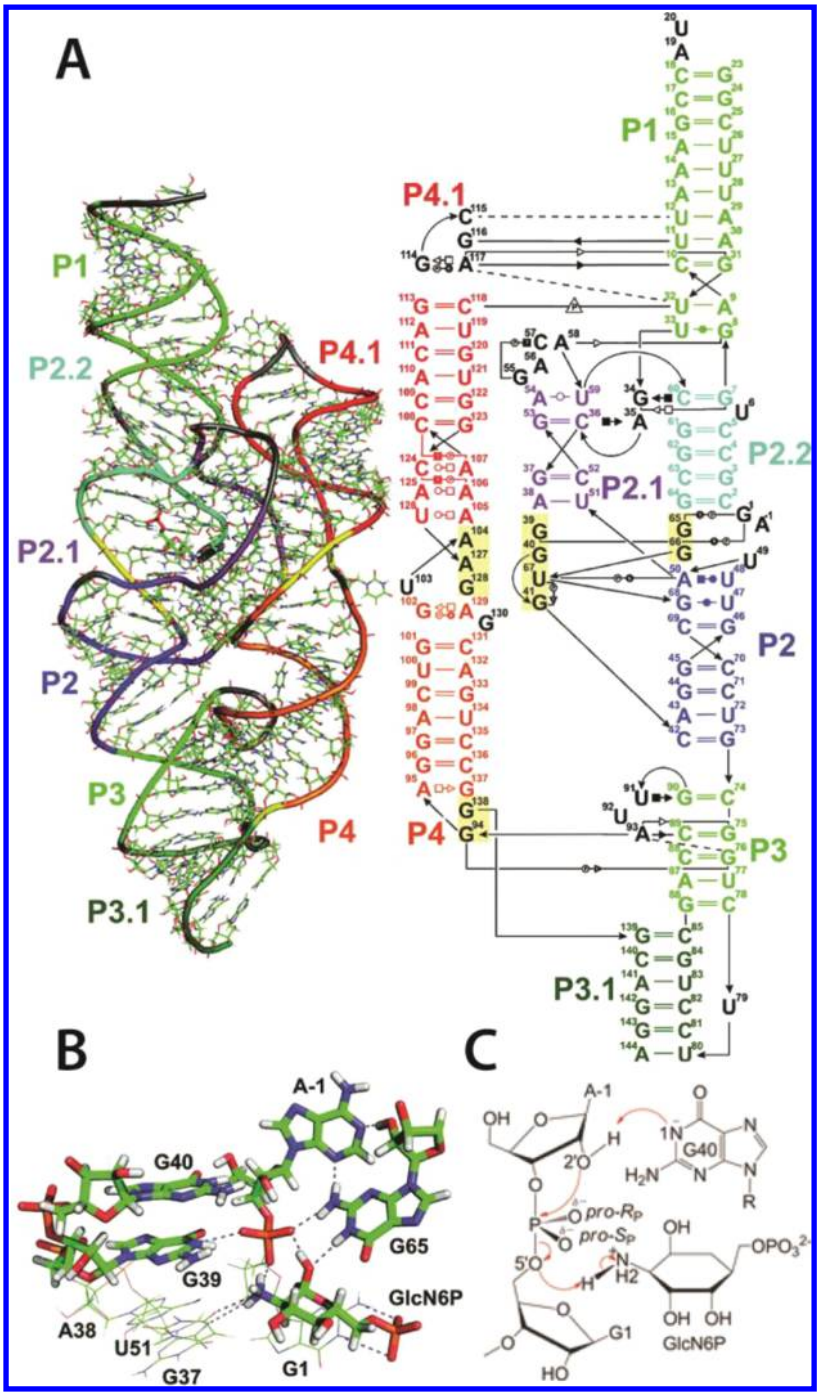


Figure 65. (A) 2D and 3D structure of glmS ribozyme, (B) detail of the active site involving glucosamine-6-phosphate as a cofactor, and (C) suggested mechanism of the glmS ribozyme self-cleavage. Panels (A) and (B) were adapted with permission from ref 1266. Copyright 2010 American Chemical Society. Panel (C) is reprinted with permission from ref 1267. Copyright 2015 American Chemical Society.

deprotonated by an exogenous base. To facilitate the deprotonation of G8, the nucleotide would have to temporarily leave its direct interaction with A-1(2′-OH). However, given that a deprotonated G8$^-$ was found to be spontaneously and rapidly expelled from the active site,[99] it was unlikely that G8$^-$ could return back to the active site after deprotonation. Subsequent MD simulations using the χ_{OL3}[94] RNA AMBER force field in combination with the $\varepsilon\zeta_{OL1}$ reparameterization of the ε/ζ torsions[146] (originally suggested for DNA, see section 3.1.2.3) for the first time revealed a significant population of the in-line attack conformation required for cleavage with G8 in its canonical neutral form.[147] However, the authors still found repulsion of the deprotonated G8$^-$ general base from the active site. A comparison of MM energy scans with the equivalent QM/MM benchmarks indicated that the repulsion of G8$^-$ in the MD simulations may be caused by an imbalance of the relative stabilities of the A-1(2′-OH)···G+1(*pro*-R_P) and A-1(2′-OH)···G8$^-$(N1) hydrogen bonds in the force-field description; the latter would be both required for general base catalysis by G8$^-$ and expected to help anchor G8$^-$ in the active site.[147] A direct verification of this hypothesis by MD simulation was, however, not possible due to a lack of a suitable

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

force-field correction. In addition, while the addition of the $\varepsilon\zeta_{OL1}$ backbone correction rendered the conformational behavior of the catalytic center of the hairpin ribozyme more consistent with existing biochemical data, subsequent analyses of other systems showed that the $\varepsilon\zeta_{OL1}$ force-field modification sometimes deteriorates the description of the RNA structural dynamics.[132] Thus, a full computational resolution of the functional role of G8$^-$ and the mechanism of its potential deprotonation remains elusive.

Increased population of the reactive in-line attack conformation along with the strengthening of the hydrogen bonding of scissile phosphate with A38H+ and G8 was revealed by T-REMD simulations (see section 3.2.4) at high hydrostatic pressure.[1259] This observation explains increased cleavage rate of the chemical reaction step observed by stop-flow FRET experiments at high hydrostatic pressure, which supported the hypothesis that ribozymes could have evolved in extreme conditions including deep sea during the RNA world era.[1259]

The above-discussed inconsistency between proposed dominant protonation states of active site residues based on structural data (and MD simulations) and catalytic mechanisms supported by biochemical analyses is a common problem for ribozymes.[1008] This indeed may reflect involvement of rare protonation states in the catalysis. Typically, when the simulations start assuming the rare protonation states (such as the deprotonated G8$^-$ of hairpin ribozyme) based on starting structures derived from the available structural data, the residue in the rare protonation state tends to quite vigorously depart from the catalytic conformation.[99,147] In addition, it is fair to assume that the residue is not (de)protonated to its rare protonation state while being in some catalytically active-like ground-state geometry, because in this geometry it is involved in a specific hydrogen bond and thus not available for an exchange with solvent or other groups (irrespective of the protonation state). Thus, it is likely that it adopts the rare protonation state when being in a different conformation and only then moves toward the catalytic active geometry. However, such a movement has so far, to our knowledge, never been reported in MD simulations. Note that force-field MD simulations themselves do not directly involve transient (de)protonations and can only explore the conformational dynamics for one specific ionization state at a time. The inability of MD simulations to adopt the catalytically competent geometry with the rare protonation states may on one side reflect some general limitations of the force field to accurately describe the catalytic center in its catalytically active conformation and chemical state. However, it may also be another indication that the conformational space of the active site (and its vicinity) of the small ribozymes is very rugged (see section 4.8.3.2.2) and is not sufficiently well characterized by single ground-state structures.[1019] Assuming that the crystal structures reflect the inactive ground state with major protonation states, a subsequent MD attempted with a rare protonation state may be unable to capture some conformational changes that would be needed to obtain a catalytically competent geometry with the rare protonation state. In any case, an exact description of the mechanisms of the formation of the rare protonation states remains one of the open issues for further computational studies of small ribozymes.[1008] The catalysis may proceed by some rarely populated but very reactive conformation, and the catalysis may require not only the formation of the rare protonation states, but also some subsequent rearrangement as compared to the ground state presumably seen in the X-ray structures. The inability of MD simulations to accommodate the rare protonation states at the active site[99,147] thus may indicate that the catalytic process is more complex than usually assumed on the basis of the structural data. This should stimulate further experimental testing.

4.8.3.4. glmS Ribozyme. The *glmS* ribozyme is the only naturally occurring ribozyme for which self-cleavage depends entirely on the binding of a specific exogenous cofactor, glucosamine-6-phosphate (GlcN6P) (Figure 65). The ribozyme is found in the 5′-UTR region of numerous Gram-positive bacteria, and its GlcN6P-induced cleavage results in down-regulation of the downstream *glmS* gene responsible for GlcN6P synthesis. Thus, the *glmS* ribozyme also acts as a riboswitch. The first crystal structures were resolved independently by the groups of Ferre-D'Amare and Strobel in 2006.[1260,1261] These and subsequent structures revealed that the *glmS* ribozyme does not rearrange upon cofactor binding or self-cleavage, and that the active site adopts an almost identical architecture in a variety of functional states and when including mutations that inhibit cleavage.[1260−1265] On the basis of this structural information, two mechanisms of self-cleavage were proposed. The first mechanism assumes that the conserved active site guanine (G40 and G33 according to *T. tengcongensis* and *B. anthracis* numbering, respectively) is deprotonated prior to the reaction and consequently involved in activation of the A-1(2′-OH) nucleophile, acting as the general base; the amino group of GlcN6P cofactor is protonated before the cleavage reaction to in turn protonate the O5′ leaving group, thus acting as the general acid.[1260] In contrast, the second mechanistic proposal suggests that the A-1(2′-OH) nucleophile instead is activated by two tightly bound water molecules transferring a proton to the GlcN6P amino group via proton hopping; this protonated amino group then in turn protonates the O5′ leaving group so that GlcN6P acts as both indirect general base and direct general acid.[1261]

The first theoretical studies on the *glmS* ribozyme reported MD simulations aiming to define the most likely protonation states of catalytically important moieties, that is, GlcN6P and the conserved active site guanine.[1266,1268] Xin and Hamelberg used MD simulations and free-energy calculations to predict that the pK_a of the GlcN6P amino group was significantly decreased from an unperturbed value of 8.2 to a pK_a near 6 by its binding to the active site of the wild-type *glmS* ribozyme, while a slight increase of this pK_a was predicted upon binding to a G40A mutant.[1268] Consistent with this notion, MD simulations by Banas et al. suggested the protonated ammonium of GlcN6P to be the form most consistent with the crystal structure.[1266] These MD simulations also revealed weakened binding of GlcN6P upon protonation of its phosphate group,[1266] which rationalizes the experimentally observed increase (weakening) of its binding affinity K_m with decreasing pH.[1260] In addition, the authors suggested that the canonical form of G40 is dominantly populated under these conditions, and plays a structural role by stabilizing the in-line-attack conformation.[1266] On the basis of these observations, a third possible reaction mechanism was proposed, in which the 2′-OH group is activated by the G1(*pro*-R_P) nonbridging oxygen, while the ammonium group of GlcN6P protonates the O5′ leaving group.[1266] Hammes-Schiffer, Bevilacqua, and co-workers used QM/MM free-energy simulations to probe the latter mechanism, as well as the original mechanism (Figure

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

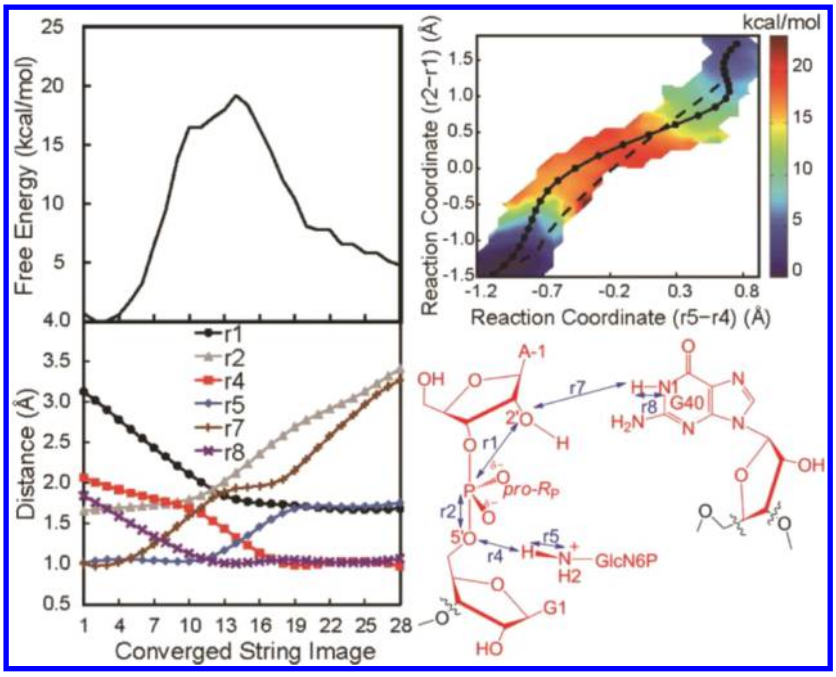


Figure 66. 1D QM/MM free-energy profile (PMF, top left) and 2D free-energy landscape (top right) obtained by combined umbrella sampling and finite temperature string method (see section 3.2.5) by Hammes-Schiffer and co-workers for self-cleavage of *glmS* ribozyme.[1267] Catalytically important distances together with their evolution along the reaction coordinate are depicted in the bottom right and left panels, respectively. Adapted with permission from ref 1267. Copyright 2015 American Chemical Society.

65C) involving the deprotonated $G40^-$ and the ammonium form of $GlcN6PH^+$ acting as the general base and general acid, respectively.[1267] They concluded that the former mechanism should be considered less likely due to a significantly higher free-energy barrier.[1267] In contrast, the latter mechanism involving $G40^-$ and $GlcN6PH^+$ as general base and general acid, respectively, was considered chemically feasible both by the Hammes-Schiffer, Bevilacqua, and co-workers team[1267] as well as by Banas and co-workers based on QM/MM scanning of the reaction path (Figure 66).[1238] Both studies reported the proton transfer from $GlcN6PH^+$ to the O5′ leaving group associated with exocyclic cleavage as the rate-determining step of the self-cleavage reaction.[1238,1267] This observation suggests that *glmS* ribozyme cleavage critically depends on the GlcN6P cofactor acting as the general acid.

On the basis of pK_a calculations, Bevilacqua, Hammes-Schiffer, and co-workers suggested that an increased pK_a of the catalytically important guanine is required for optimal catalysis because a guanine with a pK_a shifted further from neutrality rather than toward neutrality is expected to activate the 2′-OH nucleophile more efficiently.[1267] This notion warranted further investigation due to its broader implications, because use of an active site guanine in catalysis arises as a general catalytic strategy of multiple small ribozymes such as the hammerhead, hairpin, VS, and twister ribozymes. Further classical MD as well as QM/MM free-energy simulations by Hammes-Schiffer, Bevilacqua, and co-workers revealed that formation of the intramolecular hydrogen bond between the A-1(2′-OH) nucleophile and the G1(*pro*-R_P) nonbridging oxygen may compete with activation of the nucleophile and nucleophilic attack.[1141,1267] The authors suggested that, in addition to the general acid role of GlcN6P, binding of the cofactor may compete with this intramolecular hydrogen bond and thus liberate the A-1(2′-OH) nucleophile for activation and nucleophile attack.[1141] Interestingly, both roles may be assumed by an active site Mg^{2+} ion in the ligand-free apo-ribozyme under high ionic concentrations.[1141] Conversely, the presence of an active site Mg^{2+} ion in the ligand-bound holo-ribozyme may suppress catalysis due to competition with cofactor binding and disruption of H-bonding interactions within the active site.[1269]

For the sake of completeness, Xue and co-workers performed QM calculations to compare the $G40^-/GlcN6PH^+$ general base/acid mechanism with the second mechanism suggested by crystallographers, involving proton transfer via two tightly bound water molecules.[1270] The authors favored the latter mechanism as chemically more feasible.[1270] However, these calculations involved a rather small truncated model of the active site; in addition, the chosen reactant, transition, and product states were incompatible with the active site arrangement observed in the crystal structures,[1260−1265] compromising the applicability of these calculations to the full-length *glmS* ribozyme (see also section 4.8.3.2.1 for a discussion of the relevance of the truncated models).

4.8.3.5. Twister Ribozyme. A new class of small self-cleaving ribozymes denoted as twister and related ribozymes was discovered by bioinformatics.[1180] Several crystallographic studies have depicted the structure of the twister ribozyme[1271−1273] providing, however, ambiguous information about groups possibly acting as general base and general acid. Bertran and co-workers demonstrated by QM/MM MD calculations that the catalytical effect of the twister ribozyme cannot be assigned to a substrate assisted mechanism, in which the 2′-OH nucleophile is activated by one of the nonbridging oxygens of the scissile phosphate that in turn shuttles this proton toward the leaving O5′ group. Thus, some general acid and general base present at the active site should be involved in the self-cleavage mechanism.[1274] Notably, in some of the available crystal structures, the U-1(2′-OH) nucleophile was not positioned in an in-line attack conformation relative to the

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

scissile phosphate.[1271,1272] The absence of this conformation was explained by crystal packing artifacts, leading to extrusion of the U-1 nucleotide from the active site and rearrangements of the sugar−phosphate backbone around the scissile phosphate. Gaines and York showed by MD simulations that U-1 indeed can sample several conformational states, with one of them involving the U-1 stacked on G33 and the U-1(2′-OH) positioned in an in-line attack conformation.[1275] Such a conformation is additionally compatible with a model suggested by mutagenesis, in which the N1-deprotonated $G33^{-}$ and the N3-protonated $A1H^{+}$ act as general base and general acid, respectively.[1275,1276] Notably, suggestion that the active site adenine A1 acts as the general acid via its N3 instead of N1 nitrogen presents a novel mechanistic strategy for RNA-catalyzed self-cleavage.[1275,1276] An alternative mechanism was suggested by Ucisik et al. using MD simulations to probe structural dynamics of the env22 twister ribozyme.[1277] They suggested that an active site Mg^{2+} ion may play important roles during catalysis, in particular through a new ion-binding site in proximity of A1(O5′) and A1(N3) consistent with the idea that this ion acts as the general acid instead of the N3-protonated A1.[1277] However, mechanistic data suggest that, while metal ions play an important role in folding, they do not directly participate in catalysis.[1180] In summary, an unambiguous theoretical description of this ribozyme is not yet available.

4.8.4. Ribosomal Peptidyl-Transfer Center. Determination of the X-ray structures of ribosomal subunits with sufficient resolution revealed that the peptidyl-transfer center, where new peptide bonds are synthesized during translation of a mRNA into protein, is solely composed by RNA.[15] This observation supported the now widely accepted view that the ribosome acts as a ribozyme; that is, the translation process is catalyzed only by RNA without participation of ribosomal proteins. Later higher-resolution crystal structures revealed the presence of the N-terminus of protein L27 near the peptidyl-transfer center.[1278] However, the ribosome is able to catalyze transpeptidation even when lacking the L27 protein. As suggested by Trobro and Åqvist, L27 does not directly participate in peptidyl-transfer but rather assists the positioning of the A-site tRNA by its interaction with the A76 phosphate.[1279] Thus, the ribosome must indeed be considered a ribozyme.

The mechanism of peptidyl transfer attracted many theoretical studies.[1020] The structural together with many biochemical data provided evidence that the unprotonated form of the A-site α-amino group attacks the ester carbon of the peptidyl chain attached to the P-site peptidyl-tRNA, resulting in stereospecific formation of the tetrahedral reaction intermediate with S-configuration. A detailed mechanism of the peptidyl-transfer reaction was reported independently by Warshel and co-workers[1280] as well as Trobro and Åqvist[1281,1282] using MD simulations in combination with empirical valence bond methods to probe several potential mechanisms. In particular, the authors predicted that the reaction does not rely on direct participation of any rRNA group that would act as a general acid−base catalyst. Instead, the proton of the attacking α-amino group is shuttled via the P-site A76(2′-OH) group toward A76(O3′).[1280−1282] A crucial role was originally proposed for the A76(2′-OH) moiety by Weinger et al., who demonstrated that removal of this functional group from the P-site tRNA significantly reduces the rate of peptidyl transfer.[1283] On the basis of this observation, the authors suggested that the peptidyl-transfer reaction is subject to substrate-assisted catalysis.[1283] However, as noted by Warshel and co-workers, such a mechanism cannot solely explain ribosomal catalysis because the same A76(2′-OH) group would be expected to catalyze the reaction even without the ribosome.[1280] Instead, Warshel and co-workers[1280] and Trobro and Åqvist[1281,1282] jointly interpreted the main catalytic effect of the ribosome as a reduction of the exceptionally large activation entropy associated with the uncatalyzed peptidyl-transfer reaction in agreement with previously reported experimental data.[1284] This entropy reduction was found to be associated primarily with a reduction in the necessary solvent reorganization rather than with conformational alignment of the substrates.[1280−1282] Finally, Ehrenberg and co-workers explained the different pH-sensitivities of peptidyl transfer involving various A-site aminoacyl-tRNAs as correlated with the distinct pK_a shifts of their α-amino groups under conditions where reaction chemistry is rate-limiting.[1285] Note that the peptidyl-transferase center has been studied also by pure QM approaches, for example, in ref 1286.

4.8.5. Other Computations Related to the Ribosome Function. In addition to the peptidyl-transfer mechanism, computational studies were also directed at the process of translation termination. The first structural insight into the peptidyl-transfer center with release factor (RF) bound at the A-site was based on low-resolution crystal structures of the ribosome bound to RF1 or RF2,[1287] as well as corresponding cryo-EM maps.[1288,1289] These data revealed that RF directly interacts with both the peptidyl-transfer center and the decoding center. It adopts the conformation resolved in solution using SAXS[1290] rather than the conformation seen in the crystal structure of isolated RF.[1291] However, the resolution was not sufficient to provide atomistic details of the peptidyl-transfer center with RF bound at the A-site. Trobro and Åqvist thus used peptide docking constrained by the available experimental data to model the peptidyl-transfer center with RF and probed the mechanism of peptidyl-tRNA ester bond hydrolysis.[1292] They found that the methylated glutamine of the conserved GGQ loop of the RF positions the catalytic water for attack on the peptidyl-tRNA ester carbon, while the proton of this water molecule is shuttled toward A76(O3′) via A76(2′-OH).[1292] Thus, the peptidyl-tRNA ester hydrolysis facilitated by RF during translation termination and peptide release shares its mechanism with the peptidyl-transfer reaction.[1292] Subsequently, Trobro and Åqvist[1293] further confirmed these observations by simulations initiated from the crystal structures that later resolved the interaction interface of RF with the peptidyl-transfer center.[14,1294,1295] The crucial role of the A76(2′-OH) group acting as a proton shuttle was subsequently supported by experimental and simulation approaches showing that certain amino acid substitutions of the conserved RF GGQ loop rescue the inhibition caused by a A76 deoxy-modification of the peptidyl-tRNA.[1296]

4.8.6. Artificial Ribozymes. Naturally occurring small ribozymes catalyze only the cleavage and ligation of their own sugar−phosphate backbone. However, the discovery of RNA catalysis strongly supported the "RNA world" hypothesis, which posits that RNA enzymes catalyzed a much wider set of chemical processes in a preprotein world. This idea motivated the evolution of artificial ribozymes in vitro as proofs-of-concept for this RNA world chemistry. Some of these artificial ribozymes were studied by computational methods.

York and co-workers studied the mechanism of RNA backbone ligation catalyzed by the L1 ligase.[1297] The crystal

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

structure of the L1 ligase revealed two crystallographically independent conformations within an asymmetric unit cell.[1298] The authors used MD simulations and enhanced sampling techniques to probe the conformational pathway for interconverting the two observed states.[1297] They found the interconversion to include two steps: a hinge-like motion of the C-stem, followed by allosteric activation of the active site. Thus, the crystallographically observed conformations may correspond to active and inactive forms of the ligase.[1297]

While RNA ligases catalyze ligation of two RNA oligonucleotides, RNA polymerases catalyze the addition of multiple nucleoside triphosphates to one end of an RNA in sequence-specific manner; that is, polymerase ribozymes represent hypothetical catalysts that enable RNA self-replication to fuel the proposed RNA world. Sgrignani and Magistrato focused on characterizing the active site of the in vitro evolved class I RNA polymerase ribozyme, in particular the positions and roles of active site Mg^{2+} ions.[483] As the corresponding crystal structure did not reveal such active site Mg^{2+} ions,[1299] they aimed to model their positions using MD simulations and QM/MM calculations. They proposed two plausible binding sites where the Mg^{2+} ions may assist in preorganizing the active site.[483]

In addition to ligase and polymerase activity, artificial ribozymes were also developed to catalyze a broad range of other chemical reactions. Imhof, Smith, and co-workers studied the mechanism of a Diels–Alderase ribozyme catalyzing a [4+2] cycloaddition between an anthracene diene and a maleimide dienophile using classical MD simulations.[1300,1301] They reported that the catalytic pocket of this ribozyme is highly dynamic, fluctuating between catalytically inactive closed and active open conformations. The open state was found to be stabilized by specific binding of an Mg^{2+} ion in the active site pocket as well as binding of the two substrates. In addition, they proposed that the enantioselectivity of the Diels–Alderase ribozyme may be controlled by the entrance of the substrates into the active site.[1300,1301] However, because the dynamic fluctuations were quite large, it cannot easily be ruled out that force field or starting structure issues affected the results, as the studies lack a broader discussion of the potential limitations (cf., sections 3.4.1, 4.7.1, 4.8.3.2, and 4.8.7).

4.8.7. Related Protein-Catalyzed RNA O2′-Transphosphorylation Reactions. Although the RNA O2′-transphosphorylation catalyzed by protein enzymes is not formally considered RNA catalysis, the mechanisms of backbone cleavage catalyzed by ribozymes are often compared to their counterparts among protein enzymes. In particular, RNase A has been a favored model system for extensive experimental as well as theoretical studies. RNase A catalyzes the cleavage of an RNA oligonucleotide in two chemical steps. The first step involves nucleophilic attack of the 2′-OH group on the adjacent scissile phosphate, resulting in generation of the 2′,3′-cyclic phosphate and 5′-OH termini observed also in small ribozyme self-cleavage. This step is followed by hydrolysis of the cyclic phosphate. In both steps, the active site amino acids His12, His119, and Lys41 play crucial roles.

The first theoretical study performed by Karplus and co-workers aimed to reveal the mechanism and the reaction profile of the cleavage reaction. The authors used QM/MM calculations with semiempirical Hamiltonian and gas-phase DFT calculations to probe the reaction profile of cleavage.[1302] They proposed that the rate-determining step corresponds to the hydrolysis of the cyclic phosphate, in particular the activation of the attacking water molecule by His119. In addition, they showed that the increasing negative charge of the phosphorane transition state of the hydrolysis step appears to be stabilized by the positively charged His12 and Lys41 residues, the latter of which was suggested to also act as the general acid in hydrolysis.[1302] This general acid role of Lys41 to protonate the departing O2′ during the second chemical step was also supported by subsequent MD simulations performed by York and co-workers.[1303] MD simulations of a transition state mimic and QM calculations in conjugation with kinetic isotope effect measurements suggested that the cleavage reaction proceeds through a late transition state associated with exocyclic cleavage; this state was proposed to be electrostatically stabilized by H-bonding between His12, Lys41, and the nonbridging oxygens of the scissile phosphate,[1304] and by protonation of the O5′ leaving group from His119.[1305] RNase A is thus a key example of a protein enzyme that uses its amino-acid side chains to cleave an RNA sugar–phosphate backbone.

In contrast, RNase H represents an example of a protein enzyme that alternatively uses two active-site Mg^{2+} ions for the same purpose. High-resolution crystal structures of RNase H, which cleaves the RNA strand in a DNA:RNA hybrid, revealed an active site that forms a pocket of highly negative surface charge potential that becomes occupied by two Mg^{2+} ions.[1306−1309] De Vivo et al. used QM/MM Car–Parrinello MD simulations to model the cleavage mechanism.[1310] They suggested that the scissile phosphate is attacked either by a water molecule or by a hydroxide ion coordinated to one of the active site Mg^{2+} ions, forming then a phosphorane intermediate. The second Mg^{2+} ion was proposed to stabilize the negative charge developing during cleavage.[1310] In case of a water molecule acting as the nucleophile, it is thought to be activated by transfer of its proton to the nonbridging oxygen of the scissile phosphate. Subsequently, the proton is shuttled to the O3′ leaving group and the phosphorane intermediate is opened.[1310] This mechanism was independently supported by a QM/MM-constrained energy minimization by Elsässer and Fels.[1311] A slightly different mechanism was proposed by Hummer and co-workers, who suggested that the active site water molecule is activated by proton transfer to the *pro*-R_P nonbridging oxygen of the phosphate downstream of the scissile phosphate,[1312] which would explain the strong thio-effect of this particular nonbridging oxygen and the lack of a thio-effect of the scissile phosphate.[1313,1314] The reaction was proposed to proceed through the phosphorane intermediate that is subsequently cleaved as the O3′ leaving group is protonated by the neutral form of Asp132.[1312] The same mechanism was also suggested by Sgrignani and Magistrato for the cleavage catalyzed by human flap endonuclease,[1315] and thus may be common for structurally similar enzymes utilizing two active site Mg^{2+} ions for cleavage of an RNA sugar–phosphate backbone.

Classical MD simulations were also used to probe the most plausible reaction mechanisms of the CRISPR/Csy4 endonuclease cleaving the 3′-end of a small CRISPR-derived RNA.[1107] The study was performed on the basis of the available crystal structures of product and precursor,[1108,1316] and reported 55 individual simulations with a total length of almost 4 μs. The simulations identified a double-protonated His29 and deprotonated terminal phosphate as the likely dominant protonation states consistent with the product structure. Furthermore, potential substates consistent with Ser148 and His29 acting as

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX–XXX

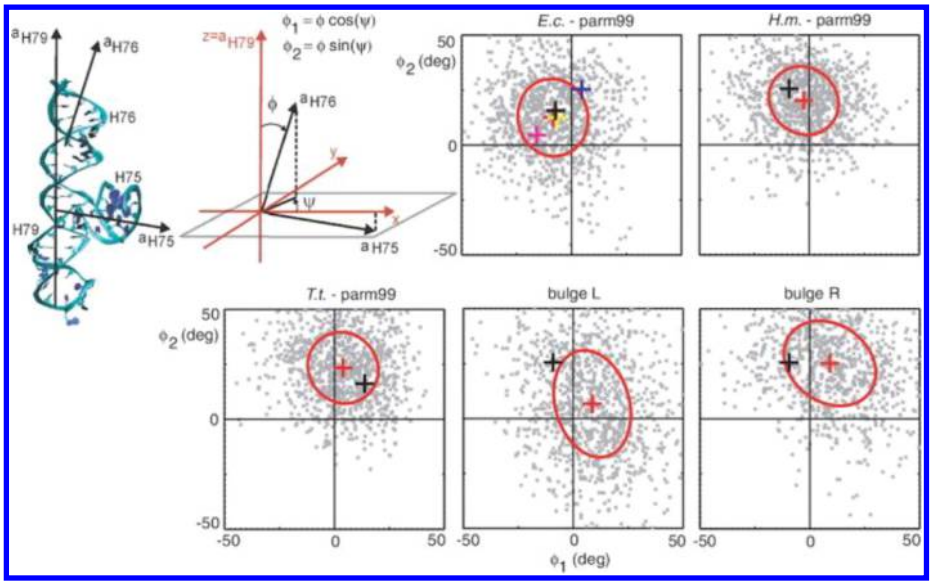


Figure 67. Elastic model of the L1 stalk rRNA parameterized using atomistic MD.[824] (Diagrams) Schematic representation of the coarse-grained description. The global reference system is defined by the helical axes of H79 and H75, a segment is then chosen within H76. The segment helical axis (αH76) serves to measure the bending magnitude ϕ and the bending direction ψ, from which the global coordinates ϕ_1 and ϕ_2 are computed as shown. (Panels) Scatter plots of ϕ_1 and ϕ_2 for some of the simulated systems. Gray points: Instantaneous conformations in 100 ps intervals. The coordinate fluctuations are interpreted using a harmonic, anisotropic elastic model. Red crosses, equilibrium conformations; ellipses, energy levels of $k_BT/2$; black crosses, starting X-ray conformations. In the *E.c.* panel, additional X-ray structures are shown: PDB codes 2I2 V (yellow), 3I1N (magenta), and 3I1P (blue). The fluctuations are almost isotropic for systems without a bulge. The initial half-closed conformation (*T.t.*, black cross) lies within the $k_BT/2$ contour and is thus thermally accessible. The *bulgeL* structure observed in eukaryotes shows biologically relevant fluctuations toward the main ribosome body and toward the 30S subunit (compare with the *T.t.* half-closed starting conformation). Inserting the bulge at the same location but in the opposite strand (*bulgeR*) results in biologically less relevant movement where H76 stays away from the 30S subunit. Adapted with permission from ref 824. Copyright 2012 Oxford University Press.

the general base and acid, respectively, were observed transiently. The authors suggested that Ser148 easily could be deprotonated by solvent to a rare protonation state and could, without further structural rearrangement, immediately deprotonate the nucleophile. Such a hypothesis is very interesting, because it contrasts with the similar MD studies investigating the involvement of rare protonation states of nucleobases in small ribozymes, where the mechanisms of formation of rare protonation states and their subsequent involvement in catalytically competent geometries remain an entire mystery (see particularly section 4.8.3.3). The Csy4 simulations, however, were not able to locate geometries consistent with His29 acting as the general base. The authors also analyzed in detail the performance of the MD technique for the protein/RNA complexes studied. This is rarely done in studies of catalytic mechanisms that often rush to propose possible catalytic pathways without analyzing potential limitations of the available experimental structures and computational methods. The authors reported severe difficulties in obtaining stable trajectories fully compatible with the available crystal structures, even for the proposed dominant protonation states. It was not possible to clarify whether the problems were related to the force field or rather to shortcomings of the data and refinement of the crystal structures. They were nevertheless reminiscent of those detailed above for the HDV ribozyme (see section 4.8.3.2). Thus, the authors cautioned that crystal structures do not always confidently capture, for a variety of reasons, catalytically active geometries and can be affected by uncertainties (such as data and refinement errors) that are only rarely resolved by the MD simulations. Consequently, the true reactive architecture may be unreachable using simulations starting from the available structural data. These uncertainties then may preclude reliable analysis of the reaction mechanism by QM/MM approaches.

MD simulations combined with extensive biochemical experiments also analyzed complexes of HIV-1 reverse transcriptase with its RNA/DNA substrates; for more details, see section 4.7.9.

4.8.8. Origin of Life Studies. MD simulations of TL-like single-strand overhangs of A-RNA stems were used to explain a series of experiments in which slow spontaneous ligation and cleavage reactions were detected in complementary oligoC/oligoG RNA constructs.[699] The simulations revealed that strand slippage (shifted pairing) in such duplexes may produce four-nucleotide overhangs that sample TL-like geometries (see section 4.2) with chemically reactive conformations suitable for transphosphorylation reactions. Such reactions allow simple cleavage and ligation processes, which are nevertheless sufficient to introduce sequence variations in the RNA. These constructs were suggested to resemble the very first RNA enzymes that emerged on the Earth following spontaneous RNA oligomerizations from precursors. They could play an important role in the formamide-based origin of life scenario, which assumes a key role for cyclic 3′,5′ nucleotide precursors in spontaneous synthesis of first RNAs.[1317] The formamide → 3′,5′ cyclic nucleotide → RNA route is currently the only known prebiotic pathway that selectively leads to nontemplated oligomerization of RNA chains with the correct 5′,3′ RNA linkage. This pathway thus requires no cumbersome chemical steps to eliminate incorrectly linked RNA chains a posteriori.[38] For a review of computational methods relevant to prebiotic chemistry, see ref 1318.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

4.9. MD Simulations on Some Other Types of RNA Systems

The preceding sections overviewed a wide range of atomistic MD simulations of RNA systems drawn from the recent literature. We selected a spectrum of systems, from the smallest to the largest, to illustrate every kind of result and behavior that can be obtained or observed through simulations, their relationship to experiments, and their general limitations. Of course, there have been many simulation studies on other RNA systems that were not included in the overview in full detail. This paragraph includes some additional studies, which were not reviewed in the specific sections above, while acknowledging that the literature in this field is already so expansive that it would be impossible to include every relevant work in full detail.

MD simulations have yielded very good results for well-defined RNA pseudoknot[1319−1321] and junction[431,824,1322,1323] structures, although longer simulations (100 ns or more) of these structures require the use of a force field that correctly describes the χ dihedral potential to achieve stable trajectories (see the discussion in ref 824). Simulations of the three-way junction of the ribosomal L1 stalk RNA[824] revealed that this RNA exhibits high isotropic flexibility that is sufficient to support the known range of RNA movements in this region (Figure 67).[1324] Although it has been suggested that the L1 stalk RNA may possess a localized hinge to enable bending at the junction,[9,1325] MD simulations strongly suggest that its flexibility is evenly distributed around the junction and along helix 76,[824] which is consistent with some later analyses of X-ray structures.[823] The eukaryotic helix 76 enhances its flexibility anisotropically via a single bulge in its upper part. MD simulations revealed that the L1 stalk rRNA junction allows the stalk to spontaneously sample various conformations at minimal energy cost, making all of the observed conformations accessible within the harmonic flexibility regime of thermal motion.[824] Similar distributed flexibility has been observed for junctions present at the L7/L12 stalk and the 5S rRNA, although the flexibility of the L7/L12 stalk RNA appears to be quite anisotropic (see discussion in section 4.3.7).[431] The L1 stalk RNA simulations were used to construct an elastic model of the RNA three-way junction parameterized by the atomistic MD,[824] and a similar approach has been applied also to the C-loop RNA motif.[989] Simulations were also used to probe the transitions between the parallel and antiparallel conformations of the four-way junction from domain 3 of the foot-and-mouth disease virus internal ribosome entry site. The interconversion appears to be almost barrierless, occurring via a rotation between the coaxially stacked helices through a perpendicular intermediate.[1323] MD simulations were applied to study internal dynamics of an RNA hairpin from hepatitis B virus[1326] and internal RNA bulge loops with 3−5 nucleotides.[1327]

A combined analysis of cryo-EM density maps and MD simulations was used to highlight the structural dynamics of the complex formed between the L11 protein and the adjacent helices 43−44 of the rRNA.[1096] This study observed strong mobility of the key A1067 nucleotide and the protein's N-terminal domain, suggesting the existence of two alternating conformations: one corresponding to the free ribosome and the other to the ribosome bound in a ternary complex in which the incoming tRNA is in the A/T position. This was one of the first studies to use MD techniques in the interpretation of cryo-EM data. A subsequent paper made a clear connection between the system's structural flexibility and its cryo-EM density maps.[1328] This idea was developed more explicitly and quantitatively by showing how fluctuations among MD flexible fitting (MDFF) snapshots can be used to determine B-factors for map sharpening.[1329] In general, MD simulations have served as instrumental components of refinement protocols for many cryo-EM experiments because they offer ways to flexibly fit atomic structures onto EM maps.[1330−1332] The simulations incorporate the EM data as an external potential that is added to the MD force field, allowing all internal features present in the EM map to be used in the fitting process, while the model remains fully flexible and stereochemically correct. MD-based approaches could in principle be even used to predict conformations corresponding to marginally populated states (minor conformations) that may be critically important for the system's function, for example, as parts of transition states or intermediate ensembles.[1034] Nevertheless, one should be always careful when using MD simulations to effectively improve the resolution of (raw) experimental data or to detect rare conformations because such applications can be particularly sensitive to force-field and sampling issues (see also section 4.3.8).

One of the most exciting functional centers of the ribosome is the exit tunnel. It is nowadays becoming widely accepted that the synthesized protein actively modulates the work of the ribosome via functional interactions with sensors in the tunnel wall.[1333] The most studied example of such processes is protein-induced ribosome stalling. The exit tunnel is also one of the most important ribosomal sites targeted by antibiotics. For a long time, it was not possible to visualize the protein in the ribosomal exit tunnel, but recent advances in experimental methods made it possible to capture a protein in this state.[1334] Data gathered from these experiments were used to enable MD simulations of the protein's interactions with the exit tunnel.[1043,1335−1337] Besides that, atomistic MD coupled with the free-energy perturbation method was used to study the binding of sparmomycin and several related compounds to a small ∼4100 atom model of the binding site near the peptidyl-transferase center.[1338] Additional studies were done using various model systems that were designed to address diverse aspects of ribosome dynamics and function, such as the monitoring of the codon−anticodon double helix[1339,1340] and other issues.[1341,1342] MD was also utilized to study telomerase RNA hairpin structures.[1343]

As some additional examples, atomistic simulations have been used to characterize RNAs of repeating transcripts that cause genetic diseases, such as expanded CAG and CUG repeats.[1344−1348] Three-dimensional WT-metadynamics (see section 3.2.5) analyzed the binding of two potent ligands to a CAG repeat with a noncanonical A/A base pair.[1346] The collective variables used in this work were the RNA−ligand distance and the numbers of H-bonds and hydrophobic contacts, and information on specific interactions was drawn from preliminary MD simulations. The WT-metadynamics investigation identified binding patterns that differed from those predicted by docking and plain MD, with estimated free energies of binding consistent with the experimental values. The ligand force field used in this work was parameterized using data from QM computations.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

5. SUMMARY AND OUTLOOK

5.1. Present State-of-the-Art

The relevance of RNA in modern day biology can hardly be overstated and was perhaps only eclipsed by its even broader role in a primordial RNA world. A tidal wave of recent fundamental discoveries is also increasing the importance of RNA in medicine. Computational treatments of RNA structure−dynamics−function relationships are playing an ever-expanding role in these developments, fueled by both the rapidly multiplying biological questions concerning RNA function that experimental tools cannot fully answer and the constantly growing power of computers. This Review provides an extensive description of the computational-chemistry methods presently used to study RNA structure and dynamics, focusing particularly on atomistic molecular dynamics. We first discuss the features (and problems) of current atomistic force fields, because their accuracy determines the success of realistically modeling and simulating RNA. Another critical factor is accessing biochemically relevant time scales of simulations to correctly characterize not only the structural dynamics, but also the thermodynamics and kinetics of various processes involving RNA molecules. Documenting efforts to extend these time scales, we next provide an extensive overview of the fast-developing field of enhanced sampling atomistic MD simulation methods and coarse-grained approaches. Many of the noted methods have already been successfully employed to study RNA systems, while we anticipate applications of others to RNA molecules in the imminent future.

After this survey of key methodological parameters, we present a comprehensive overview of their application to RNA systems, including RNA−ion interactions, conformational dynamics ranging from the smallest RNA molecules to the entire ribosome, protein/RNA complexes, and RNA catalysis. For each class of studied systems, we aim to provide a broader discussion relating the computed results to experimental findings, and vice versa. Where possible, we highlight not only the successes but also problems and limitations, often hidden in the literature. It is our deeply held belief that sharing problems and explicitly discussing limitations can only benefit the community and foster suggestions for overcoming current obstacles, thus ultimately improving the reliability of computations and their capability to predict and reproduce experimental data. For the same reason, we suggest that groups working in this field share as much detailed information as possible about their work, including all simulation parameters and, where possible, the resulting simulated trajectories. As we have extensively discussed, the accuracy of current force fields is not yet sufficient for blind RNA structure prediction. Some coarse-grained models can provide reasonable performance in this area, but they still frequently need some form of experimental input. However, in many other applications, simulation methods provide unique results that, when properly interpreted, substantially enrich our insights into the structural dynamics and evolution of RNA molecules.

5.2. Perspectives

Because of the fundamental roles of structural dynamics in RNA function and evolution, there is an urgent need for further expansion of MD simulation techniques into the RNA field. We predict that the importance of RNA simulations will organically grow in the future, and MD will become a versatile tool used pervasively to complement experimental investigations of RNA structural dynamics. However, an important challenge is to identify optimal approaches to exploit the advantages of MD methods while eliminating their inherent limitations. To continue the successful expansion of MD-based methods within the RNA field, in addition to the expected steady development of hardware, we highlight three major priorities. The first is to improve the basic molecular mechanical model of RNA, that is, the force field. Refinement of current force fields based on pairwise-additive terms may be reaching its natural limit, although reparameterization of the nonbonded terms may provide a new momentum. However, the key challenge for long-term expansion of atomistic simulations will be successful parameterization of more physically realistic polarizable force fields (although it is presently very difficult to predict the outcome of such efforts). As was also discussed in our review, we think that in parallel with a push to improve the general force field, efforts should be made to adjust force-field performance for specific goals, in a manner reminiscent of the development and deployment of coarse-grained approaches.

The second major priority for future expansion of simulation methods is their proper integration with multiscale approaches, that is, a smart combination of atomistic force-field descriptions with the fast-growing fields of enhanced sampling and coarse-grained approaches. The key for success of such complex approaches is an in-depth understanding of the principles, advantages, and limitations of these methods. In addition, enhanced sampling methods nowadays can be invaluable for validating force-field improvements. This synergy is why we spent substantial effort in our overview on explaining enhanced sampling and coarse-grained methods. We also expect significant innovations to arise from the development of QM methods applicable to nucleic acids, but we could not review these methods in more detail due to space limitations.

Finally, as close as possible direct collaborations between computational chemists and experimentalists are desirable. We believe that they offer the most straightforward and efficient opportunities to immediately expand the field and maximize gains from MD simulations, even without further methodological advances. The most sophisticated simulations can provide misleading results when applied without a sufficiently intimate understanding of the studied systems. Similarly, supplementing experiments with suboptimally executed and superficially analyzed calculations, without understanding their scope and limitations, can backfire. Recognition of the need for synergistic collaboration, and the pitfalls associated with lack of collaborative efforts, motivated us to write such a comprehensive review. We hope that experimental researchers will use it to better understand the complexity of computational methods, while computational researchers can gain insight into the complexity of the structural principles of RNA. We suggest that the ongoing efforts to integrate MD simulation approaches with solution-state NMR and cryo-EM represent exemplary areas of research where major progress can be expected over the next decade, as was already illustrated by many of the studies reviewed here.

We also would like to note that there are two main approaches to the utilization of simulation tools. In one, experiments are augmented by computational interpretation, often revealing atomistic details of dynamic events that may be challenging to capture experimentally. In this approach, which implicitly or explicitly underlies many of the ideas expressed in this Review, accuracy in the description of RNA structural dynamics and catalytic transformations is most highly valued. In

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

the second approach, quantitative accuracy is regarded as less important in calculations, as long as new experimentally testable hypotheses can be formulated, novel interpretations of experiments are provided, or new phenomena that cannot be accessed experimentally are discovered. Our Review provides examples of both approaches; they coexist and fulfill complementary purposes that are collectively increasing the importance of computational treatments in RNA analyses.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: sponer@ncbr.muni.cz.
*E-mail: bussi@sissa.it.

ORCID

Jiří Šponer: 0000-0001-6558-6186
Giovanni Bussi: 0000-0001-9216-5782
Pavel Banáš: 0000-0002-7137-8225
Petr Jurečka: 0000-0002-3741-3672
Michal Otyepka: 0000-0002-1066-5677

Notes

The authors declare no competing financial interest.

Biographies

Jiří Šponer (1964) is currently Head of the Department of Structure and Dynamics of Nucleic Acids, Institute of Biophysics of the Czech Academy of Sciences, Brno, Czech Republic, and Professor and Head of the Nucleic Acids group at the Regional Centre of Advanced Technologies and Materials, Palacky University, Olomouc, Czech Republic. He is also a Professor at the Masaryk University, Brno. He earned his M.Sc. and Ph.D. degrees at the Masaryk University, Brno. Since 1993, he has been primarily associated with the Czech Academy of Sciences. His primary research interests are applications of modern quantum-chemical and molecular simulation methods in studies of the structure, dynamics, function, and evolution of nucleic acids.

Giovanni Bussi (1977) earned his Ph.D. in physics at the University of Modena and Reggio Emilia (Italy) under the supervision of Prof. Elisa Molinari. After postdoctoral work at the ETH Zurich (Switzerland) in the lab of Prof. Michele Parrinello, he started an independent group at the Scuola Internazionale Superiore di Studi Avanzati (SISSA, Trieste, Italy). Since 2014 he has a tenured position. His research interests include the development of methods for enhanced sampling simulations and their application to the characterization of the structural dynamics of RNA systems.

Miroslav Krepl (1988) received his undergraduate degree in biochemistry and his master degree in biomolecular chemistry from Masaryk University, Brno, Czech Republic. He is now a Ph.D. student under the supervision of Prof. Jiri Sponer and a junior researcher at the Institute of Biophysics of the Czech Academy of Sciences, Brno, and at the Regional Centre of Advanced Technologies and Materials, Palacky University, Olomouc, Czech Republic. His research is focused on molecular dynamics simulations of small RNA molecules and protein/RNA complexes.

Pavel Banáš (1980) obtained his Ph.D. from Palacky University in Olomouc, Czech Republic, in 2009 (with Michal Otyepka). He is currently an Associate Professor at the Department of Physical Chemistry at Palacky University. His main interests are related to the theoretical description of RNA catalysis, structural dynamics of nucleic acids, and folding of small RNA motifs.

Sandro Bottaro (1983) received his Ph.D. from the Danish Technical University (DTU) in 2012 for his work on Monte Carlo methods for simulations of proteins. He then joined the group of Prof. Giovanni Bussi at the Scuola Internazionale Superiore di Studi Avanzati (SISSA, Trieste, Italy), where he focused on developing molecular modeling and structural bioinformatics approaches to study RNA. In 2016, he moved to a postdoctoral position with Kresten Lindorff-Larsen at the University of Copenhagen (Denmark), where he is working on studying RNA dynamics by integrating molecular dynamics simulations with NMR experiments.

Richard André Cunha (1989) received his undergraduate degree in chemistry and his master degree in physical chemistry from the Federal University of Uberlandia (UFU, Minas Gerais, Brazil). In his master, under the supervision of Prof. Eduardo Franca, he studied the interaction of polysaccharides with membranes employing molecular dynamics. He is now a graduate student under the supervision of Prof. Giovanni Bussi at the Scuola Internazionale Superiore di Studi Avanzati (SISSA, Trieste, Italy), where he is focused on describing the interactions between cations and RNA using enhanced sampling simulations.

Alejandro Gil-Ley (1984) received his undergraduate degree in biochemistry from the University of La Habana. As a graduate student in the group of Prof. Giovanni Bussi, he focused on the development of new enhancing sampling techniques and on the combination of these techniques with experimental data to reliably explore RNA structural ensembles. He is currently a postdoctoral fellow in Prof. Janice Robertson's Lab at the University of Iowa, under the cosupervision of Prof. Jose Faraldo at NHLBI/NIH.

Giovanni Pinamonti (1988) received his bachelor and master degrees in physics from the University of Trieste (Italy). He later obtained his Ph.D. at the Scuola Internazionale Superiore di Studi Avanzati (SISSA, Trieste, Italy), under the supervision of Prof. Giovanni Bussi. As a graduate student he focused on studying the biophysical properties of RNA molecules by means of different computational techniques. Currently, he works as a postdoctoral fellow at the Freie Universitaet Berlin (Germany), in the group of Prof. Cecilia Clementi.

Simón Poblete (1983) received his Ph.D. in the group of Prof. Kurt Kremer at the Max Planck Institute for Polymer Research, Mainz, Germany. A postdoctoral experience followed in the Institute of Complex Systems in the Juelich Research Center, under the supervision of Prof. Gerhard Gompper. During the last three years, he has been devoted to RNA simulations in the group of Prof. Giovanni Bussi as a postdoctoral researcher. Along his career, Dr. Poblete has struggled with problems related to the development of multiscale simulation methods, the development of coarse-grained models of liquids, polymers, proteins, and nucleic acids, and their subsequent software implementation.

Petr Jurečka (1976) obtained his Ph.D. from the Charles University in Prague, Czech Republic, and the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, in 2004 (with Pavel Hobza). He did his postdoctoral study with Professor D. R. Salahub at the University of Calgary, Canada. He is currently an Associate Professor at the Department of Physical Chemistry at the Palacký University, Olomouc, Czech Republic. His main interests are related to the empirical force-field development for nucleic acids, simulations of conformational equilibria in RNA and DNA, theoretical description of noncovalent interactions in biomolecules and nanomaterials, highly accurate quantum chemistry calculations, and density functional theory.

Nils G. Walter (1966) received his "Vordiplom" (B.S.) and "Diploma" (Masters) from the Darmstadt University of Technology after performing research with Hans-Günther Gassen on the physiochemical characterization of a protein dehyrogenase enzyme. Inspired by his

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

interest in the origin of life, he earned his Dr. rer. nat. while studying molecular in vitro evolution of DNA and RNA using fluorescence techniques with Nobel laureate Manfred Eigen at the Max-Planck-Institute for Biophysical Chemistry, Göttingen. For his postdoctoral studies, he turned to RNA enzymes as a possible driver of early evolution under the guidance of John M. Burke at the University of Vermont in Burlington, Vermont. He is currently the Francis S. Collins Collegiate Professorship in Chemistry, Biophysics, and Biological Chemistry, College of Literature, Science, and the Arts at the University of Michigan in Ann Arbor, Michigan, where he founded and now directs the Single Molecule Analysis in Real-Time (SMART) Center, as well as cofounded and codirects the Center for RNA Biomedicine. His research interests focus on noncoding RNA through the lens of single-molecule fluorescence techniques.

Michal Otyepka (1975) received his Ph.D. degree in Physical Chemistry at the Palacký University, Olomouc, Czech Republic (2004). Currently, he is a head of the Department of Physical Chemistry at the Palacký University, Olomouc, and vice-director of the Regional Centre of Advanced Technologies and Materials, Palacký University, Olomouc. His research is focused on modeling of biomacromolecules, complex molecular systems, and nanomaterials.

ACKNOWLEDGMENTS

This work was supported by grant P305/12/G034 (J.S. and M.K.) from the Czech Science Foundation. G.B., R.A.C., A.G.-L., G.P., and S.P. received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 306662, S-RNA-S. N.G.W. acknowledges funding from NIH grants GM062357, GM118524, and GM122803. Further funding was provided by project LO1305 of the Ministry of Education, Youth and Sports of the Czech Republic (P.B., P.J., and M.O.). J.S. acknowledges support by Praemium Academiae. S.B. acknowledges the Velux foundations for funding.

ABBREVIATIONS

RNA	ribonucleic acid
tRNA	transfer RNA
mRNA	messenger RNA
rRNA	ribosomal RNA
MD	molecular dynamics
DNA	deoxyribonucleic acid
WC	Watson−Crick
2D	secondary structure
3D	tertiary structure
QM	quantum mechanical
MM	molecular mechanical
vdW	van der Waals
ESP	electrostatic potential
RESP	restrained electrostatic potential (fitting)
NMR	nuclear magnetic resonance
NBfix	nonbonded fix
HBfix	hydrogen-bond fix
MC	Monte Carlo
MSM	Markov state model
CV	collective variable
TICA	time-lagged independent component analysis
PCA	principal component analysis
HMM	hidden Markov model
PT	parallel tempering
REMD	replica exchange molecular dynamics
WT	well-tempered (metadynamics)
T-REMD	temperature replica exchange molecular dynamics
H-REMD	Hamiltonian replica exchange molecular dynamics
LES	locally enhanced sampling
WHAM	weighted histogram analysis method
PMF	potential of mean force
US	umbrella sampling
TI	thermodynamic integration
RMSD	root-mean-square deviation
eRMSD	epsilon-RMSD
BE	bias exchange (metadynamics)
RAM	replica averaged metadynamics
FEP	free-energy perturbation
GBSA	generalized Born/surface area
PBSA	Poisson−Boltzmann/surface area
MM-GBSA	molecular mechanics − GBSA
MM-PBSA	molecular mechanics − PBSA
CG	coarse-grained
MMTV	mouse mammary tumor virus
ENM	elastic network model
cryo-EM	cryo-electron microscopy
NN	nearest neighbor
NMA	normal-mode analysis
SHAPE	selective 2′-hydroxyl acylation analyzed by primer extension
SIBFA	Sum of Interactions Between Fragments Ab initio computed
PME	particle Mesh Ewald
HDV	Hepatitis delta virus
LJ	Lennard-Jones
SLV	stem-loop V
TL	tetraloop
RDC	residual dipolar coupling
SRL	Sarcin-Ricin loop
SRD	Sarcin-Ricin domain
TAR (element)	trans-activation response (element)
DIS	dimerization initiation site
UTRs	untranslated regions
FRET	fluorescence resonance energy transfer
SAM	*S*-adenosylmethionine
SAH	*S*-adenosylhomocystein
AME	aminoglycoside modifying enzymes
BD	Brownian dynamics
EF-G	elongation factor G
GAFF	general AMBER force field
MDFF	molecular dynamics flexible fitting
RRM	RNA recognition motif
PTB	polypyrimidine tract binding
SRSF1	serine/arginine-rich splicing factor
ITC	isothermal titration calorimetry
TRBP	TAR RNA binding protein
DCL1	dicer-like 1
ADAR2	adenosine deaminase acting on RNA 2
RBD	RNA binding domain
snRNA	small nucleolar RNA
AARS	aminoacyl tRNA synthetase
CRISPR	clustered regularly interspaced short palindromic repeat
RT	reverse transcriptase
RRE	Rev-Responsive Element
FMRP	fragile X mental retardation protein
RF	release factor

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

KIE kinetic isotope effects

REFERENCES

(1) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227*, 561−563.
(2) Sharp, S. J.; Schaack, J.; Cooley, L.; Burke, D. J.; Soil, D. Structure and Transcription of Eukaryotic tRNA Gene. *CRC Crit. Rev. Biochem.* **1985**, *19*, 107−144.
(3) Ogle, J. M.; Ramakrishnan, V. Structural Insights into Translational Fidelity. *Annu. Rev. Biochem.* **2005**, *74*, 129−177.
(4) Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. Insights Into the Decoding Mechanism From Recent Ribosome Structures. *Trends Biochem. Sci.* **2003**, *28*, 259−266.
(5) Pereira, M. J. B.; Harris, D. A.; Rueda, D.; Walter, N. G. Reaction Pathway of the Trans-acting Hepatitis Delta Virus Ribozyme: A Conformational Change Accompanies Catalysis. *Biochemistry* **2002**, *41*, 730−740.
(6) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**, *31*, 147−157.
(7) Emilsson, G. M.; Nakamura, S.; Roth, A.; Breaker, R. R. Ribozyme Speed Limits. *RNA* **2003**, *9*, 907−918.
(8) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional Insights from the Structure of the 30S Ribosomal Subunit and its Interactions with Antibiotics. *Nature* **2000**, *407*, 340−348.
(9) Korostelev, A.; Ermolenko, D. N.; Noller, H. F. Structural Dynamics of the Ribosome. *Curr. Opin. Chem. Biol.* **2008**, *12*, 674−683.
(10) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. D. Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science* **2005**, *310*, 827−834.
(11) Ben-Shem, A.; de Loubresse, N. G.; Melnikov, S.; Jenner, L.; Yusupova, G.; Yusupov, M. The Structure of the Eukaryotic Ribosome at 3.0 angstrom Resolution. *Science* **2011**, *334*, 1524−1529.
(12) Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. A New Understanding of the Decoding Principle on the Ribosome. *Nature* **2012**, *484*, 256−259.
(13) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. *Cell* **2001**, *107*, 679−688.
(14) Laurberg, M.; Asahara, H.; Korostelev, A.; Zhu, J. Y.; Trakhanov, S.; Noller, H. F. Structural Basis for Translation Termination on the 70S Ribosome. *Nature* **2008**, *454*, 852−857.
(15) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 angstrom Resolution. *Science* **2000**, *289*, 905−920.
(16) Cech, T. R. The Ribosome Is a Ribozyme. *Science* **2000**, *289*, 878−879.
(17) Crick, F. H. C. The Origin of the Genetic Code. *J. Mol. Biol.* **1968**, *38*, 367−379.
(18) Orgel, L. E. Evolution of the Genetic Apparatus. *J. Mol. Biol.* **1968**, *38*, 381−393.
(19) Woese, C. R. *The Genetic Code: The Molecular Basis for Genetic Expression*; Harper & Row: New York, 1967.
(20) Serganov, A.; Nudler, E. A Decade of Riboswitches. *Cell* **2013**, *152*, 17−24.
(21) Wilson, R. C.; Doudna, J. A. Molecular Mechanisms of RNA Interference. In *Annual Review of Biophysics*; Dill, K. A., Ed.; Annual Reviews: Palo Alto, CA, 2013; Vol. *42*, pp 217−239.
(22) Pitchiaya, S.; Heinicke, L. A.; Custer, T. C.; Walter, N. G. Single Molecule Fluorescence Approaches Shed Light on Intracellular RNAs. *Chem. Rev.* **2014**, *114*, 3224−3265.
(23) Djebali, S.; Davis, C. A.; Merkel, A.; Dobin, A.; Lassmann, T.; Mortazavi, A.; Tanzer, A.; Lagarde, J.; Lin, W.; Schlesinger, F.; et al. Landscape of Transcription in Human Cells. *Nature* **2012**, *489*, 101−108.
(24) Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M. F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B. W.; Cassady, J. P.; et al. Chromatin Signature Reveals over a Thousand Highly Conserved Large Non-coding RNAs in Mammals. *Nature* **2009**, *458*, 223−227.
(25) Belinky, F.; Bahir, I.; Stelzer, G.; Zimmerman, S.; Rosen, N.; Nativ, N.; Dalah, I.; Iny Stein, T.; Rappaport, N.; Mituyama, T.; et al. Non-redundant Compendium of Human ncRNA Genes in GeneCards. *Bioinformatics* **2013**, *29*, 255−261.
(26) Huarte, M. The Emerging Role of lncRNAs in Cancer. *Nat. Med.* **2015**, *21*, 1253−1261.
(27) Diederichs, S.; Bartsch, L.; Berkmann, J. C.; Fröse, K.; Heitmann, J.; Hoppe, C.; Iggena, D.; Jazmati, D.; Karschnia, P.; Linsenmeier, M.; et al. The Dark Matter of the Cancer Genome: Aberrations in Regulatory Elements, Untranslated Regions, Splice Sites, Non-coding RNA and Synonymous Mutations. *EMBO Mol. Med.* **2016**, *8*, 442−457.
(28) Ranum, L. P. W.; Day, J. W. Myotonic Dystrophy: RNA Pathogenesis Comes into Focus. *Am. J. Hum. Genet.* **2004**, *74*, 793−804.
(29) Cooper, T. A.; Wan, L.; Dreyfuss, G. RNA and Disease. *Cell* **2009**, *136*, 777−793.
(30) Cheetham, S. W.; Gruhl, F.; Mattick, J. S.; Dinger, M. E. Long Noncoding RNAs and the Genetics of Cancer. *Br. J. Cancer* **2013**, *108*, 2419−2425.
(31) Bracken, C. P.; Scott, H. S.; Goodall, G. J. A Network-biology Perspective of MicroRNA Function and Dysfunction in Cancer. *Nat. Rev. Genet.* **2016**, *17*, 719−732.
(32) Khurana, E.; Fu, Y.; Chakravarty, D.; Demichelis, F.; Rubin, M. A.; Gerstein, M. Role of Non-coding Sequence Variants in Cancer. *Nat. Rev. Genet.* **2016**, *17*, 93−108.
(33) Lünse, C. E.; Schüller, A.; Mayer, G. The Promise of Riboswitches as Potential Antibacterial Drug Targets. *Int. J. Med. Microbiol.* **2014**, *304*, 79−92.
(34) Colameco, S.; Elliot, M. A. Non-coding RNAs as Antibiotic Targets. *Biochem. Pharmacol.* **2017**, *133*, 29−42.
(35) Blount, K. F.; Breaker, R. R. Riboswitches as Antibacterial Drug Targets. *Nat. Biotechnol.* **2006**, *24*, 1558−1564.
(36) Blount, K. F.; Megyola, C.; Plummer, M.; Osterman, D.; O'Connell, T.; Aristoff, P.; Quinn, C.; Chrusciel, R. A.; Poel, T. J.; Schostarez, H. J.; et al. Novel Riboswitch-Binding Flavin Analog That Protects Mice against Clostridium difficile Infection without Inhibiting Cecal Flora. *Antimicrob. Agents Chemother.* **2015**, *59*, 5736−5746.
(37) Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; et al. Selective Small-molecule Inhibition of an RNA Structural Element. *Nature* **2015**, *526*, 672−677.
(38) Šponer, J. E.; Šponer, J.; Mauro, E. D. New Evolutionary Insights Into the Non-enzymatic Origin of RNA Oligomers. *Wiley Interdiscip. Rev.: RNA* **2017**, *8*, e1400.
(39) Leontis, N. B.; Stombaugh, J.; Westhof, E. The Non-Watson-Crick Base Pairs and their Associated Isostericity Matrices. *Nucleic Acids Res.* **2002**, *30*, 3497−3531.
(40) Leontis, N. B.; Westhof, E. Geometric Nomenclature and Classification of RNA Base Pairs. *RNA* **2001**, *7*, 499−512.
(41) Sponer, J.; Sponer, J. E.; Petrov, A. I.; Leontis, N. B. Quantum Chemical Studies of Nucleic Acids Can We Construct a Bridge to the RNA Structural Biology and Bioinformatics Communities? *J. Phys. Chem. B* **2010**, *114*, 15723−15741.
(42) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B. Frequency and Isostericity of RNA Base Pairs. *Nucleic Acids Res.* **2009**, *37*, 2294−2312.
(43) Zhao, B. S.; Roundtree, I. A.; He, C. Post-transcriptional Gene Regulation by mRNA Modifications. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 31−42.
(44) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* **1999**, *288*, 911−940.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(45) Mathews, D. H.; Turner, D. H. Prediction of RNA Secondary Structure by Free Energy Minimization. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270−278.
(46) Turner, D. H.; Sugimoto, N.; Freier, S. M. RNA Structure Prediction. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167−192.
(47) Mokdad, A.; Krasovska, M. V.; Sponer, J.; Leontis, N. B. Structural and Evolutionary Classification of G/U Wobble Basepairs in the Ribosome. *Nucleic Acids Res.* **2006**, *34*, 1326−1341.
(48) Hull, C. M.; Bevilacqua, P. C. Discriminating Self and Non-Self by RNA: Roles for RNA Structure, Misfolding, and Modification in Regulating the Innate Immune Sensor PKR. *Acc. Chem. Res.* **2016**, *49*, 1242−1249.
(49) Sweeney, B. A.; Roy, P.; Leontis, N. B. An Introduction to Recurrent Nucleotide Interactions in RNA. *Wiley Interdiscip. Rev.: RNA* **2015**, *6*, 17−45.
(50) Zirbel, C. L.; Sponer, J. E.; Sponer, J.; Stombaugh, J.; Leontis, N. B. Classification and Energetics of the Base-Phosphate Interactions in RNA. *Nucleic Acids Res.* **2009**, *37*, 4898−4918.
(51) Leontis, N. B.; Westhof, E. The 5S rRNA Loop E: Chemical Probing and Phylogenetic Data Versus Crystal Structure. *RNA* **1998**, *4*, 1134−1153.
(52) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. FR3D: Finding Local and Composite Recurrent Structural Motifs in RNA 3D Structures. *J. Math. Biol.* **2008**, *56*, 215−252.
(53) Petrov, A. I.; Zirbel, C. L.; Leontis, N. B. Automated Classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* **2013**, *19*, 1327−1340.
(54) Leontis, N. B.; Westhof, E. Analysis of RNA Motifs. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300−308.
(55) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. Recurrent Structural RNA Motifs, Isostericity Matrices and Sequence Alignments. *Nucleic Acids Res.* **2005**, *33*, 2395−2409.
(56) Leontis, N. B.; Westhof, E. A Common Motif Organizes the Structure of Multi-helix Loops in 16 and 23 S Ribosomal RNAs. *J. Mol. Biol.* **1998**, *283*, 571−583.
(57) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; et al. RNA backbone: Consensus All-angle Conformers and Modular String Nomenclature (An RNA Ontology Consortium Contribution). *RNA* **2008**, *14*, 465−481.
(58) Sponer, J.; Mladek, A.; Sponer, J. E.; Svozil, D.; Zgarbova, M.; Banas, P.; Jurecka, P.; Otyepka, M. The DNA and RNA Sugar-Phosphate Backbone Emerges as the Key Player. An Overview of Quantum-chemical, Structural Biology and Simulation Studies. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15257−15277.
(59) Zirbel, C. L.; Roll, J.; Sweeney, B. A.; Petrov, A. I.; Pirrung, M.; Leontis, N. B. Identifying Novel Sequence Variants of RNA 3D Motifs. *Nucleic Acids Res.* **2015**, *43*, 7504−7520.
(60) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. Single Nucleotide RNA Choreography. *Nucleic Acids Res.* **2006**, *34*, 1481−1491.
(61) Parlea, L. G.; Sweeney, B. A.; Hosseini-Asanjan, M.; Zirbel, C. L.; Leontis, N. B. The RNA 3D Motif Atlas: Computational Methods for Extraction, Organization and Evaluation of RNA Motifs. *Methods* **2016**, *103*, 99−119.
(62) Aboul-ela, F.; Karn, J.; Varani, G. Structure of HIV-1 TAR RNA in the Absence of Ligands Reveals a Novel Conformation of the Trinucleotide Bulge. *Nucleic Acids Res.* **1996**, *24*, 3974−3981.
(63) Shen, L. X.; Tinoco, J. I. The Structure of an RNA Pseudoknot that Causes Efficient Frameshifting in Mouse Mammary Tumor Virus. *J. Mol. Biol.* **1995**, *247*, 963−978.
(64) Gultyaev, A. P.; Olsthoorn, R. C. L.; Pleij, C. W. A.; Westhof, E. RNA Structure: Pseudoknots. *eLS*; John Wiley & Sons, Ltd.: Chichester, 2012.
(65) Jaeger, L.; Verzemnieks, E. J.; Geary, C. The UA_handle: A Versatile Submotif in Stable RNA Architectures. *Nucleic Acids Res.* **2009**, *37*, 215−230.
(66) Koculi, E.; Cho, S. S.; Desai, R.; Thirumalai, D.; Woodson, S. A. Folding Path of P5abc RNA Involves Direct Coupling of Secondary and Tertiary Structures. *Nucleic Acids Res.* **2012**, *40*, 8011−8020.
(67) Brion, P.; Westhof, E. Hierarchy and Dynamics of RNA Folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113−137.
(68) Tinoco, I.; Bustamante, C. How RNA Folds. *J. Mol. Biol.* **1999**, *293*, 271−281.
(69) Woodson, S. A. Compact Intermediates in RNA Folding. *Annu. Rev. Biophys.* **2010**, *39*, 61−77.
(70) Rinnenthal, J.; Buck, J.; Ferner, J.; Wacker, A.; Fürtig, B.; Schwalbe, H. Mapping the Landscape of RNA Dynamics with NMR Spectroscopy. *Acc. Chem. Res.* **2011**, *44*, 1292−1301.
(71) Watters, K. E.; Strobel, E. J.; Yu, A. M.; Lis, J. T.; Lucks, J. B. Cotranscriptional Folding of a Riboswitch at Nucleotide Resolution. *Nat. Struct. Mol. Biol.* **2016**, *23*, 1124−1131.
(72) Al-Hashimi, H. M.; Walter, N. G. RNA Dynamics: It Is about Time. *Curr. Opin. Struct. Biol.* **2008**, *18*, 321−329.
(73) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; et al. Anton, a Special-purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91−97.
(74) Shirts, M.; Pande, V. S. Screen Savers of the World Unite! *Science* **2000**, *290*, 1903−1904.
(75) Pande, V. S.; Baker, I.; Chapman, J.; Elmer, S. P.; Khaliq, S.; Larson, S. M.; Rhee, Y. M.; Shirts, M. R.; Snow, C. D.; Sorin, E. J.; et al. Atomistic Protein Folding Simulations on the Submillisecond Time Scale Using Worldwide Distributed Computing. *Biopolymers* **2003**, *68*, 91−109.
(76) Islam, B.; Stadlbauer, P.; Gil Ley, A.; Pérez-Hernández, G.; Haider, S.; Neidle, S.; Bussi, G.; Banáš, P.; Otyepka, M.; Sponer, J. Exploring the Dynamics of Propeller Loops in Human Telomeric DNA Quadruplexes Using Atomistic Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 2458−2480.
(77) Fadrna, E.; Spackova, N.; Sarzynska, J.; Koca, J.; Orozco, M.; Cheatham, T. E.; Kulinski, T.; Sponer, J. Single Stranded Loops of Quadruplex DNA As Key Benchmark for Testing Nucleic Acids Force Fields. *J. Chem. Theory Comput.* **2009**, *5*, 2514−2530.
(78) Smith, L. G.; Zhao, J.; Mathews, D. H.; Turner, D. H. Physics-based All-atom Modeling of RNA Energetics and Structure. *Wiley Interdiscip. Rev.: RNA* **2017**, *8*, No. e1422.
(79) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(80) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(81) Savelyev, A.; MacKerell, A. D. All-atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.
(82) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Refinement Using Quantum Mechanical Base Stacking and Conformational Energetics. *J. Chem. Theory Comput.* **2017**, *13*, 2053−2071.
(83) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. *J. Chem. Theory Comput.* **2017**, *13*, 2072−2085.
(84) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(85) Kruse, H.; Mladek, A.; Gkionis, K.; Hansen, A.; Grimme, S.; Sponer, J. Quantum Chemical Benchmark Study on 46 RNA Backbone Families Using a Dinucleotide Unit. *J. Chem. Theory Comput.* **2015**, *11*, 4972−4991.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(86) Kruse, H.; Havrila, M.; Sponer, J. QM Computations on Complete Nucleic Acids Building Blocks: Analysis of the Sarcin-Ricin RNA Motif Using DFT-D3, HF-3c, PM6-D3H, and MM Approaches. *J. Chem. Theory Comput.* **2014**, *10*, 2615−2629.
(87) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-bonded base pairs in chloroform. *J. Am. Chem. Soc.* **1991**, *113*, 2810−2819.
(88) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(89) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2000**, *56*, 257−265.
(90) Wang, J. M.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(91) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821−7839.
(92) Sponer, J.; Cang, X. H.; Cheatham, T. E. Molecular Dynamics Simulations of G-DNA and Perspectives on the Simulation of Nucleic Acid Structures. *Methods* **2012**, *57*, 25−39.
(93) Cheatham, T. E.; Case, D. A. Twenty-five Years of Nucleic Acid Simulations. *Biopolymers* **2013**, *99*, 969−977.
(94) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P.; et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886−2902.
(95) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinenement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(96) MacKerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(97) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; MacKerell, A. D., Jr Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme. *Biophys. J.* **2006**, *90*, L36−L38.
(98) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D. Impact of 2 '-Hydroxyl Sampling on the Conformational Properties of RNA: Update of the CHARMM All-Atom Additive Force Field for RNA. *J. Comput. Chem.* **2011**, *32*, 1929−1943.
(99) Mlynsky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Sponer, J.; Otyepka, M. Extensive Molecular Dynamics Simulations Showing That Canonical G8 and Protonated A38H(+) Forms Are Most Consistent with Crystal Structures of Hairpin Ribozyme. *J. Phys. Chem. B* **2010**, *114*, 6642−6652.
(100) Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koča, J.; Cheatham, T. E.; Jurečka, P.; Šponer, J. Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012**, *8*, 2506−2520.
(101) Zgarbova, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P. Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theory Comput.* **2014**, *10*, 3177−3189.
(102) Zgarbová, M.; Rosnik, A. M.; Luque, F. J.; Curutchet, C.; Jurečka, P. Transferability and Additivity of Dihedral Parameters in Polarizable and Nonpolarizable Empirical Force Fields. *J. Comput. Chem.* **2015**, *36*, 1874−1884.
(103) MacKerell, A. D.; Feig, M.; Brooks, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(104) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. Performance of Empirical Potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), Semiempirical Quantum Chemical Methods (AM1, MNDO/M, PM3), and Ab Initio Hartree−Fock Method for Interaction of DNA Bases: Comparison with Non-empirical Beyond Hartree−Fock Results. *J. Comput. Chem.* **1997**, *18*, 1136−1150.
(105) Sponer, J.; Hobza, P. Nonplanar Geometries of DNA Bases. Ab Initio Second-order Moeller-Plesset Study. *J. Phys. Chem.* **1994**, *98*, 3161−3164.
(106) Šponer, J.; Mokdad, A.; Šponer, J. E.; Špačková, N.; Leszczynski, J.; Leontis, N. B. Unique Tertiary and Neighbor Interactions Determine Conservation Patterns of Cis Watson−Crick A/G Base-pairs. *J. Mol. Biol.* **2003**, *330*, 967−978.
(107) Halgren, T. A. Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(108) Hobza, P.; Šponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. C−H···O Contacts in the Adenine···Uracil Watson−Crick and Uracil··· Uracil Nucleic Acid Base Pairs: Nonempirical ab Initio Study with Inclusion of Electron Correlation Effects. *J. Phys. Chem. B* **2000**, *104*, 6286−6292.
(109) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(110) Pokorna, P.; Krepl, M.; Kruse, H.; Sponer, J. MD and QM/MM Study of the Quaternary HutP Homohexamer Complex with mRNA, L-histidine Ligand and Mg2+. *J. Chem. Theory Comput.* **2017**, *13*, 5658−5670.
(111) Zgarbova, M.; Otyepka, M.; Sponer, J.; Hobza, P.; Jurecka, P. Large-scale Compensation of Errors in Pairwise-additive Empirical Force Fields: Comparison of AMBER Intermolecular Terms with Rigorous DFT-SAPT Calculations. *Phys. Chem. Chem. Phys.* **2010**, *12*, 10476−10493.
(112) Sponer, J.; Mladek, A.; Spackova, N.; Cang, X.; Cheatham, T. E., III; Grimme, S. Relative Stability of Different DNA Guanine Quadruplex Stem Topologies Derived Using Large-Scale Quantum-Chemical Computations. *J. Am. Chem. Soc.* **2013**, *135*, 9785−9796.
(113) Szabla, R.; Havrila, M.; Kruse, H.; Šponer, J. Comparative Assessment of Different RNA Tetranucleotides from the DFT-D3 and Force Field Perspective. *J. Phys. Chem. B* **2016**, *120*, 10635−10648.
(114) Gkionis, K.; Kruse, H.; Šponer, J. Derivation of Reliable Geometries in QM Calculations of DNA Structures: Explicit Solvent QM/MM and Restrained Implicit Solvent QM Optimizations of G-Quadruplexes. *J. Chem. Theory Comput.* **2016**, *12*, 2000−2016.
(115) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the Current State of AMBER Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114−4127.
(116) Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. *J. Chem. Theory Comput.* **2013**, *9*, 5430−5449.
(117) Ryde, U. How Many Conformations Need To Be Sampled To Obtain Converged QM/MM Energies? The Curse of Exponential Averaging. *J. Chem. Theory Comput.* **2017**, *13*, 5745−5752.
(118) Šponer, J.; Šponer, J. E.; Mládek, A.; Banáš, P.; Jurečka, P.; Otyepka, M. How to Understand Quantum Chemical Computations on DNA and RNA Systems? A Practical Guide for Non-specialists. *Methods* **2013**, *64*, 3−11.
(119) Sponer, J.; Sponer, J. E.; Mladek, A.; Jurecka, P.; Banas, P.; Otyepka, M. Nature and Magnitude of Aromatic Base Stacking in DNA and RNA: Quantum Chemistry, Molecular Mechanics, and Experiment. *Biopolymers* **2013**, *99*, 978−988.
(120) Chen, A. A.; García, A. E. High-resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820−16825.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(121) Maffeo, C.; Luan, B.; Aksimentiev, A. End-to-end Attraction of Duplex DNA. *Nucleic Acids Res.* **2012**, *40*, 3812−3821.
(122) Brown, R. F.; Andrews, C. T.; Elcock, A. H. Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 2315−2328.
(123) Jafilan, S.; Klein, L.; Hyun, C.; Florián, J. Intramolecular Base Stacking of Dinucleoside Monophosphate Anions in Aqueous Solution. *J. Phys. Chem. B* **2012**, *116*, 3613−3618.
(124) Banáš, P.; Mládek, A.; Otyepka, M.; Zgarbová, M.; Jurečka, P.; Svozil, D.; Lankaš, F.; Šponer, J. Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J. Chem. Theory Comput.* **2012**, *8*, 2448−2460.
(125) Häse, F.; Zacharias, M. Free Energy Analysis and Mechanism of Base Pair Stacking in Nicked DNA. *Nucleic Acids Res.* **2016**, *44*, 7100−7108.
(126) Bergonzo, C.; Cheatham, T. E. Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11*, 3969−3972.
(127) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863−3871.
(128) Havrila, M.; Stadlbauer, P.; Islam, B.; Otyepka, M.; Sponer, J. Effect of Monovalent Ion Parameters on Molecular Dynamics Simulations of G-Quadruplexes. *J. Chem. Theory Comput.* **2017**, *13*, 3911−3926.
(129) Kuhrova, P.; Best, R.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banas, P. Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12*, 4534−4548.
(130) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.
(131) Réblová, K.; Lankaš, F.; Rázga, F.; Krasovska, M. V.; Koča, J.; Šponer, J. Structure, Dynamics, and Elasticity of Free 16s rRNA Helix 44 Studied by Molecular Dynamics Simulations. *Biopolymers* **2006**, *82*, 504−520.
(132) Zgarbová, M.; Jurečka, P.; Banáš, P.; Havrila, M.; Šponer, J.; Otyepka, M. Noncanonical α/γ Backbone Conformations in RNA and the Accuracy of Their Description by the AMBER Force Field. *J. Phys. Chem. B* **2017**, *121*, 2420−2433.
(133) Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E.; Sponer, J.; Otyepka, M. Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theory Comput.* **2010**, *6*, 3836−3849.
(134) Banas, P.; Sklenovsky, P.; Wedekind, J. E.; Sponer, J.; Otyepka, M. Molecular Mechanism of preQ(1) Riboswitch Action: A Molecular Dynamics Study. *J. Phys. Chem. B* **2012**, *116*, 12721−12734.
(135) Zgarbová, M.; Luque, F. J.; Šponer, J.; Otyepka, M.; Jurečka, P. A Novel Approach for Deriving Force Field Torsion Angle Parameters Accounting for Conformation-Dependent Solvation Effects. *J. Chem. Theory Comput.* **2012**, *8*, 3232−3242.
(136) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. Reparameterization of RNA chi Torsion Parameters for the AMBER Force Field and Comparison to NMR Spectra for Cytidine and Uridine. *J. Chem. Theory Comput.* **2010**, *6*, 1520−1531.
(137) Besseova, I.; Banas, P.; Kuhrova, P.; Kosinova, P.; Otyepka, M.; Sponer, J. Simulations of A-RNA Duplexes. The Effect of Sequence, Solute Force Field, Water Model, and Salt Concentration. *J. Phys. Chem. B* **2012**, *116*, 9899−9916.
(138) Yildirim, I.; Kennedy, S. D.; Stern, H. A.; Hart, J. M.; Kierzek, R.; Turner, D. H. Revision of AMBER Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iGiC Base Pairs. *J. Chem. Theory Comput.* **2012**, *8*, 172−181.
(139) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common RNA Force Fields. *RNA* **2015**, *21*, 1578−1590.
(140) Havrila, M.; Zgarbová, M.; Jurečka, P.; Banáš, P.; Krepl, M.; Otyepka, M.; Šponer, J. Microsecond-Scale MD Simulations of HIV-1 DIS Kissing-Loop Complexes Predict Bulged-In Conformation of the Bulged Bases and Reveal Interesting Differences between Available Variants of the AMBER RNA Force Fields. *J. Phys. Chem. B* **2015**, *119*, 15176−15190.
(141) Aytenfisu, A. H.; Spasic, A.; Grossfield, A.; Stern, H. A.; Mathews, D. H. Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 900−915.
(142) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C. ff14SB: Improving The Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(143) Gil-Ley, A.; Bottaro, S.; Bussi, G. Empirical Corrections to the Amber RNA Force Field with Target Metadynamics. *J. Chem. Theory Comput.* **2016**, *12*, 2790−2798.
(144) Cesari, A.; Gil-Ley, A.; Bussi, G. Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J. Chem. Theory Comput.* **2016**, *12*, 6192−6200.
(145) Wales, D. J.; Yildirim, I. Improving Computational Predictions of Single-Stranded RNA Tetramers with Revised α/γ Torsional Parameters for the Amber Force Field. *J. Phys. Chem. B* **2017**, *121*, 2989−2999.
(146) Zgarbova, M.; Javier-Luque, F.; Sponer, J.; Cheatham, T. E., III; Otyepka, M.; Jurecka, P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339−2354.
(147) Mlynsky, V.; Kuhrova, P.; Zgarbova, M.; Jurecka, P.; Walter, N. G.; Otyepka, M.; Sponer, J.; Banas, P. Reactive Conformation of the Active Site in the Hairpin Ribozyme Achieved by Molecular Dynamics Simulations with epsilon/zeta Force Field Reparametrizations. *J. Phys. Chem. B* **2015**, *119*, 4220−4229.
(148) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, *13*, 55−58.
(149) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E.; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar−Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723−5736.
(150) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405−4412.
(151) Darré, L.; Ivani, I.; Dans, P. D.; Gómez, H.; Hospital, A.; Orozco, M. Small Details Matter: The 2′-Hydroxyl as a Conformational Switch in RNA. *J. Am. Chem. Soc.* **2016**, *138*, 16355−16363.
(152) Yang, C.; Lim, M.; Kim, E.; Pak, Y. Predicting RNA Structures via a Simple van der Waals Correction to an All-Atom Force Field. *J. Chem. Theory Comput.* **2017**, *13*, 395−399.
(153) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. *PLoS Comput. Biol.* **2016**, *12*, No. e1004794.
(154) Šponer, J.; Krepl, M.; Banáš, P.; Kührová, P.; Zgarbová, M.; Jurečka, P.; Havrila, M.; Otyepka, M. How to Understand Atomistic Molecular Dynamics Simulations of RNA and Protein−RNA Complexes? *Wiley Interdiscip. Rev.: RNA* **2017**, *8*, No. e1405.
(155) Foloppe, N.; MacKerell, A. D. All-atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86−104.
(156) MacKerell, A. D.; Banavali, N. K. All-atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21*, 105−120.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(157) MacKerell, A. D.; Wiorkiewicz-Kuczera, J.; Karplus, M. An All-atom Empirical Energy Function for the Simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995**, *117*, 11946−11975.
(158) Van Wynsberghe, A. W.; Cui, Q. Comparison of Mode Analyses at Different Resolutions Applied to Nucleic Acid Systems. *Biophys. J.* **2005**, *89*, 2939−2949.
(159) Deng, N.-J.; Cieplak, P. Free Energy Profile of RNA Hairpins: A Molecular Dynamics Simulation Study. *Biophys. J.* **2010**, *98*, 627−636.
(160) Faustino, I.; Pérez, A.; Orozco, M. Toward a Consensus View of Duplex RNA Flexibility. *Biophys. J.* **2010**, *99*, 1876−1885.
(161) Lee, T. S.; Lopez, C. S.; Martick, M.; Scott, W. G.; York, D. M. Insight into the Role of Mg in Hammerhead Ribozyme Catalysis from X-ray Crystallography and Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2007**, *3*, 325−327.
(162) Martick, M.; Lee, T. S.; York, D. M.; Scott, W. G. Solvent Structure and Hammerhead Ribozyme Catalysis. *Chem. Biol.* **2008**, *15*, 332−342.
(163) Bottaro, S.; Banáš, P.; Sponer, J.; Bussi, G. Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J. Phys. Chem. Lett.* **2016**, *7*, 4032−4038.
(164) Gray, P. G.; Kish, L. Review: Survey Sampling. *J. R. Stat. Soc. Ser. A-G.* **1969**, *132*, 272−274.
(165) Dethoff, E. A.; Petzold, K.; Chugh, J.; Casiano-Negroni, A.; Al-Hashimi, H. M. Visualizing Transient Low-populated Structures of RNA. *Nature* **2012**, *491*, 724−728.
(166) Bokinsky, G.; Zhuang, X. Single-molecule RNA Folding. *Acc. Chem. Res.* **2005**, *38*, 566−573.
(167) Haller, A.; Souliére, M. F.; Micura, R. The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms. *Acc. Chem. Res.* **2011**, *44*, 1339−1348.
(168) Zhuang, X.; Kim, H.; Pereira, M. J.; Babcock, H. P.; Walter, N. G.; Chu, S. Correlating Structural Dynamics and Function in Single Ribozyme Molecules. *Science* **2002**, *296*, 1473−1476.
(169) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1799−1805.
(170) Šponer, J.; Bussi, G.; Stadlbauer, P.; Kührová, P.; Banáš, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of Guanine Quadruplex Molecules−funnel-like Mechanism or Kinetic Partitioning? An Overview from MD Simulation Studies. *Biochim. Biophys. Acta, Gen. Subj.* **2017**, *1861*, 1246−1263.
(171) Chen, S.-J.; Dill, K. A. RNA Folding Energy Landscapes. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 646−651.
(172) Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M. Exploring the Energy Landscape of a Small RNA Hairpin. *J. Am. Chem. Soc.* **2006**, *128*, 1523−1530.
(173) Zhang, Q.; Al-Hashimi, H. M. Extending the NMR Spatial Resolution Limit for RNA by Motional Couplings. *Nat. Methods* **2008**, *5*, 243−245.
(174) Chen, S. J. RNA Folding: Conformational Statistics, Folding Kinetics, and Ion Electrostatics. *Annu. Rev. Biophys.* **2008**, *37*, 197−214.
(175) Ditzler, M. A.; Otyepka, M.; Sponer, J.; Walter, N. G. Molecular Dynamics and Quantum Mechanics of RNA: Conformational and Chemical Change We Can Believe. *Acc. Chem. Res.* **2010**, *43*, 40−47.
(176) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular Dynamics Simulations of Nucleic Acids. From Tetranucleotides to the Ribosome. *J. Phys. Chem. Lett.* **2014**, *5*, 1771−1782.
(177) Hashem, Y.; Auffinger, P. A Short Guide for Molecular Dynamics Simulations of RNA Systems. *Methods* **2009**, *47*, 187−197.
(178) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulation. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.
(179) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16*, 163−199.
(180) Dellago, C.; Hummer, G. Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy* **2014**, *16*, 41−61.
(181) Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(182) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. In *Annual Review of Physical Chemistry*; Johnson, M. A., Martinez, T. J., Eds.; Annual Reviews: Palo Alto, CA, 2016; Vol. *67*, pp 159−184.
(183) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(184) Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L.-P.; Shukla, D.; et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2012**, *9*, 461−469.
(185) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, 19−25.
(186) Case, D. A.; B, R. M.; Botello-Smith, W.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 16*; University of California: San Francisco, CA, 2016.
(187) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.
(188) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542−1555.
(189) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184*, 374−380.
(190) Páll, S.; Hess, B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Comput. Phys. Commun.* **2013**, *184*, 2641−2650.
(191) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(192) Pan, A. C.; Weinreich, T. M.; Piana, S.; Shaw, D. E. Demonstrating an Order-of-magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J. Chem. Theory Comput.* **2016**, *13*, 1360−1367.
(193) Voelz, V. A.; Jager, M.; Zhu, L.; Yao, S. H.; Bakajin, O.; Weiss, S.; Lapidus, L. J.; Pande, V. S. Markov State Models of Millisecond Folder ACBP Reveals New Views of the Folding Reaction. *Biophys. J.* **2011**, *100*, 515.
(194) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-equilibrium Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19011−19016.
(195) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184−10189.
(196) Qiao, Q.; Bowman, G. R.; Huang, X. Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

That Potentially Seed Aggregation. *J. Am. Chem. Soc.* **2013**, *135*, 16092−16101.
(197) Shukla, D.; Meng, Y.; Roux, B.; Pande, V. S. Activation Pathway of Src Kinase Reveals Intermediate States as Targets for Drug Design. *Nat. Commun.* **2014**, *5*, No. e3397, DOI: 10.1038/ncomms4397.
(198) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-based Simulations on Google Exacycle Reveal Ligand Modulation of GPCR Activation Pathways. *Nat. Chem.* **2014**, *6*, 15−21.
(199) Sadiq, S. K.; Noé, F.; De Fabritiis, G. Kinetic Characterization of the Critical Step in HIV-1 Protease Maturation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20449−20454.
(200) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, No. e174105.
(201) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; Fabritiis, G. D.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139*, No. e015102.
(202) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(203) Bowman, G. R.; Pande, V. S.; Noé, F. An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation. *Adv. Exp. Med. Biol.* **2014**, *797*, 1.
(204) Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057−6069.
(205) Pérez-Hernández, G.; Noé, F. Hierarchical Time-Lagged Independent Component Analysis: Computing Slow Modes and Reaction Coordinates for Large Molecular Systems. *J. Chem. Theory Comput.* **2016**, *12*, 6118−6129.
(206) Naritomi, Y.; Fuchigami, S. Slow Dynamics in Protein Fluctuations Revealed by Time-structure Based Independent Component Analysis: The Case of Domain Motions. *J. Chem. Phys.* **2011**, *134*, No. e065101.
(207) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(208) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146−168.
(209) Deuflhard, P.; Weber, M. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra Appl.* **2005**, *398*, 161−184.
(210) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7*, 147−179.
(211) Noé, F.; Wu, H.; Prinz, J.-H.; Plattner, N. Projected and Hidden Markov models for Calculating Kinetics and Metastable States of Complex Molecules. *J. Chem. Phys.* **2013**, *139*, No. e184114.
(212) Sinitskiy, A. V.; Pande, V. S. Theoretical Restrictions on Longest Implicit Timescales in Markov State Models of Biomolecular Dynamics. e-Print archive, https://arxiv.org/abs/1708.03011 (accessed Nov 8, 2017).
(213) Doerr, S.; De Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2064−2069.
(214) Swope, W. C.; Pitera, J. W.; Suits, F. Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. *J. Phys. Chem. B* **2004**, *108*, 6571−6581.
(215) Thirumalai, D.; Klimov, D. K.; Woodson, S. A. Kinetic Partitioning Mechanism as a Unifying Theme in the Folding of Biomolecules. *Theor. Chem. Acc.* **1997**, *96*, 14−22.
(216) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. Theoretical Perspectives on Protein Folding. In *Annual Review of Biophysics*; Rees, D. C., Dill, K. A., Williamson, J. R., Eds.; Annual Reviews: Palo Alto, CA, 2010; Vol. *39*, pp 159−183.
(217) Guo, Z.; Thirumalai, D. Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways. *Biopolymers* **1995**, *36*, 83−102.
(218) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. K. Early Events in RNA Folding. *Annu. Rev. Phys. Chem.* **2001**, *52*, 751−762.
(219) Trendelkamp-Schroer, B.; Wu, H.; Paul, F.; Noé, F. Estimation and Uncertainty of Reversible Markov Models. *J. Chem. Phys.* **2015**, *143*, No. e174101.
(220) Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. Rapid Equilibrium Sampling Initiated from Nonequilibrium Data. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 19765−19769.
(221) DePaul, A. J.; Thompson, E. J.; Patel, S. S.; Haldeman, K.; Sorin, E. J. Equilibrium Conformational Dynamics in an RNA Tetraloop from Massively Parallel Molecular Dynamics. *Nucleic Acids Res.* **2010**, *38*, 4856−4867.
(222) Pinamonti, G.; Zhao, J.; Condon, D. E.; Paul, F.; Noè, F.; Turner, D. H.; Bussi, G. Predicting the Kinetics of RNA Oligonucleotides Using Markov State Models. *J. Chem. Theory Comput.* **2017**, *13*, 926−934.
(223) Bottaro, S.; Di Palma, F.; Bussi, G. The Role of Nucleobase Interactions in RNA Structure and Dynamics. *Nucleic Acids Res.* **2014**, *42*, 13306−13314.
(224) Pörschke, D. Molecular States in Single-stranded Adenylate Chains by Relaxation Analysis. *Biopolymers* **1978**, *17*, 315−323.
(225) Xu, X.; Yu, T.; Chen, S.-J. Understanding the Kinetic Mechanism of RNA Single Base Pair Formation. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 116−121.
(226) Sugita, Y.; Okamoto, Y. Replica-exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(227) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. others, Optimization by simmulated annealing. *Science* **1983**, *220*, 671−680.
(228) Sorin, E. J.; Engelhardt, M. A.; Herschlag, D.; Pande, V. S. RNA Simulations: Probing Hairpin Unfolding and the Dynamics of a GNRA Tetraloop. *J. Mol. Biol.* **2002**, *317*, 493−506.
(229) Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhys. Lett.* **1992**, *19*, 451−458.
(230) Park, S.; Pande, V. S. Choosing weights for simulated tempering. *Phys. Rev. E* **2007**, *76*, No. e016703, DOI: 10.1103/PhysRevE.76.016703.
(231) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(232) Hansmann, U. H. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **1997**, *281*, 140−150.
(233) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. Modified Replica Exchange Simulation Methods for Local Structure Refinement. *J. Phys. Chem. B* **2005**, *109*, 8220−8230.
(234) Zhuang, Z.; Jaeger, L.; Shea, J. E. Probing the Structural Hierarchy and Energy Landscape of an RNA T-loop Hairpin. *Nucleic Acids Res.* **2007**, *35*, 6995−7002.
(235) Deng, N.-J.; Cieplak, P. Molecular Dynamics and Free Energy Study of the Conformational Equilibria in the UUUU RNA Hairpin. *J. Chem. Theory Comput.* **2007**, *3*, 1435−1450.
(236) Villa, A.; Widjajakusuma, E.; Stock, G. Molecular Dynamics Simulation of the Structure, Dynamics, and Thermostability of the RNA Hairpins uCACGg and cUUCGg. *J. Phys. Chem. B* **2008**, *112*, 134−142.
(237) Garcia, A. E.; Paschek, D. Simulation of the Pressure and Temperature Folding/Unfolding Equilibrium of a Small RNA Hairpin. *J. Am. Chem. Soc.* **2008**, *130*, 815−817.
(238) Luckow, A.; Jha, S.; Kim, J.; Merzky, A.; Schnor, B. Adaptive Distributed Replica-exchange Simulations. *Philos. Trans. R. Soc., A* **2009**, *367*, 2595−2606.
(239) Vaiana, A. C.; Sanbonmatsu, K. Y. Stochastic Gating and Drug−Ribosome Interactions. *J. Mol. Biol.* **2009**, *386*, 648−661.
(240) Kührová, P.; Banáš, P.; Best, R. B.; Šponer, J.; Otyepka, M. Computer Folding of RNA Tetraloops? Are We There Yet? *J. Chem. Theory Comput.* **2013**, *9*, 2115−2125.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(241) Miner, J. C.; Chen, A. A.; García, A. E. Free-energy Landscape of a Hyperstable RNA Tetraloop. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 6665−6670.
(242) Bottaro, S.; Gil-Ley, A.; Bussi, G. RNA Folding Pathways in Stop Motion. *Nucleic Acids Res.* **2016**, *44*, 5883−5891.
(243) Zhang, Y.; Zhao, X.; Mu, Y. Conformational Transition Map of an RNA GCAA Tetraloop Explored by Replica-exchange Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2009**, *5*, 1146−1154.
(244) Zuo, G.; Li, W.; Zhang, J.; Wang, J.; Wang, W. Folding of a Small RNA Hairpin Based on Simulation with Replica Exchange Molecular Dynamics. *J. Phys. Chem. B* **2010**, *114*, 5835−5839.
(245) Beck, D. A. C.; White, G. W. N.; Daggett, V. Exploring the Energy Landscape of Protein Folding Using Replica-exchange and Conventional Molecular Dynamics Simulations. *J. Struct. Biol.* **2007**, *157*, 514−523.
(246) Sindhikara, D.; Meng, Y.; Roitberg, A. E. Exchange Frequency in Replica Exchange Molecular Dynamics. *J. Chem. Phys.* **2008**, *128*, No. e024103.
(247) Bussi, G. A Simple Asynchronous Replica-exchange Implementation. *Nuovo Cimento C* **2009**, *32*, 61−65.
(248) Yu, T.-Q.; Lu, J.; Abrams, C. F.; Vanden-Eijnden, E. Multiscale Implementation of Infinite-swap Replica Exchange Molecular Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 11744−11749.
(249) Kofke, D. A. On the Acceptance Probability of Replica-exchange Monte Carlo Trials. *J. Chem. Phys.* **2002**, *117*, 6911−6914.
(250) Trebst, S.; Troyer, M.; Hansmann, U. H. Optimized Parallel Tempering Simulations of Proteins. *J. Chem. Phys.* **2006**, *124*, No. e174903.
(251) Patriksson, A.; van der Spoel, D. A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073−2077.
(252) Prakash, M. K.; Barducci, A.; Parrinello, M. Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2025−2027.
(253) Stelzl, L. S.; Hummer, G. Kinetics from Replica Exchange Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 3927−3935.
(254) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(255) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007**, *3*, 26−41.
(256) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461−466.
(257) Grossfield, A.; Zuckerman, D. M. Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. *Annu. Rep. Comput. Chem.* **2009**, *5*, 23−48.
(258) Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117*, 4014−4027.
(259) Gil-Ley, A.; Bussi, G. Enhanced Conformational Sampling Using Replica Exchange with Collective-variable Tempering. *J. Chem. Theory Comput.* **2015**, *11*, 1077−1085.
(260) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104*, No. e190601, DOI: 10.1103/PhysRevLett.104.190601.
(261) Ballard, A. J.; Jarzynski, C. Replica Exchange with Non-equilibrium Switches. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12224−12229.
(262) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *J. Chem. Theory Comput.* **2012**, *8*, 2189−2192.
(263) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, No. e9058.
(264) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749−13754.
(265) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431−9438.
(266) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2*, 217−228.
(267) Bussi, G. Hamiltonian Replica Exchange in GROMACS: A Flexible Implementation. *Mol. Phys.* **2014**, *112*, 379−384.
(268) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919−11929.
(269) Roe, D. R.; Bergonzo, C.; Cheatham, T. E. Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *J. Phys. Chem. B* **2014**, *118*, 3543−3552.
(270) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11*, 3584−3595.
(271) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Swails, J. M.; Roitberg, A. E.; Cheatham, T. E. Multidimensional Replica Exchange Molecular Dynamics Yields a Converged Ensemble of an RNA Tetranucleotide. *J. Chem. Theory Comput.* **2014**, *10*, 492−499.
(272) Min, D.; Li, H.; Li, G.; Berg, B. A.; Fenley, M. O.; Yang, W. Efficient Sampling of Ion Motions in Molecular Dynamics Simulations on DNA: Variant Hamiltonian Replica Exchange Method. *Chem. Phys. Lett.* **2008**, *454*, 391−395.
(273) Curuksu, J.; Zacharias, M. Enhanced Conformational Sampling of Nucleic Acids by a New Hamiltonian Replica Exchange Molecular Dynamics Approach. *J. Chem. Phys.* **2009**, *130*, No. e104110.
(274) Simmerling, C.; Miller, J. L.; Kollman, P. A. Combined Locally Enhanced Sampling and Particle Mesh Ewald as a Strategy to Locate the Experimental Structure of a Nonhelical Nucleic Acid. *J. Am. Chem. Soc.* **1998**, *120*, 7149−7155.
(275) Cheng, X. L.; Hornak, V.; Simmerling, C. Improved Conformational Sampling Through an Efficient Combination of Mean-field Simulation Approaches. *J. Phys. Chem. B* **2004**, *108*, 426−437.
(276) Reblova, K.; Fadrna, E.; Sarzynska, J.; Kulinski, T.; Kulhanek, P.; Ennifar, E.; Koca, J.; Sponer, J. Conformations of Flanking Bases in HIV-1 RNA DIS Kissing Complexes Studied by Molecular Dynamics. *Biophys. J.* **2007**, *93*, 3932−3949.
(277) Koplin, J.; Mu, Y.; Richter, C.; Schwalbe, H.; Stock, G. Structure and Dynamics of an RNA Tetraloop: A Joint Molecular Dynamics and NMR Study. *Structure* **2005**, *13*, 1255−1267.
(278) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(279) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(280) Norberg, J.; Nilsson, L. Solvent Influence on Base Stacking. *Biophys. J.* **1998**, *74*, 394−402.
(281) Hart, K.; Nystrom, B.; Ohman, M.; Nilsson, L. Molecular Dynamics Simulations and Free Energy Calculations of Base Flipping in dsRNA. *RNA* **2005**, *11*, 609−618.
(282) Curuksu, J.; Sponer, J.; Zacharias, M. Elbow Flexibility of the Kt38 RNA Kink-Turn Motif Investigated by Free-Energy Molecular Dynamics Simulations. *Biophys. J.* **2009**, *97*, 2004−2013.
(283) Di Palma, F.; Bottaro, S.; Bussi, G. Kissing Loop Interaction in Adenine Riboswitch: Insights from Umbrella Sampling Simulations. *BMC Bioinf.* **2015**, *16*, 1−9.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(284) Allnér, O.; Nilsson, L.; Villa, A. Loop−loop Interaction in an Adenine-sensing Riboswitch: A Molecular Dynamics Study. *RNA* **2013**, *19*, 916−926.
(285) Sun, Z.; Wang, X.; Zhang, J. Z. H. Protonation-dependent Base Flipping in the Catalytic Triad of a Small RNA. *Chem. Phys. Lett.* **2017**, *684*, 239−244.
(286) Kästner, J.; Thiel, W. Bridging the Gap Between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella integration. *J. Chem. Phys.* **2005**, *123*, No. e144104.
(287) Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **1989**, *17*, 1217−1241.
(288) Neale, C.; Rodinger, T.; Pomes, R. Equilibrium Exchange Enhances the Convergence Rate of Umbrella Sampling. *Chem. Phys. Lett.* **2008**, *460*, 375−381.
(289) Zhu, F.; Hummer, G. Convergence and Error Estimation in Free Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **2012**, *33*, 453−465.
(290) Murata, K.; Sugita, Y.; Okamoto, Y. Free Energy Calculations for DNA Base Stacking by Replica-exchange Umbrella Sampling. *Chem. Phys. Lett.* **2004**, *385*, 1−7.
(291) Radak, B. K.; Romanus, M.; Lee, T.-S.; Chen, H.; Huang, M.; Treikalis, A.; Balasubramanian, V.; Jha, S.; York, D. M. Characterization of the Three-Dimensional Free Energy Manifold for the Uracil Ribonucleoside from Asynchronous Replica Exchange Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 373−377.
(292) Zeller, F.; Zacharias, M. Adaptive Biasing Combined with Hamiltonian Replica Exchange to Improve Umbrella Sampling Free Energy Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 703−710.
(293) Huang, M.; Giese, T. J.; Lee, T.-S.; York, D. M. Improvement of DNA and RNA Sugar Pucker Profiles from Semiempirical Quantum Methods. *J. Chem. Theory Comput.* **2014**, *10*, 1538−1545.
(294) Stofer, E.; Chipot, C.; Lavery, R. Free Energy Calculations of Watson-Crick Base Pairing in Aqueous Solution. *J. Am. Chem. Soc.* **1999**, *121*, 9503−9508.
(295) Grubmüller, H.; Heymann, B.; Tavan, P. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-biotin Rupture Force. *Science* **1996**, *271*, 997−999.
(296) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(297) Colizzi, F.; Bussi, G. RNA Unwinding from Reweighted Pulling Simulations. *J. Am. Chem. Soc.* **2012**, *134*, 5173−5179.
(298) Di Palma, F.; Colizzi, F.; Bussi, G. Ligand-induced Stabilization of the Aptamer Terminal Helix in the Add Adenine Riboswitch. *RNA* **2013**, *19*, 1517−1524.
(299) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. Targeted Molecular Dynamics Simulation of Conformational Change-application to the T ↔ R Transition in Insulin. *Mol. Simul.* **1993**, *10*, 291−308.
(300) Aci, S.; Mazier, S.; Genest, D. Conformational Pathway for the Kissing Complex−>Extended Dimer Transition of the SL1 Stem-loop from Genomic HIV-1 RNA as Monitored by Targeted Molecular Dynamics Techniques. *J. Mol. Biol.* **2005**, *351*, 520−530.
(301) Gore, J.; Ritort, F.; Bustamante, C. Bias and Error in Estimates of Equilibrium Free-energy Differences from Nonequilibrium Measurements. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12564−12569.
(302) Minh, D. D. L.; Adib, A. B. Optimized Free Energies from Bidirectional Single-molecule Force Spectroscopy. *Phys. Rev. Lett.* **2008**, *100*, No. e180602, DOI: 10.1103/PhysRevLett.100.180602.
(303) Do, T. N.; Carloni, P.; Varani, G.; Bussi, G. RNA/Peptide Binding Driven by Electrostatics—Insight from Bidirectional Pulling Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 1720−1730.
(304) Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.
(305) Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, No. e2893.
(306) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115*, 9169−9183.
(307) Laio, A.; Parrinello, M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(308) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, No. e020603, DOI: 10.1103/PhysRevLett.100.020603.
(309) Valsson, O.; Parrinello, M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys. Rev. Lett.* **2014**, *113*, No. e090601, DOI: 10.1103/PhysRevLett.113.090601.
(310) Babin, V.; Roland, C.; Sagui, C. Adaptively Biased Molecular Dynamics for Free Energy Calculations. *J. Chem. Phys.* **2008**, *128*, 134101.
(311) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247−2254.
(312) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(313) Haldar, S.; Kuhrova, P.; Banas, P.; Spiwok, V.; Sponer, J.; Hobza, P.; Otyepka, M. Insights into Stability and Folding of GNRA and UNCG Tetra loops Revealed by Microsecond Molecular Dynamics and Well-Tempered Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 3866−3877.
(314) Bussi, G.; Branduardi, D. Free-Energy Calculations with Metadynamics: Theory and Practice. In Reviews in Computational Chemistry Volume 28. *John Wiley & Sons, Inc* **2015**, 1−49.
(315) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. *J. Chem. Theory Comput.* **2014**, *10*, 1420−1425.
(316) Tiwary, P.; Berne, B. J. Spectral Gap Optimization of Order Parameters for Sampling Complex Molecular Systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2839−2844.
(317) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119*, 736−742.
(318) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A Portable Plugin for Free-energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(319) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(320) Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111*, 3345−3362.
(321) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533−3539.
(322) Hošek, P.; Toulcová, D.; Bortolato, A.; Spiwok, V. Altruistic Metadynamics: Multisystem Biased Simulation. *J. Phys. Chem. B* **2016**, *120*, 2209−2215.
(323) Mlýnský, V.; Bussi, G. Understanding In-line Probing Experiments by Modeling Cleavage of Nonreactive RNA Nucleotides. *RNA* **2017**, *23*, 712−720.
(324) White, A. D.; Dama, J. F.; Voth, G. A. Designing Free Energy Surfaces that Match Experimental Data with Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2451−2460.
(325) Marinelli, F.; Faraldo-Gómez, J. D. Ensemble-Biased Metadynamics: A Molecular Simulation Method to Sample Experimental Distributions. *Biophys. J.* **2015**, *108*, 2779−2782.
(326) Stadlbauer, P.; Mazzanti, L.; Cragnolini, T.; Wales, D. J.; Derreumaux, P.; Pasquali, S.; Šponer, J. Coarse-Grained Simulations Complemented by Atomistic Molecular Dynamics Provide New Insights into Folding and Unfolding of Human Telomeric G-Quadruplexes. *J. Chem. Theory Comput.* **2016**, *12*, 6077−6097.
(327) Geissler, P. L.; Dellago, C.; Chandler, D. Chemical Dynamics of the Protonated Water Trimer Analyzed by Transition Path Sampling. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1317−1322.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(328) Palazzesi, F.; Valsson, O.; Parrinello, M. Conformational Entropy as Collective Variable for Proteins. *J. Phys. Chem. Lett.* **2017**, *8*, 4752−4756.
(329) Sultan, M. M.; Pande, V. S. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017**, *13*, 2440−2447.
(330) Ferrarotti, M. J.; Bottaro, S.; Pérez-Villa, A.; Bussi, G. Accurate Multiple Time Step in Biased Molecular Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 139−146.
(331) Kabsch, W. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32*, 922−923.
(332) Parisien, M.; Cruz, J. A.; Westhof, E.; Major, F. New Metrics for Comparing and Assessing Discrepancies between RNA 3D Structures and Models. *RNA* **2009**, *15*, 1875−1885.
(333) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in Free Energy Space. *J. Chem. Phys.* **2007**, *126*, 054103.
(334) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197−2201.
(335) Cunha, R. A.; Bussi, G. Unravelling Mg2+-RNA Binding with Atomistic Molecular Dynamics. *RNA* **2017**, *23*, 628−638.
(336) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234−2246.
(337) Woo, H.-J.; Roux, B. Calculation of Absolute Protein−ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6825−6830.
(338) Allen, T. W.; Andersen, O. S.; Roux, B. Energetics of Ion Conduction through the Gramicidin Channel. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 117−122.
(339) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358−6363.
(340) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.
(341) Yang, C.; Kulkarni, M.; Lim, M.; Pak, Y. In silico direct folding of thrombin-binding aptamer G-quadruplex at all-atom level. *Nucleic Acids Res.* **2017**, DOI: 10.1093/nar/gkx1079.
(342) Piana, S.; Laio, A. A Bias-exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.
(343) Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2015**, *11*, 5062−5067.
(344) Marinelli, F.; Fabio, P.; Alessandro, L.; Stefano, P. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5*, 1−18.
(345) Camilloni, C.; Cavalli, A.; Vendruscolo, M. Replica-Averaged Metadynamics. *J. Chem. Theory Comput.* **2013**, *9*, 5610−5617.
(346) Borkar, A. N.; Bardaro, M. F.; Camilloni, C.; Aprile, F. A.; Varani, G.; Vendruscolo, M. Structure of a Low-population Binding Intermediate in Protein-RNA Recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 7171−7176.
(347) Borkar, A. N.; Vallurupalli, P.; Camilloni, C.; Kay, L. E.; Vendruscolo, M. Simultaneous NMR Characterisation of Multiple Minima in the Free Energy Landscape of an RNA UUCG Tetraloop. *Phys. Chem. Chem. Phys.* **2017**, *19*, 2797−2804.
(348) Camilloni, C.; Vendruscolo, M. A Tensor-Free Method for the Structural and Dynamical Refinement of Proteins using Residual Dipolar Couplings. *J. Phys. Chem. B* **2015**, *119*, 653−661.
(349) Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F. Multiensemble Markov Models of Molecular Thermodynamics and Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E3221−E3230.
(350) Stelzl, L. S.; Kells, A.; Rosta, E.; Hummer, G. Dynamic Histogram Analysis to Determine Free Energies and Rates from Biased Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 6328−6342.
(351) Sulc, P.; Romano, F.; Ouldridge, T. E.; Doye, J. P.; Louis, A. A. A Nucleotide-level Coarse-grained Model of RNA. *J. Chem. Phys.* **2014**, *140*, 235102.
(352) Matek, C.; Sulc, P.; Randisi, F.; Doye, J. P.; Louis, A. A. Coarse-grained Modelling of Supercoiled RNA. *J. Chem. Phys.* **2015**, *143*, 243122.
(353) Straatsma, T. P.; Berendsen, H. J. C. Free Energy of Ionic Hydration: Analysis of a Thermodynamic Integration Technique to Evaluate Free Energy Differences by Molecular Dynamics Simulations. *J. Chem. Phys.* **1988**, *89*, 5876−5886.
(354) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Alchemical Free Energy Calculations: Ready for Prime Time? In *Annual Reports in Computational Chemistry*; Spellmeyer, D. C., Wheeler, R., Eds.; Elsevier: New York, 2007; Vol. 3, Chapter 4, pp 41−59.
(355) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(356) Pearlman, D. A.; Kollman, P. A. The Lag between the Hamiltonian and the System Configuration in Free Energy Perturbation Calculations. *J. Chem. Phys.* **1989**, *91*, 7831−7839.
(357) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(358) Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(359) Mark, A. E. Free Energy Perturbation Calculations. *Encyclopedia of computational chemistry* **2002**, *2*, 1 DOI: 10.1002/0470845015.cfa010.
(360) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(361) Meng, Y.; Sabri Dashti, D.; Roitberg, A. E. Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2721−2727.
(362) Sakuraba, S.; Asai, K.; Kameda, T. Predicting RNA Duplex Dimerization Free-Energy Changes upon Mutations Using Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2015**, *6*, 4348−4351.
(363) Simonson, T. Free Energy of Particle Insertion. *Mol. Phys.* **1993**, *80*, 441−447.
(364) Beutler, T. C.; Mark, A. E.; Vanschaik, R. C.; Gerber, P. R.; Vangunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(365) Yildirim, I.; Stern, H. A.; Sponer, J.; Spackova, N.; Turner, D. H. Effects of Restrained Sampling Space and Nonplanar Amino Groups on Free-Energy Predictions for RNA with Imino and Sheared Tandem GA Base Pairs Flanked by GC, CG, iGiC or iCIG Base Pairs. *J. Chem. Theory Comput.* **2009**, *5*, 2088−2100.
(366) Sarzynska, J.; Nilsson, L.; Kulinski, T. Effects of Base Substitutions in an RNA Hairpin from Molecular Dynamics and Free Energy Simulations. *Biophys. J.* **2003**, *85*, 3445−3459.
(367) Krepl, M.; Otyepka, M.; Banas, P.; Sponer, J. Effect of Guanine to Inosine Substitution on Stability of Canonical DNA and RNA Duplexes: Molecular Dynamics Thermodynamics Integration Study. *J. Phys. Chem. B* **2013**, *117*, 1872−1879.
(368) Krepl, M.; Cléry, A.; Blatter, M.; Allain, F. H. T.; Sponer, J. Synergy between NMR Measurements and MD Simulations of Protein/RNA Complexes: Application to the RRMs, the Most Common RNA Recognition Motifs. *Nucleic Acids Res.* **2016**, *44*, 6452−6470.
(369) Krepl, M.; Blatter, M.; Cléry, A.; Damberger, F. F.; Allain, F. H. T.; Sponer, J. Structural Study of the Fox-1 RRM Protein Hydration Reveals a Role for Key Water Molecules in RRM-RNA Recognition. *Nucleic Acids Res.* **2017**, *45*, 8046−8063.
(370) Kleinjung, J.; Fraternali, F. Design and Application of Implicit Solvent Models in Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2014**, *25*, 126−134.
(371) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; et al. Calculating

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(372) Špačková, N.; Cheatham, T. E.; Ryjáček, F.; Lankaš, F.; van Meervelt, L.; Hobza, P.; Šponer, J. Molecular Dynamics Simulations and Thermodynamics Analysis of DNA−Drug Complexes. Minor Groove Binding between 4′,6-Diamidino-2-phenylindole and DNA Duplexes in Solution. *J. Am. Chem. Soc.* **2003**, *125*, 1759−1769.
(373) Islam, B.; Stadlbauer, P.; Neidle, S.; Haider, S.; Sponer, J. Can We Execute Reliable MM-PBSA Free Energy Computations of Relative Stabilities of Different Guanine Quadruplex Folds? *J. Phys. Chem. B* **2016**, *120*, 2899−2912.
(374) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449−461.
(375) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999−3094.
(376) Klamt, A. The COSMO and COSMO-RS Solvation Models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 699−709.
(377) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51*, 69−82.
(378) Khabiri, M.; Freddolino, P. L. Deficiencies in Molecular Dynamics Simulation-Based Prediction of Protein−DNA Binding Free Energy Landscapes. *J. Phys. Chem. B* **2017**, *121*, 5151−5161.
(379) Noid, W. G. Perspective: Coarse-grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139*, 090901.
(380) Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of Polymer Melts. I. Coarse-graining Procedure for Polycarbonates. *Acta Polym.* **1998**, *49*, 61−74.
(381) Müller-Plathe, F. Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back. *ChemPhysChem* **2002**, *3*, 754−769.
(382) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving Effective Mesoscale Potentials from Atomistic Simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(383) Lyubartsev, A. P.; Laaksonen, A. Calculation of Effective Interaction Potentials from Radial Distribution Functions: A Reverse Monte Carlo Approach. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 3730−3737.
(384) Lyubartsev, A. P.; Laaksonen, A. Osmotic and Activity Coefficients from Effective Potentials for Hydrated Ions. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1997**, *55*, 5689−5696.
(385) Sippl, M. J. Calculation of Conformational Ensembles from Potentials of Mean Force: An Approach to the Knowledge-based Prediction of Local Structures in Globular Proteins. *J. Mol. Biol.* **1990**, *213*, 859−883.
(386) Sippl, M. J. Boltzmann's Principle, Knowledge-based Mean Fields and Protein Folding. An Approach to the Computational Determination of Protein Structures. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 473−501.
(387) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.
(388) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.
(389) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the Representability Problem and the Physical Meaning of Coarse-grained Models. *J. Chem. Phys.* **2016**, *145*, 044108.
(390) Denesyuk, N. A.; Thirumalai, D. Coarse-grained Model for Predicting RNA Folding Thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901−4911.
(391) Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M. SimRNA: A Coarse-grained Method for RNA Folding Simulations and 3D Structure Prediction. *Nucleic Acids Res.* **2016**, *44*, e63.
(392) Dawson, W. K.; Maciejczyk, M.; Jankowska, E. J.; Bujnicki, J. M. Coarse-grained Modeling of RNA 3D Structure. *Methods* **2016**, *103*, 138−156.
(393) Cao, S.; Chen, S.-J. Predicting Structures and Stabilities for H-type Pseudoknots with Interhelix Loops. *RNA* **2009**, *15*, 696−706.
(394) Cao, S.; Chen, S.-J. Predicting RNA Folding Thermodynamics with a Reduced Chain Representation Model. *RNA* **2005**, *11*, 1884−1897.
(395) Cao, S.; Chen, S.-J. Predicting RNA Pseudoknot Folding Thermodynamics. *Nucleic Acids Res.* **2006**, *34*, 2634−2652.
(396) Cao, S.; Chen, S.-J. Structure and Stability of RNA/RNA Kissing Complex: With Application to HIV Dimerization Initiation Signal. *RNA* **2011**, *17*, 2130−2143.
(397) Cao, S.; Chen, S.-J. Physics-based de novo Prediction of RNA 3D Structures. *J. Phys. Chem. B* **2011**, *115*, 4216−4226.
(398) Denesyuk, N. A.; Thirumalai, D. Crowding Promotes the Switch from Hairpin to Pseudoknot Conformation in Human Telomerase RNA. *J. Am. Chem. Soc.* **2011**, *133*, 11858−11861.
(399) Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. Ab Initio RNA Folding by Discrete Molecular Dynamics: From Structure Prediction to Folding Mechanisms. *RNA* **2008**, *14*, 1164−1173.
(400) Cragnolini, T.; Laurin, Y.; Derreumaux, P.; Pasquali, S. Coarse-Grained HiRE-RNA Model for ab Initio RNA Folding beyond Simple Molecules, Including Noncanonical and Multiple Base Pairings. *J. Chem. Theory Comput.* **2015**, *11*, 3510−3522.
(401) Cragnolini, T.; Derreumaux, P.; Pasquali, S. Coarse-grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047−8060.
(402) Pasquali, S.; Derreumaux, P. HiRE-RNA: A High Resolution Coarse-grained Energy Model for RNA. *J. Phys. Chem. B* **2010**, *114*, 11957−11966.
(403) Jost, D.; Everaers, R. Prediction of RNA Multiloop and Pseudoknot Conformations from a Lattice-based, Coarse-grain Tertiary Structure Model. *J. Chem. Phys.* **2010**, *132*, 095101.
(404) He, Y.; Maciejczyk, M.; Oldziej, S.; Scheraga, H. A.; Liwo, A. Mean-field Interactions Between Nucleic-acid-base Dipoles Can Drive the Formation of a Double Helix. *Phys. Rev. Lett.* **2013**, *110*, No. e098101, DOI: 10.1103/PhysRevLett.110.098101.
(405) He, Y.; Liwo, A.; Scheraga, H. A. Optimization of a Nucleic Acids United-RESidue 2-Point model (NARES-2P) with a Maximum-likelihood Approach. *J. Chem. Phys.* **2015**, *143*, 243111.
(406) Mustoe, A. M.; Al-Hashimi, H. M.; Brooks, C. L., III Coarse Grained Models Reveal Essential Contributions of Topological Constraints to the Conformational Free Energy of RNA Bulges. *J. Phys. Chem. B* **2014**, *118*, 2615−2627.
(407) Bell, D. R.; Cheng, S. Y.; Salazar, H.; Ren, P. Capturing RNA Folding Free Energy with Coarse-Grained Molecular Dynamics Simulations. *Sci. Rep.* **2017**, *7*, 45812.
(408) Xia, Z.; Bell, D. R.; Shi, Y.; Ren, P. RNA 3D Structure Prediction by Using a Coarse-grained Model and Experimental Data. *J. Phys. Chem. B* **2013**, *117*, 3135−3144.
(409) Xia, Z.; Gardner, D. P.; Gutell, R. R.; Ren, P. Coarse-grained Model for Simulation of RNA Three-dimensional Structures. *J. Phys. Chem. B* **2010**, *114*, 13497−13506.
(410) Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, *11*, 3932−3945.
(411) Uusitalo, J. J.; Ingólfsson, H. I.; Marrink, S. J.; Faustino, I. Martini Coarse-Grained Force Field: Extension to RNA. *Biophys. J.* **2017**, *113*, 246−256.
(412) Malhotra, A.; Tan, R. K.; Harvey, S. C. Modeling Large RNAs and Ribonucleoprotein Particles Using Molecular Mechanics Techniques. *Biophys. J.* **1994**, *66*, 1777−1795.
(413) Tan, R. K. Z.; Petrov, A. S.; Harvey, S. C. YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models. *J. Chem. Theory Comput.* **2006**, *2*, 529−540.
(414) Jonikas, M. A.; Radmer, R. J.; Laederach, A.; Das, R.; Pearlman, S.; Herschlag, D.; Altman, R. B. Coarse-grained Modeling of Large

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

RNA Molecules with Knowledge-based Potentials and Structural Filters. *RNA* **2009**, *15*, 189−199.
(415) Kerpedjiev, P.; Höner zu Siederdissen, C.; Hofacker, I. L. Predicting RNA 3D Structure Using a Coarse-grain Helix-centered Model. *RNA* **2015**, *21*, 1110−1121.
(416) Kim, N.; Zahran, M.; Schlick, T. Computational Prediction of Riboswitch Tertiary Structures Including Pseudoknots by RAGTOP. *Methods Enzymol.* **2015**, *553*, 115−135.
(417) Bernauer, J.; Huang, X.; Sim, A. Y.; Levitt, M. Fully Differentiable Coarse-grained and All-atom Knowledge-based Potentials for RNA Structure Evaluation. *RNA* **2011**, *17*, 1066−1075.
(418) Li, J.; Zhang, J.; Wang, J.; Li, W.; Wang, W. Structure Prediction of RNA Loops with a Probabilistic Approach. *PLoS Comput. Biol.* **2016**, *12*, e1005032.
(419) Capriotti, E.; Norambuena, T.; Marti-Renom, M. A.; Melo, F. All-atom Knowledge-based Potential for RNA Structure Prediction and Assessment. *Bioinformatics* **2011**, *27*, 1086−1093.
(420) Poblete, S.; Bottaro, S.; Bussi, G. A Nucleobase-centered Coarse-grained Representation for Structure Prediction of RNA Motifs. *Nucleic Acids Res.* **2017**, DOI: 10.1093/nar/gkx1269.
(421) Setny, P.; Zacharias, M. Elastic Network Models of Nucleic Acids Flexibility. *J. Chem. Theory Comput.* **2013**, *9*, 5460−5470.
(422) Pinamonti, G.; Bottaro, S.; Micheletti, C.; Bussi, G. Elastic Network Models for RNA: A Comparative Assessment with Molecular Dynamics and SHAPE Experiments. *Nucleic Acids Res.* **2015**, *43*, 7260−7269.
(423) Bahar, I.; Jernigan, R. L. Vibrational Dynamics of Transfer RNAs: Comparison of the Free and Synthetase-bound Forms. *J. Mol. Biol.* **1998**, *281*, 871−884.
(424) Wang, Y.; Rader, A. J.; Bahar, I.; Jernigan, R. L. Global Ribosome Motions Revealed with Elastic Network Model. *J. Struct. Biol.* **2004**, *147*, 302−314.
(425) Kurkcuoglu, O.; Kurkcuoglu, Z.; Doruker, P.; Jernigan, R. L. Collective Dynamics of the Ribosomal Tunnel Revealed by Elastic Network Modeling. *Proteins: Struct., Funct., Genet.* **2009**, *75*, 837−845.
(426) Tama, F.; Valle, M.; Frank, J.; Brooks, C. L. Dynamic Reorganization of the Functionally Active Ribosome Explored by Normal Mode Analysis and Cryo-Electron Microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9319−9323.
(427) Wang, Y.; Jernigan, R. L. Comparison of tRNA Motions in the Free and Ribosomal Bound Structures. *Biophys. J.* **2005**, *89*, 3399−3409.
(428) Zimmermann, M. T.; Jernigan, R. L. Elastic Network Models Capture the Motions Apparent within Ensembles of RNA Structures. *RNA* **2014**, *20*, 792−804.
(429) Li, C.; Lv, D.; Zhang, L.; Yang, F.; Wang, C.; Su, J.; Zhang, Y. Approach to the Unfolding and Folding Dynamics of add A-riboswitch upon Adenine Dissociation Using a Coarse-grained Elastic Network Model. *J. Chem. Phys.* **2016**, *145*, 014104.
(430) González, R. L.; Teixidó, J.; Borrell, J. I.; Estrada-Tejedor, R., On the Applicability of Elastic Network Models for the Study of RNA CUG Trinucleotide Repeat Overexpansion. *PLoS One* **2016**, *11*, e0152049.
(431) Beššeová, I.; Réblová, K.; Leontis, N. B.; Šponer, J. Molecular Dynamics Simulations Suggest that RNA Three-way Junctions Can Act as Flexible RNA Structural Elements in the Ribosome. *Nucleic Acids Res.* **2010**, *38*, 6247−6264.
(432) Durand, P.; Trinquier, G.; Sanejouand, Y.-H. A New Approach for Determining Low-frequency Normal Modes in Macromolecules. *Biopolymers* **1994**, *34*, 759−771.
(433) Fuglebakk, E.; Reuter, N.; Hinsen, K. Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions. *J. Chem. Theory Comput.* **2013**, *9*, 5618−5628.
(434) Hori, N.; Takada, S. Coarse-Grained Structure-Based Model for RNA-Protein Complexes Developed by Fluctuation Matching. *J. Chem. Theory Comput.* **2012**, *8*, 3384−3394.
(435) Benítez, A. A.; Cifre, J. G. H.; Bańos, F. G. D.; de la Torre, J. G. Prediction of Solution Properties and Dynamics of RNAs by Means of Brownian Dynamics Simulation of Coarse-grained Models: Ribosomal 5S RNA and Phenylalanine Transfer RNA. *BMC Biophys.* **2015**, *8*, No. e11, DOI: 10.1186/s13628-015-0025-7.
(436) Tinoco, I.; Uhlenbeck, O. C.; Levine, M. D. Estimation of Secondary Structure in Ribonucleic Acids. *Nature* **1971**, *230*, 362−367.
(437) Serra, M. J.; Turner, D. H. Predicting Thermodynamic Properties of RNA. *Methods Enzymol.* **1995**, *259*, 242−261.
(438) Xia, T.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson−Crick Base Pairs. *Biochemistry* **1998**, *37*, 14719−14735.
(439) Mathews, D. H.; Disney, M. D.; Childs, J. L.; Schroeder, S. J.; Zuker, M.; Turner, D. H. Incorporating Chemical Modification Constraints into a Dynamic Programming Algorithm for Prediction of RNA Secondary Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7287−7292.
(440) Walter, A. E.; Turner, D. H. Sequence Dependence of Stability for Coaxial Stacking of RNA Helixes with Watson-Crick Base Paired Interfaces. *Biochemistry* **1994**, *33*, 12715−12719.
(441) Walter, A. E.; Turner, D. H.; Kim, J.; Lyttle, M. H.; Müller, P.; Mathews, D. H.; Zuker, M. Coaxial Stacking of Helixes Enhances Binding of Oligoribonucleotides and Improves Predictions of RNA Folding. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9218−9222.
(442) Lu, Z. J.; Turner, D. H.; Mathews, D. H. A Set of Nearest Neighbor Parameters for Predicting the Enthalpy Change of RNA Secondary Structure Formation. *Nucleic Acids Res.* **2006**, *34*, 4912−4924.
(443) Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **2003**, *31*, 3406−3415.
(444) Ueda, Y.; Taketomi, H.; Gō, N. Studies on Protein Folding, Unfolding, and Fluctuations by Computer Simulation. II. A. Three-dimensional Lattice Model of Lysozyme. *Biopolymers* **1978**, *17*, 1531−1548.
(445) Hills, R.; Brooks, C. Insights from Coarse-Grained Gō Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889−905.
(446) Hyeon, C.; Thirumalai, D. Mechanical Unfolding of RNA Hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6789−6794.
(447) Hyeon, C.; Thirumalai, D. Forced-Unfolding and Force-Quench Refolding of RNA Hairpins. *Biophys. J.* **2006**, *90*, 3410−3427.
(448) Hyeon, C.; Thirumalai, D. Mechanical Unfolding of RNA: From Hairpins to Structures with Internal Multiloops. *Biophys. J.* **2007**, *92*, 731−743.
(449) Lin, J.-C.; Thirumalai, D. Relative Stability of Helices Determines the Folding Landscape of Adenine Riboswitch Aptamers. *J. Am. Chem. Soc.* **2008**, *130*, 14080−14081.
(450) Manning, G. S. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions I. Colligative Properties. *J. Chem. Phys.* **1969**, *51*, 924−933.
(451) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.
(452) SantaLucia, J., Jr; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415−440.
(453) Daher, M.; Mustoe, A. M.; Morriss-Andrews, A.; Brooks Iii, C. L.; Walter, N. G. Tuning RNA Folding and Function through Rational Design of Junction Topology. *Nucleic Acids Res.* **2017**, *45*, 9706−9715.
(454) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.
(455) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.
(456) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(457) Bayrak, C. S.; Kim, N.; Schlick, T. Using Sequence Signatures and Kink-turn Motifs in Knowledge-based Statistical Potentials for RNA Structure Prediction. *Nucleic Acids Res.* **2017**, *45*, 5414−5422.
(458) Das, R.; Karanicolas, J.; Baker, D. Atomic Accuracy in Predicting and Designing Noncanonical RNA. *Nat. Methods* **2010**, *7*, 291−294.
(459) Magnus, M.; Boniecki, M. J.; Dawson, W.; Bujnicki, J. M. SimRNAweb: A Web Server for RNA 3D Structure Modeling with Optional Restraints. *Nucleic Acids Res.* **2016**, *44*, W315−W319.
(460) Frellsen, J.; Moltke, I.; Thiim, M.; Mardia, K. V.; Ferkinghoff-Borg, J.; Hamelryck, T. A Probabilistic Model of RNA Conformational Space. *PLoS Comput. Biol.* **2009**, *5*, e1000406.
(461) Krokhotin, A.; Houlihan, K.; Dokholyan, N. V. iFoldRNA v2: Folding RNA with Constraints. *Bioinformatics* **2015**, *31*, 2891−2893.
(462) Xu, X.; Zhao, P.; Chen, S.-J. Vfold: A web server for RNA Structure and Folding Thermodynamics Prediction. *PLoS One* **2014**, *9*, e107504.
(463) Cruz, J. A.; Blanchet, M.-F.; Boniecki, M.; Bujnicki, J. M.; Chen, S.-J.; Cao, S.; Das, R.; Ding, F.; Dokholyan, N. V.; Flores, S. C.; et al. RNA-Puzzles: A CASP-like Evaluation of RNA Three-dimensional Structure Prediction. *RNA* **2012**, *18*, 610−625.
(464) Miao, Z.; Adamiak, R. W.; Antczak, M.; Batey, R. T.; Becka, A. J.; Biesiada, M.; Boniecki, M. J.; Bujnicki, J.; Chen, S.-J.; Cheng, C. Y.; et al. RNA-Puzzles Round III: 3D RNA Structure Prediction of Five Riboswitches and One Ribozyme. *RNA* **2017**, *23*, 655−672.
(465) Miao, Z.; Adamiak, R. W.; Blanchet, M.-F.; Boniecki, M.; Bujnicki, J. M.; Chen, S.-J.; Cheng, C.; Chojnowski, G.; Chou, F.-C.; Cordero, P.; et al. RNA-Puzzles Round II: Assessment of RNA Structure Prediction Programs Applied to Three Large RNA Structures. *RNA* **2015**, *21*, 1066−1084.
(466) Flores, S. C.; Sherman, M. A.; Bruns, C. M.; Eastman, P.; Altman, R. B. Fast Flexible Modeling of RNA Structure Using Internal Coordinates. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2011**, *8*, 1247−1257.
(467) Parisien, M.; Major, F. The MC-Fold and MC-Sym Pipeline Infers RNA Structure from Sequence Data. *Nature* **2008**, *452*, 51−55.
(468) Sijenyi, F.; Saro, P.; Ouyang, Z.; Damm-Ganamet, K.; Wood, M.; Jiang, J.; SantaLucia, J. The RNA Folding Problems: Different Levels of sRNA Structure Prediction. In *RNA 3D Structure Analysis and Prediction*; Leontis, N., Westhof, E., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp 91−117.
(469) Rother, M.; Rother, K.; Puton, T.; Bujnicki, J. M. RNA Tertiary Structure Prediction with ModeRNA. *Briefings Bioinf.* **2011**, *12*, 601−613.
(470) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905−1908.
(471) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. Global Dynamics of Proteins: Bridging Between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.
(472) Fuglebakk, E.; Tiwari, S. P.; Reuter, N. Comparing the Intrinsic Dynamics of Multiple Protein Structures Using Elastic Network Models. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 911−922.
(473) Hinsen, K. Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins: Struct., Funct., Genet.* **1998**, *33*, 417−429.
(474) Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophys. J.* **2001**, *80*, 505−515.
(475) Micheletti, C.; Carloni, P.; Maritan, A. Accurate and Efficient Description of Protein Vibrational Dynamics: Comparing Molecular Dynamics and Gaussian Models. *Proteins: Struct., Funct., Genet.* **2004**, *55*, 635−645.
(476) Delarue, M.; Sanejouand, Y. H. Simplified Normal Mode Analysis of Conformational Transitions in DNA-dependent Polymerases: the Elastic Network Model. *J. Mol. Biol.* **2002**, *320*, 1011−1024.
(477) McGinnis, J. L.; Dunkle, J. A.; Cate, J. H.; Weeks, K. M. The Mechanisms of RNA SHAPE Chemistry. *J. Am. Chem. Soc.* **2012**, *134*, 6617−6624.
(478) Chu, J. W.; Izveko, S.; Voth, G. A. The Multiscale Challenge for Biomolecular Systems: Coarse-grained Modeling. *Mol. Simul.* **2006**, *32*, 211−218.
(479) Ermak, D. L.; McCammon, J. A. Brownian Dynamics with Hydrodynamic Interactions. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(480) Rotne, J.; Prager, S. Variational Treatment of Hydrodynamic Interaction in Polymers. *J. Chem. Phys.* **1969**, *50*, 4831−4837.
(481) Auffinger, P.; Westhof, E. Simulations of the Molecular Dynamics of Nucleic Acids. *Curr. Opin. Struct. Biol.* **1998**, *8*, 227−236.
(482) Hermann, T.; Auffinger, P.; Scott, W. G.; Westhof, E. Evidence for a Hydroxide Ion Bridging Two Magnesium Ions at the Active Site of the Hammerhead Ribozyme. *Nucleic Acids Res.* **1997**, *25*, 3421−3427.
(483) Sgrignani, J.; Magistrato, A. The Structural Role of Mg2+ Ions in a Class I RNA Polymerase Ribozyme: A Molecular Simulation Study. *J. Phys. Chem. B* **2012**, *116*, 2259−2268.
(484) Reblova, K.; Spackova, N.; Stefl, R.; Csaszar, K.; Koca, J.; Leontis, N. B.; Sponer, J. Non-Watson-Crick Basepairing and Hydration in RNA Motifs: Molecular Dynamics of 5S rRNA Loop E. *Biophys. J.* **2003**, *84*, 3564−3582.
(485) Auffinger, P.; Bielecki, L.; Westhof, E. Symmetric K+ and Mg2+ Ion-Binding Sites in the 5 S rRNA Loop E Inferred from Molecular Dynamics Simulations. *J. Mol. Biol.* **2004**, *335*, 555−571.
(486) Auffinger, P.; Bielecki, L.; Westhof, E. The Mg2+ Binding Sites of the 5S rRNA Loop E Motif as Investigated by Molecular Dynamics Simulations. *Chem. Biol.* **2003**, *10*, 551−561.
(487) Reblova, K.; Spackova, N.; Koca, J.; Leontis, N. B.; Sponer, J. Long-residency Hydration, Cation Binding, and Dynamics of Loop E/helix IV rRNA-L25 Protein Complex. *Biophys. J.* **2004**, *87*, 3397−3412.
(488) Casalino, L.; Palermo, G.; Abdurakhmonova, N.; Rothlisberger, U.; Magistrato, A. Development of Site-specific Mg2+-RNA Force Field Parameters: A Dream or Reality? Guidelines from Combined Molecular Dynamics and Quantum Mechanics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 340−352.
(489) Reblova, K.; Spackova, N.; Sponer, J. E.; Koca, J.; Sponer, J. Molecular Dynamics Simulations of RNA Kissing-Loop Motifs Reveal Structural Dynamics and Formation of Cation-Binding Pockets. *Nucleic Acids Res.* **2003**, *31*, 6942−6952.
(490) Bergonzo, C.; Hall, K. B.; Cheatham, T. E. Divalent Ion Dependent Conformational Changes in an RNA Stem-Loop Observed by Molecular Dynamics. *J. Chem. Theory Comput.* **2016**, *12*, 3382−3389.
(491) Bergonzo, C.; Hall, K. B.; Cheatham, T. E. Stem-Loop V of Varkud Satellite RNA Exhibits Characteristics of the Mg2+ Bound Structure in the Presence of Monovalent Ions. *J. Phys. Chem. B* **2015**, *119*, 12355−12364.
(492) Sun, L. Z.; Zhang, D.; Chen, S. J. Theory and Modeling of RNA Structure and Interactions with Metal Ions and Small Molecules. In *Annual Review of Biophysics*; Dill, K. A., Ed.; Annual Review: Palo Alto, CA, 2017; Vol. *46*, pp 227−246.
(493) Rangan, P.; Woodson, S. A. Structural Requirement for Mg2+ Binding in the Group I Intron Core. *J. Mol. Biol.* **2003**, *329*, 229−238.
(494) Draper, D. E.; Grilley, D.; Soto, A. M. Ions and RNA Folding. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 221−243.
(495) Woodson, S. A.; Ions, Metal; Folding, R. N. A. A Highly Charged Topic With a Dynamic Future. *Curr. Opin. Chem. Biol.* **2005**, *9*, 104−109.
(496) Chu, V. B.; Bai, Y.; Lipfert, J.; Herschlag, D.; Doniach, S. A Repulsive Field: Advances in the Electrostatics of the Ion Atmosphere. *Curr. Opin. Chem. Biol.* **2008**, *12*, 619−625.
(497) Auffinger, P.; Grover, N.; Westhof, E. Metal Ion Binding to RNA. In Structural and Catalytic Roles of Metal Ions in RNA. *Royal Society of Chemistry* **2011**, *Vol. 9*, 1−36.
(498) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in Charge: Magnesium Ions in RNA Folding and Catalysis. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262−272.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(499) Auffinger, P.; D'Ascenzo, L.; Ennifar, E. Sodium and Potassium Interactions with Nucleic Acids. In *The Alkali Metal Ions: Their Role for Life*; Sigel, A., Sigel, H., Sigel, O. R. K., Eds.; Springer International Publishing: Cham, 2016; pp 167−201.
(500) Heilman-Miller, S. L.; Thirumalai, D.; Woodson, S. A. Role of Counterion Condensation in Folding of the Tetrahymena Ribozyme. I. Equilibrium Stabilization by Cations. *J. Mol. Biol.* **2001**, *306*, 1157−1166.
(501) Heilman-Miller, S. L.; Pan, J.; Thirumalai, D.; Woodson, S. A. Role of Counterion Condensation in Folding of the Tetrahymena Ribozyme II. Counterion-dependence of Folding Kinetics. *J. Mol. Biol.* **2001**, *309*, 57−68.
(502) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic Acid-Ion Interactions. In *Annual Review of Biochemistry*; Kornberg, R. D., Ed.; Annual Reviews: Palo Alto, CA, 2014; Vol. *83*, pp 813−841.
(503) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Sponer, J. Cations and Hydration in Catalytic RNA: Molecular Dynamics of the Hepatitis Delta Virus Ribozyme. *Biophys. J.* **2006**, *91*, 626−638.
(504) Allner, O.; Nilsson, L.; Villa, A. Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493−1502.
(505) Bleuzen, A.; Pittet, P.-A.; Helm, L.; Merbach, A. E. Water Exchange on Magnesium(II) in Aqueous Solution: A Variable Temperature and Pressure 17O NMR Study. *Magn. Reson. Chem.* **1997**, *35*, 765−773.
(506) Cowan, J. A. Coordination Chemistry of Magnesium Ions and 5S rRNA (Escherichia coli): Binding Parameters, Ligand Symmetry, and Implications for Activity. *J. Am. Chem. Soc.* **1991**, *113*, 675−676.
(507) Krepl, M.; Reblova, K.; Koca, J.; Sponer, J. Bioinformatics and Molecular Dynamics Simulation Study of L1 Stalk Non-Canonical rRNA Elements: Kink-Turns, Loops, and Tetraloops. *J. Phys. Chem. B* **2013**, *117*, 5540−5555.
(508) Krepl, M.; Havrila, M.; Stadlbauer, P.; Banas, P.; Otyepka, M.; Pasulka, J.; Stefl, R.; Sponer, J. Can We Execute Stable Microsecond-Scale Atomistic Simulations of Protein-RNA Complexes? *J. Chem. Theory Comput.* **2015**, *11*, 1220−1243.
(509) Pan, F.; Roland, C.; Sagui, C. Ion Distributions Around Left- and Right-handed DNA and RNA Duplexes: A Comparative Study. *Nucleic Acids Res.* **2014**, *42*, 13981−13996.
(510) Nakano, S.-i.; Cerrone, A. L.; Bevilacqua, P. C. Mechanistic Characterization of the HDV Genomic Ribozyme: Classifying the Catalytic and Structural Metal Ion Sites within a Multichannel Reaction Mechanism. *Biochemistry* **2003**, *42*, 2982−2994.
(511) Nakano, S.-i.; Proctor, D. J.; Bevilacqua, P. C. Mechanistic Characterization of the HDV Genomic Ribozyme: Assessing the Catalytic and Structural Contributions of Divalent Metal Ions within a Multichannel Reaction Mechanism. *Biochemistry* **2001**, *40*, 12022−12038.
(512) Klein, D. J.; Moore, P. B.; Steitz, T. A. The Contribution of Metal Ions to the Structural Stability of the Large Ribosomal Subunit. *RNA* **2004**, *10*, 1366−1379.
(513) Leonarski, F.; D'Ascenzo, L.; Auffinger, P. Binding of Metals to Purine N7 Nitrogen Atoms and Implications for Nucleic Acids: A CSD Survey. *Inorg. Chim. Acta* **2016**, *452*, 82−89.
(514) Leonarski, F.; D'Ascenzo, L.; Auffinger, P. Mg2+ Ions: Do They Bind to Nucleobase Nitrogens? *Nucleic Acids Res.* **2017**, *45*, 987−1004.
(515) Murray, J. B.; Seyhan, A. A.; Walter, N. G.; Burke, J. M.; Scott, W. G. The Hammerhead, Hairpin and VS Ribozymes Are Catalytically Proficient In Monovalent Cations Alone. *Chem. Biol.* **1998**, *5*, 587−595.
(516) Fedor, M. J. Structure and Function of the Hairpin Ribozyme. *J. Mol. Biol.* **2000**, *297*, 269−291.
(517) Johnson-Buck, A. E.; McDowell, S. E.; Walter, N. G. Metal Ions: Supporting Actors in the Playbook of Small Ribozymes. *Metal ions in life sciences* **2011**, *9*, 175−196.
(518) Steitz, T. A.; Steitz, J. A. A General Two-metal-ion Mechanism for Catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6498−6502.
(519) Toor, N.; Keating, K. S.; Taylor, S. D.; Pyle, A. M. Crystal Structure of a Self-Spliced Group II Intron. *Science* **2008**, *320*, 77−82.
(520) Hsiao, C.; Williams, L. D. A Recurrent Magnesium-binding Motif Provides a Framework For the Ribosomal Peptidyl Transferase Center. *Nucleic Acids Res.* **2009**, *37*, 3134−3142.
(521) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. The Kink-Turn: A New RNA Secondary Structure Motif. *EMBO J.* **2001**, *20*, 4214−4221.
(522) Razga, F.; Koca, J.; Sponer, J.; Leontis, N. B. Hinge-Like Motions in RNA Kink-Turns: The Role of the Second A-Minor Motif and Nominally Unpaired Bases. *Biophys. J.* **2005**, *88*, 3466−3485.
(523) Matsumura, S.; Ikawa, Y.; Inoue, T. Biochemical Characterization of the Kink-turn RNA Motif. *Nucleic Acids Res.* **2003**, *31*, 5544−5551.
(524) Goody, T. A.; Melcher, S. E.; Norman, D. G.; Lilley, D. M. J. The Kink-Turn Motif in RNA is Dimorphic, and Metal Ion-Dependent. *RNA* **2004**, *10*, 254−264.
(525) Dallas, A.; Moore, P. B. The Loop E-loop D Region of Escherichia coli 5S rRNA: The Solution Structure Reveals an Unusual Loop That May Be Important For Binding Ribosomal Proteins. *Structure* **1997**, *5*, 1639−1653.
(526) Stoldt, M.; Wöhnert, J.; Ohlenschläger, O.; Görlach, M.; Brown, L. R. The NMR Structure of the 5S rRNA E-domain−protein L25 Complex Shows Preformed and Induced Recognition. *EMBO J.* **1999**, *18*, 6508−6521.
(527) Correll, C. C.; Freeborn, B.; Moore, P. B.; Steitz, T. A. Metals, Motifs, and Recognition in the Crystal Structure of a 5S rRNA Domain. *Cell* **1997**, *91*, 705−712.
(528) Correll, C. C.; Wool, I. G.; Munishkin, A. The Two Faces of the Escherichia coli 23 S rRNA Sarcin/ricin Domain: The Structure at 1.11 Å Resolution. *J. Mol. Biol.* **1999**, *292*, 275−287.
(529) Szewczak, A. A.; Moore, P. B. The Sarcin/Ricin Loop, a Modular RNA. *J. Mol. Biol.* **1995**, *247*, 81−98.
(530) Spackova, N.; Sponer, J. Molecular Dynamics Simulations of Sarcin-ricin rRNA Motif. *Nucleic Acids Res.* **2006**, *34*, 697−708.
(531) Orlov, S. N.; Hamet, P. Intracellular Monovalent Ions as Second Messengers. *J. Membr. Biol.* **2006**, *210*, 161−172.
(532) Ku, D.; Akera, T.; Tobin, T.; Brody, T. M. Effects of Monovalent Cations on Cardiac Na+, K+-ATPase Activity and on Contractile Force. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1975**, *290*, 113−131.
(533) Hurwitz, C.; Rosano, C. L. The Intracellular Concentration of Bound and Unbound Magnesium Ions in Escherichia coli. *J. Biol. Chem.* **1967**, *242*, 3719−3722.
(534) Vink, R.; McIntosh, T. K.; Demediuk, P.; Weiner, M. W.; Faden, A. I. Decline in Intracellular Free Mg2+ is Associated with Irreversible Tissue Injury after Brain Trauma. *J. Biol. Chem.* **1988**, *263*, 757−761.
(535) Sun, L.-Z.; Chen, S.-J. Monte Carlo Tightly Bound Ion Model: Predicting Ion-Binding Properties of RNA with Ion Correlations and Fluctuations. *J. Chem. Theory Comput.* **2016**, *12*, 3370−3381.
(536) Denesyuk, N. A.; Thirumalai, D. How do Metal Ions Direct Ribozyme Folding? *Nat. Chem.* **2015**, *7*, 793−801.
(537) Šponer, J.; Sabat, M.; Gorb, L.; Leszczynski, J.; Lippert, B.; Hobza, P. The Effect of Metal Binding to the N7 Site of Purine Nucleotides on Their Structure, Energy, and Involvement in Base Pairing. *J. Phys. Chem. B* **2000**, *104*, 7535−7544.
(538) Gkionis, K.; Kruse, H.; Platts, J. A.; Mladek, A.; Koca, J.; Sponer, J. Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 1326−1340.
(539) Muñoz, J.; Sponer, J.; Hobza, P.; Orozco, M.; Luque, F. J. Interactions of Hydrated Mg2+ Cation with Bases, Base Pairs, and Nucleotides. Electron Topology, Natural Bond Orbital, Electrostatic, and Vibrational Study. *J. Phys. Chem. B* **2001**, *105*, 6051−6060.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(540) Gresh, N.; Šponer, J. E.; Špačková, N. a.; Leszczynski, J.; Šponer, J., Theoretical Study of Binding of Hydrated Zn(II) and Mg(II) Cations to 5'-Guanosine Monophosphate. Toward Polarizable Molecular Mechanics for DNA and RNA. *J. Phys. Chem. B* **2003**, *107*, 8669−8681.
(541) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand−Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.
(542) Li, P.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564−1686.
(543) Babu, C. S.; Dudev, T.; Casareno, R.; Cowan, J. A.; Lim, C. A Combined Experimental and Theoretical Study of Divalent Metal Ion Selectivity and Function in Proteins: Application to E. coli Ribonuclease H1. *J. Am. Chem. Soc.* **2003**, *125*, 9318−9328.
(544) Rulíšek, L.; Šponer, J. Outer-Shell and Inner-Shell Coordination of Phosphate Group to Hydrated Metal Ions (Mg2+, Cu2+, Zn2+, Cd2+) in the Presence and Absence of Nucleobase. The Role of Nonelectrostatic Effects. *J. Phys. Chem. B* **2003**, *107*, 1913−1923.
(545) Šponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. Interaction between the Guanine−Cytosine Watson−Crick DNA Base Pair and Hydrated Group IIa (Mg2+, Ca2+, Sr2+, Ba2+) and Group IIb (Zn2+, Cd2+, Hg2+) Metal Cations. *J. Phys. Chem. A* **1998**, *102*, 5951−5957.
(546) Mackerell, A. D. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584−1604.
(547) Lamoureux, G.; Roux, B. t., Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(548) Bock, C. W.; Katz, A. K.; Markham, G. D.; Glusker, J. P. Manganese as a Replacement for Magnesium and Zinc: Functional Comparison of the Divalent Ions. *J. Am. Chem. Soc.* **1999**, *121*, 7360−7372.
(549) Lamoureux, G.; Roux, B. Absolute Hydration Free Energy Scale for Alkali and Halide Ions Established from Simulations with a Polarizable Force Field. *J. Phys. Chem. B* **2006**, *110*, 3308−3322.
(550) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125*, 15671−15682.
(551) Perera, L.; Berkowitz, M. L. Many-body Effects in Molecular Dynamics Simulations of Na+(H2O)n and Cl−(H2O)n Clusters. *J. Chem. Phys.* **1991**, *95*, 1954−1963.
(552) Sponer, J.; Spackova, N. Molecular Dynamics Simulations and Their Application to Four-stranded DNA. *Methods* **2007**, *43*, 278−290.
(553) Gresh, N.; Naseem-Khan, S.; Lagardère, L.; Piquemal, J.-P.; Sponer, J. E.; Sponer, J. Channeling through Two Stacked Guanine Quartets of One and Two Alkali Cations in the Li+, Na+, K+, and Rb+ Series. Assessment of the Accuracy of the SIBFA Anisotropic Polarizable Molecular Mechanics Potential. *J. Phys. Chem. B* **2017**, *121*, 3997−4014.
(554) Spackova, N.; Berger, I.; Sponer, J. Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules. *J. Am. Chem. Soc.* **1999**, *121*, 5519−5534.
(555) Anwander, E. H. S.; Probst, M. M.; Rode, B. M. The Influence of Li+, Na+, Mg2+, Ca2+, and Zn2+ Ions on the Hydrogen Bonds of the Watson−Crick Base Pairs. *Biopolymers* **1990**, *29*, 757−769.
(556) Sigel, R. K. O.; Freisinger, E.; Lippert, B. Effects of N7-methylation, N7-platination, and C8-hydroxylation of Guanine on H-bond Formation with Cytosine: Platinum Coordination Strengthens the Watson-Crick Pair. *JBIC, J. Biol. Inorg. Chem.* **2000**, *5*, 287−299.
(557) K. O. Sigel, R.; Lippert, B. PtII Coordination to Guanine-N7: Enhancement of the Stability of the Watson-Crick Base Pair with Cytosine. *Chem. Commun.* **1999**, *0*, 2167−2168.
(558) Nakano, S.-i.; Fujimoto, M.; Hara, H.; Sugimoto, N. Nucleic Acid Duplex Stability: Influence of Base Composition on Cation Effects. *Nucleic Acids Res.* **1999**, *27*, 2957−2965.
(559) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 774−786.
(560) Li, P.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10*, 289−297.
(561) Li, P.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12−6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883−895.
(562) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Force Field for Mg2+, Mn2+, Zn2+, and Cd2+ Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **2015**, *119*, 15460−15470.
(563) Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Åqvist, J.; Kamerlin, S. C. L. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Phys. Chem. B* **2014**, *118*, 4351−4362.
(564) Oelschlaeger, P.; Klahn, M.; Beard, W. A.; Wilson, S. H.; Warshel, A. Magnesium-cationic Dummy Atom Molecules Enhance Representation of DNA Polymerase β in Molecular Dynamics Simulations: Improved Accuracy in Studies of Structural Features and Mutational Effects. *J. Mol. Biol.* **2007**, *366*, 687−701.
(565) Jiang, Y.; Zhang, H.; Tan, T. Rational Design of Methodology-Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Chem. Theory Comput.* **2016**, *12*, 3250−3260.
(566) Saxena, A.; García, A. E. Multisite Ion Model in Concentrated Solutions of Divalent Cations (MgCl2 and CaCl2): Osmotic Pressure Calculations. *J. Phys. Chem. B* **2015**, *119*, 219−227.
(567) Fyta, M.; Kalcher, I.; Dzubiella, J.; Vrbka, L.; Netz, R. R. Ionic Force Field Optimization Based on Single-ion and Ion-pair Solvation Properties. *J. Chem. Phys.* **2010**, *132*, 024911.
(568) Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. Spontaneous Formation of KCl Aggregates in Biomolecular Simulations: A Force Field Issue? *J. Chem. Theory Comput.* **2007**, *3*, 1851−1859.
(569) Ponomarev, S. Y.; Thayer, K. M.; Beveridge, D. L. Ion Motions in Molecular Dynamics Simulations on DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14771−14775.
(570) Várnai, P.; Zakrzewska, K. DNA and Its Counterions: A Molecular Dynamics Study. *Nucleic Acids Res.* **2004**, *32*, 4269−4280.
(571) Lemkul, J. A.; Lakkaraju, S. K.; MacKerell, A. D. Characterization of Mg2+ Distributions around RNA in Solution. *ACS Omega* **2016**, *1*, 680−688.
(572) Aqvist, J. Ion Water Interaction Potentials Derived from Free-Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021−8024.
(573) Kumar, P. K. R.; Kumarevel, T.; Mizuno, H. Structural Basis of HutP-mediated Transcription Anti-termination. *Curr. Opin. Struct. Biol.* **2006**, *16*, 18−26.
(574) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for + 2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733−2748.
(575) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(576) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(577) Chen, A. A.; Pappu, R. V. Parameters of Monovalent Ions in the AMBER-99 Forcefield: Assessment of Inaccuracies and Proposed Improvements. *J. Phys. Chem. B* **2007**, *111*, 11884−11887.
(578) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg2+ Ion Models Commonly Used in Biomolecular Simulations. *J. Comput. Chem.* **2015**, *36*, 970−982.
(579) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(580) Chen, A. A.; Draper, D. E.; Pappu, R. V. Molecular Simulation Studies of Monovalent Counterion-Mediated Interactions in a Model RNA Kissing Loop. *J. Mol. Biol.* **2009**, *390*, 805−819.
(581) Chen, A. A.; Marucho, M.; Baker, N. A.; Pappu, R. V. Simulations of RNA Interactions with Monovalent Ions. *Methods in Enzymology*; Academic Press: New York, 2009; Vol. *469*, Chapter 20, pp 411−432.
(582) Kirmizialtin, S.; Pabit, Suzette A.; Meisburger, Steve P.; Pollack, L.; Elber, R. RNA and Its Ionic Cloud: Solution Scattering Experiments and Atomically Detailed Simulations. *Biophys. J.* **2012**, *102*, 819−828.
(583) Kirmizialtin, S.; Elber, R. Computational Exploration of Mobile Ion Distributions around RNA Duplex. *J. Phys. Chem. B* **2010**, *114*, 8207−8220.
(584) Siegfried, N. A.; Kierzek, R.; Bevilacqua, P. C. Role of Unsatisfied Hydrogen Bond Acceptors in RNA Energetics and Specificity. *J. Am. Chem. Soc.* **2010**, *132*, 5342−5344.
(585) Hünenberger, P. H.; McCammon, J. A. Ewald Artifacts in Computer Simulations of Ionic Solvation and Ion−ion Interaction: A Continuum Electrostatics Study. *J. Chem. Phys.* **1999**, *110*, 1856−1872.
(586) Rook, M. S.; Treiber, D. K.; Williamson, J. R. An Optimal Mg2+ Concentration for Kinetic Folding of the Tetrahymena Ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 12471−12476.
(587) Islam, B.; Stadlbauer, P.; Krepl, M.; Koca, J.; Neidle, S.; Haider, S.; Sponer, J. Extended Molecular Dynamics of a C-kit Promoter Quadruplex. *Nucleic Acids Res.* **2015**, *43*, 8673−8693.
(588) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular Dynamics Simulations of Solutions at Constant Chemical Potential. *J. Chem. Phys.* **2015**, *142*, 144113.
(589) Fiala, R.; Špačková, N.; Foldynová-Trantírková, S.; Šponer, J.; Sklenář, V.; Trantírek, L. NMR Cross-Correlated Relaxation Rates Reveal Ion Coordination Sites in DNA. *J. Am. Chem. Soc.* **2011**, *133*, 13790−13793.
(590) Sychrovský, V.; Šponer, J.; Hobza, P. Theoretical Calculation of the NMR Spin−Spin Coupling Constants and the NMR Shifts Allow Distinguishability between the Specific Direct and the Water-Mediated Binding of a Divalent Metal Cation to Guanine. *J. Am. Chem. Soc.* **2004**, *126*, 663−672.
(591) Kowerko, D.; König, S. L. B.; Skilandat, M.; Kruschel, D.; Hadzic, M. C. A. S.; Cardo, L.; Sigel, R. K. O. Cation-induced Kinetic Heterogeneity of the Intron−exon Recognition in Single Group II Introns. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 3403−3408.
(592) Auffinger, P.; Bielecki, L.; Westhof, E. Anion Binding to Nucleic Acids. *Structure* **2004**, *12*, 379−388.
(593) Ferre-D'Amare, A. R.; Zhou, K.; Doudna, J. A. Crystal Structure of a Hepatitis Delta Virus Ribozyme. *Nature* **1998**, *395*, 567−574.
(594) Ke, A.; Zhou, K.; Ding, F.; Cate, J. H. D.; Doudna, J. A. A Conformational Switch Controls Hepatitis Delta Virus Ribozyme Catalysis. *Nature* **2004**, *429*, 201−205.
(595) D'Ascenzo, L.; Auffinger, P. Anions in Nucleic Acid Crystallography. In *Nucleic Acid Crystallography: Methods and Protocols*; Ennifar, E., Ed.; Springer: New York, 2016; pp 337−351.
(596) Ennifar, E.; Walter, P.; Dumas, P. A Crystallographic Study of the Binding of 13 Metal Ions to Two Related RNA Duplexes. *Nucleic Acids Res.* **2003**, *31*, 2671−2682.
(597) Ennifar, E.; Walter, P.; Ehresmann, B.; Ehresmann, C.; Dumas, P. Crystal Structures of Coaxially Stacked Kissing Complexes of the HIV-1 RNA Dimerization Initiation Site. *Nat. Struct. Biol.* **2001**, *8*, 1064−1068.
(598) Ennifar, E.; Dumas, P. Polymorphism of Bulged-out Residues in HIV-1 RNA DIS Kissing Complex and Structure Comparison with Solution Studies. *J. Mol. Biol.* **2006**, *356*, 771−782.
(599) Ke, A.; Ding, F.; Batchelor, J. D.; Doudna, J. A. Structural Roles of Monovalent Cations in the HDV Ribozyme. *Structure* **2007**, *15*, 281−287.
(600) Zheng, H.; Shabalin, I. G.; Handing, K. B.; Bujnicki, J. M.; Minor, W. Magnesium-binding Architectures in RNA Crystal Structures: Validation, Binding Preferences, Classification and Motif Detection. *Nucleic Acids Res.* **2015**, *43*, 3789−3801.
(601) Jacobson, D. R.; Saleh, O. A. Quantifying the Ion Atmosphere of Unfolded, Single-stranded Nucleic Acids Using Equilibrium Dialysis and Single-molecule Methods. *Nucleic Acids Res.* **2016**, *44*, 3763−3771.
(602) Giambaşu, George M.; Luchko, T.; Herschlag, D.; York, Darrin M.; Case, David A Ion Counting from Explicit-Solvent Simulations and 3D-RISM. *Biophys. J.* **2014**, *106*, 883−894.
(603) Razga, F.; Zacharias, M.; Reblova, K.; Koca, J.; Sponer, J. RNA Kink-Turns as Molecular Elbows: Hydration, Cation Binding, and Large-Scale Dynamics. *Structure* **2006**, *14*, 825−835.
(604) Sklenovsky, P.; Florova, P.; Banas, P.; Reblova, K.; Lankas, F.; Otyepka, M.; Sponer, J. Understanding RNA Flexibility Using Explicit Solvent Simulations: The Ribosomal and Group I Intron Reverse Kink-Turn Motifs. *J. Chem. Theory Comput.* **2011**, *7*, 2963−2980.
(605) Kuehrova, P.; Otyepka, M.; Sponer, J.; Banas, P. Are Waters around RNA More than Just a Solvent? - An Insight from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 401−411.
(606) Bergonzo, C.; Cheatham, T. E. Mg2+ Binding Promotes SLV as a Scaffold in Varkud Satellite Ribozyme SLI-SLV Kissing Loop Junction. *Biophys. J.* **2017**, *113*, 313−320.
(607) Campbell, D. O.; Bouchard, P.; Desjardins, G.; Legault, P. NMR Structure of Varkud Satellite Ribozyme Stem−Loop V in the Presence of Magnesium Ions and Localization of Metal-Binding Sites. *Biochemistry* **2006**, *45*, 10591−10605.
(608) Hayatshahi, H. S.; Bergonzo, C. Cheatham Iii, T. E., Investigating the Ion Dependence of the First Unfolding Step of GTPase-Associating Center Ribosomal RNA. *J. Biomol. Struct. Dyn.* **2017**, 1−11.
(609) Hayatshahi, H. S.; Roe, D. R.; Galindo-Murillo, R.; Hall, K. B.; Cheatham, T. E. Computational Assessment of Potassium and Magnesium Ion Binding to a Buried Pocket in GTPase-Associating Center RNA. *J. Phys. Chem. B* **2017**, *121*, 451−462.
(610) Drozdetski, A. V.; Tolokh, I. S.; Pollack, L.; Baker, N.; Onufriev, A. V. Opposing Effects of Multivalent Ions on the Flexibility of DNA and RNA. *Phys. Rev. Lett.* **2016**, *117*, No. e028101, DOI: 10.1103/PhysRevLett.117.028101.
(611) Siegbahn, P. E. M.; Blomberg, M. R. A. Transition-Metal Systems in Biochemistry Studied by High-Accuracy Quantum Chemical Methods. *Chem. Rev.* **2000**, *100*, 421−438.
(612) Cramer, C. J.; Truhlar, D. G. Density Functional Theory for Transition Metals and Transition Metal Chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757−10816.
(613) Blomberg, M. R. A.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. M. Quantum Chemical Studies of Mechanisms for Metalloenzymes. *Chem. Rev.* **2014**, *114*, 3601−3658.
(614) Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H. Benchmarking AMBER Force Fields for RNA: Comparisons to NMR Spectra for Single-Stranded r(GACC) Are Improved by Revised chi Torsions. *J. Phys. Chem. B* **2011**, *115*, 9261−9270.
(615) Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H. The Nuclear Magnetic Resonance of CCCC RNA Reveals a Right-handed Helix, and Revised Parameters for AMBER Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry* **2013**, *52*, 996−1010.
(616) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H. Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **2015**, *11*, 2729−2742.
(617) Zagrovic, B.; Van Gunsteren, W. F. Comparing Atomistic Simulation Data with the NMR Experiment: How Much Can NOEs Actually Tell Us? *Proteins* **2006**, *63*, 210−218.
(618) Best, R. B.; Lindorff-Larsen, K.; DePristo, M. A.; Vendruscolo, M. Relation Between Native Ensembles and Experimental Structures of Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10901−10906.
(619) Schrodt, M. V.; Andrews, C. T.; Elcock, A. H. Large-Scale Analysis of 48 DNA and 48 RNA Tetranucleotides Studied by 1 μs Explicit-Solvent Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2015**, *11*, 5906−5917.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(620) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(621) Woese, C. R.; Winker, S.; Gutell, R. R. Architecture of Ribosomal-RNA - Constraints on the Sequence of Tetra-Loops. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8467−8471.
(622) Thapar, R.; Denmon, A. P.; Nikonowicz, E. P. Recognition Modes of RNA Tetraloops and Tetraloop-like Motifs by RNA-binding Proteins. *Wiley Interdiscip. Rev.: RNA* **2014**, *5*, 49−67.
(623) Svoboda, P.; Cara, A. D. Hairpin RNA: A Secondary Structure of Primary Importance. *Cell. Mol. Life Sci.* **2006**, *63*, 901−908.
(624) Uhlenbeck, O. C.; Tetraloops; Folding, R. N. A. *Nature* **1990**, *346*, 613−614.
(625) Quigley, G. J.; Rich, A. Structural Domains of Transfer RNA Molecules. *Science* **1976**, *194*, 796−806.
(626) Jucker, F. M.; Pardi, A. GNRA Tetraloops Make a U-turn. *RNA* **1995**, *1*, 219−222.
(627) Pley, H. W.; Flaherty, K. M.; McKay, D. B. Three-dimensional Structure of a Hammerhead Ribozyme. *Nature* **1994**, *372*, 68−74.
(628) Stallings, S. C.; Moore, P. B. The Structure of an Essential Splicing Element: Stem Loop IIa from Yeast U2 snRNA. *Structure* **1997**, *5*, 1173−1185.
(629) Puglisi, E. V.; Puglisi, J. D. HIV-1 A-rich RNA Loop Mimics the tRNA Anticodon Structure. *Nat. Struct. Biol.* **1998**, *5*, 1033−1036.
(630) Puglisi, J.; Tan, R.; Calnan, B.; Frankel, A. Williamson, Conformation of the TAR RNA-arginine Complex by NMR Spectroscopy. *Science* **1992**, *257*, 76−80.
(631) Varani, G. Exceptionally Stable Nucleic-Acid Hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 379−404.
(632) Chauhan, S.; Woodson, S. A. Tertiary Interactions Determine the Accuracy of RNA Folding. *J. Am. Chem. Soc.* **2008**, *130*, 1296−1303.
(633) Cheong, C.; Varani, G.; Tinoco, I. Solution Structure of an Unusually Stable RNA Hairpin, 5GGAC (UUCG) GUCC. *Nature* **1990**, *346*, 680−682.
(634) Varani, G.; Cheong, C.; Tinoco, I., Jr Structure of an Unusually Stable RNA Hairpin. *Biochemistry* **1991**, *30*, 3280−3289.
(635) Hyeon, C.; Thirumalai, D. Multiple Probes Are Required to Explore and Control the Rugged Energy Landscape of RNA Hairpins. *J. Am. Chem. Soc.* **2008**, *130*, 1538−1539.
(636) Antao, V. P.; Lai, S. Y.; Tinoco, I. A Thermodynamic Study of Unusually Stable RNA and DNA Hairpins. *Nucleic Acids Res.* **1991**, *19*, 5901−5905.
(637) SantaLucia, J., Jr; Kierzek, R.; Turner, D. H. Context Dependence of Hydrogen Bond Free Energy Revealed by Substitutions in an RNA Hairpin. *Science* **1992**, *256*, 217−219.
(638) Antao, V. P.; Tinoco, I. Thermodynamic Parameters for Loop Formation in RNA and DNA Hairpin Tetraloops. *Nucleic Acids Res.* **1992**, *20*, 819−824.
(639) Leulliot, N.; Abdelkafi, M.; Turpin, P.-Y.; Ghomi, M.; Baumruk, V.; Namane, A.; Gouyette, C.; Huynh-Dinh, T. Unusual Nucleotide Conformations in GNRA and UNCG Type Tetraloop Hairpins: Evidence from Raman Markers Assignments. *Nucleic Acids Res.* **1999**, *27*, 1398−1404.
(640) Menger, M.; Eckstein, F.; Porschke, D. Dynamics of the RNA Hairpin GNRA Tetraloop. *Biochemistry* **2000**, *39*, 4500−4507.
(641) Proctor, D. J.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C. Folding Thermodynamics and Kinetics of YNMG RNA Hairpins: Specific Incorporation of 8-bromoguanosine Leads to Stabilization by Enhancement of the Folding Rate. *Biochemistry* **2004**, *43*, 14004−14014.
(642) Zhao, L.; Xia, T. Direct Revelation of Multiple Conformations in RNA by Femtosecond Dynamics. *J. Am. Chem. Soc.* **2007**, *129*, 4118−4119.
(643) Johnson, J. E., Jr; Hoogstraten, C. G. Extensive Backbone Dynamics in the GCAA RNA Tetraloop Analyzed Using 13C NMR Spin Relaxation and Specific Isotope Labeling. *J. Am. Chem. Soc.* **2008**, *130*, 16757−16769.
(644) Stancik, A. L.; Brauns, E. B. Rearrangement of Partially Ordered Stacked Conformations Contributes to the Rugged Energy Landscape of a Small RNA Hairpin. *Biochemistry* **2008**, *47*, 10834−10840.
(645) Sheehy, J. P.; Davis, A. R.; Znosko, B. M. Thermodynamic Characterization of Naturally Occurring RNA Tetraloops. *RNA* **2010**, *16*, 417−429.
(646) Mohan, S.; Hsiao, C.; Bowman, J. C.; Wartell, R.; Williams, L. D. RNA Tetraloop Folding Reveals Tension between Backbone Restraints and Molecular Interactions. *J. Am. Chem. Soc.* **2010**, *132*, 12679−12689.
(647) Fürtig, B.; Schnieders, R.; Richter, C.; Zetzsche, H.; Keyhani, S.; Helmling, C.; Kovacs, H.; Schwalbe, H. Direct 13C-detected NMR Experiments for Mapping and Characterization of Hydrogen Bonds in RNA. *J. Biomol. NMR* **2016**, *64*, 207−221.
(648) Sripakdeevong, P.; Cevec, M.; Chang, A. T.; Erat, M. C.; Ziegeler, M.; Zhao, Q.; Fox, G. E.; Gao, X.; Kennedy, S. D.; Kierzek, R.; et al. Structure Determination of Noncanonical RNA Motifs Guided by 1H NMR Chemical Shifts. Nat. *Nat. Methods* **2014**, *11*, 413−416.
(649) Zhang, W.; Chen, S.-J. RNA Hairpin-folding Kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1931−1936.
(650) Zichi, D. A. Molecular Dynamics of RNA with the OPLS Force Field. Aqueous Simulation of a Hairpin Containing a Tetranucleotide Loop. *J. Am. Chem. Soc.* **1995**, *117*, 2957−2969.
(651) Sorin, E. J.; Rhee, Y. M.; Nakatani, B. J.; Pande, V. S. Insights into Nucleic Acid Conformational Dynamics from Massively Parallel Stochastic Simulations. *Biophys. J.* **2003**, *85*, 790−803.
(652) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. Does Water Play a Structural Role in the Folding of Small Nucleic Acids? *Biophys. J.* **2005**, *88*, 2516−2524.
(653) Woese, C. R.; Gutell, R.; Gupta, R.; Noller, H. F. Detailed Analysis of the Higher-order Structure of 16S-like Ribosomal Ribonucleic Acids. *Microbiol. Mol. Biol. Rev.* **1983**, *47*, 621−669.
(654) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; Daubentoncarafa, Y.; Uhlenbeck, O. C.; Tinoco, I.; et al. CUUCGG Hairpins - Extraordinarily Stable RNA Secondary Structures Associated with Various Biochemical Processes. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1364−1368.
(655) Keating, K. S.; Toor, N.; Pyle, A. M. The GANC Tetraloop: A Novel Motif in the Group IIC Intron Structure. *J. Mol. Biol.* **2008**, *383*, 475−481.
(656) Butcher, S. E.; Dieckmann, T.; Feigon, J. Solution Structure of the Conserved 16 S-like Ribosomal RNA UGAA Tetraloop. *J. Mol. Biol.* **1997**, *268*, 348−358.
(657) Wu, H.; Yang, P. K.; Butcher, S. E.; Kang, S.; Chanfreau, G.; Feigon, J. A Novel Family of RNA Tetraloop Structure Forms the Recognition Site for Saccharomyces Cerevisiae RNase III. *EMBO J.* **2001**, *20*, 7240−7249.
(658) DeJong, E. S.; Marzluff, W. F.; Nikonowicz, E. P. NMR Structure and Ddynamics of the RNA-binding Site for the Histone mRNA Stem-loop Binding Protein. *RNA* **2002**, *8*, 83−96.
(659) Zanier, K.; Luyten, I.; Crombie, C.; Muller, B.; Schümperli, D.; Linge, J. P.; Nilges, M.; Sattler, M. Structure of the Histone mRNA Hairpin Required for Cell Cycle Regulation of Histone Gene Expression. *RNA* **2002**, *8*, 29−46.
(660) Ennifar, E.; Nikulin, A.; Tishchenko, S.; Serganov, A.; Nevskaya, N.; Garber, M.; Ehresmann, B.; Ehresmann, C.; Nikonov, S.; Dumas, P. The Crystal Structure of UUCG Tetraloop. *J. Mol. Biol.* **2000**, *304*, 35−42.
(661) Heus, H. A.; Pardi, A. Structural Features that Give Rise to the Unusual Stability of RNA Hairpins Containing GNRA Loops. *Science* **1991**, *253*, 191−194.
(662) D'Ascenzo, L.; Leonarski, F.; Vicens, Q.; Auffinger, P. Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops. *RNA* **2017**, *23*, 259−269.
(663) Hall, K. B. Mighty Tiny. *RNA* **2015**, *21*, 630−631.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(664) Auffinger, P.; Westhof, E. Molecular Dynamics Simulations of Nucleic Acids. *Encyclopedia of computational chemistry* **1997**, *8*, 227−236.
(665) Zacharias, M. Simulation of the Structure and Dynamics of Nonhelical RNA Motifs. *Curr. Opin. Struct. Biol.* **2000**, *10*, 311−317.
(666) Kara, M.; Zacharias, M. Theoretical Studies of Nucleic Acids Folding. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 116−126.
(667) Jucker, F. M.; Heus, H. A.; Yip, P. F.; Moors, E. H.; Pardi, A. A Network of Heterogeneous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol.* **1996**, *264*, 968−980.
(668) Bothe, J. R.; Nikolova, E. N.; Eichhorn, C. D.; Chugh, J.; Hansen, A. L.; Al-Hashimi, H. M. Characterizing RNA Dynamics at Atomic Resolution Using Solution-state NMR Spectroscopy. Nat. *Nat. Methods* **2011**, *8*, 919−931.
(669) Zhao, B.; Zhang, Q. Characterizing Excited Conformational States of RNA by NMR Spectroscopy. *Curr. Opin. Struct. Biol.* **2015**, *30*, 134−146.
(670) Bottaro, S.; Lindorff-Larsen, K. Mapping the Universe of RNA Tetraloop Folds. *Biophys. J.* **2017**, *113*, 257−267.
(671) Sarkar, K.; Meister, K.; Sethi, A.; Gruebele, M. Fast Folding of an RNA Tetraloop on a Rugged Energy Landscape Detected by a Stacking-Sensitive Probe. *Biophys. J.* **2009**, *97*, 1418−1427.
(672) Ewald, P. P. Ewald Summation. *Ann. Phys.* **1921**, *369*, 253−287.
(673) Auffinger, P.; LouiseMay, S.; Westhof, E. Molecular Dynamics Simulations of the Anticodon Hairpin of tRNA(Asp): Structuring Effects of C-H Center Dot Center Dot Center Dot O Hydrogen Bonds and of Long-range Hydration Forces. *J. Am. Chem. Soc.* **1996**, *118*, 1181−1189.
(674) Auffinger, P.; Louisemay, S.; Westhof, E. Multiple Molecular-Dynamics Simulations of the Anticodon Loop of tRNA(Asp) in Aqueous-Solution with Counterious. *J. Am. Chem. Soc.* **1995**, *117*, 6720−6726.
(675) Cheatham, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193−4194.
(676) Allain, F. H. T.; Varani, G. Structure of the P1 Helix from Group-I Self-Splicing Introns. *J. Mol. Biol.* **1995**, *250*, 333−353.
(677) Miller, J. L.; Kollman, P. A. Theoretical Studies of an Exceptionally Stable RNA Tetraloop: Observation of Convergence from an Incorrect NMR Structure to the Correct One Using Unrestrained Molecular Dynamics. *J. Mol. Biol.* **1997**, *270*, 436−450.
(678) Williams, D. J.; Hall, K. B. Unrestrained Stochastic Dynamics Simulations of the UUCG Tetraloop Using an Implicit Solvation Model. *Biophys. J.* **1999**, *76*, 3192−3205.
(679) Williams, D. J.; Hall, K. B. Experimental and Theoretical Studies of the Effects of Deoxyribose Substitutions on the Stability of the UUCG Tetraloop. *J. Mol. Biol.* **2000**, *297*, 251−265.
(680) Williams, D. J.; Hall, K. B. Experimental and Computational Studies of the G[UUCG]C RNA Tetraloop. *J. Mol. Biol.* **2000**, *297*, 1045−1061.
(681) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(682) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401−9409.
(683) Vorobjev, Y. N.; Almagro, J. C.; Hermans, J. Discrimination between Native and Intentionally Misfolded Conformations of Proteins: ES/IS, A New Method for Calculating Conformational Free Energy that Uses Both Dynamics Simulations with an Explicit Solvent and an Implicit Solvent Continuum Model. *Proteins* **1998**, *32*, 399−413.
(684) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. Free Energy Analysis of the Conformational Preferences of A and B Forms of DNA in Solution. *J. Am. Chem. Soc.* **1998**, *120*, 10629−10633.
(685) Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of RNA Hairpin Loops and Helices. *J. Biomol. Struct. Dyn.* **1998**, *16*, 671−682.
(686) van Gunsteren, W. F.; Mark, A. E. On the Interpretation of Biochemical Data by Molecular Dynamics Computer Simulation. *Eur. J. Biochem.* **1992**, *204*, 947−961.
(687) Singh, S. B.; Kollman, P. A. Understanding the Thermodynamic Stability of an RNA Hairpin and Its Mutant. *Biophys. J.* **1996**, *70*, 1940−1948.
(688) Villa, A.; Stock, G. What NMR Relaxation Can Tell Us about the Internal Motion of an RNA Hairpin: A Molecular Dynamics Simulation Study. *J. Chem. Theory Comput.* **2006**, *2*, 1228−1236.
(689) Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wöhnert, J.; Stock, G.; Schwalbe, H.; NMR, M. D. Studies of the Temperature-dependent Dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res.* **2008**, *36*, 1928−1940.
(690) Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S. Structural Insight into RNA Hairpin Folding Intermediates. *J. Am. Chem. Soc.* **2008**, *130*, 9676−9678.
(691) Chakraborty, D.; Collepardo-Guevara, R.; Wales, D. J. Energy Landscapes, Folding Mechanisms, and Kinetics of RNA Tetraloop Hairpins. *J. Am. Chem. Soc.* **2014**, *136*, 18052−18061.
(692) Giambaşu, G. M.; York, D. M.; Case, D. A. Structural Fidelity and NMR Relaxation Analysis in a Prototype RNA Hairpin. *RNA* **2015**, *21*, 963−974.
(693) Duchardt, E.; Schwalbe, H. Residue Specific Ribose and Nucleobase Dynamics of the cUUCGg RNA Tetraloop Motif by MNMR 13C Relaxation. *J. Biomol. NMR* **2005**, *32*, 295−308.
(694) Juneja, A.; Villa, A.; Nilsson, L. Elucidating the Relation between Internal Motions and Dihedral Angles in an RNA Hairpin Using Molecular Dynamics. *J. Chem. Theory Comput.* **2014**, *10*, 3532−3540.
(695) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. NMR Spectroscopy of RNA. *ChemBioChem* **2003**, *4*, 936−962.
(696) Ode, H.; Matsuo, Y.; Neya, S.; Hoshino, T. Force Field Parameters for Rotation Around chi Torsion Axis in Nucleic Acids. *J. Comput. Chem.* **2008**, *29*, 2531−2542.
(697) Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002**, *100*, 3285−3305.
(698) Nozinovic, S.; Furtig, B.; Jonker, H. R. A.; Richter, C.; Schwalbe, H. High-Resolution NMR Structure of an RNA Model System: The 14-mer cUUCGg Tetraloop Hairpin RNA. *Nucleic Acids Res.* **2010**, *38*, 683−694.
(699) Stadlbauer, P.; Sponer, J.; Costanzo, G.; Di Mauro, E.; Pino, S.; Sponer, J. E. Tetraloop-like Geometries Could Form the Basis of the Catalytic Activity of the Most Ancient Ribooligonucleotides. *Chem. - Eur. J.* **2015**, *21*, 3596−3604.
(700) Major, F.; Turcotte, M.; Gautheret, D.; Lapalme, G.; Fillion, E.; Cedergren, R. The Combination of Symbolic and Numerical Computation for Three-dimensional Modeling of RNA. *Science* **1991**, *253*, 1255−1260.
(701) Maier, A.; Sklenar, H.; Kratky, H. F.; Renner, A.; Schuster, P. Force Field Based Conformational Analysis of RNA Structural Motifs: GNRA Tetraloops and Their Pyrimidine Relatives. *Eur. Biophys. J.* **1999**, *28*, 564−573.
(702) Rhee, Y. M.; Pande, V. S. Multiplexed-replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys. J.* **2003**, *84*, 775−786.
(703) Moazed, D.; Stern, S.; Noller, H. F. Rapid Chemical Probing of Conformation in 16 S Ribosomal RNA and 30 S Ribosomal Subunits Using Primer Extension. *J. Mol. Biol.* **1986**, *187*, 399−416.
(704) Michel, F.; Westhof, E. Modelling of the Three-dimensional Architecture of Group I Catalytic Introns Based on Comparative Sequence Analysis. *J. Mol. Biol.* **1990**, *216*, 585−610.
(705) Havrila, M.; Reblova, K.; Zirbel, C. L.; Leontis, N. B.; Sponer, J. Isosteric and Nonisosteric Base Pairs in RNA Motifs: Molecular Dynamics and Bioinformatics Study of the Sarcin Ricin Internal Loop. *J. Phys. Chem. B* **2013**, *117*, 14302−14319.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(706) Lilley, D. M. J. Structures of Helical Junctions in Nucleic Acids. *Q. Rev. Biophys.* **2000**, *33*, 109−159.
(707) Mathews, D. H.; Turner, D. H. Experimentally Derived Nearest-Neighbor Parameters for the Stability of RNA Three- and Four-Way Multibranch Loops. *Biochemistry* **2002**, *41*, 869−880.
(708) Lescoute, A.; Westhof, E. Topology of Three-way Junctions in Folded RNAs. *RNA* **2006**, *12*, 83−93.
(709) Laing, C.; Schlick, T. Analysis of Four-Way Junctions in RNA Structures. *J. Mol. Biol.* **2009**, *390*, 547−559.
(710) Bindewald, E.; Hayes, R.; Yingling, Y. G.; Kasprzak, W.; Shapiro, B. A. RNAJunction: A Database of RNA Junctions and Kissing Loops for Three-dimensional Structural Analysis and Nano-design. *Nucleic Acids Res.* **2008**, *36*, D392−D397.
(711) Laing, C.; Jung, S.; Iqbal, A.; Schlick, T. Tertiary Motifs Revealed in Analyses of Higher-Order RNA Junctions. *J. Mol. Biol.* **2009**, *393*, 67−82.
(712) Laing, C.; Jung, S.; Kim, N.; Elmetwaly, S.; Zahran, M.; Schlick, T. Predicting Helical Topologies in RNA Junctions as Tree Graphs. *PLoS One* **2013**, *8*, e71947.
(713) Shu, D.; Shu, Y.; Haque, F.; Abdelmawla, S.; Guo, P. X. Thermodynamically Stable RNA Three-way Junction for Constructing Multifunctional Nanoparticles for Delivery of Therapeutics. *Nat. Nanotechnol.* **2011**, *6*, 658−667.
(714) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. RNA Tertiary Structure Mediation by Adenosine Platforms. *Science* **1996**, *273*, 1696−1699.
(715) Basu, S.; P. Rambo, R.; Strauss-Soukup, J.; H.Cate, J.; R. Ferre-Damare, A.; Strobel, S. A.; Doudna, J. A. A Specific Monovalent Metal Ion Integral to the AA Platform of the RNA Tetraloop Receptor. *Nat. Struct. Biol.* **1998**, *5*, 986−992.
(716) Correll, C. C.; Beneken, J.; Plantinga, M. J.; Lubbers, M.; Chan, Y.-L. The Common and the Distinctive Features of the Bulged-G Motif Based on a 1.04 Å Resolution RNA Structure. *Nucleic Acids Res.* **2003**, *31*, 6806−6818.
(717) Lu, X.-J.; Olson, W. K.; Bussemaker, H. J. The RNA Backbone Plays a Crucial Role in Mediating the Intrinsic Stability of the GpU Dinucleotide Platform and the GpUpA/GpA Miniduplex. *Nucleic Acids Res.* **2010**, *38*, 4868−4876.
(718) Mladek, A.; Sponer, J. E.; Kulhanek, P.; Lu, X. J.; Olson, W. K.; Sponer, J. Understanding the Sequence Preference of Recurrent RNA Building Blocks Using Quantum Chemistry: The Intrastrand RNA Dinucleotide Platform. *J. Chem. Theory Comput.* **2012**, *8*, 335−347.
(719) Vallurupalli, P.; Moore, P. B. The Solution Structure of the Loop E Region of the 5 S rRNA from Spinach Chloroplasts. *J. Mol. Biol.* **2003**, *325*, 843−856.
(720) Nomenclature Committee of the International Union of Biochemistry (NC-IUB). Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences. Recommendations 1984. *Biochem. J.* **1985**, *229*, 281−286.
(721) Šponer, J.; Leszczyński, J.; Hobza, P. Nature of Nucleic Acid−Base Stacking: Nonempirical ab Initio and Empirical Potential Characterization of 10 Stacked Base Dimers. Comparison of Stacked and H-Bonded Base Pairs. *J. Phys. Chem.* **1996**, *100*, 5590−5596.
(722) Šponer, J. E.; Špačková, N. a.; Kulhánek, P.; Leszczynski, J.; Šponer, J. Non-Watson−Crick Base Pairing in RNA. Quantum Chemical Analysis of the cis Watson−Crick/Sugar Edge Base Pair Family. *J. Phys. Chem. A* **2005**, *109*, 2292−2301.
(723) Pan, Y.; MacKerell, A. D., Jr Altered Structural Fluctuations in Duplex RNA Versus DNA: A Conformational Switch Involving Base Pair Opening. *Nucleic Acids Res.* **2003**, *31*, 7131−7140.
(724) Snoussi, K.; Leroy, J. L. Imino Proton Exchange and Base-Pair Kinetics in RNA Duplexes. *Biochemistry* **2001**, *40*, 8898−8904.
(725) Pan, Y.; Priyakumar, U. D.; MacKerell, A. D. Conformational Determinants of Tandem GU Mismatches in RNA: Insights from Molecular Dynamics Simulations and Quantum Mechanical Calculations. *Biochemistry* **2005**, *44*, 1433−1443.
(726) Halder, S.; Bhattacharyya, D. Structural Stability of Tandemly Occurring Noncanonical Basepairs within Double Helical Fragments: Molecular Dynamics Studies of Functional RNA. *J. Phys. Chem. B* **2010**, *114*, 14028−14040.
(727) Halder, S.; Bhattacharyya, D. Structural Variations of Single and Tandem Mismatches in RNA Duplexes: A Joint MD Simulation and Crystal Structure Database Analysis. *J. Phys. Chem. B* **2012**, *116*, 11845−11856.
(728) Freier, S. M.; Kierzek, R.; Jaeger, J. A.; Sugimoto, N.; Caruthers, M. H.; Neilson, T.; Turner, D. H. Improved Free-energy Parameters for Predictions of RNA Duplex Stability. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 9373−9377.
(729) Jaeger, J. A.; Turner, D. H.; Zuker, M. Improved Predictions of Secondary Structures for RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 7706−7710.
(730) Turner, D. H.; Mathews, D. H. NNDB: The Nearest Neighbor Parameter Database for Predicting Stability of Nucleic Acid Secondary Structure. *Nucleic Acids Res.* **2010**, *38*, D280−D282.
(731) McDowell, J. A.; Turner, D. H. Investigation of the Structural Basis for Thermodynamic Stabilities of Tandem GU Mismatches: Solution Structure of (rGAGGUCUC)2 by Two-Dimensional NMR and Simulated Annealing. *Biochemistry* **1996**, *35*, 14077−14089.
(732) Chen, G.; Znosko, B. M.; Jiao, X.; Turner, D. H. Factors Affecting Thermodynamic Stabilities of RNA 3 × 3 Internal Loops. *Biochemistry* **2004**, *43*, 12865−12876.
(733) Chen, G.; Znosko, B. M.; Kennedy, S. D.; Krugh, T. R.; Turner, D. H. Solution Structure of an RNA Internal Loop with Three Consecutive Sheared GA Pairs. *Biochemistry* **2005**, *44*, 2845−2856.
(734) Chen, G.; Kierzek, R.; Yildirim, I.; Krugh, T. R.; Turner, D. H.; Kennedy, S. D. Stacking Effects on Local Structure in RNA: Changes in the Structure of Tandem GA Pairs when Flanking GC Pairs Are Replaced by isoG−isoC Pairs. *J. Phys. Chem. B* **2007**, *111*, 6718−6727.
(735) Shankar, N.; Xia, T.; Kennedy, S. D.; Krugh, T. R.; Mathews, D. H.; Turner, D. H. NMR Reveals the Absence of Hydrogen Bonding in Adjacent UU and AG Mismatches in an Isolated Internal Loop from Ribosomal RNA. *Biochemistry* **2007**, *46*, 12665−12678.
(736) Shankar, N.; Kennedy, S. D.; Chen, G.; Krugh, T. R.; Turner, D. H. The NMR Structure of an Internal Loop from 23S Ribosomal RNA Differs from Its Structure in Crystals of 50S Ribosomal Subunits. *Biochemistry* **2006**, *45*, 11776−11789.
(737) Réblová, K.; Střelcová, Z.; Kulhánek, P.; Beššeová, I.; Mathews, D. H.; Van Nostrand, K.; Yildirim, I.; Turner, D. H.; Šponer, J. An RNA Molecular Switch: Intrinsic Flexibility of 23S rRNA Helices 40 and 68 5′-UAA/5′-GAN Internal Loops Studied by Molecular Dynamics Methods. *J. Chem. Theory Comput.* **2010**, *6*, 910−929.
(738) Hammond, N. B.; Tolbert, B. S.; Kierzek, R.; Turner, D. H.; Kennedy, S. D. RNA Internal Loops with Tandem AG Pairs: The Structure of the 5′GAGU/3′UGAG Loop Can Be Dramatically Different from Others, Including 5′AAGU/3′UGAA. *Biochemistry* **2010**, *49*, 5817−5827.
(739) Lerman, Y. V.; Kennedy, S. D.; Shankar, N.; Parisien, M.; Major, F.; Turner, D. H. NMR Structure of a 4 × 4 Nucleotide RNA Internal Loop from an R2 Retrotransposon: Identification of a Three Purine−purine Sheared Pair Motif and Comparison to MC-SYM Predictions. *RNA* **2011**, *17*, 1664−1677.
(740) Yildirim, I.; Turner, D. H. RNA Challenges for Computational Chemists. *Biochemistry* **2005**, *44*, 13225−13234.
(741) Van Nostrand, K. P.; Kennedy, S. D.; Turner, D. H.; Mathews, D. H. Molecular Mechanics Investigation of an Adenine−Adenine Non-Canonical Pair Conformational Change. *J. Chem. Theory Comput.* **2011**, *7*, 3779−3792.
(742) Morgado, C. A.; Svozil, D.; Turner, D. H.; Sponer, J. Understanding the Role of Base Stacking in Nucleic Acids. MD and QM Analysis of Tandem GA Base Pairs in RNA Duplexes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12580−12591.
(743) Aytenfisu, A. H.; Spasic, A.; Seetin, M. G.; Serafini, J.; Mathews, D. H. Modified Amber Force Field Correctly Models the Conformational Preference for Tandem GA pairs in RNA. *J. Chem. Theory Comput.* **2014**, *10*, 1292−1301.
(744) Huguet, J. M.; Bizarro, C. V.; Forns, N.; Smith, S. B.; Bustamante, C.; Ritort, F. Single-molecule Derivation of Salt

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Dependent Base-pair Free Energies in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 15431−15436.
(745) Mosayebi, M.; Romano, F.; Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. The Role of Loop Stacking in the Dynamics of DNA Hairpin Formation. *J. Phys. Chem. B* **2014**, *118*, 14326−14335.
(746) Sponer, J.; Morgado, C. A.; Svozil, D. Comment on "Computational Model for Predicting Experimental RNA and DNA Nearest-Neighbor Free Energy Rankings. *J. Phys. Chem. B* **2012**, *116*, 8331−8332.
(747) Siegfried, N. A.; Metzger, S. L.; Bevilacqua, P. C. Folding Cooperativity in RNA and DNA is Dependent on Position in the Helix. *Biochemistry* **2007**, *46*, 172−181.
(748) Spasic, A.; Serafini, J.; Mathews, D. H. The Amber ff99 Force Field Predicts Relative Free Energy Changes for RNA Helix Formation. *J. Chem. Theory Comput.* **2012**, *8*, 2497−2505.
(749) Westhof, E. Water: An Integral Part of Nucleic Acid Structure. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 125−144.
(750) Schneider, B.; Cohen, D. M.; Schleifer, L.; Srinivasan, A. R.; Olson, W. K.; Berman, H. M. A Systematic Method for Studying the Spatial Distribution of Water Molecules Around Nucleic Acid Bases. *Biophys. J.* **1993**, *65*, 2291−2303.
(751) Schneider, B.; Berman, H. M. Hydration of the DNA Bases Is Local. *Biophys. J.* **1995**, *69*, 2661−2669.
(752) Egli, M.; Portmann, S.; Usman, N. RNA Hydration: A Detailed Look. *Biochemistry* **1996**, *35*, 8489−8494.
(753) Auffinger, P.; Hashem, Y. SwS: A Solvation Web Service for Nucleic Acids. *Bioinformatics* **2007**, *23*, 1035−1037.
(754) Auffinger, P.; Westhof, E. Hydration of RNA Base Pairs. *J. Biomol. Struct. Dyn.* **1998**, *16*, 693−707.
(755) Kirillova, S.; Carugo, O. Hydration Sites of Unpaired RNA Bases: A Statistical Analysis of the PDB Structures. *BMC Struct. Biol.* **2011**, *11*, 1−12.
(756) Das, U.; Chen, S.; Fuxreiter, M.; Vaguine, A. A.; Richelle, J.; Berman, H. M.; Wodak, S. J. Checking Nucleic Acid Crystal Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 813−828.
(757) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; et al. MolProbity: All-atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Res.* **2007**, *35*, W375−W383.
(758) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein Crystallography for Non-crystallographers, or How to Get the Best (but not more) From Published Macromolecular Structures. *FEBS J.* **2008**, *275*, 1−21.
(759) Chen, X.; Weber, I.; Harrison, R. W. Hydration Water and Bulk Water in Proteins Have Distinct Properties in Radial Distributions Calculated from 105 Atomic Resolution Crystal Structures. *J. Phys. Chem. B* **2008**, *112*, 12073−12080.
(760) Carugo, O.; Argos, P. Reliability of Atomic Displacement Parameters in Protein Crystal Structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 473−478.
(761) Conte, M. R.; Conn, G. L.; Brown, T.; Lane, A. N. Hydration of the RNA Duplex r(CGCAAAUUUGCG) 2 Determined by NMR. *Nucleic Acids Res.* **1996**, *24*, 3693−3699.
(762) LouiseMay, S.; Auffinger, P.; Westhof, E. Calculations of Nucleic Acid Conformations. *Curr. Opin. Struct. Biol.* **1996**, *6*, 289−298.
(763) Auffinger, P.; Westhof, E. RNA Hydration: Three Nanoseconds of Multiple Molecular Dynamics Simulations of the Solvated tRNA(Asp) Anticodon Hairpin. *J. Mol. Biol.* **1997**, *269*, 326−341.
(764) Auffinger, P.; Westhof, E. Rules Governing the Orientation of the 2'-hydroxyl Group in RNA. *J. Mol. Biol.* **1997**, *274*, 54−63.
(765) Auffinger, P.; Louise-May, S.; Westhof, E. Molecular Dynamics Simulations of Solvated Yeast tRNA(Asp). *Biophys. J.* **1999**, *76*, 50−64.
(766) Nagan, M. C.; Kerimo, S. S.; Musier-Forsyth, K.; Cramer, C. J. Wild-Type RNA MicrohelixAla and 3:70 Variants: Molecular Dynamics Analysis of Local Helical Structure and Tightly Bound Water. *J. Am. Chem. Soc.* **1999**, *121*, 7310−7317.
(767) Auffinger, P.; Westhof, E. Singly and Bifurcated Hydrogen-bonded Base-pairs in tRNA Anticodon Hairpins and Ribozymes. *J. Mol. Biol.* **1999**, *292*, 467−483.
(768) Brandl, M.; Meyer, M.; Sühnel, J. Water-Mediated Base Pairs in RNA: A Quantum-Chemical Study. *J. Phys. Chem. A* **2000**, *104*, 11177−11187.
(769) Auffinger, P.; Westhof, E. Water and Ion Binding Around RNA and DNA (C,G) Oligomers. *J. Mol. Biol.* **2000**, *300*, 1113−1131.
(770) Csaszar, K.; Špačková, N. a.; Štefl, R.; Šponer, J.; Leontis, N. B. Molecular Dynamics of the Frame-shifting Pseudoknot from Beet Western Yellows Virus: The Role of Non-Watson-Crick Base-pairing, Ordered Hydration, Cation Binding and Base Mutations on Stability and Unfolding. *J. Mol. Biol.* **2001**, *313*, 1073−1091.
(771) Auffinger, P.; Westhof, E. Water and Ion Binding Around r(UpA)(12) and d(TpA)(12) Oligomers - Comparison with RNA and DNA (CpG)(12) Duplexes. *J. Mol. Biol.* **2001**, *305*, 1057−1072.
(772) Auffinger, P.; Westhof, E. An Extended Structural Signature for the tRNA Anticodon Loop. *RNA* **2001**, *7*, 334−341.
(773) Vaiana, A. C.; Westhof, E.; Auffinger, P. A Molecular Dynamics Simulation Study of an Aminoglycoside/A-site RNA Complex: Conformational and Hydration Patterns. *Biochimie* **2006**, *88*, 1061−1073.
(774) Auffinger, P.; Hashem, Y. Nucleic Acid Solvation: From Outside to Insight. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325−333.
(775) Auffinger, P. Molecular Dynamics Simulations of RNA Systems. In *Handbook of RNA Biochemistry*; Hartmann, R. K., Bindereif, A., Schön, A., Westhof, E., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2014; pp 687−718.
(776) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*, 044101.
(777) Ramsey, S.; Nguyen, C.; Salomon-Ferrer, R.; Walker, R. C.; Gilson, M. K.; Kurtzman, T. Solvation Thermodynamic Mapping of Molecular Surfaces in AmberTools: GIST. *J. Comput. Chem.* **2016**, *37*, 2029−2037.
(778) Shanker, S.; Bandyopadhyay, P. How Mg2+ Ion and Water Network Affect the Stability and Structure of Non-Watson−Crick Base Pairs in E. coli Loop E of 5S rRNA: A Molecular Dynamics and Reference Interaction Site Model (RISM) Study. *J. Biomol. Struct. Dyn.* **2017**, *35*, 2103−2122.
(779) Lee, J. C.; Gutell, R. R.; Russell, R. The UAA/GAN Internal Loop Motif: A New RNA Structural Element that Forms a Cross-strand AAA Stack and Long-range Tertiary Interactions. *J. Mol. Biol.* **2006**, *360*, 978−988.
(780) Gutell, R. R.; Schnare, M. N.; Gray, M. W. A Compilation of Large Subunit (23s-like and 23s-like) Ribosomal-rna Structures. *Nucleic Acids Res.* **1992**, *20*, 2095−2109.
(781) Endo, Y.; Mitsui, K.; Motizuki, M.; Tsurugi, K. The Mechanism of Action of Ricin and Related Toxic Lectins on Eukaryotic Ribosomes. The Site and the Characteristics of the Modification in 28 S Ribosomal RNA Caused By the Toxins. *J. Biol. Chem.* **1987**, *262*, 5908−5912.
(782) Qin, S.; Zhou, H.-X. Dissection of the High Rate Constant for the Binding of a Ribotoxin to the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6974−6979.
(783) Lacadena, J.; Álvarez-García, E.; Carreras-Sangrà, N.; Herrero-Galán, E.; Alegre-Cebollada, J.; García-Ortega, L.; Oñaderra, M.; Gavilanes, J. G.; Martínez del Pozo, Á. Fungal Ribotoxins: Molecular Dissection of a Family of Natural Killers. *FEMS Microbiol. Rev.* **2007**, *31*, 212−237.
(784) Schindler, D. G.; Davies, J. E. Specific Cleavage of Ribosomal RNA Caused by Alpha Sarcin. *Nucleic Acids Res.* **1977**, *4*, 1097−1110.
(785) Endo, Y.; Tsurugi, K. RNA N-glycosidase Activity of Ricin A-chain. Mechanism of Action of the Toxic Lectin Ricin on Eukaryotic Ribosomes. *J. Biol. Chem.* **1987**, *262*, 8128−8130.
(786) Hausner, T.-P.; Atmadja, J.; Nierhaus, K. H. Evidence that the G2661 Region of 23S rRNA is Located at the Ribosomal Binding Sites of Both Elongation Factors. *Biochimie* **1987**, *69*, 911−923.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(787) Lancaster, L.; Lambert, N. J.; Maklan, E. J.; Horan, L. H.; Noller, H. F. The Sarcin−ricin Loop of 23S rRNA Is Essential for Assembly of the Functional Core of the 50S Ribosomal Subunit. *RNA* **2008**, *14*, 1999−2012.
(788) Macbeth, M. R.; Wool, I. G. The Phenotype of Mutations of G2655 in the Sarcin/Ricin Domain of 23S Ribosomal RNA. *J. Mol. Biol.* **1999**, *285*, 965−975.
(789) Chan, Y.-L.; Sitikov, A. S.; Wool, I. G. The Phenotype of Mutations of the Base-pair C2658·G2663 That Closes the Tetraloop in the Sarcin/ricin Domain of Escherichia coli 23 S Ribosomal RNA. *J. Mol. Biol.* **2000**, *298*, 795−805.
(790) Garst, A. D.; Héroux, A.; Rambo, R. P.; Batey, R. T. Crystal Structure of the Lysine Riboswitch Regulatory mRNA Element. *J. Biol. Chem.* **2008**, *283*, 22347−22351.
(791) Serganov, A.; Huang, L.; Patel, D. J. Structural Insights Into Amino Acid Binding and Gene Control by a Lysine Riboswitch. *Nature* **2008**, *455*, 1263−1267.
(792) Garst, A. D.; Porter, E. B.; Batey, R. T. Insights into the Regulatory Landscape of the Lysine Riboswitch. *J. Mol. Biol.* **2012**, *423*, 17−33.
(793) Yang, X.; Gerczei, T.; Glover, L.; Correll, C. C. Crystal Structures of Restrictocin-inhibitor Complexes with Implications for RNA Recognition and Base Flipping. *Nat. Struct. Biol.* **2001**, *8*, 968−973.
(794) Szewczak, A. A.; Moore, P. B.; Chang, Y. L.; Wool, I. G. The Conformation of the Sarcin/ricin Loop from 28S Ribosomal RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 9581−9585.
(795) Duarte, C. M.; Wadley, L. M.; Pyle, A. M. RNA Structure Comparison, Motif Search and Discovery Using a Reduced Representation of RNA Conformational Space. *Nucleic Acids Res.* **2003**, *31*, 4755−4761.
(796) Tian, S.; Das, R. RNA Structure Through Multidimensional Chemical Mapping. *Q. Rev. Biophys.* **2016**, *49*, No. e7, DOI: 10.1017/S0033583516000020.
(797) Vidovic, I.; Nottrott, S.; Hartmuth, K.; Lührmann, R.; Ficner, R. Crystal Structure of the Spliceosomal 15.5kD Protein Bound to a U4 snRNA Fragment. Mol. *Mol. Cell* **2000**, *6*, 1331−1342.
(798) Schroeder, K. T.; McPhee, S. A.; Ouellet, J.; Lilley, D. M. J. A Structural Database for K-Turn Motifs in RNA. *RNA* **2010**, *16*, 1463−1468.
(799) Schroeder, Kersten T.; Daldrop, P.; Lilley, David M. J. RNA Tertiary Interactions in a Riboswitch Stabilize the Structure of a Kink Turn. *Structure* **2011**, *19*, 1233−1240.
(800) Huang, L.; Lilley, D. M. J. The Kink Turn, a Key Architectural Element in RNA Structure. *J. Mol. Biol.* **2016**, *428*, 790−801.
(801) Szewczak, L. B. W.; Gabrielsen, J. S.; DeGregorio, S. J.; Strobel, S. A.; Steitz, J. A. Molecular Basis for RNA Kink-Turn Recognition by the h15.5K Small RNP Protein. *RNA* **2005**, *11*, 1407−1419.
(802) Razga, F.; Spackova, N.; Reblova, K.; Koca, J.; Leontis, N. B.; Sponer, J. Ribosomal RNA Kink-Turn Motif - A Flexible Molecular Hinge. *J. Biomol. Struct. Dyn.* **2004**, *22*, 183−193.
(803) Cojocaru, V.; Klement, R.; Jovin, T. M. Loss of G-A Base Pairs is Insufficient for Achieving a Large Opening of U4 snRNA K-Turn Motif. *Nucleic Acids Res.* **2005**, *33*, 3435−3446.
(804) Cojocaru, V.; Nottrott, S.; Klement, R.; Jovin, T. M. The snRNP 15.5K Protein Folds its Cognate K-turn RNA: A Combined Theoretical and Biochemical Study. *RNA* **2005**, *11*, 197−209.
(805) Razga, F.; Koca, J.; Mokdad, A.; Sponer, J. Elastic Properties of Ribosomal RNA Building Blocks: Molecular Dynamics of the GTPase-Associated Center rRNA. *Nucleic Acids Res.* **2007**, *35*, 4007−4017.
(806) Reblova, K.; Razga, F.; Li, W.; Gao, H. X.; Frank, J.; Sponer, J. Dynamics of the Base of Ribosomal A-Site Finger Revealed by Molecular Dynamics Simulations and Cryo-EM. *Nucleic Acids Res.* **2010**, *38*, 1325−1340.
(807) Spackova, N.; Reblova, K.; Sponer, J. Structural Dynamics of the Box C/D RNA Kink-Turn and Its Complex with Proteins: The Role of the A-Minor 0 Interaction, Long-Residency Water Bridges, and Structural Ion-Binding Sites Revealed by Molecular Simulations. *J. Phys. Chem. B* **2010**, *114*, 10581−10593.
(808) Ye, W.; Yang, J.; Yu, Q.; Wang, W.; Hancy, J.; Luo, R.; Chen, H.-F. Kink Turn sRNA Folding upon L7Ae Binding Using Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2013**, *15*, 18510−18522.
(809) Frank, J.; Spahn, C. M. T. The Ribosome and the Mechanism of Protein Synthesis. *Rep. Prog. Phys.* **2006**, *69*, 1383−1417.
(810) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. RNA Tertiary Interactions in the Large Ribosomal Subunit: The A-Minor Motif. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4899−4903.
(811) Falb, M.; Amata, I.; Gabel, F.; Simon, B.; Carlomagno, T. Structure of the K-turn U4 RNA: A Combined NMR and SANS Study. *Nucleic Acids Res.* **2010**, *38*, 6274−6285.
(812) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H. D.; Noller, H. F. Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science* **2001**, *292*, 883−896.
(813) Liu, Q.; Fredrick, K. Intersubunit Bridges of the Bacterial Ribosome. *J. Mol. Biol.* **2016**, *428*, 2146−2164.
(814) Shaikh, T. R.; Yassin, A. S.; Lu, Z.; Barnard, D.; Meng, X.; Lu, T.-M.; Wagenknecht, T.; Agrawal, R. K. Initial Bridges Between Two Ribosomal Subunits Are Formed Within 9.4 ms, As Studied By Time-resolved Cryo-EM. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9822−9827.
(815) Weis, F.; Bron, P.; Giudice, E.; Rolland, J.-P.; Thomas, D.; Felden, B.; Gillet, R. tmRNA−SmpB: A Journey to the Centre of the Bacterial Ribosome. *EMBO J.* **2010**, *29*, 3810−3818.
(816) Frank, J.; Agrawal, R. K. A Ratchet-like Inter-subunit Reorganization of the Ribosome During Translocation. *Nature* **2000**, *406*, 318−322.
(817) Valle, M.; Zavialov, A.; Sengupta, J.; Rawat, U.; Ehrenberg, M.; Frank, J. Locking and Unlocking of Ribosomal Motions. *Cell* **2003**, *114*, 123−134.
(818) Liiv, A.; O'Connor, M. Mutations in the Intersubunit Bridge Regions of 23 S rRNA. *J. Biol. Chem.* **2006**, *281*, 29850−29862.
(819) Piekna-Przybylska, D.; Przybylski, P.; Baudin-Baillieu, A.; Rousset, J.-P.; Fournier, M. J. Ribosome Performance Is Enhanced by a Rich Cluster of Pseudouridines in the A-site Finger Region of the Large Subunit. *J. Biol. Chem.* **2008**, *283*, 26026−26036.
(820) Komoda, T.; Sato, N. S.; Phelps, S. S.; Namba, N.; Joseph, S.; Suzuki, T. The A-site Finger in 23 S rRNA Acts as a Functional Attenuator for Translocation. *J. Biol. Chem.* **2006**, *281*, 32303−32309.
(821) Rakauskaite, R.; Dinman, J. D. An Arc of Unpaired "Hinge Bases" Facilitates Information Exchange among Functional Centers of the Ribosome. *Mol. Cell. Biol.* **2006**, *26*, 8992−9002.
(822) Sergiev, P. V.; Kiparisov, S. V.; Burakovsky, D. E.; Lesnyak, D. V.; Leonov, A. A.; Bogdanov, A. A.; Dontsova, O. A. The Conserved A-site Finger of the 23 S rRNA: Just One of the Intersubunit Bridges or a Part of the Allosteric Communication Pathway? *J. Mol. Biol.* **2005**, *353*, 116−123.
(823) Paci, M.; Fox, G. E. Major Centers of Motion in the Large Ribosomal RNAs. *Nucleic Acids Res.* **2015**, *43*, 4640−4649.
(824) Reblova, K.; Sponer, J.; Lankas, F. Structure and Mechanical Properties of the Ribosomal L1 Stalk Three-way Junction. *Nucleic Acids Res.* **2012**, *40*, 6290−6303.
(825) Strobel, S. A.; Adams, P. L.; Stahley, M. R.; Wang, J. RNA Kink Turns to the Left and to the Right. *RNA* **2004**, *10*, 1852−1854.
(826) Zhong, C.; Tang, H.; Zhang, S. RNAMotifScan: Automatic Identification of RNA Structural Motifs Using Secondary Structural Alignment. *Nucleic Acids Res.* **2010**, *38*, e176.
(827) Dethoff, E. A.; Hansen, A. L.; Musselman, C.; Watt, E. D.; Andricioaei, I.; Al-Hashimi, H. M. Characterizing Complex Dynamics in the Transactivation Response Element Apical Loop and Motional Correlations with the Bulge by NMR, Molecular Dynamics, and Mutagenesis. *Biophys. J.* **2008**, *95*, 3906−3915.
(828) Fulle, S.; Christ, N. A.; Kestner, E.; Gohlke, H. HIV-1 TAR RNA Spontaneously Undergoes Relevant Apo-to-Holo Conformational Transitions in Molecular Dynamics and Constrained Geometrical Simulations. *J. Chem. Inf. Model.* **2010**, *50*, 1489−1501.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(829) Sethaphong, L.; Singh, A.; Marlowe, A. E.; Yingling, Y. G. The Sequence of HIV-1 TAR RNA Helix Controls Cationic Distribution. *J. Phys. Chem. C* **2010**, *114*, 5506−5512.
(830) Do, T. N.; Ippoliti, E.; Carloni, P.; Varani, G.; Parrinello, M. Counterion Redistribution upon Binding of a Tat-Protein Mimic to HIV-1 TAR RNA. *J. Chem. Theory Comput.* **2012**, *8*, 688−694.
(831) Musiani, F.; Rossetti, G.; Capece, L.; Gerger, T. M.; Micheletti, C.; Varani, G.; Carloni, P. Molecular Dynamics Simulations Identify Time Scale of Conformational Changes Responsible for Conformational Selection in Molecular Recognition of HIV-1 Transactivation Responsive RNA. *J. Am. Chem. Soc.* **2014**, *136*, 15631−15637.
(832) Andralojc, W.; Ravera, E.; Salmon, L.; Parigi, G.; Al-Hashimi, H. M.; Luchinat, C. Inter-helical Conformational Preferences of HIV-1 TAR-RNA from Maximum Occurrence Analysis of NMR Data and Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5743−5752.
(833) Aboul-ela, F.; Karn, J.; Varani, G. The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* **1995**, *253*, 313−332.
(834) Ippolito, J. A.; Steitz, T. A. A 1.3-resolution Crystal Structure of the HIV-1 Trans-activation Response Region RNA Stem Reveals a Metal Ion-dependent Bulge Conformation. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 9819−9824.
(835) Bardaro, M. F.; Shajani, Z.; Patora-Komisarska, K.; Robinson, J. A.; Varani, G. How Binding of Small Molecule and Peptide Ligands to HIV-1 TAR Alters the RNA Motional Landscape. *Nucleic Acids Res.* **2009**, *37*, 1529−1540.
(836) Zhang, Q.; Sun, X.; Watt, E. D.; Al-Hashimi, H. M. Resolving the Motional Modes That Code for RNA Adaptation. *Science* **2006**, *311*, 653−656.
(837) Long, K. S.; Crothers, D. M. Characterization of the Solution Conformations of Unbound and Tat Peptide-Bound Forms of HIV-1 TAR RNA. *Biochemistry* **1999**, *38*, 10059−10069.
(838) Hamy, F.; Felder, E. R.; Heizmann, G.; Lazdins, J.; Aboul-ela, F.; Varani, G.; Karn, J.; Klimkait, T. An Inhibitor of the Tat/TAR RNA Interaction that Effectively Suppresses HIV-1 Replication. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 3548−3553.
(839) Davidson, A.; Leeper, T. C.; Athanassiou, Z.; Patora-Komisarska, K.; Karn, J.; Robinson, J. A.; Varani, G. Simultaneous Recognition of HIV-1 TAR RNA Bulge and Loop Sequences by Cyclic Peptide Mimics of Tat Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11931−11936.
(840) Puglisi, J. D.; Chen, L.; Frankel, A. D.; Williamson, J. R. Role of RNA Structure in Arginine Recognition of TAR RNA. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 3680−3684.
(841) Frank, A. T.; Stelzer, A. C.; Al-Hashimi, H. M.; Andricioaei, I. Constructing RNA Dynamical Ensembles by Combining MD and Motionally Decoupled NMR RDCs: New Insights into RNA Dynamics and Adaptive Ligand Recognition. *Nucleic Acids Res.* **2009**, *37*, 3670−3679.
(842) Salmon, L.; Bascom, G.; Andricioaei, I.; Al-Hashimi, H. M. A General Method for Constructing Atomic-Resolution RNA Ensembles using NMR Residual Dipolar Couplings: The Basis for Interhelical Motions Revealed. *J. Am. Chem. Soc.* **2013**, *135*, 5457−5466.
(843) Lee, J.; Dethoff, E. A.; Al-Hashimi, H. M. Invisible RNA State Dynamically Couples Distant Motifs. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9485−9490.
(844) Clay, M. C.; Ganser, L. R.; Merriman, D. K.; Al-Hashimi, H. M. Resolving Sugar Puckers in RNA Excited States Exposes Slow Modes of Repuckering Dynamics. *Nucleic Acids Res.* **2017**, *45*, e134.
(845) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106−116.
(846) Zhang, Q.; Throolin, R.; Pitt, S. W.; Serganov, A.; Al-Hashimi, H. M. Probing Motions between Equivalent RNA Domains Using Magnetic Field Induced Residual Dipolar Couplings: Accounting for Correlations between Motions and Alignment. *J. Am. Chem. Soc.* **2003**, *125*, 10530−10531.
(847) Yang, S.; Salmon, L.; Al-Hashimi, H. M. Measuring Similarity between Dynamic Ensembles of Biomolecules. *Nat. Methods* **2014**, *11*, 552−554.
(848) Salmon, L.; Yang, S.; Al-Hashimi, H. M. Advances in the Determination of Nucleic Acid Conformational Ensembles. *Annu. Rev. Phys. Chem.* **2014**, *65*, 293−316.
(849) Salmon, L.; Giambaşu, G. M.; Nikolova, E. N.; Petzold, K.; Bhattacharya, A.; Case, D. A.; Al-Hashimi, H. M. Modulating RNA Alignment Using Directional Dynamic Kinks: Application in Determining an Atomic-Resolution Ensemble for a Hairpin using NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* **2015**, *137*, 12954−12965.
(850) Li, H.; Lee, T.; Dziubla, T.; Pi, F.; Guo, S.; Xu, J.; Li, C.; Haque, F.; Liang, X.-J.; Guo, P. RNA as a Stable Polymer to Build Controllable and Defined Nanostructures for Material and Biomedical Applications. *Nano Today* **2015**, *10*, 631−655.
(851) Pasi, M.; Maddocks, J. H.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dans, P. D.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. μABC: A Systematic Microsecond Molecular Dynamics Study of Tetranucleotide Sequence Effects in B-DNA. *Nucleic Acids Res.* **2014**, *42*, 12272−12283.
(852) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T.; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; et al. A Systematic Molecular Dynamics Study of Nearest-neighbor Effects on Base Pair and Base Pair Step Conformations and Fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38*, 299−313.
(853) Calladine, C. R. Mechanics of Sequence-dependent Stacking of Bases in B-DNA. *J. Mol. Biol.* **1982**, *161*, 343−352.
(854) Dickerson, R. E.; Klug, A. Base Sequence and Helix Structure Variation in B and A DNA. *J. Mol. Biol.* **1983**, *166*, 419−441.
(855) Šponer, J.; Kypr, J. Different Intrastrand and Interstrand Contributions to Stacking Account for Roll Variations at the Alternating Purine-pyrimidine Sequences in A-DNA and A-RNA. *J. Mol. Biol.* **1991**, *221*, 761−764.
(856) Wahl, M. C.; Sundaralingam, M. Crystal Structures of A-DNA Duplexes. *Biopolymers* **1997**, *44*, 45−63.
(857) Besseova, I.; Otyepka, M.; Reblova, K.; Sponer, J. Dependence of A-RNA Simulations on the Choice of the Force Field and Salt Strength. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10701−10711.
(858) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D. Crystallographic Structure of an RNA Helix: [U(UA)6A]2. *J. Mol. Biol.* **1989**, *209*, 459−474.
(859) Liu, C.; Janowski, P. A.; Case, D. A. All-atom Crystal Simulations of DNA and RNA Duplexes. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1059−1071.
(860) Liebl, K.; Drsata, T.; Lankas, F.; Lipfert, J.; Zacharias, M. Explaining the Striking Difference in Twist-stretch Coupling between DNA and RNA: A Comparative Molecular Dynamics Analysis. *Nucleic Acids Res.* **2015**, *43*, 10143−10156.
(861) Bao, L.; Zhang, X.; Shi, Y.-Z.; Wu, Y.-Y.; Tan, Z.-J. Understanding the Relative Flexibility of RNA and DNA Duplexes: Stretching and Twist-Stretch Coupling. *Biophys. J.* **2017**, *112*, 1094−1104.
(862) Lankaš, F.; Šponer, J.; Hobza, P.; Langowski, J. Sequence-dependent Elastic Properties of DNA. *J. Mol. Biol.* **2000**, *299*, 695−709.
(863) Lankaš, F.; Šponer, J.; Langowski, J.; Cheatham, T. E., III DNA Basepair Step Deformability Inferred from Molecular Dynamics Simulations. *Biophys. J.* **2003**, *85*, 2872−2883.
(864) Dršata, T.; Lankaš, F. Theoretical Models of DNA Flexibility. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 355−363.
(865) Herrero-Galán, E.; Fuentes-Perez, M. E.; Carrasco, C.; Valpuesta, J. M.; Carrascosa, J. L.; Moreno-Herrero, F.; Arias-Gonzalez, J. R. Mechanical Identities of RNA and DNA Double Helices Unveiled at the Single-Molecule Level. *J. Am. Chem. Soc.* **2013**, *135*, 122−131.
(866) Lipfert, J.; Skinner, G. M.; Keegstra, J. M.; Hensgens, T.; Jager, T.; Dulin, D.; Köber, M.; Yu, Z.; Donkers, S. P.; Chou, F.-C.; et al.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Double-stranded RNA under Force and Torque: Similarities to and Striking Differences from Double-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15408−15413.
(867) Marin-Gonzalez, A.; Vilhena, J. G.; Perez, R.; Moreno-Herrero, F. Understanding the Mechanical Response of Double-stranded DNA and RNA under Constant Stretching Forces Using All-atom Molecular Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7049−7054.
(868) Parsch, J.; Engels, J. W. C−F···H−C Hydrogen Bonds in Ribonucleic Acids. *J. Am. Chem. Soc.* **2002**, *124*, 5664−5672.
(869) Zacharias, M.; Engels, J. W. Influence of a Fluorobenzene Nucleobase Analogue on the Conformational Flexibility of RNA Studied by Molecular Dynamics Simulations. *Nucleic Acids Res.* **2004**, *32*, 6304−6311.
(870) Kopitz, H.; Zivkovic, A.; Engels, J. W.; Gohlke, H. Determinants of the Unexpected Stability of RNA Fluorobenzene Self Pairs. *ChemBioChem* **2008**, *9*, 2619−2622.
(871) Koller, A. N.; Bozilovic, J.; Engels, J. W.; Gohlke, H. Aromatic N versus Aromatic F: Bioisosterism Discovered in RNA Base Pairing Interactions Leads to a Novel Class of Universal Base Analogs. *Nucleic Acids Res.* **2010**, *38*, 3133−3146.
(872) Jaeger, L.; Leontis, N. B. Tecto-RNA: One-Dimensional Self-Assembly through Tertiary Interactions. *Angew. Chem., Int. Ed.* **2000**, *39*, 2521−2524.
(873) Afonin, K. A.; Kasprzak, W.; Bindewald, E.; Puppala, P. S.; Diehl, A. R.; Hall, K. T.; Kim, T. J.; Zimmermann, M. T.; Jernigan, R. L.; Jaeger, L.; et al. Computational and Experimental Characterization of RNA Cubic Nanoscaffolds. *Methods* **2014**, *67*, 256−265.
(874) Yingling, Y. G.; Shapiro, B. A. Computational Design of an RNA Hexagonal Nanoring and an RNA. *Nano Lett.* **2007**, *7*, 2328−2334.
(875) Singh, A.; Sethaphong, L.; Yingling, Y. G. Interactions of Cations with RNA Loop-Loop Complexes. *Biophys. J.* **2011**, *101*, 727−735.
(876) Golebiowski, J.; Antonczak, S.; Fernandez-Carmona, J.; Condom, R.; Cabrol-Bass, D. Closing Loop Base Pairs in RNA Loop−loop Complexes: Structural Behavior, Interaction Energy and Solvation Analysis through Molecular Dynamics Simulations. *J. Mol. Model.* **2004**, *10*, 408−417.
(877) Pattabiraman, N.; Martinez, H. M.; Shapiro, B. A. Molecular Modeling and Dynamics Studies of HIV-1 Kissing Loop Structures. *J. Biomol. Struct. Dyn.* **2002**, *20*, 397−411.
(878) Beaurain, F.; Laguerre, M. MD Studies of the DIS/DIS Kissing Complex Solution and X-Ray Structures. *Oligonucleotides* **2003**, *13*, 501−514.
(879) Aci, S.; Gangneux, L.; Paoletti, J.; Genest, D. On the Stability of Different Experimental Dimeric Structures of the SL1 Sequence from the Genomic RNA of HIV-1 in Solution: A Molecular Dynamics Simulation and Electrophoresis Study. *Biopolymers* **2004**, *74*, 177−188.
(880) Mazier, S.; Genest, D. Molecular Dynamics Simulation for Probing the Flexibility of the 35 Nucleotide SL1 Sequence Kissing Complex from HIV-1Lai Genomic RNA. *J. Biomol. Struct. Dyn.* **2007**, *24*, 471−479.
(881) Kim, T.; Shapiro, B. A. The Role of Salt Concentration and Magnesium Binding in HIV-1 Subtype-A and Subtype-B Kissing Loop Monomer Structures. *J. Biomol. Struct. Dyn.* **2013**, *31*, 495−510.
(882) Sarzyńska, J.; Réblová, K.; Šponer, J.; Kuliński, T. Conformational Transitions of Flanking Purines in HIV-1 RNA Dimerization Initiation Site Kissing Complexes Studied by CHARMM Explicit Solvent Molecular Dynamics. *Biopolymers* **2008**, *89*, 732−746.
(883) Mujeeb, A.; Clever, J. L.; Billeci, T. M.; James, T. L.; Parslow, T. G. Structure of the Dimer a Initiation Complex of HIV-1 Genomic RNA. *Nat. Struct. Biol.* **1998**, *5*, 432−436.
(884) Baba, S.; Takahashi, K.-i.; Noguchi, S.; Takaku, H.; Koyanagi, Y.; Yamamoto, N.; Kawai, G. Solution RNA Structures of the HIV-1 Dimerization Initiation Site in the Kissing-Loop and Extended-Duplex Dimers. *J. Biochem.* **2005**, *138*, 583−592.
(885) Kieken, F.; Paquet, F.; Brulé, F.; Paoletti, J.; Lancelot, G. A New NMR Solution Structure of the SL1 HIV-1Lai Loop−loop Dimer. *Nucleic Acids Res.* **2006**, *34*, 343−352.
(886) Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. Understanding of Assembly Phenomena by Aromatic−Aromatic Interactions: Benzene Dimer and the Substituted Systems. *J. Phys. Chem. A* **2007**, *111*, 3446−3457.
(887) Řezáč, J.; Hobza, P. Benchmark Calculations of Interaction Energies in Noncovalent Complexes and Their Applications. *Chem. Rev.* **2016**, *116*, 5038−5071.
(888) Chen, A. A.; García, A. E. Mechanism of Enhanced Mechanical Stability of a Minimal RNA Kissing Complex Elucidated by Nonequilibrium Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1530−E1539.
(889) Li, P. T. X.; Bustamante, C.; Tinoco, I. Unusual Mechanical Stability of a Minimal RNA Kissing Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15847−15852.
(890) Stephenson, W.; Asare-Okai, P. N.; Chen, A. A.; Keller, S.; Santiago, R.; Tenenbaum, S. A.; Garcia, A. E.; Fabris, D.; Li, P. T. X. The Essential Role of Stacking Adenines in a Two-Base-Pair RNA Kissing Complex. *J. Am. Chem. Soc.* **2013**, *135*, 5602−5611.
(891) Winkler, W. C.; Breaker, R. R. Regulation of Bacterial Gene Expression by Riboswitches. *Annu. Rev. Microbiol.* **2005**, *59*, 487−517.
(892) Winkler, W. C.; Breaker, R. R. Genetic Control by Metabolite-binding Riboswitches. *ChemBioChem* **2003**, *4*, 1024−1032.
(893) Barrick, J. E.; Breaker, R. R. The Power of Riboswitches. *Sci. Am.* **2007**, *296*, 36−43.
(894) Nudler, E.; Mironov, A. S. The Riboswitch Control of Bacterial Metabolism. Trends Biochem. *Trends Biochem. Sci.* **2004**, *29*, 11−17.
(895) Nahvi, A.; Sudarsan, N.; Ebert, M. S.; Zou, X.; Brown, K. L.; Breaker, R. R. Genetic Control by a Metabolite Binding mRNA. *Chem. Biol.* **2002**, *9*, 1043−1049.
(896) Weinberg, Z.; Wang, J. X.; Bogue, J.; Yang, J. Y.; Corbino, K.; Moy, R. H.; Breaker, R. R. Comparative Genomics Reveals 104 Candidate Structured RNAs from Bacteria, Archaea, and their Metagenomes. *Genome Biol.* **2010**, *11*, R31.
(897) Wickiser, J. K.; Winkler, W. C.; Breaker, R. R.; Crothers, D. M. The Speed of RNA Transcription and Metabolite Binding Kinetics Operate an FMN Riboswitch. *Mol. Cell* **2005**, *18*, 49−60.
(898) Chauvier, A.; Picard-Jean, F.; Berger-Dancause, J.-C.; Bastet, L.; Naghdi, M. R.; Dubé, A.; Turcotte, P.; Perreault, J.; Lafontaine, D. A. Transcriptional Pausing at the Translation Start Site Operates as a Critical Checkpoint for Riboswitch Regulation. *Nat. Commun.* **2017**, *8*, 13892.
(899) Stoddard, C. D.; Widmann, J.; Trausch, J. J.; Marcano-Velazquez, J. G.; Knight, R.; Batey, R. T. Nucleotides Adjacent to the Ligand-Binding Pocket are Linked to Activity Tuning in the Purine Riboswitch. *J. Mol. Biol.* **2013**, *425*, 1596−1611.
(900) Serganov, A.; Yuan, Y. R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-sensing mRNAs. *Chem. Biol.* **2004**, *11*, 1729−1741.
(901) Gilbert, S. D.; Mediatore, S. J.; Batey, R. T. Modified Pyrimidines Specifically Bind the Purine Riboswitch. *J. Am. Chem. Soc.* **2006**, *128*, 14214−14215.
(902) Gilbert, S. D.; Reyes, F. E.; Edwards, A. L.; Batey, R. T. Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. *Structure* **2009**, *17*, 857−868.
(903) Gilbert, S. D.; Stoddard, C. D.; Wise, S. J.; Batey, R. T. Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain. *J. Mol. Biol.* **2006**, *359*, 754−768.
(904) Gilbert, S. D.; Love, C. E.; Edwards, A. L.; Batey, R. T. Mutational Analysis of the Purine Riboswitch Aptamer Domain. *Biochemistry* **2007**, *46*, 13297−13309.
(905) Batey, R. T.; Gilbert, S. D.; Montange, R. K. Structure of a Natural Guanine-responsive Riboswitch Complexed with the Metabolite Hypoxanthine. *Nature* **2004**, *432*, 411−415.
(906) Stagno, J. R.; Liu, Y.; Bhandari, Y. R.; Conrad, C. E.; Panja, S.; Swain, M.; Fan, L.; Nelson, G.; Li, C.; Wendel, D. R.; et al. Structures

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

of Riboswitch RNA Reaction States by Mix-and-inject XFEL Serial Crystallography. *Nature* **2017**, *541*, 242−246.
(907) Liu, Y.; Holmstrom, E.; Zhang, J. W.; Yu, P.; Wang, J. B.; Dyba, M. A.; Chen, D.; Ying, J. F.; Lockett, S.; Nesbitt, D. J.; et al. Synthesis and Applications of RNAs with Position-selective Labelling and Mosaic Composition. *Nature* **2015**, *522*, 368−372.
(908) Mandal, M.; Boese, B.; Barrick, J. E.; Winkler, W. C.; Breaker, R. R. Riboswitches Control Fundamental Biochemical Pathways in Bacillus Subtilis and Other Bacteria. *Cell* **2003**, *113*, 577−586.
(909) Mandal, M.; Breaker, R. R. Adenine Riboswitches and Gene Activation by Disruption of a Transcription Terminator. *Nat. Struct. Mol. Biol.* **2004**, *11*, 29−35.
(910) Rieder, R.; Lang, K.; Graber, D.; Micura, R. Ligand-induced Folding of the Adenosine Deaminase A-riboswitch and Implications on Riboswitch Translational Control. *ChemBioChem* **2007**, *8*, 896−902.
(911) Jain, N.; Zhao, L.; Liu, J. D.; Xia, T. Heterogeneity and Dynamics of the Ligand Recognition Mode in Purine-Sensing Riboswitches. *Biochemistry* **2010**, *49*, 3703−3714.
(912) Priyakumar, U. D.; MacKerell, A. D. Role of the Adenine Ligand on the Stabilization of the Secondary and Tertiary Interactions in the Adenine Riboswitch. *J. Mol. Biol.* **2010**, *396*, 1422−1438.
(913) Villa, A.; Wöhnert, J.; Stock, G. Molecular Dynamics Simulation Study of the Binding of Purine Bases to the Aptamer Domain of the Guanine Sensing Riboswitch. *Nucleic Acids Res.* **2009**, *37*, 4774−4786.
(914) Sharma, M.; Bulusu, G.; Mitra, A. MD Simulations of Ligand-bound and Ligand-free Aptamer: Molecular Level Insights into the Binding and Switching Mechanism of the add A-riboswitch. *RNA* **2009**, *15*, 1673−1692.
(915) Gong, Z.; Zhao, Y. J.; Chen, C. J.; Xiao, Y. Role of Ligand Binding in Structural Organization of Add A-riboswitch Aptamer: A Molecular Dynamics Simulation. *J. Biomol. Struct. Dyn.* **2011**, *29*, 403−416.
(916) Greenleaf, W. J.; Frieda, K. L.; Foster, D. A.; Woodside, M. T.; Block, S. M. Direct Observation of Hierarchical Folding in Single Riboswitch Aptamers. *Science* **2008**, *319*, 630−633.
(917) Lin, J. C.; Hyeon, C.; Thirumalai, D. Sequence-dependent Folding Landscapes of Adenine Riboswitch Aptamers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6376−6382.
(918) Nguyen, P. H.; Derreumaux, P.; Stock, G. Energy Flow and Long-Range Correlations in Guanine-Binding Riboswitch: A Non-equilibrium Molecular Dynamics Study. *J. Phys. Chem. B* **2009**, *113*, 9340−9347.
(919) McDowell, S. E.; Jun, J. M.; Walter, N. G. Long-range Tertiary Interactions in Single Hammerhead Ribozymes Bias Motional Sampling Toward Catalytically Active Conformations. *RNA* **2010**, *16*, 2414−2426.
(920) Leipply, D.; Draper, D. E. Effects of Mg2+ on the Free Energy Landscape for Folding a Purine Riboswitch RNA. *Biochemistry* **2011**, *50*, 2790−2799.
(921) Ling, B. P.; Zhang, R.; Wang, Z. G.; Dong, L. H.; Liu, Y. J.; Zhang, C. Q.; Liu, C. B. Theoretical Studies on the Interaction of Guanine Riboswitch with Guanine and its Closest Analogues. *Mol. Simul.* **2010**, *36*, 929−938.
(922) Ling, B. P.; Wang, Z. G.; Zhang, R.; Meng, X. H.; Liu, Y. J.; Zhang, C. Q.; Liu, C. B. Theoretical Studies on the Interaction of Modified Pyrimidines and Purines with Purine Riboswitch. *J. Mol. Graphics Modell.* **2009**, *28*, 37−45.
(923) Sund, J.; Lind, C.; Aqvist, J. Binding Site Preorganization and Ligand Discrimination in the Purine Riboswitch. *J. Phys. Chem. B* **2015**, *119*, 773−782.
(924) Hu, G.; Ma, A.; Wang, J. Ligand Selectivity Mechanism and Conformational Changes in Guanine Riboswitch by Molecular Dynamics Simulations and Free Energy Calculations. *J. Chem. Inf. Model.* **2017**, *57*, 918−928.
(925) Spitale, R. C.; Torelli, A. T.; Krucinska, J.; Bandarian, V.; Wedekind, J. E. The Structural Basis for Recognition of the PreQ(0) Metabolite by an Unusually Small Riboswitch Aptamer Domain. *J. Biol. Chem.* **2009**, *284*, 11012−11016.
(926) Liberman, J. A.; Suddala, K. C.; Aytenfisu, A.; Chan, D. L.; Belashov, I. A.; Salim, M.; Mathews, D. H.; Spitaled, R. C.; Walter, N. G.; Wedekind, J. E. Structural Analysis of a Class III PreQ(1) Riboswitch Reveals an Aptamer Distant from a Ribosome-binding Site Regulated by Fast Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E3485−E3494.
(927) Liberman, J. A.; Salim, M.; Krucinska, J.; Wedekind, J. E. Structure of a Class II PreQ(1) Riboswitch Reveals Ligand Recognition by a New Fold. *Nat. Chem. Biol.* **2013**, *9*, 353−355.
(928) Klein, D. J.; Edwards, T. E.; Ferre-D'Amare, A. R. Cocrystal Structure of a Class I PreQ(1) Riboswitch Reveals a Pseudoknot Recognizing an Essential Hypermodified Nucleobase. *Nat. Struct. Mol. Biol.* **2009**, *16*, 343−344.
(929) Jenkins, J. L.; Krucinska, J.; McCarty, R. M.; Bandarian, V.; Wedekind, J. E. Comparison of a PreQ(1) Riboswitch Aptamer in Metabolite-bound and Free States with Implications for Gene Regulation. *J. Biol. Chem.* **2011**, *286*, 24626−24637.
(930) Kang, M. J.; Eichhorn, C. D.; Feigon, J. Structural Determinants for Ligand Capture by a Class II PreQ(1) Riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E663−E671.
(931) Kang, M.; Peterson, R.; Feigon, J. Structural Insights into Riboswitch Control of the Biosynthesis of Queuosine, a Modified Nucleotide Found in the Anticodon of tRNA. *Mol. Cell* **2009**, *33*, 784−790.
(932) Gong, Z.; Zhao, Y. J.; Chen, C. J.; Xiao, Y. Computational Study of Unfolding and Regulation Mechanism of preQ(1) Riboswitches. *PLoS One* **2012**, *7*, e45239.
(933) Feng, J.; Walter, N. G.; Brooks, C. L. Cooperative and Directional Folding of the preQ(1) Riboswitch Aptamer Domain. *J. Am. Chem. Soc.* **2011**, *133*, 4196−4199.
(934) Yoon, J.; Thirumalai, D.; Hyeon, C. Urea-Induced Denaturation of PreQ(1)-Riboswitch. *J. Am. Chem. Soc.* **2013**, *135*, 12112−12121.
(935) Petrone, P. M.; Dewhurst, J.; Tommasi, R.; Whitehead, L.; Pomerantz, A. K. Atomic-scale Characterization of Conformational Changes in the PreQ(1) Riboswitch Aptamer upon Ligand Binding. *J. Mol. Graphics Modell.* **2011**, *30*, 179−185.
(936) Gong, Z.; Zhao, Y. J.; Chen, C. J.; Duan, Y.; Xiao, Y. Insights into Ligand Binding to PreQ(1) Riboswitch Aptamer from Molecular Dynamics Simulations. *PLoS One* **2014**, *9*, e92247.
(937) Eichhorn, C. D.; Feng, J.; Suddala, K. C.; Walter, N. G.; Brooks, C. L.; Al-Hashimi, H. M. Unraveling the Structural Complexity in a Single-stranded RNA Tail: Implications for Efficient Ligand Binding in the Prequeuosine Riboswitch. *Nucleic Acids Res.* **2012**, *40*, 1345−1355.
(938) Aytenfisu, A. H.; Liberman, J. A.; Wedekind, J. E.; Mathews, D. H. Molecular Mechanism for preQ(1)-II Riboswitch Function Revealed by Molecular Dynamics. *RNA* **2015**, *21*, 1898−1907.
(939) Wang, W.; Jiang, C.; Zhang, J. M.; Ye, W.; Luo, R.; Chen, H. F. Dynamics Correlation Network for Allosteric Switching of PreQ(1) Riboswitch. *Sci. Rep.* **2016**, *6*, No. e31005, DOI: 10.1038/srep31005.
(940) Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R. An mRNA Structure that Controls Gene Expression by Binding S-adenosylmethionine. *Nat. Struct. Biol.* **2003**, *10*, 701−707.
(941) Wang, J. X.; Breaker, R. R. Riboswitches that Sense S-adenosylmethionine and S-adenosylhomocysteine. *Biochem. Cell Biol.* **2008**, *86*, 157−168.
(942) McDaniel, B. A. M.; Grundy, F. J.; Artsimovitch, I.; Henkin, T. M. Transcription Termination Control of the S Box System: Direct Measurement of S-adenosylmethionine by the Leader RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3083−3088.
(943) Epshtein, V.; Mironov, A. S.; Nudler, E. The Riboswitch-mediated Control of Sulfur Metabolism in Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5052−5056.
(944) Grundy, F. J.; Henkin, T. M. The S Box Regulon: A New Global Transcription Termination Control System for Methionine and

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Cysteine Biosynthesis Genes in Gram-positive Bacteria. *Mol. Microbiol.* **1998**, *30*, 737−749.
(945) Montange, R. K.; Mondragon, E.; van Tyne, D.; Garst, A. D.; Ceres, P.; Batey, R. T. Discrimination between Closely Related Cellular Metabolites by the SAM-I Riboswitch. *J. Mol. Biol.* **2010**, *396*, 761−772.
(946) Montange, R. K.; Batey, R. T. Structure of the S-adenosylmethionine Riboswitch Regulatory mRNA Element. *Nature* **2006**, *441*, 1172−1175.
(947) Lu, C. R.; Ding, F.; Chowdhury, A.; Pradhan, V.; Tomsic, J.; Holmes, W. M.; Henkin, T. M.; Ke, A. L. SAM Recognition and Conformational Switching Mechanism in the Bacillus subtilis yitJ S Box/SAM-I Riboswitch. *J. Mol. Biol.* **2010**, *404*, 803−818.
(948) Weinberg, Z.; Regulski, E. E.; Hammond, M. C.; Barrick, J. E.; Yao, Z.; Ruzzo, W. L.; Breaker, R. R. The Aptamer Core of SAM-IV Riboswitches Mimics the Ligand-binding Site of SAM-I Riboswitches. *RNA* **2008**, *14*, 822−828.
(949) Trausch, J. J.; Xu, Z. J.; Edwards, A. L.; Reyes, F. E.; Ross, P. E.; Knight, R.; Batey, R. T. Structural Basis for Diversity in the SAM Clan of Riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 6624−6629.
(950) Lim, J.; Winkler, W. C.; Nakamura, S.; Scott, V.; Breaker, R. R. Molecular-recognition Characteristics of SAM-binding Riboswitches. *Angew. Chem., Int. Ed.* **2006**, *45*, 964−968.
(951) Gilbert, S. D.; Rambo, R. P.; Van Tyne, D.; Batey, R. T. Structure of the SAM-II Riboswitch Bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* **2008**, *15*, 177−182.
(952) Corbino, K. A.; Barrick, J. E.; Lim, J.; Welz, R.; Tucker, B. J.; Puskarz, I.; Mandal, M.; Rudnick, N. D.; Breaker, R. R. Evidence for a Second Class of S-adenosylmethionine Riboswitches and Other Regulatory RNA Motifs in Alpha-proteobacteria. *Genome Biol.* **2005**, *6*, R70.
(953) Poiata, E.; Meyer, M. M.; Ames, T. D.; Breaker, R. R. A Variant Riboswitch Aptamer Class for S-adenosylmethionine Common in Marine Bacteria. *RNA* **2009**, *15*, 2046−2056.
(954) Lu, C.; Smith, A. M.; Fuchs, R. T.; Ding, F.; Rajashankar, K.; Henkin, T. M.; Ke, A. Crystal Structures of the SAM-III/S-MK Riboswitch Reveal the SAM-dependent Translation Inhibition Mechanism. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1076−1083.
(955) Fuchs, R. T.; Grundy, F. J.; Henkin, T. M. S-adenosylmethionine Directly Inhibits Binding of 30S Ribosornal Subunits to the S-MK Box Translational Riboswitch RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4876−4880.
(956) Fuchs, R. T.; Grundy, F. J.; Henkin, T. M. The S-MK Box is a New SAM-binding RNA for Translational Regulation of SAM Synthetase. *Nat. Struct. Mol. Biol.* **2006**, *13*, 226−233.
(957) Stoddard, C. D.; Montange, R. K.; Hennelly, S. P.; Rambo, R. P.; Sanbonmatsu, K. Y.; Batey, R. T. Free State Conformational Sampling of the SAM-I Riboswitch Aptamer Domain. *Structure* **2010**, *18*, 787−797.
(958) Huang, W.; Kim, J.; Jha, S.; Aboul-ela, F. The Impact of a Ligand Binding on Strand Migration in the SAM-I Riboswitch. *PLoS Comput. Biol.* **2013**, *9*, e1003069.
(959) Huang, W.; Kim, J.; Jha, S.; Aboul-Ela, F. A Mechanism for S-adenosyl Methionine Assisted Formation of a Riboswitch Conformation: A Small Molecule with a Strong Arm. *Nucleic Acids Res.* **2009**, *37*, 6528−6539.
(960) Hayes, R. L.; Noel, J. K.; Mohanty, U.; Whitford, P. C.; Hennelly, S. P.; Onuchic, J. N.; Sanbonmatsu, K. Y. Magnesium Fluctuations Modulate RNA Dynamics in the SAM-I Riboswitch. *J. Am. Chem. Soc.* **2012**, *134*, 12043−12053.
(961) Kelley, J. M.; Hamelberg, D. Atomistic Basis for the On−off Signaling Mechanism in SAM-II Riboswitch. *Nucleic Acids Res.* **2010**, *38*, 1392−1400.
(962) Doshi, U.; Kelley, J. M.; Hamelberg, D. Atomic-level Insights Into Metabolite Recognition and Specificity of the SAM-II Riboswitch. *RNA* **2012**, *18*, 300−307.
(963) Priyakumar, U. D. Atomistic Details of the Ligand Discrimination Mechanism of S-MK/SAM-III Riboswitch. *J. Phys. Chem. B* **2010**, *114*, 9920−9925.
(964) Suresh, G.; Srinivasan, H.; Nanda, S.; Priyakumar, U. D. Ligand-Induced Stabilization of a Duplex-like Architecture Is Crucial for the Switching Mechanism of the SAM-III Riboswitch. *Biochemistry* **2016**, *55*, 3349−3360.
(965) Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M.; Morgan-Warren, R. J.; Carter, A. P.; Vonrhein, C.; Hartsch, T.; Ramakrishnan, V. Structure of the 30S Ribosomal Subunit. *Nature* **2000**, *407*, 327−339.
(966) Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. Structural Insights into the Translational Infidelity Mechanism. *Nat. Commun.* **2015**, *6*, 7251.
(967) Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. New Structural Insights into Translational Miscoding. Trends Biochem. *Trends Biochem. Sci.* **2016**, *41*, 798−814.
(968) Rozov, A.; Westhof, E.; Yusupov, M.; Yusupova, G. The Ribosome Prohibits the G•U Wobble Geometry at the First Position of the Codon−anticodon Helix. *Nucleic Acids Res.* **2016**, *44*, 6434−6441.
(969) Ogle, J. M.; Murphy, F. V. I. V.; Tarry, M. J.; Ramakrishnan, V. Selection of tRNA by the Ribosome Requires a Transition from an Open to a Closed Form. *Cell* **2002**, *111*, 721−732.
(970) Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. Structure of the A Site of Escherichia coli 16S Ribosomal RNA Complexed with an Aminoglycoside Antibiotic. *Science* **1996**, *274*, 1367−1371.
(971) Vicens, Q.; Westhof, E. Crystal Structure of Paromomycin Docked into the Eubacterial Ribosomal Decoding A Site. *Structure* **2001**, *9*, 647−658.
(972) François, B.; Russell, R. J. M.; Murray, J. B.; Aboul-ela, F.; Masquida, B.; Vicens, Q.; Westhof, E. Crystal Structures of Complexes Between Aminoglycosides and Decoding A Site Oligonucleotides: Role of the Number of Rings and Positive Charges in the Specific Binding Leading to Miscoding. *Nucleic Acids Res.* **2005**, *33*, 5677−5690.
(973) Kondo, J.; Francois, B.; Urzhumtsev, A.; Westhof, E. Crystal Structure of the Homo Sapiens Cytoplasmic Ribosomal Decoding Site Complexed with Apramycin. *Angew. Chem., Int. Ed.* **2006**, *45*, 3310−3314.
(974) Kondo, J.; Urzhumtsev, A.; Westhof, E. Two Conformational States in the Crystal Structure of the Homo Sapiens Cytoplasmic Ribosomal Decoding A Site. *Nucleic Acids Res.* **2006**, *34*, 676−685.
(975) Kondo, J.; Westhof, E. The Bacterial and Mitochondrial Ribosomal A-site Molecular Switches Possess Different Conformational Substates. *Nucleic Acids Res.* **2008**, *36*, 2654−2666.
(976) Sanbonmatsu, K. Y. Energy Landscape of the Ribosomal Decoding Center. *Biochimie* **2006**, *88*, 1053−1059.
(977) Długosz, M.; Antosiewicz, J. M.; Trylska, J. Association of Aminoglycosidic Antibiotics with the Ribosomal A-Site Studied with Brownian Dynamics. *J. Chem. Theory Comput.* **2008**, *4*, 549−559.
(978) Romanowska, J.; Setny, P.; Trylska, J. Molecular Dynamics Study of the Ribosomal A-Site. *J. Phys. Chem. B* **2008**, *112*, 15227−15243.
(979) Romanowska, J.; McCammon, J. A.; Trylska, J. Understanding the Origins of Bacterial Resistance to Aminoglycosides through Molecular Dynamics Mutational Study of the Ribosomal A-Site. *PLoS Comput. Biol.* **2011**, *7*, e1002099.
(980) Panecka, J.; Havrila, M.; Reblova, K.; Sponer, J.; Trylska, J. Role of S-turn2 in the Structure, Dynamics, and Function of Mitochondrial Ribosomal A-Site. A Bioinformatics and Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2014**, *118*, 6687−6701.
(981) Panecka, J.; Mura, C.; Trylska, J. Interplay of the Bacterial Ribosomal A-Site, S12 Protein Mutations and Paromomycin Binding: A Molecular Dynamics Study. *PLoS One* **2014**, *9*, e111811.
(982) Zeng, X.; Chugh, J.; Casiano-Negroni, A.; Al-Hashimi, H. M.; Brooks Iii, C. L. Flipping of the Ribosomal A-Site Adenines Provides a Basis for tRNA Selection. *J. Mol. Biol.* **2014**, *426*, 3201−3213.
(983) Panecka, J.; Sponer, J.; Trylska, J. Conformational Dynamics of Bacterial and Human Cytoplasmic Models of the Ribosomal A-site. *Biochimie* **2015**, *112C*, 96−110.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(984) Wadley, L. M.; Pyle, A. M. The Identification of Novel RNA Structural Motifs Using COMPADRES: An Automated Approach to Structural Discovery. *Nucleic Acids Res.* **2004**, *32*, 6650−6659.
(985) Romanowska, J.; Reuter, N.; Trylska, J. Comparing Aminoglycoside Binding Sites in Bacterial Ribosomal RNA and Aminoglycoside Modifying Enzymes. *Proteins* **2013**, *81*, 63−80.
(986) Długosz, M.; Huber, G. A.; McCammon, J. A.; Trylska, J. Brownian Dynamics Study of the Association between the 70S Ribosome and Elongation Factor G. *Biopolymers* **2011**, *95*, 616−627.
(987) Innis, C. A.; Blaha, G.; Bulkley, D.; Steitz, T. A. Structural Studies of Complexes of the 70S Ribosome. In *Ribosomes: Structure, Function, and Dynamics*; Rodnina, M. V., Wintermeyer, W., Green, R., Eds.; Springer Vienna: Vienna, 2011; pp 31−43.
(988) VanLoock, M. S.; Agrawal, R. K.; Gabashvili, I. S.; Qi, L.; Frank, J.; Harvey, S. C. Movement of the Decoding Region of the 16S Ribosomal RNA Accompanies tRNA Translocation. *J. Mol. Biol.* **2000**, *304*, 507−515.
(989) Drsata, T.; Reblova, K.; Beššeová, I.; Sponer, J.; Lankas, F. rRNA C-loops: Mechanical Properties of a Recurrent Structural Motif. *J. Chem. Theory Comput.* **2017**, *13*, 3359−3371.
(990) Barthel, A.; Zacharias, M. Conformational Transitions in RNA Single Uridine and Adenosine Bulge Structures: A Molecular Dynamics Free Energy Simulation Study. *Biophys. J.* **2006**, *90*, 2450−2462.
(991) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(992) Nivedha, A. K.; Makeneni, S.; Foley, B. L.; Tessier, M. B.; Woods, R. J. Importance of Ligand Conformational Energies in Carbohydrate Docking: Sorting the Wheat from the Chaff. *J. Comput. Chem.* **2014**, *35*, 526−539.
(993) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622−655.
(994) Satpati, P.; Åqvist, J. Why Base Tautomerization Does Not Cause Errors in mRNA Decoding on the Ribosome. *Nucleic Acids Res.* **2014**, *42*, 12876−12884.
(995) Vries, M. S. d.; Hobza, P. Gas-Phase Spectroscopy of Biomolecular Building Blocks. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585−612.
(996) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. Guanine Tautomerism Revealed by UV−UV and IR−UV Hole Burning Spectroscopy. *J. Chem. Phys.* **2001**, *115*, 4604−4611.
(997) Mons, M.; Dimicoli, I.; Piuzzi, F.; Tardivel, B.; Elhanine, M. Tautomerism of the DNA Base Guanine and Its Methylated Derivatives as Studied by Gas-Phase Infrared and Ultraviolet Spectroscopy. *J. Phys. Chem. A* **2002**, *106*, 5088−5094.
(998) Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Orozco, M.; Luque, F. J.; Sponer, J.; Slavicek, P.; Hobza, P. Correlated Ab Initio Study of Nucleic Acid Bases and Their Tautomers in the Gas Phase, in a Microhydrated Environment and in Aqueous Solution Part 1. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4192−4203.
(999) Gorb, L.; Leszczynski, J. Intramolecular Proton Transfer in Mono- and Dihydrated Tautomers of Guanine: An ab Initio Post Hartree−Fock Study. *J. Am. Chem. Soc.* **1998**, *120*, 5024−5032.
(1000) Orozco, M.; Hernández, B.; Luque, F. J. Tautomerism of 1-Methyl Derivatives of Uracil, Thymine, and 5-Bromouracil. Is Tautomerism the Basis for the Mutagenicity of 5-Bromouridine? *J. Phys. Chem. B* **1998**, *102*, 5228−5233.
(1001) Bebenek, K.; Pedersen, L. C.; Kunkel, T. A. Replication Infidelity via a Mismatch with Watson−Crick Geometry. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1862−1867.
(1002) Wang, W.; Hellinga, H. W.; Beese, L. S. Structural Evidence for the Rare Tautomer Hypothesis of Spontaneous Mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 17644−17648.
(1003) Kimsey, I. J.; Petzold, K.; Sathyamoorthy, B.; Stein, Z. W.; Al-Hashimi, H. M. Visualizing Transient Watson-Crick-like Mispairs in DNA and RNA Duplexes. *Nature* **2015**, *519*, 315−320.
(1004) Colominas, C.; Luque, F. J.; Orozco, M. Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes. *J. Am. Chem. Soc.* **1996**, *118*, 6811−6821.
(1005) Watson, J. D.; Crick, F. H. C. Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953**, *171*, 964−967.
(1006) Topal, M. D.; Fresco, J. R. Complementary Base Pairing and the Origin of Substitution Mutations. *Nature* **1976**, *263*, 285−289.
(1007) Florián, J.; Leszczyński, J. Spontaneous DNA Mutations Induced by Proton Transfer in the Guanine·Cytosine Base Pairs: An Energetic Perspective. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017.
(1008) Wilcox, J. L.; Ahluwalia, A. K.; Bevilacqua, P. C. Charged Nucleobases and Their Potential for RNA Catalysis. *Acc. Chem. Res.* **2011**, *44*, 1270−1279.
(1009) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *J. Chem. Theory Comput.* **2009**, *5*, 3105−3113.
(1010) Priyakumar, U. D.; MacKerell, A. D. Computational Approaches for Investigating Base Flipping in Oligonucleotides. *Chem. Rev.* **2006**, *106*, 489−505.
(1011) Sanbonmatsu, K. Y. Flipping through the Genetic Code: New Developments in Discrimination between Cognate and Near-Cognate tRNAs and the Effect of Antibiotics. *J. Mol. Biol.* **2014**, *426*, 3197−3200.
(1012) Pavlov, M. Y.; Liljas, A.; Ehrenberg, M. A Recent Intermezzo at the Ribosome Club. *Philos. Trans. R. Soc., B* **2017**, *372*, 20160185.
(1013) Fischer, N.; Neumann, P.; Bock, L. V.; Maracci, C.; Wang, Z.; Paleskava, A.; Konevega, A. L.; Schröder, G. F.; Grubmüller, H.; Ficner, R.; et al. The Pathway to GTPase Activation of Elongation Factor SelB on the Ribosome. *Nature* **2016**, *540*, 80−85.
(1014) Rodnina, M. V.; Fischer, N.; Maracci, C.; Stark, H. Ribosome Dynamics During Decoding. *Philos. Trans. R. Soc., B* **2017**, *372*, 20160182.
(1015) Fagan, C. E.; Dunkle, J. A.; Maehigashi, T.; Dang, M. N.; Devaraj, A.; Miles, S. J.; Qin, D.; Fredrick, K.; Dunham, C. M. Reorganization of an Intersubunit Bridge Induced by Disparate 16S Ribosomal Ambiguity Mutations Mimics an EF-Tu-bound State. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 9716−9721.
(1016) Maximoff, S. N.; Kamerlin, S. C. L.; Florián, J. DNA Polymerase λ Active Site Favors a Mutagenic Mispair between the Enol Form of Deoxyguanosine Triphosphate Substrate and the Keto Form of Thymidine Template: A Free Energy Perturbation Study. *J. Phys. Chem. B* **2017**, *121*, 7813−7822.
(1017) Krueger, A. T.; Kool, E. T. Model systems for understanding DNA base pairing. *Curr. Opin. Chem. Biol.* **2007**, *11*, 588−594.
(1018) Oertell, K.; Harcourt, E. M.; Mohsen, M. G.; Petruska, J.; Kool, E. T.; Goodman, M. F. Kinetic Selection vs. Free Energy of DNA Base Pairing in Control of Polymerase Fidelity. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E2277−E2285.
(1019) Sripathi, K. N.; Tay, W. W.; Banas, P.; Otyepka, M.; Sponer, J.; Walter, N. G. Disparate HDV Ribozyme Crystal Structures Represent Intermediates on a Rugged Free-energy Landscape. *RNA* **2014**, *20*, 1112−1128.
(1020) Åqvist, J.; Lind, C.; Sund, J.; Wallin, G. Bridging the Gap between Ribosome Structure and Biochemistry by Mechanistic Computations. *Curr. Opin. Struct. Biol.* **2012**, *22*, 815−823.
(1021) Almlof, M.; Ander, M.; Aqvist, J. Energetics of Codon-anticodon Recognition on the Small Ribosomal Subunit. *Biochemistry* **2007**, *46*, 200−209.
(1022) Sund, J.; Ander, M.; Aqvist, J. Principles of Stop-codon Reading on the Ribosome. *Nature* **2010**, *465*, 947−950.
(1023) Satpati, P.; Sund, J.; Aqvist, J. Structure-Based Energetics of mRNA Decoding on the Ribosome. *Biochemistry* **2014**, *53*, 1714−1722.
(1024) Satpati, P.; Bauer, P.; Aqvist, J. Energetic Tuning by tRNA Modifications Ensures Correct Decoding of Isoleucine and Methionine on the Ribosome. *Chem. - Eur. J.* **2014**, *20*, 10271−10275.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(1025) Voorhees, R. M.; Schmeing, T. M.; Kelley, A. C.; Ramakrishnan, V. The Mechanism for Activation of GTP Hydrolysis on the Ribosome. *Science* **2010**, *330*, 835−838.
(1026) Wallin, G.; Kamerlin, S. C. L.; Åqvist, J. Energetics of Activation of GTP Hydrolysis on the Ribosome. *Nat. Commun.* **2013**, *4*, 1733.
(1027) Åqvist, J.; Kamerlin, S. C. L. The Conformation of a Catalytic Loop Is Central to GTPase Activity on the Ribosome. *Biochemistry* **2015**, *54*, 546−556.
(1028) Åqvist, J.; Kamerlin, S. C. L. Exceptionally Large Entropy Contributions Enable the High Rates of GTP Hydrolysis on the Ribosome. *Sci. Rep.* **2015**, *5*, No. e15817, DOI: 10.1038/srep15817.
(1029) Åqvist, J.; Kamerlin, S. C. L. Conserved Motifs in Different Classes of GTPases Dictate their Specific Modes of Catalysis. *ACS Catal.* **2016**, *6*, 1737−1743.
(1030) Carvalho, A. T. P.; Szeler, K.; Vavitsas, K.; Åqvist, J.; Kamerlin, S. C. L. Modeling the Mechanisms of Biological GTP Hydrolysis. *Arch. Biochem. Biophys.* **2015**, *582*, 80−90.
(1031) Whitford, P. C.; Sanbonmatsu, K. Y. Simulating Movement of tRNA Through the Ribosome During Hybrid-state Formation. *J. Chem. Phys.* **2013**, *139*, 121919.
(1032) Whitford, P. C.; Blanchard, S. C.; Cate, J. H. D.; Sanbonmatsu, K. Y. Connecting the Kinetics and Energy Landscape of tRNA Translocation on the Ribosome. *PLoS Comput. Biol.* **2013**, *9*, e1003003.
(1033) Sanbonmatsu, K. Y. Computational Studies of Molecular Machines: The Ribosome. *Curr. Opin. Struct. Biol.* **2012**, *22*, 168−174.
(1034) Whitford, P. C.; Ahmed, A.; Yu, Y. A.; Hennelly, S. P.; Tama, F.; Spahn, C. M. T.; Onuchic, J. N.; Sanbonmatsu, K. Y. Excited States of Ribosome Translocation Revealed Through Integrative Molecular Modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18943−18948.
(1035) Whitford, P. C.; Geggier, P.; Altman, R. B.; Blanchard, S. C.; Onuchic, J. N.; Sanbonmatsu, K. Y. Accommodation of Aminoacyl-tRNA Into the Ribosome Involves Reversible Excursions Along Multiple Pathways. *RNA* **2010**, *16*, 1196−1204.
(1036) Munro, J. B.; Sanbonmatsu, K. Y.; Spahn, C. M. T.; Blanchard, S. C. Navigating the Ribosome's Metastable Energy Landscape. Trends Biochem. *Trends Biochem. Sci.* **2009**, *34*, 390−400.
(1037) Sanbonmatsu, K. Y.; Tung, C. S. High Performance Computing in Biology: Multimillion Atom Simulations of Nanoscale Systems. *J. Struct. Biol.* **2007**, *157*, 470−480.
(1038) Sanbonmatsu, K. Y.; Joseph, S.; Tung, C. S. Simulating Movement of tRNA Into the Ribosome During Decoding. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15854−15859.
(1039) Trabuco, L. G.; Schreiner, E.; Eargle, J.; Cornish, P.; Ha, T.; Luthey-Schulten, Z.; Schulten, K. The Role of L1 Stalk-tRNA Interaction in the Ribosome Elongation Cycle. *J. Mol. Biol.* **2010**, *402*, 741−760.
(1040) Bock, L. V.; Blau, C.; Schröder, G. F.; Davydov, I. I.; Fischer, N.; Stark, H.; Rodnina, M. V.; Vaiana, A. C.; Grubmüller, H. Energy Barriers and Driving Forces in tRNA Translocation through the Ribosome. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1390−1396.
(1041) Brandman, R.; Brandman, Y.; Pande, V. S. A-Site Residues Move Independently from P-Site Residues in all-Atom Molecular Dynamics Simulations of the 70S Bacterial Ribosome. *PLoS One* **2012**, *7*, e29377.
(1042) Sothiselvam, S.; Liu, B.; Han, W.; Ramu, H.; Klepacki, D.; Atkinson, G. C.; Brauer, A.; Remm, M.; Tenson, T.; Schulten, K.; et al. Macrolide Antibiotics Allosterically Predispose the Ribosome for Translation Arrest. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9804−9809.
(1043) Gumbart, J.; Schreiner, E.; Wilson, Daniel N.; Beckmann, R.; Schulten, K. Mechanisms of SecM-Mediated Stalling in the Ribosome. *Biophys. J.* **2012**, *103*, 331−341.
(1044) Tung, C. S.; Sanbonmatsu, K. Y. Atomic Model of the Thermus thermophilus 70S Ribosome Developed in silico. *Biophys. J.* **2004**, *87*, 2714−2722.
(1045) Yang, H.; Noel, J. K.; Whitford, P. C. Anisotropic Fluctuations in the Ribosome Determine tRNA Kinetics. *J. Phys. Chem. B* **2017**, *121*, 10593−10601.
(1046) Noel, J. K.; Whitford, P. C. How EF-Tu Can Contribute to Efficient Proofreading of aa-tRNA by the Ribosome. *Nat. Commun.* **2016**, *7*, 13314.
(1047) Shalev-Benami, M.; Zhang, Y.; Matzov, D.; Halfon, Y.; Zackay, A.; Rozenberg, H.; Zimmerman, E.; Bashan, A.; Jaffe, C. L.; Yonath, A.; et al. 2.8-Å Cryo-EM Structure of the Large Ribosomal Subunit from the Eukaryotic Parasite Leishmania. *Cell Rep.* **2016**, *16*, 288−294.
(1048) Fischer, N.; Neumann, P.; Konevega, A. L.; Bock, L. V.; Ficner, R.; Rodnina, M. V.; Stark, H. Structure of the E. coli Ribosome−EF-Tu complex at < 3 Å Resolution by Cs-corrected Cryo-EM. *Nature* **2015**, *520*, 567−570.
(1049) Von Loeffelholz, O.; Natchiar, S. K.; Djabeur, N.; Myasnikov, A. G.; Kratzat, H.; Ménétret, J.-F.; Hazemann, I.; Klaholz, B. P. Focused Classification and Refinement in High-resolution Cryo-EM Structural Analysis of Ribosome Complexes. *Curr. Opin. Struct. Biol.* **2017**, *46*, 140−148.
(1050) Chen, Y.; Varani, G. Protein Families and RNA Recognition. *FEBS J.* **2005**, *272*, 2088−2097.
(1051) Daubner, G. M.; Cléry, A.; Allain, F. H. T. RRM−RNA Recognition: NMR or Crystallography···and New Findings. *Curr. Opin. Struct. Biol.* **2013**, *23*, 100−108.
(1052) Chen, Y.-C. Beware of docking! *Trends Pharmacol. Sci.* **2015**, *36*, 78−95.
(1053) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.
(1054) Yadav, D. K.; Lukavsky, P. J. NMR Solution Structure Determination of Large RNA-protein Complexes. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *97*, 57−81.
(1055) Schlundt, A.; Tants, J.-N.; Sattler, M. Integrated Structural Biology to Unravel Molecular Mechanisms of Protein-RNA Recognition. *Methods* **2017**, *118*, 119−136.
(1056) Auweter, S. D.; Fasan, R.; Reymond, L.; Underwood, J. G.; Black, D. L.; Pitsch, S.; Allain, F. H. T. Molecular Basis of RNA Recognition by the Human Alternative Splicing Factor Fox-1. *EMBO J.* **2006**, *25*, 163−173.
(1057) Pitera, J. W.; Chodera, J. D. On the Use of Experimental Observations to Bias Simulated Ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445−3451.
(1058) Afroz, T.; Cienikova, Z.; Cléry, A.; Allain, F. H. T. One, Two, Three, Four! How Multiple RRMs Read the Genome Sequence. In *Methods in Enzymology*; Woodson, S. A., Allain, F. H. T., Eds.; Academic Press: New York, 2015; Vol. *558*, pp 235−278.
(1059) Oubridge, C.; Ito, N.; Evans, P. R.; Teo, C. H.; Nagai, K. Crystal-Structure at 1.92 Angstrom Resolution of the RNA-Binding Domain of the U1A Spliceosomal Protein Complexed with an RNA Hairpin. *Nature* **1994**, *372*, 432−438.
(1060) Allain, F. H. T.; Howe, P. W. A.; Neuhaus, D.; Varani, G. Structural Basis of the RNA-binding Specificity of Human U1A Protein. *EMBO J.* **1997**, *16*, 5764−5774.
(1061) Reyes, C. M.; Kollman, P. A. Molecular Dynamics Studies of U1A-RNA Complexes. *RNA* **1999**, *5*, 235−244.
(1062) Reyes, C. M.; Kollman, P. A. Structure and Thermodynamics of RNA-protein Binding: Using Molecular Dynamics and Free Energy Analyses to Calculate the Free Energies of Binding and Conformational Change. *J. Mol. Biol.* **2000**, *297*, 1145−1158.
(1063) Blakaj, D. M.; McConnell, K. J.; Beveridge, D. L.; Baranger, A. M. Molecular Dynamics and Thermodynamics of Protein-RNA Interactions: Mutation of a Conserved Aromatic Residue Modifies Stacking Interactions and Structural Adaptation in the U1A-stem Loop 2 RNA Complex. *J. Am. Chem. Soc.* **2001**, *123*, 2548−2551.
(1064) Kormos, B. L.; Benitex, Y.; Baranger, A. M.; Beveridge, D. L. Affinity and Specificity of Protein U1A-RNA Complex Formation

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Based on an Additive Component Free Energy Model. *J. Mol. Biol.* **2007**, *371*, 1405−1419.
(1065) Kormos, B. L.; Pieniazek, S. N.; Beveridge, D. L.; Baranger, A. M. U1A Protein-stem Loop 2 RNA Recognition: Prediction of Structural Differences from Protein Mutations. *Biopolymers* **2011**, *95*, 591−606.
(1066) Law, M. J.; Linde, M. E.; Chambers, E. J.; Oubridge, C.; Katsamba, P. S.; Nilsson, L.; Haworth, I. S.; Laird-Offringa, I. A. The Role of Positively Charged Amino Acids and Electrostatic Interactions in the Complex of U1A Protein and U1 Hairpin II RNA. *Nucleic Acids Res.* **2006**, *34*, 275−285.
(1067) Showalter, S. A.; Hall, K. B. Altering the RNA-binding Mode of the U1A RBD1 Protein. *J. Mol. Biol.* **2004**, *335*, 465−480.
(1068) Pitici, F.; Beveridge, D. L.; Baranger, A. M. Molecular Dynamics Simulation Studies of Induced Fit and Conformational Capture in U1A−RNA Binding: Do Molecular Substates Code for Specificity? *Biopolymers* **2002**, *65*, 424−435.
(1069) Anunciado, D.; Agumeh, M.; Kormos, B. L.; Beveridge, D. L.; Knee, J. L.; Baranger, A. M. Characterization of the Dynamics of an Essential Helix in the U1A Protein by Time-Resolved Fluorescence Measurements. *J. Phys. Chem. B* **2008**, *112*, 6122−6130.
(1070) Kurisaki, I.; Takayanagi, M.; Nagaoka, M. Combined Mechanism of Conformational Selection and Induced Fit in U1A−RNA Molecular Recognition. *Biochemistry* **2014**, *53*, 3646−3657.
(1071) Guzman, I.; Ghaemi, Z.; Baranger, A.; Luthey-Schulten, Z.; Gruebele, M. Native Conformational Dynamics of the Spliceosomal U1A Protein. *J. Phys. Chem. B* **2015**, *119*, 3651−3661.
(1072) Guo, J. X.; Gmeiner, W. H. Molecular Dynamics Simulation of the Human U2B ' Protein Complex with U2 snRNA Hairpin IV in Aqueous Solution. *Biophys. J.* **2001**, *81*, 630−642.
(1073) Schmid, N.; Zagrovic, B.; van Gunsteren, W. F. Mechanism and Thermodynamics of Binding of the Polypyrimidine Tract Binding Protein to RNA. *Biochemistry* **2007**, *46*, 6500−6512.
(1074) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843−856.
(1075) Konté, N. D. d.; Krepl, M.; Damberger, F. F.; Ripin, N.; Duss, O.; Sponer, J.; Allain, F. H.-T. Aromatic Side-chain Conformational Switch on the Surface of the RNA Recognition Motif Enables RNA Discrimination. *Nat. Commun.* **2017**, *8*, No. e654, DOI: 10.1038/s41467-017-00631-3.
(1076) Wang, L.; Yan, F. Molecular Insights into the Specific Recognition between the RNA Binding Domain qRRM2 of hnRNP F and G-tract RNA: A Molecular Dynamics Study. *Biochem. Biophys. Res. Commun.* **2017**, *494*, 95−100.
(1077) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. The Double-stranded-RNA-binding motif: Interference and Much More. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 1013−1023.
(1078) Masliah, G.; Barraud, P.; Allain, F. H.-T. RNA Recognition by Double-stranded RNA Binding Domains: A Matter of Shape and Sequence. *Cell. Mol. Life Sci.* **2013**, *70*, 1875−1895.
(1079) Castrignanò, T.; Chillemi, G.; Varani, G.; Desideri, A. Molecular Dynamics Simulation of the RNA Complex of a Double-Stranded RNA-Binding Domain Reveals Dynamic Features of the Intermolecular Interface and Its Hydration. *Biophys. J.* **2002**, *83*, 3542−3552.
(1080) Yang, J.; Song, J.; Zhang, J. Z. H.; Ji, C. Effect of Mismatch on Binding of ADAR2/GluR-2 Pre-mRNA Complex. *J. Mol. Model.* **2015**, *21*, No. e222, DOI: 10.1007/s00894-015-2760-8.
(1081) Stefl, R.; Oberstrass, F. C.; Hood, J. L.; Jourdan, M.; Zimmermann, M.; Skrisovska, L.; Maris, C.; Peng, L.; Hofr, C.; Emeson, R. B.; et al. The Solution Structure of the ADAR2 dsRBM-RNA Complex Reveals a Sequence-Specific Readout of the Minor Groove. *Cell* **2010**, *143*, 225−237.
(1082) Drusin, S. I.; Suarez, I. P.; Gauto, D. F.; Rasia, R. M.; Moreno, D. M. dsRNA-protein Interactions Studied by Molecular Dynamics Techniques Unravelling dsRNA Recognition by DCL1. *Arch. Biochem. Biophys.* **2016**, *596*, 118−125.
(1083) Acevedo, R.; Evans, D.; Penrod, Katheryn A.; Showalter, Scott A Binding by TRBP-dsRBD2 Does Not Induce Bending of Double-Stranded RNA. *Biophys. J.* **2016**, *110*, 2610−2617.
(1084) Ryter, J. M.; Schultz, S. C. Molecular Basis of Double-stranded RNA-protein Interactions: Structure of a dsRNA-binding Domain Complexed with dsRNA. *EMBO J.* **1998**, *17*, 7505−7513.
(1085) Wang, X.; Vukovic, L.; Koh, H. R.; Schulten, K.; Myong, S. Dynamic Profiling of Double-stranded RNA Binding Proteins. *Nucleic Acids Res.* **2015**, *43*, 7566−7576.
(1086) Xue, Q.; Zheng, Q.-C.; Zhang, J.-L.; Cui, Y.-L.; Zhang, H.-X. Exploring the Mechanism How Marburg Virus VP35 Recognizes and Binds dsRNA by Molecular Dynamics Simulations and Free Energy Calculations. *Biopolymers* **2014**, *101*, 849−860.
(1087) Allen, W. J.; Wiley, M. R.; Myles, K. M.; Adelman, Z. N.; Bevan, D. R. Steered Molecular Dynamics Identifies Critical Residues of the Nodamura Virus B2 Suppressor of RNAi. *J. Mol. Model.* **2014**, *20*, 1−10.
(1088) Xia, Z.; Zhu, Z.; Zhu, J.; Zhou, R. Recognition Mechanism of siRNA by Viral p19 Suppressor of RNA Silencing: A Molecular Dynamics Study. *Biophys. J.* **2009**, *96*, 1761−1769.
(1089) Harikrishna, S.; Pradeepkumar, P. I. Probing the Binding Interactions between Chemically Modified siRNAs and Human Argonaute 2 Using Microsecond Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2017**, *57*, 883−896.
(1090) Reyes, C. M.; Nifosì, R.; Frankel, A. D.; Kollman, P. A. Molecular Dynamics and Binding Specificity Analysis of the Bovine Immunodeficiency Virus BIV Tat-TAR Complex. *Biophys. J.* **2001**, *80*, 2833−2842.
(1091) Mu, Y.; Stock, G. Conformational Dynamics of RNA-Peptide Binding: A Molecular Dynamics Simulation Study. *Biophys. J.* **2006**, *90*, 391−399.
(1092) Li, C. H.; Zuo, Z. C.; Su, J. G.; Xu, X. J.; Wang, C. X. The Interactions and Recognition of Cyclic Peptide Mimetics of Tat with HIV-1 TAR RNA: A Molecular Dynamics Simulation Study. *J. Biomol. Struct. Dyn.* **2013**, *31*, 276−287.
(1093) Créty, T.; Malliavin, T. E. The Conformational Landscape of the Ribosomal Protein S15 and Its Influence on the Protein Interaction with 16S RNA. *Biophys. J.* **2007**, *92*, 2647−2665.
(1094) Chen, K.; Eargle, J.; Sarkar, K.; Gruebele, M.; Luthey-Schulten, Z. Functional Role of Ribosomal Signatures. *Biophys. J.* **2010**, *99*, 3930−3940.
(1095) Wolf, A.; Baumann, S.; Arndt, H.-D.; Kirschner, K. N. Influence of Thiostrepton Binding on the Ribosomal GTPase Associated Region Characterized by Molecular Dynamics Simulation. *Bioorg. Med. Chem.* **2012**, *20*, 7194−7205.
(1096) Li, W.; Sengupta, J.; Rath, B. K.; Frank, J. Functional Conformations of the L11-ribosomal RNA Complex Revealed by Correlative Analysis of Cryo-EM and Molecular Dynamics Simulations. *RNA* **2006**, *12*, 1240−1253.
(1097) Schimmel, P.; Beebe, K. Aminoacyl tRNA Synthetases: From the RNA World to the Theater of Proteins. In *The RNA World*, 3rd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Long Island, NY, 2006; Vol. *43*, pp 227−255.
(1098) Yamasaki, S.; Nakamura, S.; Terada, T.; Shimizu, K. Mechanism of the Difference in the Binding Affinity of E. coli tRNAGln to Glutaminyl-tRNA Synthetase Caused by Noninterface Nucleotides in Variable Loop. *Biophys. J.* **2007**, *92*, 192−200.
(1099) Li, R.; Macnamara, L.; Leuchter, J.; Alexander, R.; Cho, S. MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. *Int. J. Mol. Sci.* **2015**, *16*, 15872−15902.
(1100) Ghosh, A.; Vishveshwara, S. A Study of Communication Pathways in Methionyl- tRNA Synthetase by Molecular Dynamics Simulations and Structure Network Analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15711−15716.
(1101) Bhattacharyya, M.; Ghosh, A.; Hansia, P.; Vishveshwara, S. Allostery and Conformational Free Energy Changes in Human

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Tryptophanyl-tRNA Synthetase from Essential Dynamics and Structure Networks. *Proteins* **2010**, *78*, 506−517.
(1102) Ghosh, A.; Sakaguchi, R.; Liu, C.; Vishveshwara, S.; Hou, Y.-M. Allosteric Communication in Cysteinyl tRNA Synthetase: A Network of Direct and Indirect Readout. *J. Biol. Chem.* **2011**, *286*, 37721−37731.
(1103) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. Dynamical Networks in tRNA:protein Complexes. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(1104) Bushnell, E. A. C.; Huang, W.; Llano, J.; Gauld, J. W. Molecular Dynamics Investigation into Substrate Binding and Identity of the Catalytic Base in the Mechanism of Threonyl-tRNA Synthetase. *J. Phys. Chem. B* **2012**, *116*, 5205−5212.
(1105) Minajigi, A.; Francklyn, C. S. RNA-assisted Catalysis in a Protein Enzyme: The 2′-hydroxyl of tRNAThr A76 Promotes Aminoacylation by Threonyl-tRNA Synthetase. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17748−17753.
(1106) Grant, T. D.; Luft, J. R.; Wolfley, J. R.; Snell, M. E.; Tsuruta, H.; Corretore, S.; Quartley, E.; Phizicky, E. M.; Grayhack, E. J.; Snell, E. H. The Structure of Yeast Glutaminyl-tRNA Synthetase and Modeling of Its Interaction with tRNA. *J. Mol. Biol.* **2013**, *425*, 2480−2493.
(1107) Estarellas, C.; Otyepka, M.; Koca, J.; Banas, P.; Krepl, M.; Sponer, J. Molecular Dynamic Simulations of Protein/RNA Complexes: CRISPR/Csy4 Endoribonuclease. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 1072−1090.
(1108) Haurwitz, R. E.; Sternberg, S. H.; Doudna, J. A. Csy4 Relies on an Unusual Catalytic Dyad to Position and Cleave CRISPR RNA. *EMBO J.* **2012**, *31*, 2824−2832.
(1109) Suresh, G.; Priyakumar, U. D. Atomistic Details of the Molecular Recognition of DNA-RNA Hybrid Duplex by Ribonuclease H Enzyme. *J. Chem. Sci.* **2015**, *127*, 1701−1713.
(1110) Figiel, M.; Krepl, M.; Poznański, J.; Gołąb, A.; Šponer, J.; Nowotny, M. Coordination between the Polymerase and RNase H Activity of HIV-1 Reverse Transcriptase. *Nucleic Acids Res.* **2017**, *45*, 3341−3352.
(1111) Lapkouski, M.; Tian, L.; Miller, J. T.; Le Grice, S. F. J.; Yang, W. Complexes of HIV-1 RT, NNRTI and RNA/DNA Hybrid Reveal a Structure Compatible with RNA Degradation. *Nat. Struct. Mol. Biol.* **2013**, *20*, 230−236.
(1112) Das, K.; Martinez, S. E.; Bandwar, R. P.; Arnold, E. Structures of HIV-1 RT-RNA/DNA Ternary Complexes with dATP and Nevirapine Reveal Conformational Flexibility of RNA/DNA: Insights into Requirements for RNase H Cleavage. *Nucleic Acids Res.* **2014**, *42*, 8125−8137.
(1113) Figiel, M.; Krepl, M.; Park, S.; Poznański, J.; Skowronek, K.; Gołąb, A.; Ha, T.; Šponer, J.; Nowotny, M. Mechanism of Polypurine Tract Primer Generation by HIV-1 Reverse Transcriptase. *J. Biol. Chem.* **2017**, jbc.M117.798256.
(1114) Rausch, J. W.; Le Grice, S. F. J. 'Binding, Bending and Bonding': Polypurine Tract-primed Initiation of Plus-strand DNA Synthesis in Human Immunodeficiency Virus. *Int. J. Biochem. Cell Biol.* **2004**, *36*, 1752−1766.
(1115) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent. Sci.* **2016**, *2*, 756−763.
(1116) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA−Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816−821.
(1117) Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 Conformational Activation as Elucidated from Enhanced Molecular Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 7260−7265.
(1118) Zuo, Z.; Liu, J. Cas9-catalyzed DNA Cleavage Generates Staggered Ends: Evidence from Molecular Dynamics Simulations. *Sci. Rep.* **2016**, *5*, No. e37584, DOI: 10.1038/srep37584.
(1119) Michael, L. A.; Chenault, J. A.; Miller Iii, B. R.; Knolhoff, A. M.; Nagan, M. C. Water, Shape Recognition, Salt Bridges, and Cation−Pi Interactions Differentiate Peptide Recognition of the HIV Rev-Responsive Element. *J. Mol. Biol.* **2009**, *392*, 774−786.
(1120) Fischer, U.; Huber, J.; Boelens, W. C.; Mattajt, L. W.; Lührmann, R. The HIV-1 Rev Activation Domain is a Nuclear Export Signal that Accesses an Export Pathway Used by Specific Cellular RNAs. *Cell* **1995**, *82*, 475−483.
(1121) Das, A. Control of Transcription Termination by RNA-binding Proteins. *Annu. Rev. Biochem.* **1993**, *62*, 893−930.
(1122) Bahadur, R. P.; Kannan, S.; Zacharias, M. Binding of the Bacteriophage P22 N-Peptide to the boxB RNA Motif Studied by Molecular Dynamics Simulations. *Biophys. J.* **2009**, *97*, 3139−3149.
(1123) Mori, M.; Dietrich, U.; Manetti, F.; Botta, M. Molecular Dynamics and DFT Study on HIV-1 Nucleocapsid Protein-7 in Complex with Viral Genome. *J. Chem. Inf. Model.* **2010**, *50*, 638−650.
(1124) Darnell, J. C.; Jensen, K. B.; Jin, P.; Brown, V.; Warren, S. T.; Darnell, R. B. Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function. *Cell* **2001**, *107*, 489−499.
(1125) Phan, A. T.; Kuryavyi, V.; Darnell, J. C.; Serganov, A.; Majumdar, A.; Ilin, S.; Raslin, T.; Polonskaia, A.; Chen, C.; Clain, D.; et al. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* **2011**, *18*, 796−804.
(1126) Vasilyev, N.; Polonskaia, A.; Darnell, J. C.; Darnell, R. B.; Patel, D. J.; Serganov, A. Crystal Structure Reveals Specific Recognition of a G-quadruplex RNA by a β-turn in the RGG Motif of FMRP. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E5391−E5400.
(1127) Pérez-Villa, A.; Darvas, M.; Bussi, G. ATP Dependent NS3 Helicase Interaction with RNA: Insights from Molecular Simulations. *Nucleic Acids Res.* **2015**, *43*, 8725−8734.
(1128) Appleby, T. C.; Anderson, R.; Fedorova, O.; Pyle, A. M.; Wang, R.; Liu, X.; Brendza, K. M.; Somoza, J. R. Visualizing ATP-Dependent RNA Translocation by the NS3 Helicase from HCV. *J. Mol. Biol.* **2011**, *405*, 1139−1153.
(1129) Sharma, C.; Mohanty, D. Molecular Dynamics Simulations for Deciphering the Structural Basis of Recognition of Pre-let-7 miRNAs by LIN28. *Biochemistry* **2017**, *56*, 723−735.
(1130) Zhang, L.; Pardo-Avila, F.; Unarta, I. C.; Cheung, P. P.-H.; Wang, G.; Wang, D.; Huang, X. Elucidation of the Dynamics of Transcription Elongation by RNA Polymerase II using Kinetic Network Models. *Acc. Chem. Res.* **2016**, *49*, 687−694.
(1131) Da, L.-T.; Wang, D.; Huang, X. Dynamics of Pyrophosphate Ion Release and Its Coupled Trigger Loop Motion from Closed to Open State in RNA Polymerase II. *J. Am. Chem. Soc.* **2012**, *134*, 2399−2406.
(1132) Da, L.-T.; Pardo Avila, F.; Wang, D.; Huang, X. A Two-State Model for the Dynamics of the Pyrophosphate Ion Release in Bacterial RNA Polymerase. *PLoS Comput. Biol.* **2013**, *9*, e1003020.
(1133) Da, L.-T.; E, C.; Duan, B.; Zhang, C.; Zhou, X.; Yu, J. A Jump-from-Cavity Pyrophosphate Ion Release Assisted by a Key Lysine Residue in T7 RNA Polymerase Transcription Elongation. *PLoS Comput. Biol.* **2015**, *11*, e1004624.
(1134) Silva, D.-A.; Weiss, D. R.; Pardo Avila, F.; Da, L.-T.; Levitt, M.; Wang, D.; Huang, X. Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7665−7670.
(1135) Jin, Y.; Lin-Tai, D.; Xuhui, H. Constructing Kinetic Models to Elucidate Structural Dynamics of a Complete RNA Polymerase II Elongation Cycle. *Phys. Biol.* **2015**, *12*, 016004.
(1136) Da, L.-T.; Pardo-Avila, F.; Xu, L.; Silva, D.-A.; Zhang, L.; Gao, X.; Wang, D.; Huang, X. Bridge Helix Bending Promotes RNA Polymerase II Backtracking through a Critical and Conserved Threonine Residue. *Nat. Commun.* **2016**, *7*, 11244.
(1137) Jiang, H.; Sheong, F. K.; Zhu, L.; Gao, X.; Bernauer, J.; Huang, X. Markov State Models Reveal a Two-Step Mechanism of miRNA Loading into the Human Argonaute Protein: Selective Binding followed by Structural Re-arrangement. *PLoS Comput. Biol.* **2015**, *11*, e1004404.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(1138) Banáš, P.; Jurečka, P.; Walter, N. G.; Šponer, J.; Otyepka, M. Theoretical Studies of RNA Catalysis: Hybrid QM/MM Methods and their Comparison with MD and QM. *Methods* **2009**, *49*, 202−216.
(1139) Bertran, J.; Oliva, A. Chapter 12: Ribozymes. *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis* **2017**, *2017*, 404−435.
(1140) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.-i.; Yajima, R. Catalytic Roles for Proton Transfer and Protonation in Ribozymes. *Biopolymers* **2004**, *73*, 90−109.
(1141) Bingaman, J. L.; Zhang, S.; Stevens, D. R.; Yennawar, N. H.; Hammes-Schiffer, S.; Bevilacqua, P. C. The GlcN6P Cofactor Plays Multiple Catalytic Roles in the glmS Ribozyme. *Nat. Chem. Biol.* **2017**, *13*, 439−445.
(1142) Weinberg, Z.; Kim, P. B.; Chen, T. H.; Li, S. S.; Harris, K. A.; Lunse, C. E.; Breaker, R. R. New Classes of Self-cleaving Ribozymes Revealed by Comparative Genomics Analysis. *Nat. Chem. Biol.* **2015**, *11*, 606−610.
(1143) Lee, K.-Y.; Lee, B.-J. Structural and Biochemical Properties of Novel Self-Cleaving Ribozymes. *Molecules* **2017**, *22*, No. e678.
(1144) Fedor, M. J.; Williamson, J. R. The Catalytic Diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 399−412.
(1145) Spitale, R. C.; Wedekind, J. E. Exploring Ribozyme Conformational Changes with X-ray Crystallography. *Methods* **2009**, *49*, 87−100.
(1146) Lilley, D. M. J. The Origins of RNA Catalysis in Ribozymes. Trends Biochem. *Trends Biochem. Sci.* **2003**, *28*, 495−501.
(1147) Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486−504.
(1148) Nam, K. H.; Gaot, J. L.; York, D. M. Quantum Mechanical/Molecular Mechanical Simulation Study of the Mechanism of Hairpin Ribozyme Catalysis. *J. Am. Chem. Soc.* **2008**, *130*, 4680−4691.
(1149) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210−3235.
(1150) Klahn, M.; Rosta, E.; Warshel, A. On the Mechanism of Hydrolysis of Phosphate Monoesters Dianions in Solutions and Proteins. *J. Am. Chem. Soc.* **2006**, *128*, 15310−15323.
(1151) Rosta, E.; Kamerlin, S. C.; Warshel, A. On the Interpretation of the Observed Llinear Free Energy Relationship in Phosphate Hydrolysis: A thorough Computational Study of Phosphate Diester Hydrolysis in Solution. *Biochemistry* **2008**, *47*, 3725−3735.
(1152) Florian, J.; Warshel, A. Phosphate Ester Hydrolysis in Aqueous Solution: Associative Versus Dissociative Mechanisms. *J. Phys. Chem. B* **1998**, *102*, 719−734.
(1153) Breaker, R. R.; Emilsson, G. M.; Lazarev, D.; Nakamura, S.; Puskarz, I. J.; Roth, A.; Sudarsan, N. A Common Speed Limit for RNA-cleaving Ribozymes and Deoxyribozymes. *RNA* **2003**, *9*, 949−957.
(1154) Min, D.; Xue, S.; Li, H.; Yang, W. 'In-line attack' Conformational Effect Plays a Modest Role in an Enzyme-catalyzed RNA Cleavage: A Free Energy Simulation Study. *Nucleic Acids Res.* **2007**, *35*, 4001−4006.
(1155) Lopez, X.; Dejaegere, A.; Leclerc, F.; York, D. M.; Karplus, M. Nucleophilic Attack on Phosphate Diesters: A Density Functional Study of In-line Reactivity in Dianionic, Monoanionic, and Neutral Systems. *J. Phys. Chem. B* **2006**, *110*, 11525−11539.
(1156) Radak, B. K.; Harris, M. E.; York, D. M. Molecular Simulations of RNA 2'-O-transesterification Reaction Models in Solution. *J. Phys. Chem. B* **2013**, *117*, 94−103.
(1157) Kellerman, D. L.; York, D. M.; Piccirilli, J. A.; Harris, M. E. Altered (transition) States: Mechanisms of Solution and Enzyme Catalyzed RNA 2'-O-transphosphorylation. *Curr. Opin. Chem. Biol.* **2014**, *21*, 96−102.
(1158) Chval, Z.; Chvalova, D.; Leclerc, F. Modeling the RNA 2'OH Activation: Possible Roles of Metal Ion and Nucleobase as Catalysts in Self-cleaving Ribozymes. *J. Phys. Chem. B* **2011**, *115*, 10943−10956.
(1159) Huang, M.; York, D. M. Linear Free Energy Relationships in RNA Transesterification: Theoretical Models to Aid Experimental Interpretations. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15846−15855.
(1160) Chen, H.; Giese, T. J.; Huang, M.; Wong, K. Y.; Harris, M. E.; York, D. M. Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions. *Chem. - Eur. J.* **2014**, *20*, 14336−14343.
(1161) Chen, H.; Piccirilli, J. A.; Harris, M. E.; York, D. M. Effect of Zn2+ Binding and Enzyme Active Site on the Transition State for RNA 2'-O-transphosphorylation Interpreted through Kinetic Isotope Effects. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854*, 1795−1800.
(1162) Wong, K. Y.; Gu, H.; Zhang, S.; Piccirilli, J. A.; Harris, M. E.; York, D. M. Characterization of the Reaction Path and Transition States for RNA Transphosphorylation Models from Theory and Experiment. *Angew. Chem., Int. Ed.* **2012**, *51*, 647−651.
(1163) Zhang, S.; Gu, H.; Chen, H.; Strong, E.; Ollie, E. W.; Kellerman, D.; Liang, D.; Miyagi, M.; Anderson, V. E.; Piccirilli, J. A.; et al. Isotope Effect Analyses Provide Evidence for an Altered Transition State for RNA 2'-O-transphosphorylation Catalyzed by Zn(2+). *Chem. Commun.* **2016**, *52*, 4462−4465.
(1164) Gregersen, B. A.; Lopez, X.; York, D. M. Hybrid QM/MM Study of Thio Effects in Transphosphorylation Reactions. *J. Am. Chem. Soc.* **2003**, *125*, 7178−7179.
(1165) Gregersen, B. A.; Lopez, X.; York, D. M. Hybrid QM/MM Study of Thio Effects in Transphosphorylation Reactions: The Role of Solvation. *J. Am. Chem. Soc.* **2004**, *126*, 7504−7513.
(1166) Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. Transesterification Thio Effects of Phosphate Diesters: Free Energy Barriers and Kinetic and Equilibrium Isotope Effects from Density-functional Theory. *Biochemistry* **2006**, *45*, 10043−10053.
(1167) Liu, Y.; Lopez, X.; York, D. M. Kinetic Isotope Effects on Thio-substituted Biological Phosphoryl Transfer Reactions from Density-functional Theory. *Chem. Commun.* **2005**, *0*, 3909−3911.
(1168) Liu, Y.; Gregersen, B. A.; Lopez, X.; York, D. M. Density Functional Study of the In-line Mechanism of Methanolysis of Cyclic Phosphate and Thiophosphate Esters in Solution: Insight into Thio Effects in RNA Transesterification. *J. Phys. Chem. B* **2005**, *109*, 19987−20003.
(1169) Wong, K. Y.; Xu, Y.; York, D. M. Ab Initio Path-integral Calculations of Kinetic and Equilibrium Isotope Effects on Base-catalyzed RNA Transphosphorylation Models. *J. Comput. Chem.* **2014**, *35*, 1302−1316.
(1170) Lopez, C. S.; Faza, O. N.; de Lera, A. R.; York, D. M. Pseudorotation Barriers of Biological Oxyphosphoranes: A Challenge for Simulations of Ribozyme Catalysis. *Chem. - Eur. J.* **2005**, *11*, 2081−2093.
(1171) Lopez, C. S.; Faza, O. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. Pseudorotation of Natural and Chemically Modified Biological Phosphoranes: Implications for RNA Catalysis. *ChemPhysChem* **2004**, *5*, 1045−1049.
(1172) Mlynsky, V.; Kuhrova, P.; Jurecka, P.; Sponer, J.; Otyepka, M.; Banas, P. Mapping Chemical Space of the RNA Cleavage and its Implications for Ribozyme Catalysis. *J. Phys. Chem. B* **2017**, *121*, 10828−10840.
(1173) Soukup, G. A.; Breaker, R. R. Relationship between Internucleotide Linkage Geometry and the Stability of RNA. *RNA* **1999**, *5*, 1308−1325.
(1174) Fedor, M. J. Comparative Enzymology and Structural Biology of RNA Self-Cleavage. *Annu. Rev. Biophys.* **2009**, *38*, 271−299.
(1175) Prody, G. A.; Bakos, J. T.; Buzayan, J. M.; Schneider, I. R.; Bruening, G. Autolytic Processing of Dimeric Plant-Virus Satellite Rna. *Science* **1986**, *231*, 1577−1580.
(1176) Sharmeen, L.; Kuo, M. Y. P.; Dintergottlieb, G.; Taylor, J. Antigenomic Rna of Human Hepatitis Delta-Virus Can Undergo Self-Cleavage. *J. Virol.* **1988**, *62*, 2674−2679.
(1177) Buzayan, J. M.; Gerlach, W. L.; Bruening, G. Nonenzymatic Cleavage and Ligation of Rnas Complementary to a Plant-Virus Satellite Rna. *Nature* **1986**, *323*, 349−353.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(1178) Saville, B. J.; Collins, R. A. A Site-specific Self-cleavage Reaction Performed by a Novel RNA in Neurospora Mitochondria. *Cell* **1990**, *61*, 685−696.
(1179) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. Control of Gene Expression by a Natural Metabolite-responsive Ribozyme. *Nature* **2004**, *428*, 281−286.
(1180) Roth, A.; Weinberg, Z.; Chen, A. G. Y.; Kim, P. B.; Ames, T. D.; Breaker, R. R. A Widespread Self-cleaving Ribozyme Class is Revealed by Bioinformatics. *Nat. Chem. Biol.* **2014**, *10*, 56−60.
(1181) Harris, K. A.; Lunse, C. E.; Li, S. S.; Brewer, K. I.; Breaker, R. R. Biochemical Analysis of Pistol Self-cleaving Ribozymes. *RNA* **2015**, *21*, 1852−1858.
(1182) Li, S.; Lunse, C. E.; Harris, K. A.; Breaker, R. R. Biochemical Analysis of Hatchet Self-cleaving Ribozymes. *RNA* **2015**, *21*, 1845−1851.
(1183) Liu, Y.; Wilson, T. J.; Lilley, D. M. J. The Structure of a Nucleolytic Ribozyme that Employs a Catalytic Metal Ion. *Nat. Chem. Biol.* **2017**, *13*, 508−513.
(1184) Casalino, L.; Palermo, G.; Rothlisberger, U.; Magistrato, A. Who Activates the Nucleophile in Ribozyme Catalysis? An Answer from the Splicing Mechanism of Group II Introns. *J. Am. Chem. Soc.* **2016**, *138*, 10374−10377.
(1185) Hammann, C.; Luptak, A.; Perreault, J.; de la Peña, M. The Ubiquitous Hammerhead Ribozyme. *RNA* **2012**, *18*, 871−885.
(1186) Martick, M.; Scott, W. G. Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* **2006**, *126*, 309−320.
(1187) Lee, T. S.; York, D. M. Computational Mutagenesis Studies of Hammerhead Ribozyme Catalysis. *J. Am. Chem. Soc.* **2010**, *132*, 13505−13518.
(1188) Scott, W. G.; Finch, J. T.; Klug, A. The Crystal-Structure of an All-Rna Hammerhead Ribozyme - A Proposed Mechanism for Rna Catalytic Cleavage. *Cell* **1995**, *81*, 991−1002.
(1189) Scott, W. G.; Murray, J. B.; Arnold, J. R. P.; Stoddard, B. L.; Klug, A. Capturing the Structure of a Catalytic RNA Intermediate: The Hammerhead Ribozyme. *Science* **1996**, *274*, 2065−2069.
(1190) Dunham, C. M.; Murray, J. B.; Scott, W. G. A Helical Twist-induced Conformational Switch Activates Cleavage in the Hammerhead Ribozyme. *J. Mol. Biol.* **2003**, *332*, 327−336.
(1191) Feig, A. L.; Scott, W. G.; Uhlenbeck, O. C. Inhibition of the Hammerhead Ribozyme Cleavage Reaction by Site-specific Binding of Tb(III). *Science* **1998**, *279*, 81−84.
(1192) Murray, J. B.; Szoke, H.; Szoke, A.; Scott, W. G. Capture and Visualization of a Catalytic RNA Enzyme-product Complex Using Crystal Lattice Trapping and X-ray Holographic Reconstruction. *Mol. Cell* **2000**, *5*, 279−287.
(1193) Murray, J. B.; Terwey, D. P.; Maloney, L.; Karpeisky, A.; Usman, N.; Beigelman, L.; Scott, W. G. The Structural Basis of Hammerhead Ribozyme Self-cleavage. *Cell* **1998**, *92*, 665−673.
(1194) Murray, J. B.; Dunham, C. M.; Scott, W. G. A pH-dependent Conformational Change, Rather than the Chemical Step, Appears to be Rate-limiting in the Hammerhead Ribozyme Cleavage Reaction. *J. Mol. Biol.* **2002**, *315*, 121−130.
(1195) Hermann, T.; Auffinger, P.; Westhof, E. Molecular dynamics investigations of hammerhead ribozyme RNA. *Eur. Biophys. J.* **1998**, *27*, 153−165.
(1196) Mei, H. Y.; Kaaret, T. W.; Bruice, T. C. A Computational Approach to the Mechanism of Self-Cleavage of Hammerhead Rna. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 9727−9731.
(1197) Torres, R. A.; Bruice, T. C. Molecular Dynamics Study Displays Near In-line Attack Conformations in the Hammerhead Ribozyme Self-cleavage Reaction. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11077−11082.
(1198) Torres, R. A.; Bruice, T. C. The Mechanism of Phosphodiester Hydrolysis: Near In-line Attack Conformations in the Hammerhead Ribozyme. *J. Am. Chem. Soc.* **2000**, *122*, 781−791.
(1199) Lyne, P. D.; Karplus, M. Determination of the pK(a) of the 2 '-hydroxyl Group of a Phosphorylated Ribose: Implications for the Mechanism of Hammerhead Ribozyme Catalysis. *J. Am. Chem. Soc.* **2000**, *122*, 166−167.
(1200) Leclerc, F.; Karplus, M. Two-metal-ion Mechanism for Hammerhead-ribozyme Catalysis. *J. Phys. Chem. B* **2006**, *110*, 3395−3409.
(1201) Torres, R. A.; Himo, F.; Bruice, T. C.; Noodleman, L.; Lovell, T. Theoretical Examination of Mg2+-mediated Hydrolysis of a Phosphodiester Linkage as Proposed for the Hammerhead Ribozyme. *J. Am. Chem. Soc.* **2003**, *125*, 9861−9867.
(1202) Radhakrishnan, R. Coupling of Fast and Slow Modes in the Reaction Pathway of the Minimal Hammerhead Ribozyme Cleavage. *Biophys. J.* **2007**, *93*, 2391−2399.
(1203) Suzumura, K.; Takagi, Y.; Orita, M.; Taira, K. NMR-based Reappraisal of the Coordination of a Metal Ion at the Pro-Rp Oxygen of the A9/G10.1 Site in a Hammerhead Ribozyme. *J. Am. Chem. Soc.* **2004**, *126*, 15504−15511.
(1204) Wang, S.; Karbstein, K.; Peracchi, A.; Beigelman, L.; Herschlag, D. Identification of the Hammerhead Ribozyme Metal Ion Binding Site Responsible for Rescue of the Deleterious Effect of a Cleavage Site Phosphorothioate. *Biochemistry* **1999**, *38*, 14363−14378.
(1205) Blount, K. F.; Uhlenbeck, O. C. The Structure-function Dilemma of the Hammerhead Ribozyme. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 415−440.
(1206) Chen, H.; Giese, T. J.; Golden, B. L.; York, D. M. Divalent Metal Ion Activation of a Guanine General Base in the Hammerhead Ribozyme: Insights from Molecular Simulations. *Biochemistry* **2017**, *56*, 2985−2994.
(1207) Lee, T. S.; Giambasu, G. M.; Sosa, C. P.; Martick, M.; Scott, W. G.; York, D. M. Threshold Occupancy and Specific Cation Binding Modes in the Hammerhead Ribozyme Active Site Are Required for Active Conformation. *J. Mol. Biol.* **2009**, *388*, 195−206.
(1208) Lee, T. S.; York, D. M. Origin of Mutational Effects at the C3 and G8 Positions on Hammerhead Ribozyme Catalysis from Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2008**, *130*, 7168−7169.
(1209) Lee, T. S.; Silva Lopez, C.; Giambasu, G. M.; Martick, M.; Scott, W. G.; York, D. M. Role of Mg2+ in Hammerhead Ribozyme Catalysis from Molecular Simulation. *J. Am. Chem. Soc.* **2008**, *130*, 3053−3064.
(1210) Wong, K. Y.; Lee, T. S.; York, D. M. Active Participation of Mg Ion in the Reaction Coordinate of RNA Self-cleavage Catalyzed by the Hammerhead Ribozyme. *J. Chem. Theory Comput.* **2011**, *7*, 1−3.
(1211) Mir, A.; Golden, B. L. Two Active Site Divalent Ions in the Crystal Structure of the Hammerhead Ribozyme Bound to a Transition State Analogue. *Biochemistry* **2016**, *55*, 633−636.
(1212) Mir, A.; Chen, J.; Robinson, K.; Lendy, E.; Goodman, J.; Neau, D.; Golden, B. L. Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction. *Biochemistry* **2015**, *54*, 6369−6381.
(1213) Curtis, E. A.; Bartel, D. P. The Hammerhead Cleavage Reaction in Monovalent Cations. *RNA* **2001**, *7*, 546−552.
(1214) Anderson, M.; Schultz, E. P.; Martick, M.; Scott, W. G. Active-Site Monovalent Cations Revealed in a 1.55-Å-Resolution Hammerhead Ribozyme Structure. *J. Mol. Biol.* **2013**, *425*, 3790−3798.
(1215) Riccitelli, N.; Lupták, A. HDV Family of Self-Cleaving Ribozymes. In *Progress in Molecular Biology and Translational Science*; Garrett, A. S., Ed.; Academic Press: New York, 2013; Vol. *120*, Chapter 4, pp 123−171.
(1216) Tanner, N. K.; Schaff, S.; Thill, G.; Petitkoskas, E.; Craindenoyelle, A. M.; Westhof, E. A 3-Dimensional Model of Hepatitis-Delta Virus Ribozyme Based on Biochemical and Mutational Analyses. *Curr. Biol.* **1994**, *4*, 488−498.
(1217) Doudna, J. A.; Lorsch, J. R. Ribozyme Catalysis: Not Different, Just Worse. *Nat. Struct. Mol. Biol.* **2005**, *12*, 395−402.
(1218) Perrotta, A. T.; Shih, I.-h.; Been, M. D. Imidazole Rescue of a Cytosine Mutation in a Self-Cleaving Ribozyme. *Science* **1999**, *286*, 123−126.
(1219) Nakano, S.-i.; Chadalavada, D. M.; Bevilacqua, P. C. General Acid-Base Catalysis in the Mechanism of a Hepatitis Delta Virus Ribozyme. *Science* **2000**, *287*, 1493−1497.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(1220) Tinsley, R. A.; Harris, D. A.; Walter, N. G. Magnesium Dependence of the Amplified Conformational Switch in the Trans-acting Hepatitis Delta Virus Ribozyme. *Biochemistry* **2004**, *43*, 8935−8945.
(1221) Chen, J.-H.; Yajima, R.; Chadalavada, D. M.; Chase, E.; Bevilacqua, P. C.; Golden, B. L. A 1.9 Å Crystal Structure of the HDV Ribozyme Precleavage Suggests both Lewis Acid and General Acid Mechanisms Contribute to Phosphodiester Cleavage. *Biochemistry* **2010**, *49*, 6508−6518.
(1222) Ganguly, A.; Thaplyal, P.; Rosta, E.; Bevilacqua, P. C.; Hammes-Schiffer, S. Quantum Mechanical/Molecular Mechanical Free Energy Simulations of the Self-Cleavage Reaction in the Hepatitis Delta Virus Ribozyme. *J. Am. Chem. Soc.* **2014**, *136*, 1483−1496.
(1223) Lee, T.-S.; Radak, B. K.; Harris, M. E.; York, D. M. A Two-Metal-Ion-Mediated Conformational Switching Pathway for HDV Ribozyme Activation. *ACS Catal.* **2016**, *6*, 1853−1869.
(1224) Krasovska, M. V.; Sefcikova, J.; Špačková, N. a.; Šponer, J.; Walter, N. G. Structural Dynamics of Precursor and Product of the RNA Enzyme from the Hepatitis Delta Virus as Revealed by Molecular Dynamics Simulations. *J. Mol. Biol.* **2005**, *351*, 731−748.
(1225) Sefcikova, J.; Krasovska, M. V.; Spackova, N.; Sponer, J.; Walter, N. G. Impact of an Extruded Nucleotide on Cleavage Activity and Dynamic Catalytic Core Conformation of the Hepatitis Delta Virus Ribozyme. *Biopolymers* **2007**, *85*, 392−406.
(1226) Lee, T.-S.; Giambaşu, G. M.; Harris, M. E.; York, D. M. Characterization of the Structure and Dynamics of the HDV Ribozyme in Different Stages Along the Reaction Path. *J. Phys. Chem. Lett.* **2011**, *2*, 2538−2543.
(1227) Sefcikova, J.; Krasovska, M. V.; Šponer, J.; Walter, N. G. The Genomic HDV Ribozyme Utilizes a Previously Unnoticed U-turn Motif to Accomplish Fast Site-specific Catalysis. *Nucleic Acids Res.* **2007**, *35*, 1933−1946.
(1228) Veeraraghavan, N.; Bevilacqua, P. C.; Hammes-Schiffer, S. Long-Distance Communication in the HDV Ribozyme: Insights from Molecular Dynamics and Experiments. *J. Mol. Biol.* **2010**, *402*, 278−291.
(1229) Liu, H. N.; Robinet, J. J.; Ananvoranich, S.; Gauld, J. W. Density Functional Theory Investigation on the Mechanism of the Hepatitis Delta Virus Ribozyme. *J. Phys. Chem. B* **2007**, *111*, 439−445.
(1230) Wei, K.; Liu, L.; Cheng, Y. H.; Fu, Y.; Guo, Q. X. Theoretical Examination of Two Opposite Mechanisms Proposed for Hepatitis Delta Virus Ribozyme. *J. Phys. Chem. B* **2007**, *111*, 1514−1516.
(1231) Banáš, P.; Rulíšek, L.; Hánošová, V.; Svozil, D.; Walter, N. G.; Šponer, J.; Otyepka, M. General Base Catalysis for Cleavage by the Active-Site Cytosine of the Hepatitis Delta Virus Ribozyme: QM/MM Calculations Establish Chemical Feasibility. *J. Phys. Chem. B* **2008**, *112*, 11177−11187.
(1232) Das, S. R.; Piccirilli, J. A. General Acid Catalysis by the Hepatitis Delta Virus Ribozyme. *Nat. Chem. Biol.* **2005**, *1*, 45−52.
(1233) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. Metal Binding Motif in the Active Site of the HDV Ribozyme Binds Divalent and Monovalent Ions. *Biochemistry* **2011**, *50*, 2672−2682.
(1234) Veeraraghavan, N.; Ganguly, A.; Golden, B. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. Mechanistic Strategies in the HDV Ribozyme: Chelated and Diffuse Metal Ion Interactions and Active Site Protonation. *J. Phys. Chem. B* **2011**, *115*, 8346−8357.
(1235) Sripathi, K. N.; Banas, P.; Reblova, K.; Sponer, J.; Otyepka, M.; Walter, N. G. Wobble Pairs of the HDV Ribozyme Play Specific Roles in Stabilization of Active Site Dynamics. *Phys. Chem. Chem. Phys.* **2015**, *17*, 5887−5900.
(1236) Chen, J.; Ganguly, A.; Miswan, Z.; Hammes-Schiffer, S.; Bevilacqua, P. C.; Golden, B. L. Identification of the Catalytic Mg2+ Ion in the Hepatitis Delta Virus Ribozyme. *Biochemistry* **2013**, *52*, 557−567.
(1237) Ganguly, A.; Bevilacqua, P. C.; Hammes-Schiffer, S. Quantum Mechanical/Molecular Mechanical Study of the HDV Ribozyme: Impact of the Catalytic Metal Ion on the Mechanism. *J. Phys. Chem. Lett.* **2011**, *2*, 2906−2911.
(1238) Dubecky, M.; Walter, N. G.; Sponer, J.; Otyepka, M.; Banas, P. Chemical Feasibility of the General Acid/Base Mechanism of glmS Ribozyme Self-Cleavage. *Biopolymers* **2015**, *103*, 550−562.
(1239) Mlynsky, V.; Banas, P.; Walter, N. G.; Sponer, J.; Otyepka, M. QM/MM Studies of Hairpin Ribozyme Self-Cleavage Suggest the Feasibility of Multiple Competing Reaction Mechanisms. *J. Phys. Chem. B* **2011**, *115*, 13911−13924.
(1240) Mlynsky, V.; Walter, N. G.; Sponer, J.; Otyepka, M.; Banas, P. The Role of an Active Site Mg2+ in HDV Ribozyme Self-cleavage: Insights from QM/MM Calculations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 670−679.
(1241) Thaplyal, P.; Ganguly, A.; Golden, B. L.; Hammes-Schiffer, S.; Bevilacqua, P. C. Thio Effects and an Unconventional Metal Ion Rescue in the Genomic Hepatitis Delta Virus Ribozyme. *Biochemistry* **2013**, *52*, 6499−6514.
(1242) Thaplyal, P.; Ganguly, A.; Hammes-Schiffer, S.; Bevilacqua, P. C. Inverse Thio Effects in the Hepatitis Delta Virus Ribozyme Reveal that the Reaction Pathway Is Controlled by Metal Ion Charge Density. *Biochemistry* **2015**, *54*, 2160−2175.
(1243) Radak, B. K.; Lee, T.-S.; Harris, M. E.; York, D. M. Assessment of Metal-assisted Nucleophile Activation in the Hepatitis Delta Virus Ribozyme from Molecular Simulation and 3D-RISM. *RNA* **2015**, *21*, 1566−1577.
(1244) Shih, I. H.; Been, M. D. Kinetic Scheme for Intermolecular RNA Cleavage by a Ribozyme Derived from Hepatitis Delta Virus RNA. *Biochemistry* **2000**, *39*, 9055−9066.
(1245) Harris, D. A.; Tinsley, R. A.; Walter, N. G. Terbium-mediated Footprinting Probes a Catalytic Conformational Switch in the Antigenomic Hepatitis Delta Virus Ribozyme. *J. Mol. Biol.* **2004**, *341*, 389−403.
(1246) Kapral, G. J.; Jain, S.; Noeske, J.; Doudna, J. A.; Richardson, D. C.; Richardson, J. S. New Tools Provide a Second Look at HDV Ribozyme Structure, Dynamics and Cleavage. *Nucleic Acids Res.* **2014**, *42*, 12833−12846.
(1247) Walter, N. G.; Burker, J. M. The Hairpin Ribozyme: Structure, Assembly and Catalysis. *Curr. Opin. Chem. Biol.* **1998**, *2*, 24−30.
(1248) Müller, S.; Appel, B.; Krellenberg, T.; Petkovic, S. The Many Faces of the Hairpin Ribozyme: Structural and Functional Variants of a Small Catalytic Rna. *IUBMB Life* **2012**, *64*, 36−47.
(1249) Young, K. J.; Gill, F.; Grasby, J. A. Metal Ions Play a Passive Role in the Hairpin Ribozyme Catalysed Reaction. *Nucleic Acids Res.* **1997**, *25*, 3760−3766.
(1250) Rupert, P. B.; Ferre-D'Amare, A. R. Crystal Structure of a Hairpin Ribozyme-inhibitor Complex with Implications for Catalysis. *Nature* **2001**, *410*, 780−786.
(1251) Rupert, P. B.; Massey, A. P.; Sigurdsson, S. T.; Ferre-D'Amare, A. R. Transition State Stabilization by a Catalytic RNA. *Science* **2002**, *298*, 1421−1424.
(1252) Salter, J.; Krucinska, J.; Alam, S.; Grum-Tokars, V.; Wedekind, J. E. Water in the Active Site of an All-RNA hairpin Ribozyme and Effects of Gua8 Base Variants on the Geometry of Phosphoryl Transfer. *Biochemistry* **2006**, *45*, 686−700.
(1253) Heldenbrand, H.; Janowski, P. A.; Giambasu, G.; Giese, T. J.; Wedekind, J. E.; York, D. M. Evidence for the Role of Active Site Residues in the Hairpin Ribozyme from Molecular Simulations along the Reaction Path. *J. Am. Chem. Soc.* **2014**, *136*, 7789−7792.
(1254) Park, H.; Lee, S. Role of Solvent Dynamics in Stabilizing the Transition State of RNA Hydrolysis by Hairpin Ribozyme. *J. Chem. Theory Comput.* **2006**, *2*, 858−862.
(1255) Rhodes, M. M.; Réblová, K.; Šponer, J.; Walter, N. G. Trapped Water Molecules Are Essential to Structural Dynamics and Function of a Ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13380−13385.
(1256) Ditzler, M. A.; Sponer, J.; Walter, N. G. Molecular Dynamics Suggest Multifunctionality of an Adenine Imino Group in Acid-base Catalysis of the Hairpin Ribozyme. *RNA* **2009**, *15*, 560−575.
(1257) Nam, K.; Gao, J. L.; York, D. M. Electrostatic Interactions in the Hairpin Ribozyme Account for the Majority of the Rate

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Acceleration Without Chemical Participation by Nucleobases. *RNA* **2008**, *14*, 1501−1507.
(1258) Mlynsky, V.; Banas, P.; Sponer, J.; van der Kamp, M. W.; Mulholland, A. J.; Otyepka, M. Comparison of ab Initio, DFT, and Semiempirical QM/MM Approaches for Description of Catalytic Mechanism of Hairpin Ribozyme. *J. Chem. Theory Comput.* **2014**, *10*, 1608−1622.
(1259) Schuabb, C.; Kumar, N.; Pataraia, S.; Marx, D.; Winter, R. Pressure Modulates the Self-cleavage Step of the Hairpin Ribozyme. *Nat. Commun.* **2017**, *8*, 14661.
(1260) Cochrane, J. C.; Lipchock, S. V.; Strobel, S. A. Structural Investigation of the GlmS Ribozyme Bound to its Catalytic Cofactor. *Chem. Biol.* **2007**, *14*, 97−105.
(1261) Klein, D. J.; Ferre-D'Amare, A. R. Structural Basis of glmS Ribozyme Activation by Glucosamine-6-phosphate. *Science* **2006**, *313*, 1752−1756.
(1262) Klein, D. J.; Been, M. D.; Ferre-D'Amare, A. R. Essential Role of an Active-site Guanine in glmS Ribozyme Catalysis. *J. Am. Chem. Soc.* **2007**, *129*, 14858−14859.
(1263) Klein, D. J.; Wilkinson, S. R.; Been, M. D.; Ferre-D'Amare, A. R. Requirement of Helix p2.2 and Nucleotide g1 for Positioning the Cleavage Site and Cofactor of the glmS Ribozyme. *J. Mol. Biol.* **2007**, *373*, 178−189.
(1264) Cochrane, J. C.; Lipchock, S. V.; Smith, K. D.; Strobel, S. A. Structural and Chemical Basis for Glucosamine 6-Phosphate Binding and Activation of the glmS Ribozyme. *Biochemistry* **2009**, *48*, 3239−3246.
(1265) Lau, M. W. L.; Ferre-D'Amare, A. R. An in vitro Evolved glmS Ribozyme Has the Wild-type Fold but Loses Coenzyme Dependence. *Nat. Chem. Biol.* **2013**, *9*, 805−810.
(1266) Banáš, P.; Walter, N. G.; Šponer, J.; Otyepka, M. Protonation States of the Key Active Site Residues and Structural Dynamics of the glmS Riboswitch As Revealed by Molecular Dynamics. *J. Phys. Chem. B* **2010**, *114*, 8701−8712.
(1267) Zhang, S. X.; Ganguly, A.; Goyal, P.; Bingaman, J. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. Role of the Active Site Guanine in the glmS Ribozyme Self-Cleavage Mechanism: Quantum Mechanical/Molecular Mechanical Free Energy Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 784−798.
(1268) Xin, Y.; Hamelberg, D. Deciphering the Role of Glucosamine-6-phosphate in the Riboswitch Action of glmS Ribozyme. *RNA* **2010**, *16*, 2455−2463.
(1269) Zhang, S.; Stevens, D. R.; Goyal, P.; Bingaman, J. L.; Bevilacqua, P. C.; Hammes-Schiffer, S. Assessing the Potential Effects of Active Site Mg2+ Ions in the glmS Ribozyme-Cofactor Complex. *J. Phys. Chem. Lett.* **2016**, *7*, 3984−3988.
(1270) Dong, X.; Tian, Z. Y.; Yang, X.; Xue, Y. Theoretical Study on the Mechanism of Self-cleavage Reaction of the glmS Ribozyme. *Theor. Chem. Acc.* **2015**, *134*, No. e68, DOI: 10.1007/s00214-015-1667-x.
(1271) Eiler, D.; Wang, J. M.; Steitz, T. A. Structural Basis for the Fast Self-cleavage Reaction Catalyzed by the Twister Ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13028−13033.
(1272) Liu, Y. J.; Wilson, T. J.; McPhee, S. A.; Lilley, D. M. J. Crystal Structure and Mechanistic Investigation of the Twister Ribozyme. *Nat. Chem. Biol.* **2014**, *10*, 739−744.
(1273) Ren, A. M.; Kosutic, M.; Rajashankar, K. R.; Frener, M.; Santner, T.; Westhof, E.; Micura, R.; Patel, D. J. In-line Alignment and Mg2+ Coordination at the Cleavage Site of the Env22 Twister Ribozyme. *Nat. Commun.* **2014**, *5*, 5534.
(1274) Świderek, K.; Marti, S.; Tuñón, I.; Moliner, V.; Bertran, J. Molecular Mechanism of the Site-specific Self-cleavage of the RNA Phosphodiester Backbone by a Twister Ribozyme. *Theor. Chem. Acc.* **2017**, *136*, No. e31, DOI: 10.1007/s00214-017-2060-8.
(1275) Gaines, C. S.; York, D. M. Ribozyme Catalysis with a Twist: Active State of the Twister Ribozyme in Solution Predicted from Molecular Simulation. *J. Am. Chem. Soc.* **2016**, *138*, 3058−3065.
(1276) Wilson, T. J.; Liu, Y. J.; Domnick, C.; Kath-Schorr, S.; Liiley, D. M. J. The Novel Chemical Mechanism of the Twister Ribozyme. *J. Am. Chem. Soc.* **2016**, *138*, 6151−6162.
(1277) Ucisik, M. N.; Bevilacqua, P. C.; Hammes-Schiffer, S. Molecular Dynamics Study of Twister Ribozyme: Role of Mg2+ Ions and the Hydrogen-Bonding Network in the Active Site. *Biochemistry* **2016**, *55*, 3834−3846.
(1278) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. Structure of the 70S Ribosome Complexed with mRNA and tRNA. *Science* **2006**, *313*, 1935−1942.
(1279) Trobro, S.; Aqvist, J. Role of Ribosomal Protein L27 in Peptidyl Transfer. *Biochemistry* **2008**, *47*, 4898−4906.
(1280) Sharma, P. K.; Xiang, Y.; Kato, M.; Warshel, A. What are the Roles of Substrate-assisted Catalysis and Proximity Effects in Peptide Bond Formation by the Ribosome? *Biochemistry* **2005**, *44*, 11307−11314.
(1281) Trobro, S.; Aqvist, J. Mechanism of Peptide Bond Synthesis on the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 12395−12400.
(1282) Trobro, S.; Aqvist, J. Analysis of Predictions for the Catalytic Mechanism of Ribosomal Peptidyl Transfer. *Biochemistry* **2006**, *45*, 7049−7056.
(1283) Weinger, J. S.; Parnell, K. M.; Dorner, S.; Green, R.; Strobel, S. A. Substrate-assisted Catalysis of Peptide Bond Formation by the Ribosome. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1101−1106.
(1284) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. The Ribosome as an Entropy Trap. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7897−7901.
(1285) Johansson, M.; Ieong, K. W.; Trobro, S.; Strazewski, P.; Aqvist, J.; Pavlov, M. Y.; Ehrenberg, M. pH-sensitivity of the Ribosomal Peptidyl Transfer Reaction Dependent on the Identity of the A-site Aminoacyl-tRNA. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 79−84.
(1286) Acosta-Silva, C.; Bertran, J.; Branchadell, V.; Oliva, A. Theoretical Study of a Proton Wire Mechanism for the Peptide Bond Formation in the Ribosome. *Theor. Chem. Acc.* **2017**, *136*, No. e49, DOI: 10.1007/s00214-017-2066-2.
(1287) Petry, S.; Brodersen, D. E.; Murphy, F. V.; Dunham, C. M.; Selmer, M.; Tarry, M. J.; Kelley, A. C.; Ramakrishnan, V. Crystal Structures of the Ribosome in Complex with Release Factors RF1 and RF2 Bound to a Cognate Stop Codon. *Cell* **2005**, *123*, 1255−1266.
(1288) Klaholz, B. P.; Pape, T.; Zavialov, A. V.; Myasnikov, A. G.; Orlova, E. V.; Vestergaard, B.; Ehrenberg, M.; Van Heel, M. Structure of the Escherichia coli Ribosomal Termination Complex with Release Factor 2. *Nature* **2003**, *421*, 90−94.
(1289) Rawat, U. B. S.; Zavialov, A. V.; Sengupta, J.; Valle, M.; Grassucci, R. A.; Linde, J.; Vestergaard, B.; Ehrenberg, M.; Frank, J. A Cryo-electron Microscopic Study of Ribosome-bound Termination Factor RF2. *Nature* **2003**, *421*, 87−90.
(1290) Vestergaard, B.; Sanyal, S.; Roessle, M.; Mora, L.; Buckingham, R. H.; Kastrup, J. S.; Gajhede, M.; Svergun, D. I.; Ehrenberg, M. The SAXS Solution Structure of RF1 Differs from its Crystal Structure and Is Similar to its Ribosome Bound Cryo-EM Structure. Mol. *Mol. Cell* **2005**, *20*, 929−938.
(1291) Song, H. W.; Mugnier, P.; Das, A. K.; Webb, H. M.; Evans, D. R.; Tuite, M. F.; Hemmings, B. A.; Barford, D. The Crystal Structure of Human Eukaryotic Release Factor eRF1-Mechanism of Stop Codon Recognition and Peptidyl-tRNA Hydrolysis. *Cell* **2000**, *100*, 311−321.
(1292) Trobro, S.; Aqvist, J. A Model for How Ribosomal Release Factors Induce Peptidyl-tRNA Cleavage in Termination of Protein Synthesis. Mol. *Mol. Cell* **2007**, *27*, 758−766.
(1293) Trobro, S.; Aqvist, J. Mechanism of the Translation Termination Reaction on the Ribosome. *Biochemistry* **2009**, *48*, 11296−11303.
(1294) Korostelev, A.; Asahara, H.; Lancaster, L.; Laurberg, M.; Hirschi, A.; Zhua, J. Y.; Trakhanov, S.; Scott, W. G.; Noller, H. F. Crystal Structure of a Translation Termination Complex Formed with Release Factor RF2. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19684−19689.
(1295) Weixlbaumer, A.; Jin, H.; Neubauer, C.; Voorhees, R. M.; Petry, S.; Kelley, A. C.; Ramakrishnan, V. Insights into Translational

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

Termination from the Structure of RF2 Bound to the Ribosome. *Science* **2008**, *322*, 953−956.
(1296) Shaw, J. J.; Trobro, S.; He, S. L.; Aqvist, J.; Green, R. A Role for the 2 ' OH of Peptidyl-tRNA Substrate in Peptide Release on the Ribosome Revealed through RF-Mediated Rescue. *Chem. Biol.* **2012**, *19*, 983−993.
(1297) Giambasu, G. M.; Lee, T. S.; Scott, W. G.; York, D. M. Mapping L1 Ligase Ribozyme Conformational Switch. *J. Mol. Biol.* **2012**, *423*, 106−122.
(1298) Robertson, M. P.; Scott, W. G. The Structural Basis of Ribozyme-catalyzed RNA Assembly. *Science* **2007**, *315*, 1549−1553.
(1299) Shechner, D. M.; Grant, R. A.; Bagby, S. C.; Koldobskaya, Y.; Piccirilli, J. A.; Bartel, D. P. Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme. *Science* **2009**, *326*, 1271−1275.
(1300) Berezniak, T.; Jaschke, A.; Smith, J. C.; Imhof, P. Stereoselection in the Diels-alderase Ribozyme: A Molecular Dynamics Study. *J. Comput. Chem.* **2012**, *33*, 1603−1614.
(1301) Berezniak, T.; Zahran, M.; Imhof, P.; Jaschke, A.; Smith, J. C. Magnesium-Dependent Active-Site Conformational Selection in the Diels-Alderase Ribozyme. *J. Am. Chem. Soc.* **2010**, *132*, 12587−12596.
(1302) Lopez, X.; York, D. M.; Dejaegere, A.; Karplus, M. Theoretical Studies on the Hydrolysis of Phosphate Diesters in the Gas Phase, Solution, and RNase A. *Int. J. Quantum Chem.* **2002**, *86*, 10−26.
(1303) Formoso, E.; Matxain, J. M.; Lopez, X.; York, D. M. Molecular Dynamics Simulation of Bovine Pancreatic Ribonuclease A-CpA and Transition State-like Complexes. *J. Phys. Chem. B* **2010**, *114*, 7371−7382.
(1304) Gu, H.; Zhang, S. M.; Wong, K. Y.; Radak, B. K.; Dissanayake, T.; Kellerman, D. L.; Dai, Q.; Miyagi, M.; Anderson, V. E.; York, D. M.; et al. Experimental and Computational Analysis of the Transition State for Ribonuclease A-catalyzed RNA 2 '-O-transphosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13002−13007.
(1305) Harris, M. E.; Piccirilli, J. A.; York, D. M. Integration of Kinetic Isotope Effect Analyses to Elucidate Ribonuclease Mechanism. *Biochim. Biophys. Acta, Proteins Proteomics* **2015**, *1854*, 1801−1808.
(1306) Nowotny, M.; Cerritelli, S. M.; Ghirlando, R.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Specific Recognition of RNA/DNA Hybrid and Enhancement of Human RNase H1 Activity by HBD. *EMBO J.* **2008**, *27*, 1172−1181.
(1307) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W. Crystal Structures of RNase H Nound to an RNA/DNA Hybrid: Substrate Specificity and Metal-dependent Catalysis. *Cell* **2005**, *121*, 1005−1016.
(1308) Nowotny, M.; Gaidamakov, S. A.; Ghirlando, R.; Cerritelli, S. M.; Crouch, R. J.; Yang, W. Structure of Human RNase h1 Complexed with an RNA/DNA Hybrid: Insight into HIV Reverse Transcription. Mol. *Mol. Cell* **2007**, *28*, 264−276.
(1309) Nowotny, M.; Yang, W. Stepwise Analyses of Metal Ions in RNase H Catalysis from Substrate Destabilization to Product Release. *EMBO J.* **2006**, *25*, 1924−1933.
(1310) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester Cleavage in Ribonuclease H Occurs via an Associative Two-metal-aided Catalytic Mechanism. *J. Am. Chem. Soc.* **2008**, *130*, 10955−10962.
(1311) Elsasser, B.; Fels, G. Atomistic Details of the Associative Phosphodiester Cleavage in Human Ribonuclease H. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11081−11088.
(1312) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic Mechanism of RNA Backbone Cleavage by Ribonuclease H from Quantum Mechanics/Molecular Mechanics Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 8934−8941.
(1313) Haruki, M.; Tsunaka, Y.; Morikawa, M.; Iwai, S.; Kanaya, S. Catalysis by Escherichia coli Ribonuclease HI is Facilitated by a Phosphate Group of the Substrate. *Biochemistry* **2000**, *39*, 13939−13944.
(1314) Uchiyama, Y.; Miura, Y.; Inoue, H.; Ohtsuka, E.; Ueno, Y.; Ikehara, M.; Iwai, S. Studies of the Interactions between Escherichia-Coli Ribonuclease Hi and Its Substrate. *J. Mol. Biol.* **1994**, *243*, 782−791.
(1315) Sgrignani, J.; Magistrato, A. QM/MM MD Simulations on the Enzymatic Pathway of the Human Flap Endonuclease (hFEN1) Elucidating Common Cleavage Pathways to RNase H Enzymes. *ACS Catal.* **2015**, *5*, 3864−3875.
(1316) Haurwitz, R. E.; Jinek, M.; Wiedenheft, B.; Zhou, K.; Doudna, J. A. Sequence- and Structure-specific RNA Processing by a CRISPR Endonuclease. *Science* **2010**, *329*, 1355−1358.
(1317) Šponer, J. E.; Šponer, J.; Nováková, O.; Brabec, V.; Šedo, O.; Zdráhal, Z.; Costanzo, G.; Pino, S.; Saladino, R.; Di Mauro, E. Emergence of the First Catalytic Oligonucleotides in a Formamide-Based Origin Scenario. *Chem. - Eur. J.* **2016**, *22*, 3572−3586.
(1318) Sponer, J. E.; Szabla, R.; Gora, R. W.; Saitta, A. M.; Pietrucci, F.; Saija, F.; Di Mauro, E.; Saladino, R.; Ferus, M.; Civis, S.; et al. Prebiotic Synthesis of Nucleic Acids and their Building Blocks at the Atomic Level - Merging Models and Mechanisms from Advanced Computations and Experiments. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20047−20066.
(1319) Gupta, A.; Bansal, M. Local Structural and Environmental Factors Define the Efficiency of an RNA Pseudoknot Involved in Programmed Ribosomal Frameshift Process. *J. Phys. Chem. B* **2014**, *118*, 11905−11920.
(1320) Zhang, Y.; Zhang, J.; Wang, W. Atomistic Analysis of Pseudoknotted RNA Unfolding. *J. Am. Chem. Soc.* **2011**, *133*, 6882−6885.
(1321) Bian, Y. Q.; Zhang, J.; Wang, J.; Wang, J. H.; Wang, W. Free Energy Landscape and Multiple Folding Pathways of an H-Type RNA Pseudoknot. *PLoS One* **2015**, *10*, e0129089.
(1322) Jung, S.; Schlick, T. Candidate RNA Structures for Domain 3 of the Foot-and-mouth-disease Virus Internal Ribosome Entry Site. *Nucleic Acids Res.* **2013**, *41*, 1483−1495.
(1323) Jung, S.; Schlick, T. Interconversion between Parallel and Antiparallel Conformations of a 4H RNA junction in Domain 3 of Foot-and-Mouth Disease Virus IRES Captured by Dynamics Simulations. *Biophys. J.* **2014**, *106*, 447−458.
(1324) Frank, J.; Ruben L. Gonzalez, J. Structure and Dynamics of a Processive Brownian Motor: The Translating Ribosome. *Annu. Rev. Biochem.* **2010**, *79*, 381−412.
(1325) Munro, J. B.; Altman, R. B.; Tung, C. S.; Cate, J. H. D.; Sanbonmatsu, K. Y.; Blanchard, S. C. Spontaneous Formation of the Unlocked State of the Ribosome is a Multistep Process. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 709−714.
(1326) Juneja, A.; Nilsson, L.; Villa, A. Effect of Mutations on Internal Dynamics of an RNA Hairpin from Hepatitis B Virus. *Biophys. Chem.* **2016**, *218*, 7−13.
(1327) Macchion, B. N.; Strömberg, R.; Nilsson, L. Analysis of the Stability and Flexibility of RNA Complexes Containing Bulge Loops of Different Sizes. *J. Biomol. Struct. Dyn.* **2008**, *26*, 163−173.
(1328) Li, W.; Frank, J. Transfer RNA in the Hybrid P/E State: Correlating Molecular Dynamics Simulations with Cryo-EM Data. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16540−16545.
(1329) Singharoy, A.; Teo, I.; McGreevy, R.; Stone, J. E.; Zhao, J. H.; Schulten, K. Molecular Dynamics-based Model Refinement and Validation for Sub-5 Angstrom Cryo-electron Microscopy Maps. *eLife* **2016**, *5*, No. e16105, DOI: 10.7554/eLife.16105.
(1330) Trabuco, L. G.; Villa, E.; Mitra, K.; Frank, J.; Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* **2008**, *16*, 673−683.
(1331) Villa, E.; Sengupta, J.; Trabuco, L. G.; LeBarron, J.; Baxter, W. T.; Shaikh, T. R.; Grassucci, R. A.; Nissen, P.; Ehrenberg, M.; Schulten, K.; et al. Ribosome-induced Changes in Elongation Factor Tu Conformation Control GTP Hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 1063−1068.
(1332) Ahmed, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Tama, F. Consensus Among Flexible Fitting Approaches Improves the Interpretation of Cryo-EM Data. *J. Struct. Biol.* **2012**, *177*, 561−570.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX

(1333) Wilson, D. N.; Beckmann, R. The Ribosomal Tunnel as a Functional Environment for Nascent Polypeptide Folding and Translational Stalling. *Curr. Opin. Struct. Biol.* **2011**, *21*, 274−282.
(1334) Seidelt, B.; Innis, C. A.; Wilson, D. N.; Gartmann, M.; Armache, J.-P.; Villa, E.; Trabuco, L. G.; Becker, T.; Mielke, T.; Schulten, K.; et al. Structural Insight into Nascent Polypeptide Chain−Mediated Translational Stalling. *Science* **2009**, *326*, 1412−1415.
(1335) Fulle, S.; Gohlke, H. Statics of the Ribosomal Exit Tunnel: Implications for Cotranslational Peptide Folding, Elongation Regulation, and Antibiotics Binding. *J. Mol. Biol.* **2009**, *387*, 502−517.
(1336) Ishida, H.; Hayward, S. Path of Nascent Polypeptide in Exit Tunnel Revealed by Molecular Dynamics Simulation of Ribosome. *Biophys. J.* **2008**, *95*, 5962−5973.
(1337) Trabuco, L. G.; Harrison, C. B.; Schreiner, E.; Schulten, K. Recognition of the Regulatory Nascent Chain TnaC by the Ribosome. *Structure* **2010**, *18*, 627−637.
(1338) Ge, X.; Roux, B. Absolute Binding Free Energy Calculations of Sparsomycin Analogs to the Bacterial Ribosome. *J. Phys. Chem. B* **2010**, *114*, 9525−9539.
(1339) Sanbonmatsu, K. Y.; Joseph, S. Understanding Discrimination by the Ribosome: Stability Testing and Groove Measurement of Codon-anticodon Pairs. *J. Mol. Biol.* **2003**, *328*, 33−47.
(1340) Allnér, O.; Nilsson, L. Nucleotide Modifications and tRNA Anticodon−mRNA Codon Interactions on the Ribosome. *RNA* **2011**, *17*, 2177−2188.
(1341) Burton, B.; Zimmermann, M. T.; Jernigan, R. L.; Wang, Y. A Computational Investigation on the Connection between Dynamics Properties of Ribosomal Proteins and Ribosome Assembly. *PLoS Comput. Biol.* **2012**, *8*, e1002530.
(1342) Caulfield, T.; Devkota, B. Motion of Transfer RNA from the A/T State into the A-site using Docking and Simulations. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 2489−2500.
(1343) Yingling, Y. G.; Shapiro, B. A. Dynamic Behavior of the Telomerase RNA Hairpin Structure and its Relationship to Dyskeratosis Congenita. *J. Mol. Biol.* **2005**, *348*, 27−42.
(1344) Yildirim, I.; Park, H.; Disney, M. D.; Schatz, G. C. A Dynamic Structural Model of Expanded RNA CAG Repeats: A Refined X-ray Structure and Computational Investigations Using Molecular Dynamics and Umbrella Sampling Simulations. *J. Am. Chem. Soc.* **2013**, *135*, 3528−3538.
(1345) Yildirim, I.; Chakraborty, D.; Disney, M. D.; Wales, D. J.; Schatz, G. C. Computational Investigation of RNA CUG Repeats Responsible for Myotonic Dystrophy 1. *J. Chem. Theory Comput.* **2015**, *11*, 4943−4958.
(1346) Bochicchio, A.; Rossetti, G.; Tabarrini, O.; Krau*β*, S.; Carloni, P. Molecular View of Ligands Specificity for CAG Repeats in Anti-Huntington Therapy. *J. Chem. Theory Comput.* **2015**, *11*, 4911−4922.
(1347) Chen, J. L.; VanEtten, D. M.; Fountain, M. A.; Yildirim, I.; Disney, M. D. Structure and Dynamics of RNA Repeat Expansions That Cause Huntington's Disease and Myotonic Dystrophy Type 1. *Biochemistry* **2017**, *56*, 3463−3474.
(1348) Pan, F.; Man, V. H.; Roland, C.; Sagui, C. Structure and Dynamics of DNA and RNA Double Helices of CAG and GAC Trinucleotide Repeats. *Biophys. J.* **2017**, *113*, 19−36.

DOI: 10.1021/acs.chemrev.7b00427
Chem. Rev. XXXX, XXX, XXX−XXX