Computational approaches to the investigation of proteins involved in Parkinson's Disease

Thesis submitted for the degree of
Doctor Philosophiæ
in Statistical and Biological Physics

Candidate:

Fernando Enrique Herrera

Supervisor:

Prof. Paolo Carloni

February, 2008

to my wife MARIELA

Contents

List of Figures

List of Tables

Chapter 1

Motivation and outline

1.1 Motivation of the present work

Parkinson's disease (PD) is the second most spread neurodegenerative disease, after Alzheimer's, affecting 1-2 % of world population above the age of 65 [1, 2]. There are two forms of the disease. The so-called 'sporadic form' is caused by environmental conditions and, probably, the genetic constitution of the individuals [3]. The 'familial form' ($\sim$ 10 % of all the PD cases) is instead related only to the genetic predisposition of certain families (mutations on proteins involved in PD).

In this thesis, we have used molecular simulation tools to investigated proteins involved in the familial PD, in a joint collaborative effort with the neurobiology group headed by Prof. S. Gustincich at SISSA. Several are the proteins which have been found to play a key role for the derangement of this form (Tab. 2.1). Among them, α-synuclein (AS)- the first protein discovered to be related with PD, a decade ago [4] - is one of the most actively investigated. It is the main component, in the form of fibrils, of the so-called Lewy Bodies (LB). LB' are plaques found in *post-mortem* brains which constitute the hallmark of PD. Its role in PD is further confirmed by the fact that AS mutations are found in patients suffering from the disease.

Several are the proposed functions of AS, from regulation of dopamine release and transport [5], to lipid metabolism [6], from vesicle-mediated transport [7], to trafficking within the endoplasmic reticulum/Golgi network [8], regulation of dopamine metabolism [9, 10, 11] and chaperone activity [12, 13]. A striking feature of AS is that the fact that the protein is totally unstructured in aqueous solution: NMR investigations have provided several thousands structures compatible with the experimental data [14, 15] (Fig. 4.1). AS may form fibrils spontaneously after a couple of days *in vitro* from its aqueous solution; thus, any mean which would

interfere with the such process might impact on the derangement of the disease. In this respect, the recent discovery that non-covalent binding of ligands such as dopamine and its derivative inhibits AS fibrillation [16, 17] is exciting - it gives the opportunity for future design of ligands interfering with the fibrillation process- and intriguing at the same time: how can binding occur in such an unstructured protein? This issue even more general, as many as 400 proteins have been shown experimentally to be either completely or partially disordered [18, 19, 20](See article review of Peter Wright on the subject [21]), many of them implicated in neurodegenerative diseases [22, 23, 24]. Therefore, ligand binding to unstructured protein may impact largely on drug development [25, 26].

Surprisingly, however, neither theoretical investigations on the AS subject nor theoretical studies involving the interactions between ligands and unstructured proteins have been done before the starting of this thesis.

Here we have attempted to gain insights on the structural determinants of dopamine binding. Clearly, this is not a trivial task, as the established and standard principles which apply to docking and design of ligands targeting proteins with a defined 3D fold are challenged here. We have hence used a rather elaborated protocol which involves cluster analysis, docking and molecular dynamics simulations. Based on our calculations, we have been able to provide structural models of the AS/dopamine adducts, fully consistent with experimental data reported in the literature [16, 17] and with experiments performed within this joint project by Prof. Gustincich at SISSA and Prof. Lashuel at EPFL, Lausanne. In addition, the computational studies have been used to design new ligands which might be able to inhibit AS fibrillation.

A second, experimentally well-characterized protein in familial PD is the DJ-1 protein [27, 28]. Several DJ-1's mutations emerge in patients suffering from the disease, and they may affect some of the several functions proposed for the protein, including a redox-regulated chaperone activity [29], a regulator of transcription activity [30], and protection the cell against oxidative stress conditions -a possible cause of PD [31, 32]. A major issue relates the PD-linked DJ-1 mutants to the protein's oligomeric state. Whilst the WT is a dimer, mutations may alter the state and render the protein monomeric or multimeric. This of course might play a role for the disease [27]. Here we have applied bioinformatics and molecular simulation tools to investigate the relationships between PD-linked mutants and the protein structure. The computational methodology was validated by predicting mutations which destabilizes the dimeric form [33] and verifying the prediction from experiments performed in Prof. Gustincich's lab.

1.2 Outline of the thesis

Chapter 2 introduces Parkinson's Disease and the proteins studied in this thesis. The next, (Chapter 3) presents the computational tools used in this work: The bionformatics Section focuses on cluster analysis, molecular docking approaches, molecular screening and computational alanine scanning. The molecular simulation section outlines some of the key principles of molecular dynamics simulation and force field development.

Chapter 4 features a computational study of AS-dopamine interactions. This study can be helpful in understanding the fibrillation process and might be used to design new ligands with similar characteristics and interactions as those of dopamine.

The following results of my work are presented in Chapter 5, where new compounds with similar structural or electrostatic properties of dopamine are identified. These compounds might share some of the binding features of dopamine and its derivatives towards AS.

Chapter 6 focuses on the computational approach for the study of the DJ-1 protein. The structural properties of the DJ-1 WT protein in the reduced and oxidized state along with that of two PD-linked mutants (M26I and L166P) and one physiological variant (R98Q) are studied in order to gain insights into the mechanism involved in the disease. These studies allow us to design later a new mutant of the DJ-1 protein which affects the oligomeric state.

Most of the work presented has been validated against experimental data, most of which performed in a joint collaboration within SISSA.

The conclusions are summarized in Chapter 7.

Chapter 2

Introduction

2.1 Parkinson's Disease

2.1.1 Clinical and neuropathological features

Parkinson's disease (PD) is the second most spread neurodegenerative disease (after Alzheimer's). It is a disease for which there is currently no cure.

The average age of onset is 60 years, and the prevalence rises significantly from 0.3% in the entire population to about 1% in people over 60 years old and up to 5% in over 85 years old [1, 2]. However, about 5 to 10 % of people with PD have "early-onset" disease that begins before the age of 50 [34, 35, 36]. PD is a progressive neurodegenerative movement disorder: cardinal symptoms of the diseases include bradykinesia, resting tremor, rigidity, and postural instability, along with, at times, autonomic, cognitive, and psychiatric disturbances. These symptoms result from loss of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNc). Approximately 60-80% of DA are lost before the appearance of the first motor signs of PD. This causes a dysregulation in the motor circuits resulting in the clinical manifestations (symptoms) of PD [37].

The examination of brains from PD patients shows the depigmentation of the substantia nigra (SN). This is caused by the selective and progressive loss of DA and the presence the Lewy bodies (LBs) within the surviving neurons of the SN and other regions of the brain. The LBs - the hallmark of PD - are eosinophilic intracytoplasmic proteinaceous inclusions present in surviving neurons [38]. They enriched in fibrils of the α-synuclein protein and other (often ubiquinated) proteins (Fig. 2.1).

There are two forms of the disease, the sporadic and the familial forms. Patients with familial PD are distinguished from the ones that suffer from sporadic PD

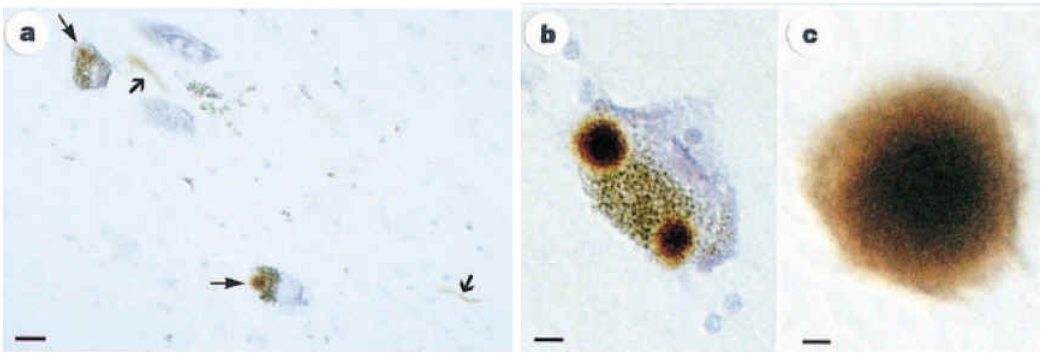


Figure 2.1: Substantia nigra from patients with Parkinson's disease immunostained for α-synuclein. a) Two pigmented nerve cells, each containing AS in the Lewy body (thin arrows) and Lewy neurites (thick arrows). b) A pigmented nerve cell with two α-synuclein-positive Lewy bodies. c) α-synuclein positive, extracellular Lewy body (Taken from [39]).

because of an early start, greater consanguinity rate and greater frequency of a similar disease in their parents.

The **sporadic form** of the disease has been suggested to spread from the interaction of humans with chemicals in the environment. There is a higher incidence of PD in developed countries. This may be due to the larger exposure to environmental toxins. Consistent with this proposal, a drug contaminant called MPTP, which bears a strong chemical similarity to many pesticides and other environmental chemicals, is able to induce PD [40].

The **familial cases** are about the 10% of the total number of cases and are based on the genetic component of the disease [41, 42, 43, 44]. The genes so far shown to be involved are 10 (Tab. 2.1). In four of these (AS, parkin, DJ-1 and pink1 genes), mutations have been identified that definitively and unambiguously cause familial forms of PD, whilst in the other 6 no mutations have been reported and they are involved in the sporadic form of PD. Age of onset tends to be younger in familial PD compared with sporadic PD, particularly in the autosomal recessive PD.

2.1.2 Known factors playing a role for PD

Misfolding and aggregation of proteins (i), mitochondrial dysfunction (ii), and oxidative stress (iii), play a key role for the progression of both forms of the disease.

(i) As in many other neurodegenerative diseases such as Alzheimer's and prion diseases, PD is characterized by neuronal loss and the formation of protein aggregates

Locus	Gene	Protein function	Phenotype
PARK1 PARK4 (AD)	SNCA	Synaptic ? Lipid binding ?	PD/DLBD. Onset from age 30 to 60, rapid course. Fulminat LB [4, 45].
PARK2 (AR)	Parkin	E3 ligase	Parkinsonism. Onset from teenage to 40s, slow course. No LB, except one case [46].
PARK3 (AD)	Unknown		PD, dementia. Onset age 50 to 60s. LB, tangles and plaques [47].
PARK5 (AD?)	UCHL1	Ubiquitin hydrolase/ligase	Typical PD. Onset at about 50 Unknown pathology [48].
PARK6 (AR)	PINK1	Protein kinase	Parkinsonism. Onset age 30 to 50s. Unknown pathology [49].
PARK7 (AR)	DJ-1	Oxidative stress response ?	Parkinsonism. Onset from age 20 to 40s, slow course. Unknown pathology [50].
PARK8 (AD)	LRRK2		Parkinsonism. Onset from age 40 to 60s. Variable pathology [51].
PARK10 (unclear)	Unknown		Typical PD. Onset at about 60, Unknown pathology [42].
PARK11 (unclear)	Unknown		Typical PD. Onset at about 60, Unknown pathology [52].

Table 2.1: Monogenic forms of Parkinson's disease. The inheritance is shown in parentheses (AD, autosomal dominant and AR, autosomal recessive). PD: Parkinson's disease. DLBD: diffuse Lewy body disease. The SNCA is the gene name for alpha-synuclein (Taken from [53]).

that are typically fibrillar. Cell death could arise from such protein aggregates [54, 55], or by ordered prefibrillar, oligomer, or protofibrils [56]. In the latter case, the fibrils could be neuroprotective [56]. Nevertheless, in any case, the molecular mechanism of fibrillation is currently largely unknown.

(ii) PD *post mortem* brains feature a reduction of mitochondrial complex I activity of the mitochondrial respiratory chain [40]. Further evidence in support of the role of mithochondrial dysfunction in PD came from the discovery of some of the familial PD genes. The discovery of PINK1, HtrA serine peptidase 2 (Omi/HtrA2) and DJ-1 mutations confirmed that mitochondrial dysfunction is a pathway to parkinsonism. In particular, the subcellular distribution of the DJ-1 protein is primarily cytoplasmic with a smaller pool of mitochondrial and nuclear associated protein [57]. Furthermore, the DJ-1 mutations disrupt the protein activity by either destabilizing the protein or affecting its subcellular localization. All DJ-1 PD-linked mutations show reduced nuclear localization in favor of mitochondrial localization [50, 58]. Nevertheless, it is not clear whether the toxicity associated with increased mitochondrial localization is due to a mitochondrial gain of function or to a loss of access to binding partners in different cellular compartments [58].

(iii) Several factors involved in PD points toward the role of the oxidative stress into the progression of the disease: A. the inhibition of complex I increases the production of reactive oxygen species (ROS) [59]; B. DA neurons may be a particularly fertile environment for the generation of ROS, as the metabolism of dopamine (DOP) produces hydrogen peroxide and superoxyde radicals (see appendix A.1); and C. the DJ-1 protein may protect the cells against oxidative stress-induced death [31, 32] and its ablation sensitizes cells to oxidative stress [60].

2.2 Alpha-synuclein and DJ-1 proteins in familial cases of PD

2.2.1 Alpha-Synuclein and its inhibitor of fibrillization dopamine

AS is the major component of the Lewy bodies and Lewy neurites in familiar or sporadic PD [39], suggesting that the aggregation and fibrillation of this protein play a central role in the development and/or progression of the disease. Therefore, the role of AS in neurodegenerative disorders has become evident, and drug discovery efforts have been focusing on targets related to AS misfolding and fibrillization. AS is also found in the LB variant of Alzheimer's Disease, dementia with LB, multiple system atrophy and related diseases, which indeed are collectively known as α-synucleinopathies [61].

AS is an abundant, natively unfolded neuronal protein. It is highly soluble and its localization is both cytosolic and nuclear. It is enriched at the synaptic terminals, where it is associated with synaptic vesicles. It is an intrinsically unstructured protein but has significant conformational plasticity (depending on the environment AS can remain unstructured, can form monomeric and oligomeric species, or can form amyloidogenic filaments [62], it exists in solution as an unstructured monomer. Inside the cell, the monomer is in equilibrium with membrane-associated forms with higher helical content. Oligomers and other intermediates are kinetically stabilized by dopamine (DA). However, these are transient species that further aggregate to form mature fibrils. The formation of Lewy bodies is presumed to be a consequence of fibrillization. Events such as the attachment of ubiquitin are thought to be secondary to the initial aggregation and deposition processes [53] (see Fig. 2.2).

α-synuclein is a member of a family of three synaptic proteins that share considerable sequence homology and include α-synuclein, β-synuclein and γ-synuclein. Synucleins have in common a highly conserved β-helical lipid-binding motif with similarity to the class-A2 lipid-binding domains of the exchangeable apolipoproteins. The α- and β-synuclein proteins are found primarily in brain tissue, where

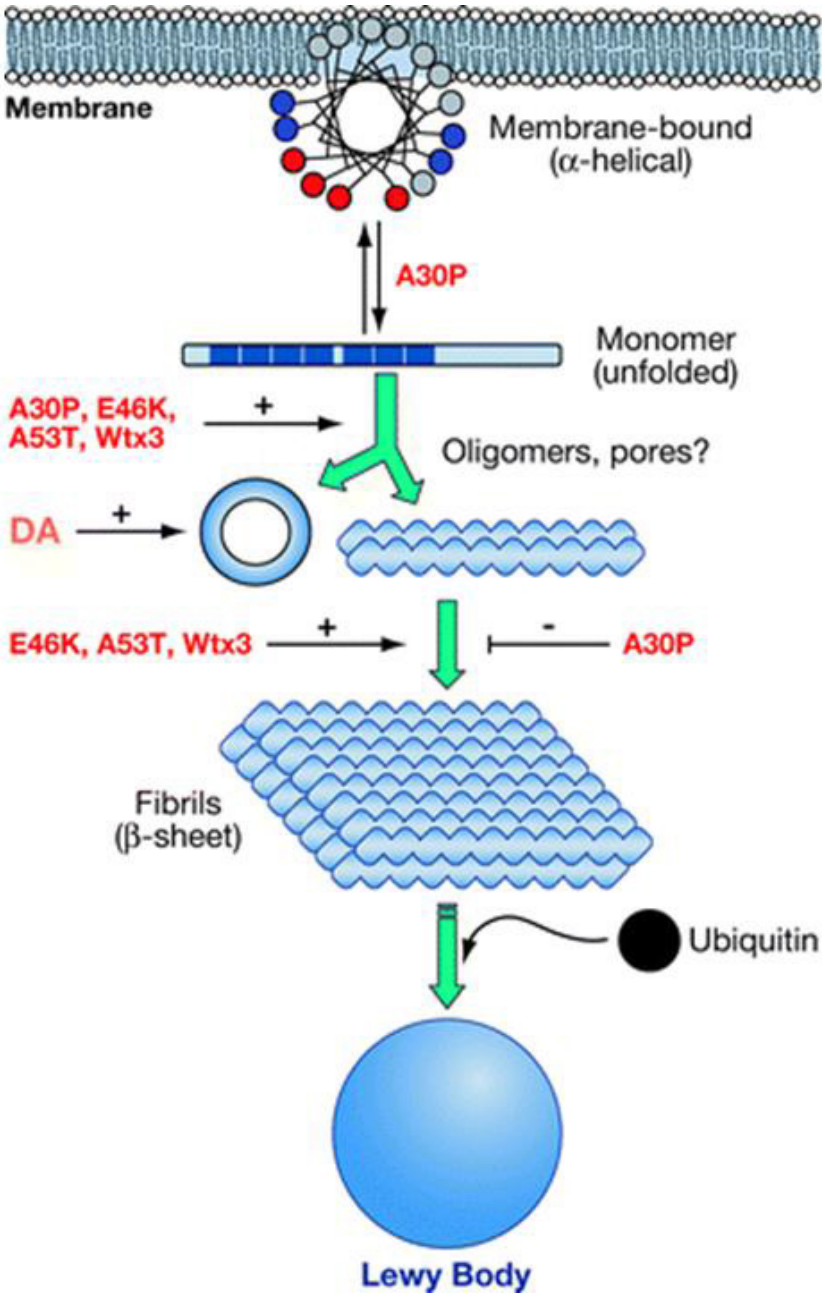


Figure 2.2: The pathogenic cascade of AS aggregation. The unstructured monomer is shown as a linear structure. In the helix (bounded to the membrane) blue and red circles indicate charged residues, gray circles are nonpolar and hydrophobic amino acids. The green arrows indicate the pathogenic formation of aggregated species. Taken from [53].

they are localized mainly in presynaptic terminals. The γ-synuclein protein is found primarily in the peripheral nervous system and in the retina; moreover its expression in breast tumors is a marker for tumor progression. Although normal cellular functions have not been determined for any of the synuclein proteins, there is evidence that the expression of β-synuclein may regulate the levels or the metabolism of α-synuclein, since β-synuclein inhibits α-synuclein aggregation in transgenic mice [63]. Furthermore, more recent studies indicate that increased expression of β-synuclein can decrease the levels of α-synuclein by mechanisms that do not appear to affect α-synuclein mRNA levels [64], but it is likely that the functions of β-synuclein extend beyond regulating the expression of α-synuclein.

The AS sequence consist of 140 amino acids, lacks both cysteine and tryptophan residues and can be divided into three different regions (Fig 2.3) [65, 66]:
(i) The positively charged N-terminal region (amino acids 1-60), comprising a seven imperfect 11 amino acids repeats containing the consensus sequence KTKEGV.
(ii) The non-amyloid-β-component (NAC) (amino acids 61-95) [67, 53, 68].
(iii) The negatively charged C-terminal region 96-140, which contains several sites for post-translational modifications and metal binding (Fig. 2.3)[69, 70].

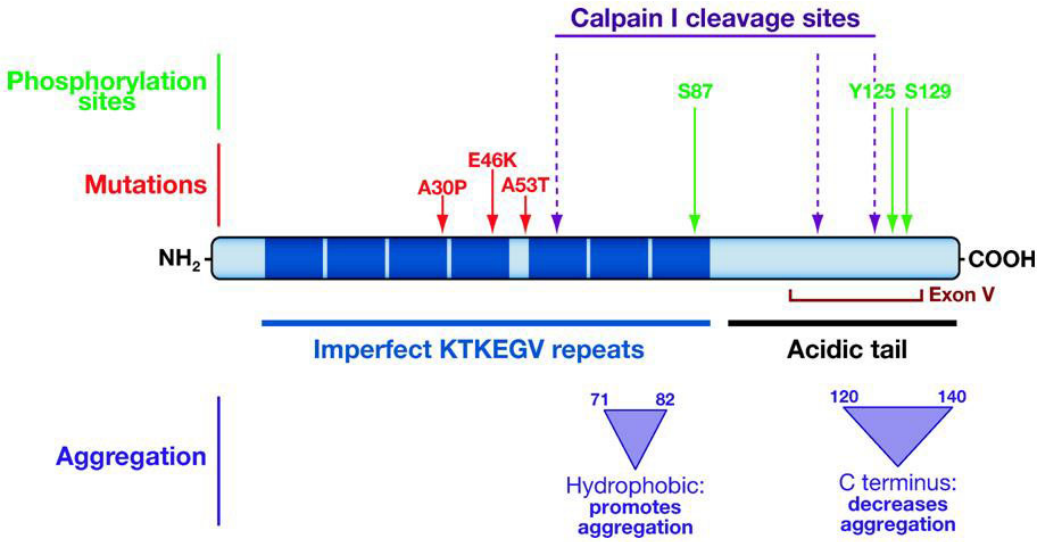


Figure 2.3: Motifs in the α-synuclein protein sequence. The natively unfolded α-synuclein protein is shown in a linear form. Shaded areas represent the imperfect KTKEGV repeats. Human mutations are shown in red and map to the repeat region. The phosphorylation sites are shown in green. Taken from [53].

The NAC region has been found in amyloid plaques associated with Alzheimer's disease, and several reports suggest that this region represents the critical determinant for the fibrillation process of AS [71] and it has been proposed that this region of the protein could promote the formation of β-amyloid *in vivo* [72, 73].

Several indications point to this region as the highly amyloidogenic part of the molecule. The acidic C-terminal tail of the protein contains several sites of phosphorylation. The C-terminus also contains the alternatively spliced exon V and a calpain I cleavage site. The acidic tail tends to decrease protein aggregation, whereas a hydrophobic region near the imperfect repeats promotes aggregation.

The AS's gene for familial PD was initially mapped to chromosome 4q21–q23 [74]. The mutations involved in the development of PD are: (i) the A53T missense mutation [4], (ii) the A30P mutation [75] and (iii) the E46K mutation [76]. In addition, the genomic triplication of a region spanning the AS gene were also found to be related with PD [45]. Patients with gene duplications are also found, but they have a less severe phenotype and a later age of onset than those with triplications, which implicates that variability of AS protein levels can predispose individuals to disease [77]. Nevertheless, familial PD caused by AS missense mutations is extremely rare.

Structure: Structural information, derived from NMR studies, is available for the monomer structure and has been increasingly used to study and model AS aggregation and interaction with other proteins [78]. In aqueous solution, AS exists as an ensemble of disordered conformations. NMR studies based on paramagnetic relaxation enhancement (PRE) effects generated an ensemble of ∼4,000 structures characterized by transient C-terminal - N-terminal interactions [15, 14]. Upon binding to SDS micelles and negatively charged synthetic vesicles [66], AS adopts partially an helical conformation: two helices (amino acids 1-38 and 44-94) are formed in the non-polar environment, and the remainder of the protein is disordered, as shown by NMR spectroscopy [66] (Fig. 2.4). The monomer is prone to aggregation into amyloid-like structures, particularly at high concentrations or upon exposure to various chemical and physical factors (*e.g.* shaking). AS fibrillation proceeds through a series of β-sheet rich aggregation intermediates, including early spherical protofibrils, pore-like, chain-like aggregates, which disappear once amyloid fibrils are formed [79, 80].

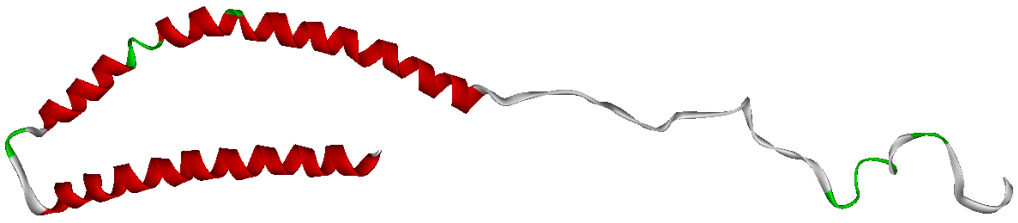


Figure 2.4: Structure of the micelle bound human alpha-synuclein (PDB code 1XQ8 [66]).

Fibrillation: During the *in vitro* fibrillization process of AS, early spherical protofibrils of AS are followed by chainlike and porelike protofibrils and finally with a rapid conversion into fibrils. Both A53T and A30P PD-associated mutations promote AS aggregation [79], but in the case of A53T, the fibril formation is very rapid, while in the case of A30P the formation of spherical protofibrils is promoted [81] with slow conversion into fibrils [79, 82]. This observation prompted the first proposal that the protofibril, rather than the fibril, may be pathogenic in PD [83, 84].

It is know that AS is able to bind membranes, this could be related with the undetermined physiological function of the protein [11, 8]. In addition, the affinity of AS for membranes is increased in the form of protofibrils and is decreased in the form of fibrils [80]. Furthermore, the protofibrillar form is capable of permeabilizing synthetic vesicles by a porelike mechanism [85, 86]. These amyloid pore structures are much more prevalent in the PD-linked mutant protofibril fractions.

All fibrils, independent of the original structure of the given amyloidogenic protein, have a common cross-β structure consisting of β-sheets in which the β-strands are perpendicular to the axis of the fibril (Fig. 2.5). In order to attain this structure, the proteins have to undergo significant conformational rearrangements to allow the needed topological changes. This suggests a simple kinetic model of fibrillation involving conversion of monomeric AS into the critical, partially folded intermediate, which leads, depending on the conditions, to the formation of soluble oligomers, insoluble amorphous aggregates, or insoluble fibrils.

Acceleration: The rate of the fibrillization is accelerated by di- and trivalent metal ions. The most effective cations *in vitro* include aluminum(III), copper(II), iron(III), and lead(II). The mechanism for this presumably involves the metal ions binding to negatively charged carboxylates, thus masking the electrostatic repulsion and facilitating collapse to the partially folded conformation.

Inhibition: A variety of flavonoids and polyphenols can instead inhibit the fibril formation of AS and in some cases also disaggregate existing fibrils [88]. These compounds bind to monomeric AS and lead to formation of stable AS oligomers. Another interesting class of compounds able to inhibit fibrillization of AS is that of dopamine and related catecholamines. Although several factors including post-translational modifications, oxidative stress and interaction with toxins and metals have been demonstrated to influence AS aggregation and toxicity [69, 70, 89, 90, 91], the exact molecular mechanisms underlying AS aggregation and toxicity *in vivo* and the links between aggregation and neurodegeneration remains poorly understood.

Dopamine as AS inhibitor of fibrillation: A role for dopamine (DOP) in mediating α-synuclein toxicity is a plausible hypothesis to account for the relatively selective degeneration of DA neurons in PD. Indeed, (i) the toxic effects of AS

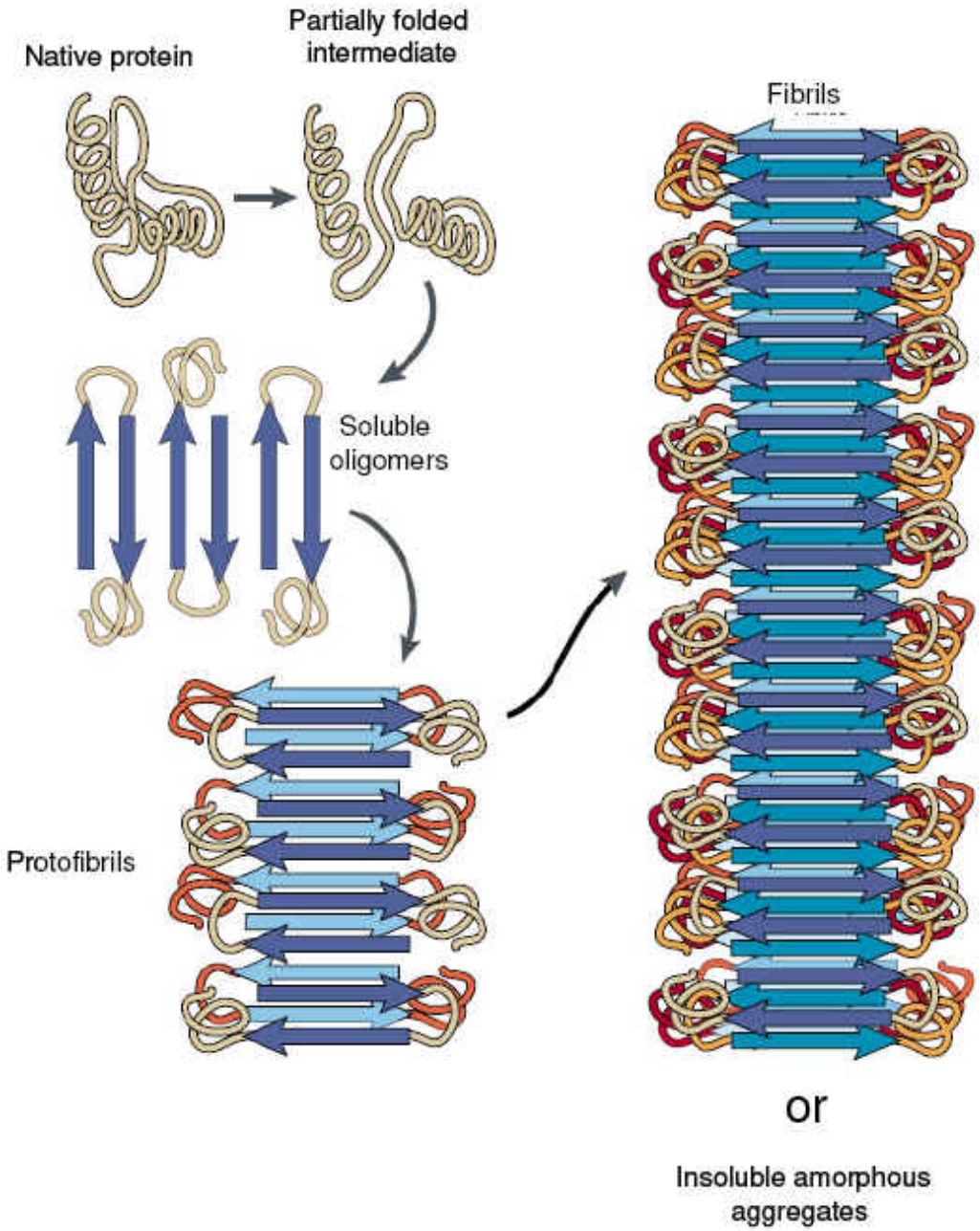


Figure 2.5: The α-synuclein fibrillization process (adapted from [87]).

appear to be selective for DA [92], (ii) AS can interact with and enhance the activity of the dopamine transporter [93] and (iii) overexpression of mutant α-synuclein can also downregulate the vesicular monoamine transporter 2 [11].

Several evidences also point to an important role of AS in regulation of the DOP metabolism at multiple levels: (i) the AS interaction with tyrosine hydroxylase (TH) inhibits the synthesis of DOP [94] (see Appendix A.1), (ii) AS is involved in vesicles formation and recycling through interaction with PLD2 [95] and binds fatty acids in vesicles with different affinities among the pathological mutants [10], (iii) VMAT2 expression is decreased by AS A53T over-expression [96] and (iv) AS binds to DAT and regulates its trafficking [93, 9].

Furthermore, *in vitro* catechol derivatives [97, 98], including dopamine [16, 99, 100], have been shown to inhibit AS fibrillogenesis causing accumulation of oligomeric species *in vitro* [16, 98]. These studies suggest that oxidation of dopamine may play a key role in modulating AS aggregation and toxicity and may be linked to the selective vulnerability of dopaminergic neurons in PD [101]. Dopamine forms covalent [98, 99] and/or non-covalent [16, 17] DOP-AS complexes. The non-covalent complexes were shown to involve interactions between dopamine and the C-terminal residues ^{125}YEMPS^{129} in the C-terminal region of the AS sequence [16].

Further information on this topic is presented in the Thesis Appendix A.2.

2.2.2 DJ-1 and its PD-linked mutants

DJ-1 is a highly conserved, ubiquitously expressed protein (Fig. 2.6) [102]. The human protein has 189 amino acids and belongs to the DJ-1/ThiJ/PfpI superfamily (Fig. 2.6).

DJ-1 is involved in multiple cellular processes including sperm maturation, fertilization in rodents and oncogenesis in humans. It is present in the nucleus, cytoplasm, mitochondria and extracellular space of mammalian cells [103, 104, 105, 106]. *In vitro* studies showed that the ectopic expression of DJ-1 protects cells from cell death induced by various toxic stimuli including oxidative stress [58, 107]. DJ-1 is indeed an indicator of oxidative stress state *in vivo* since it is converted into pI variants in response to small amounts of reactive oxygen species. Interestingly, some pI isoforms are accumulated in PD *post mortem* brains [108, 109].

It is abundantly expressed in most mammalian tissues, including the brain, where it is localized to both neurons and glia [108, 111]. DJ-1 does not appear to be localized to LBs in sporadic PD and other synucleinopathies but it does colocalize with tau-positive inclusions in a number of neurodegenerative tauopathies and with AS-positive glial inclusions in multiple system atrophy [112, 113], which suggests

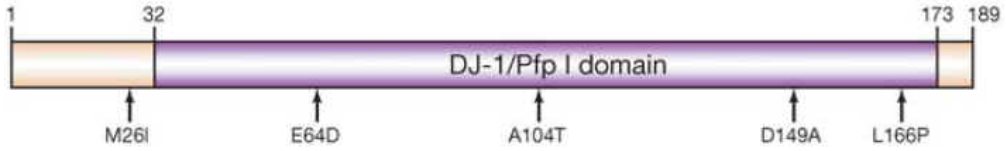


Figure 2.6: Domain architecture and familial mutations of DJ-1. The position of amino acid missense mutations (single letter code) associated with familial forms of PD are indicated (Taken from [110]).

that DJ-1 may also play a diverse role in seemingly distinct neurodegenerative diseases.

Several familiar cases prove that DJ-1 loss in humans causes PD: (i) a frame shift and a splice mutation were found in a young-onset PD patient [114]; (ii) a Dutch family showed a large homozygous genomic deletion that removed 4 Kb of promoter sequences and the first five exons of the gene [115]; (iii) additional PD cases have been described with truncating, splice-site mutations and deletions.

Structure: X-ray crystallography (Fig. 2.7) showed that WT DJ-1 in the reduced state (that is, in a state in which Cys and/or Met residues are not oxidized [116]) is a homodimer. Each subunit assumes an α/β sandwich fold, similar to the flavodoxin-like Rossman fold [116], conserved across the ThiJ-PfpI superfamily. It contains seven parallel β-strands flanked by nine α-helices. The subunit-subunit interface is stabilized by a large number of hydrophobic interactions, H-bonds and salt bridges. The dimeric state of the protein has been confirmed by both *in vivo* and *in vitro* investigations. It is believed that DJ-1 carries out its function exclusively in the dimeric state [117, 118, 119, 120], since the formation of High Molecular Weight (HMW) oligomers and/or inefficient formation of dimeric structures may affect the stability of the protein and/or its affinity for molecular partners in the cell [27, 102, 121].

Structure/function relationships in PD-linked mutations.

Some PD families present missense mutations of DJ-1 in homozygous and/or heterozygous forms (M26I, E64D, A104T, D149A and L166P) [28, 115, 50, 114, 122, 123, 124]. The mechanism by which these mutants change or abolish DJ-1 function is still a matter of debate.

L166P and M26I are the most studied DJ-1 missense mutations [121, 125, 50, 120, 111]. In particular L166P is very unstable and its expression level, both in transfection studies and in patient lymphoblasts, is lower than in the WT protein. This suggests that L166P mutation may induce a loss of DJ-1 function [111, 28, 126, 127]. The mutant does not form a dimeric structure; instead, it may assemble

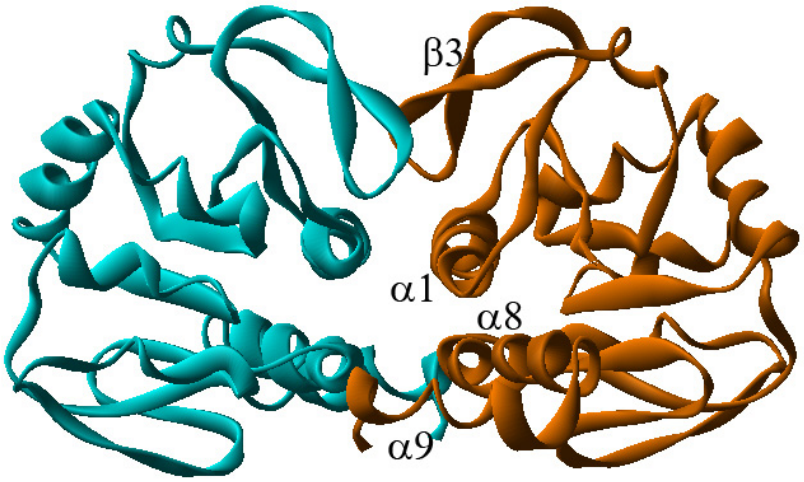


Figure 2.7: DJ-1 WT structure. X-ray structure in the reduced state (PDB code 1UCF) [116]. The two monomers are colored in light blue and brown respectively.

in HMW oligomers [121, 119, 111]. At the structural level, L166 is located in the middle of one of the helical regions of the protein (α8 helix in Fig. 2.7), thus its mutation to Pro breaks the helix [102, 50, 123, 111]. Because the α8 helix forms hydrophobic interactions with a series of residues located at the subunit-subunit interface (V181, K182, L187 of α9 helix and the C-term), such a mutation could affect the stability of the homodimer [102, 123, 28]. However, this hypothesis has not yet been proven by *in silico* and/or *in vitro* analysis of the conformational changes in the mutant.

M26I does form dimers but it is a matter of debate whether dimer formation occurs at the same rate as WT or less efficiently. Furthermore, the relevance of these differences, if any, for neurodegeneration is unclear [121, 125, 120]. This mutation is located on α1 helix at the subunit-subunit interface (Fig. 2.7), therefore it may affect subunit-subunit interactions. However, methionine is structurally and chemically very similar to isoleucine: they are similar in shape and volume and they are both non-polar residues. Thus, the interactions between M26I subunits may well be similar to those of the WT, although structural information for this mutant is lacking.

Chapter 3

Methods

This chapter describes the methodologies used in this thesis. It is divided into two parts. The first part is related to the bioinformatics and molecular docking. It includes different cluster algorithm, computational alanine scanning and molecular screening protocols used in this thesis. The basic concepts of molecular dynamics simulations are discussed in the second part. The specific features of each of the studied systems are reported in the Computational Details section of the corresponding chapters.

3.1 Bioinformatics and docking

3.1.1 Cluster Analysis

There are several different algorithms for the clustering of protein structures [128]. They may be classified as: (i) Exclusive Clustering: a structure belonging to a cluster cannot be included in another cluster (Fig. 3.1) [128, 129, 130]. (ii) Hierarchical Clustering: here the algorithm is based on the union between two nearest clusters (Fig. 3.2) [131]. The initial condition is realized by setting every structure as a cluster, and after a few iterations the algorithm reaches to the final set of clusters wanted. (ii) Overlapping Clustering: each structure may belong to two or more clusters with different degrees of occupation on each cluster [132]. (iv) Probabilistic Clustering: here the algorithm uses a completely probabilistic approach [133].

Approaches (i)-(ii) were used in the thesis and they are outlined here.

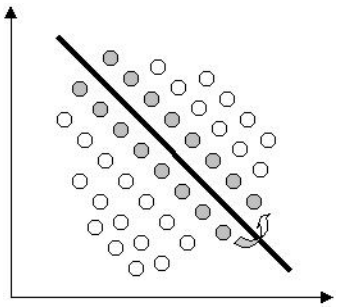


Figure 3.1: Exclusive Clustering: the separation of points is achieved by a straight line on a bi–dimensional plane

Exclusive Clustering

This clustering algorithm is based on the intuitive expectation that the best representative of each cluster is the one closest to the cluster center. The best way to identify the center of a cluster, is to find the member with the largest number of neighbors within a suitably chosen cutoff (the proximity score). In this case, the proximity score is chosen to be the Root Means Square Distance (RMSD) of atom positions between all pairs of structures. Indeed, points further from the center will have fewer neighbors.

The algorithm starts by choosing the first representative of the set as the one with the highest number of neighbors (highest proximity score). The cluster associated to this representative is defined as all of the representaitve's neighbors within a certain predefined RMSD cut-off. This starting point identifies simultaneously both the largest cluster and its representative. The next step is the remotion of the previous representative and its cluster from the set, and the proximity score of the remaining points is recalculated and, again, the member with the highest score is selected. The algorithm proceeds in this iterative manner until the set of surviving points is exhausted. In this way, a series of nonoverlapping clusters of structures is obtained [129, 130].

Hierarchical clustering

The hierarchical clustering starts from an ensemble of conformations and after several iterations it finish with the best set of clusters with their representatives. The algorithm involves 6 steps (Fig. 3.3) [131]

(i) "Distance determination": the clustering requires the calculation of the pair RMSD matrix among all the conformations in the ensemble. The RMS is calculated by superposing each member of the ensemble onto each of the other members

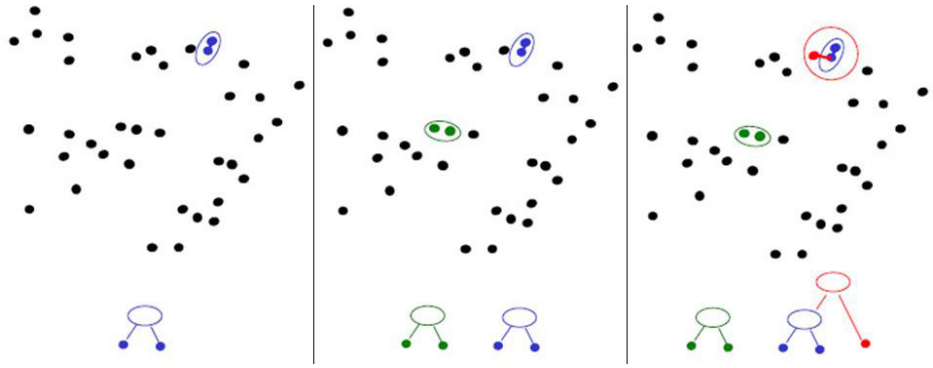


Figure 3.2: Hierarchical clustering

of the ensemble in a pairwise manner. The superposition is carried out on all non-hydrogen atoms. For N members, this results in a NxN matrix of RMSD values.

(ii) "Clustering": the RMSD matrix is used with the average linkage algorithm (based on the average distance between objects)* for a hierarchical cluster analysis. The method of average linkage takes the distance ($DIST$) between two cluster n and m to be:

$$DIST_{n,m} = \frac{\sum_{i=1,i\in n}^{N_n} \sum_{j=1,j\in m}^{N_m} dist(i,j)}{N_n N_m} \tag{3.1}$$

where cluster n contains N_n members, and cluster m contains N_m members; $dist(i,j)$ is the RMSD between the two members, i and j, of clusters n and m, respectively, after superposition. At each stage of the clustering algorithm, a search is performed for the two nearest clusters; these are merged to form a single cluster.

At each stage of clustering, the *spread* of each cluster and the average spread ($AvSp$) for the clustering are calculated. The spread of a cluster m containing N members is given by:

$$spread_m = \frac{\sum_{k=1}^{N_m} \sum_{l=1,l<k}^{N_m} dist(k,l)}{N_m(N_m-1)/2} \tag{3.2}$$

where i and k are members of cluster m. The average spread for the clustering is then given by:

*Other possible algorithms are: single linkage(based on the shortest distance between objects); complete linkage (based on the largest distance between objects); the Ward's method (based on the sum of squares between the two clusters, summed over all variables and the centroid method (based on the distance between clusters centroids.)

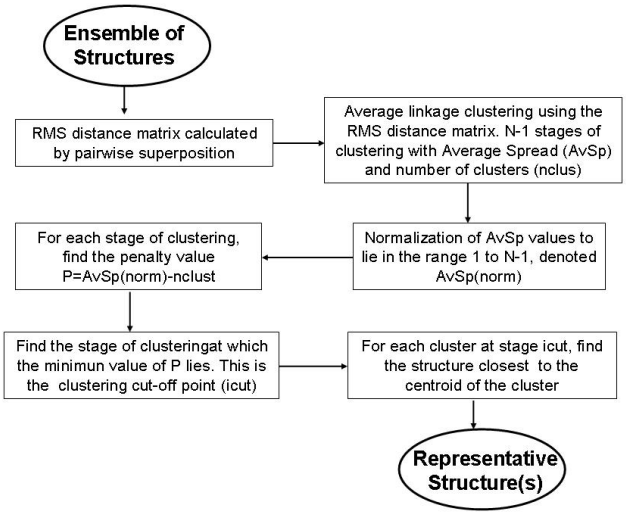


Figure 3.3: Hierarchical Clustering algorithm

$$AvSp_i = \frac{\sum_{m=1}^{cnum_i} spread_m}{cnum_i} \tag{3.3}$$

where $cnum_i$, is the number of clusters at stage i of the clustering (excluding outlying points, i.e. clusters that contain only one member).

(iii) "Normalization of the average spread": Once clustering is complete, the set of $AvSp_i$ values are normalized to lie within the range 1 to $N-1$, where N is the number of structures in the original data set. Normalization is performed to give equal weight in the penalty function (Step 4) to the number of clusters and the average spread.

$$AvSp(norm)_i = (N-2)\frac{AvSp_i - Min(AvSp)}{Max(AvSp) - Min(AvSp)} + 1 \tag{3.4}$$

where $Max(AvSp)$ and $Min(AvSp)$ are the maximum and minimum values respectively of average speed in the set $(AvSp_1, AvSp_2, ...AvSp_{N-1})$.

(iv)"Penalty function" : For each stage of clustering i, a penalty value, P_i is calculated as:

$$P_i = AvSp(norm)_i + nclus_i \tag{3.5}$$

where $nclus_i$, is the total number of clusters at step i of the clustering (including outlying points).

(v) "Cut–off "determination: The minimum penalty value in the set $(P_1, P_2, ...P_{N-1})$ is chosen as the cut–off level.

$$P_{icut} = Min(P_i) \tag{3.6}$$

Thus, the stage *icut* represents a state where the clusters are as highly populated as possible, whilst simultaneously maintaining the smallest spread. The smaller the spread of the clusters, the more similar the conformations of its members; the greater the population of a cluster, the less likely is the chance of excluding a member of similar conformation.

(vi) "Representative structures" calculation: Once a cut–off value in the clustering has been determined in this way, Eigen analysis [134] is performed on each cluster at stage *icut*. This allows for the determination of the structure within each cluster that is closest to the centroid of that cluster.

3.1.2 Computational Alanine Scanning

The alanine scanning methodology was used in order to analyze the interface of protein–protein complexes. Several studies have shown the existence of a small set of "hot spot" residues at the interface of protein–protein complexes that contribute significantly to the binding free energy [135, 136]. Many subsequent studies suggest that the presence of a few –hot spots–may be a general characteristic of most protein–protein interfaces. The alanine scanning methodology is a fast computational approach for the prediction of these energetically important amino acid residues or "hot spots " in protein–protein interfaces. The input consists of a three–dimensional structure of a protein–protein complex and the output is a list of "hot spots," or amino acid side chains that are predicted to significantly destabilize the interface when mutated to alanine, analogous to the results of experimental alanine–scanning mutagenesis.

The computed energies take into account van der Waals and the electrostatic contributions to the binding free–energy. Hot spots are identified as those residues at the interface whose mutation to alanine causes a free–energy loss greater than 1 kcal/mol.

$$\begin{aligned}\Delta G = {} & W_{attr}E_{LJattr} + W_{rep}E_{LJrep} + W_{HB(sc-bb)}E_{HB(sc-bb)} \\ & + W_{HB(sc-sc)}E_{HB(sc-sc)} + W_{sol}G_{sol} + W_{Coul}E_{Coul} \\ & + W_{\phi/\psi}E_{\phi/\psi}(aa) + \sum_{aa=1}^{20} n_{aa}E_{aa}^{ref}\end{aligned} \tag{3.7}$$

where, E_{LJattr} is the attractive part of the Leonnard–Jones potential, E_{LJrep} is a linear distance–dependent repulssive term, $E_{HB(sc-bb)}$ is the orientation dependent hydrogen bond potential (SC: side chain, BB: backbone), G_{sol} is the solvation term and W are the relative weights of the different energy terms. The $E_{\phi/\psi}$ is the amino acid type (aa) dependent backbone torsion angle propensity and E_{aa}^{ref} is the amino acid type dependent reference energy.

The effects of alanine replacement are computed both for the protein complex and for the corresponding uncomplexed partners to yield the change in binding energy $\Delta\Delta G_{bind}$:

$$\begin{aligned}\Delta\Delta G_{bind} &= \Delta G_{bind}^{MUT} - \Delta G_{bind}^{WT} \\ &= \left(\Delta G_{complex}^{MUT} - \Delta G_{partnerA}^{MUT} - \Delta G_{partnerB}^{MUT}\right) \\ &- \left(\Delta G_{complex}^{WT} - \Delta G_{partnerA}^{WT} - \Delta G_{partnerB}^{WT}\right)\end{aligned} \tag{3.8}$$

where $\Delta G_{complex}$, $\Delta G_{partnerA}$, and $\Delta G_{partnerB}$ are the stabilities of the complex and the unbound partners, and WT and MUT describe wild–type and mutant proteins.

Neutral residues and hot spots are defined as residues showing a change in the binding free energy by less or more than 1 kcal/mol when replaced by alanine ($\Delta\Delta G_{bind}$), respectively (alanine substitutions with experimentally measured stabilizing effects were rare and not larger than -0.9 kcal/mol; these were included in the neutral category.) A correctly identified hot spot is a residue with a predicted and observed $\Delta\Delta G_{bind}$ value larger than or equal to 1 kcal/mol; a correctly identified neutral residue has both predicted and observed $\Delta\Delta G_{bind}$ values less than 1 kcal/mol. Interface residues were defined as residues with a side chain having at least one atom within a sphere with 4 Å radius of an atom belonging to the other partner in the complex or as those that, upon complex formation, become significantly buried, i.e. if there is a significant increase in the number of $C\beta$ atoms located within a sphere of radius 8 Å centered on the $C\beta$ atom of the residue considered. These energies are necessary very approximate and are here used only for qualitative comparisons.

3.1.3 Molecular screening and design protocol

Screening

The molecular screening of a database of molecules against a set of template molecules performed in this thesis is based on 2 algorithms. The first algorithm is based on the shape similarity between all molecules in the database and the template molecules and the second algorithm is based on the electrostatic similarity.

Shape similarity The more the volumes of two molecules differ, the more the shapes will differ. The volumes of two molecules can be defined as:

$$\begin{aligned} V_1 &= \int f(x,y,z)dv \\ V_2 &= \int g(x,y,z)dv \end{aligned} \tag{3.9}$$

Where $f(x,y,z)$ and $g(x,y,z)$ are the characteristic functions of two different molecules. They correspond to the common definition of a volume field.

If the function S is defined as:

$$S_{f,g} = \sqrt{\int \left[f(x,y,z) - g(x,y,z)\right]^2 dV} \tag{3.10}$$

Is possible to arrive to the fundamental equation for shape comparison defined as:

$$S_{f,g}^2 = \int f(x,y,z)^2 dV + \int g(x,y,z)^2 dV - 2\int f(x,y,z)g(x,y,z)dV \tag{3.11}$$

If $\int f(x,y,z)^2 dV = I_f$ (the self-volume overlap of $f(x,y,z)$); $\int g(x,y,z)^2 dV = I_g$ (the self-volume overlap of $g(x,y,z)$) and $\int f(x,y,z)g(x,y,z)dV = O_{f,g}$ (the overlap between the two functions), then:

$$S_{f,g}^2 = I_f + I_g - 2O_{f,g} \tag{3.12}$$

The three terms in this function constitute the necessary terms to calculate and compare the shapes of two molecules. The quantity referred to as a Tanimoto coefficient may be derived by recombining I's and O so:

$$Tanimoto_{f,g} = \frac{O_{f,g}}{I_f + I_g - O_{f,g}} \tag{3.13}$$

Thus, knowing the volume of the molecules and their alignment, it is possible to calculate the Tanimoto coefficient for the shape similarity between the two molecules. The coefficient is 1 if the two molecules are the same and 0 if they are completely different.

The programs ROCS (Rapid Overlay of Chemical Structures) [137], was used to calculate the shape similarity between molecules. It calculate the Tanimoto

coefficient (eq. 3.13) using a shape-based superposition method. The molecules are aligned by a solid-body optimization process that maximizes the overlap volume between them. The volume function is calculated taken each atom as solid sphere but using a Gaussian approximation [138].

The screening of a database of molecules against a template molecule, using this algorithm, give an output that consists on a list with all the molecules in the database ranked according to their shape similarity with the template molecule

Electrostatic similarity The program EON [139] was used to calculated the electrostatic similarity. It uses a field-based measure of Tanimoto to compare the electrostatic potential of two small molecules. This electrostatic potential is calculated solving the Poisson-Boltzman (PB) equation. If A and B are the functions for the electrostatic potential of two molecules, the basic equation for a field Tanimoto for the electrostatic is:

$$Tanimoto_{A,B} = \frac{\int A(\vec{r}) \cdot B(\vec{r})}{\int A(\vec{r}) \cdot A(\vec{r}) + \int B(\vec{r}) \cdot B(\vec{r}) - \int A(\vec{r}) \cdot B(\vec{r})} \tag{3.14}$$

The two boundary cases for electrostatic Tanimoto occur when $B = A$:

$$Tanimoto = \frac{\int A(\vec{r}) \cdot A(\vec{r})}{\int A(\vec{r}) \cdot A(\vec{r}) + \int A(\vec{r}) \cdot A(\vec{r}) - \int A(\vec{r}) \cdot A(\vec{r})} = 1 \tag{3.15}$$

and the opposite case, when $B = -A$:

$$Tanimoto = \frac{\int A(\vec{r}) \cdot -A(\vec{r})}{\int A(\vec{r}) \cdot A(\vec{r}) + \int -A(\vec{r}) \cdot -A(\vec{r}) - \int A(\vec{r}) \cdot -A(\vec{r})} = -\frac{1}{3} \tag{3.16}$$

Thus, the Tanimoto coefficient for the electrostatic is between 1 and -1/3.

Atomic charges: The AM1BCC charges were calculated for all the molecules in order to use them for the calculation of the electrostatic potential. The AM1BCC method [140] starts with partial charges derived from the AM1 semiempirical wave-function and in a second stage, bond-charge corrections (BCC) are applied to the partial charges on each atom to generate the final partial charges. All the AM1BCC parameters were tuned to reproduce the HF/6-31G* RESP-fit charges [141, 142]. The MOLCHARGE program from the OpenEye software was used for the calculation of the atomic charges.

Design of new compounds

The process for the generation of new virtual compounds is based on the generation of new molecules from a "seed" molecule. The WABE program from the OpenEye software was used for this purpose. It takes the "seed" molecule and generates from it a number of new virtual compounds, changing and adding different sets of chemical groups. This process is typically used to explore the chemical space around a lead compound or drug of interest. Once the virtual structures are enumerated, they are scored using a scoring method based on their electrostatic complementarities to another template molecule. Basically it is used to generate new virtual compounds with high electrostatic similarity with a template molecule starting from a "seed" molecule (the seed molecule could be another molecule similar to the template or only a chemical group).

Filter undesirable compounds: The filter program was used to eliminate inappropriate or undesirable generated new compounds. It attempts to remove all of the compounds that are wrong candidates to suggest to a medicinal chemist as a potential hit. It is based on different molecular properties like: polar surface area, pKa normalization (pH 7.4), LogP, LogS, aggregators, lipinski and hydrogen-bonds rules, reagent selection, pharmacokinetics, structural and chemical features and functional groups. A drug-like filter was used for these purpose in this thesis (see [143] for details).

3.1.4 Docking

The molecular docking is procedure to obtain the 3D structure of protein-protein or protein-ligands complexes. It uses a scoring function to assign to each result a score that it could be related with the binding affinity. In this thesis, the AUTODOCK program [144, 145] was used for this procedure. Historically, the version 3.0.5 of the program was used, but the final results were obtained with the version 4 of the program. In this section, the algorithm and the scoring function implemented in the program are described.

Genetic Algorithm

In order to perform the docking experiments, the AUTODOCK program with the Lamarckian Genetic Algorithms (LGA) was used [144, 145]

The LGA uses ideas based on the language of natural genetics and biological evolution and it involved several step.

The first is the definition of all the variables and evaluations that will be used in the LGA:

Gene, genotype and phenotype: the particular arrangement of a ligand and a protein can be defined by a set of values describing the translation, orientation, and conformation of the ligand with respect to the protein: these are the ligand's state variables and, in the LGA, each state variable corresponds to a gene. The ligand's state corresponds to the genotype, whereas its atomic coordinates correspond to the phenotype. Each genotype is composed by one chromosome. The chromosome is composed of a string of real valued genes: three Cartesian coordinates for the ligand translation; four variables defining a quaternion specifying the ligand orientation; and one real–value for each ligand torsion, in that order.

Fitness and generation: the fitness is the total interaction energy of the ligand with the protein, and is evaluated using the energy function. Random pairs of individuals are mated using a process of crossover, in which new individuals inherit genes from either parent. In addition, some offspring undergo random mutation, in which one gene changes by a random amount. Selection of the offspring of the current generation occurs based on the individual's fitness: thus, solutions better suited to their environment reproduce, whereas poorer suited ones die.

The next step is the beginning of the Genetic Algorithm (GA). In AUTODOCK, the simplest GA is applied, this GA algorithm it is also called the "Canonical GA".

The GA starts by creating a random population of individuals, where the user defines the number of individuals. For each random individual in the initial population, each of the three translation genes for x, y, and z is given a uniformly distributed random value between the minimum and maximum x, y, and z extents of the grid maps, respectively; the four genes defining the orientation are given a random quaternion, consisting of a random unit vector and a random rotation angle between -180° and +180°; and the torsion angle genes, if any, are given random values between -180° and +180°.

In the LGA, the creation of the random initial population is followed by a loop over generations and local search on a defined proportion of the population, repeated until the maximum number of generations or the maximun number of energy evaluations is reached.

In the following steps of the generation, each individual is mapped to its coordinates (genotype to phenotype) and the individual fitness is calculated. The fitness is calculated based on a free energy evaluation that will be discussed in more detail later.

The proportional selection is the following step in the protocol. It is based on the fitness values of each individual and it helps to decide which individuals will reproduce.

The number of offspring is calculated as follows

$$n_i = \frac{f_w - f_i}{f_w - \langle f \rangle} \qquad f_w \neq \langle f \rangle \tag{3.17}$$

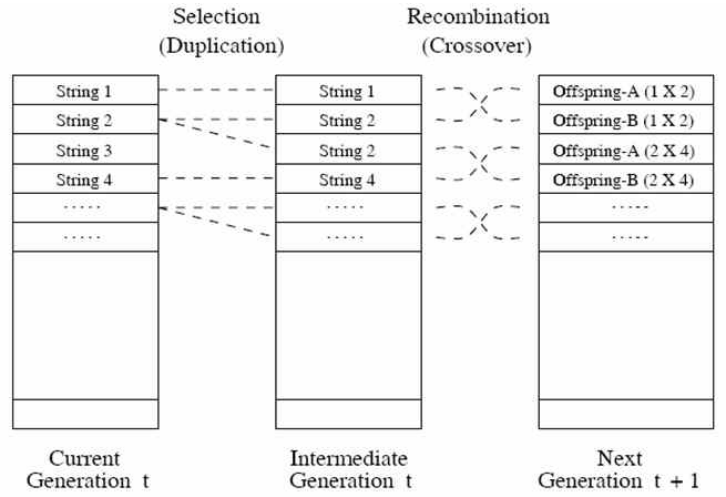


Figure 3.4: Genetic algorithm: selection and recombination.

where n_i is the number of offspring to be allocated to the individual i; f_i is the fitness of the individual (the energy of the ligand); f_w is the fitness of the worst individual, or highest energy in the last defined generations and $\langle f \rangle$ is the mean fitness of the actual population. After this stage the worse individuals ($n_i < 1$) will be replaced by the best individuals ($n_i > 1$) in a proportional way creating an intermediate population, like a normal evolutionary program.

Crossover and mutation are performed after the selection on random members of the intermediate population according to a defined rate of crossover and mutation Fig.3.4.

First, crossover is performed. Two–point crossover is used, with breaks occurring only between genes, never within a gene. Thus, both parents' chromosomes would be broken into three pieces at the same gene positions, each piece containing one or more genes; for instance, ABC and abc. The chromosomes of the resulting offspring after two–point crossover would be AbC and aBc. These offspring replace the parents in the population, keeping the population size constant. Crossover is followed by mutation; it is performed by adding a random real number that has a Cauchy distribution to the variable, the distribution being given by:

$$c(\alpha, \beta, x) = \frac{\beta}{\pi\left(\beta^2 + (x - \alpha)^2\right)} \tag{3.18}$$

where α and β are parameters that affect the mean and the spread of the distribution.

The last step in the generation is the elitism selection. The elitism is an integer parameter that determines how many of the top individuals automatically survive into the next generation. The new population that has resulted from the

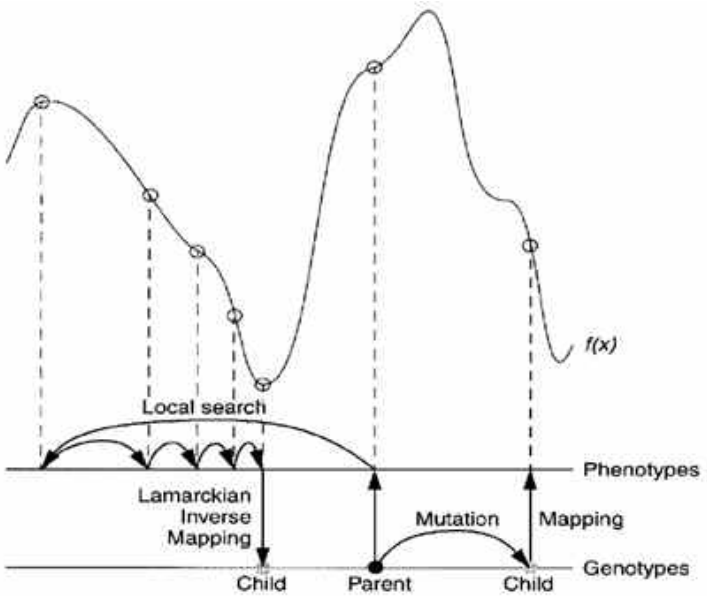


Figure 3.5: This figure illustrates genotypic and phenotypic search, and contrasts Darwinian and Lamarckian search. The space of the genotypes is represented by the lower horizontal line, and the space of the phenotypes is represented by the upper horizontal line. Genotypes are mapped to phenotypes by a developmental mapping function. The fitness function is f(x). The result of applying the genotypic mutation operator to the parent's genotype is shown on the right–hand side of the diagram, and has the corresponding phenotype shown. Local search is shown on the left–hand side. It is normally performed in phenotypic space and employs information about the fitness landscape. Sufficient iterations of the local search arrive at a local minimum, and an inverse mapping function is used to convert from its phenotype to its corresponding genotype. In the case of molecular docking, however, local search is performed by continuously converting from the genotype to the phenotype, so inverse mapping is not required. The genotype of the parent is replaced by the resulting genotype, however, in accordance with Lamarckian principles. Taken from [144].

proportional selection, crossover, and mutation is sorted according to its fitness; the fitness of new individuals having resulted from crossover and/or mutation is calculated as necessary.

In the LGA, each generation is followed by a local search algorithm on a defined proportion of the population; the indivual in the population is then replaced by the result of the local search Fig. 3.5. The local search algorithm implemented in AUTODOCK is the one proposed by Solits and Wets [146]. In particular, this algorithm searches through the genotypic space rather than the more typical phenotypic space. This means that there is no need to pass from the genotypic space to the phenotypic space like any other algorithm and then to come back again. Nonetheless, this molecular variation of the genetic algorithm still qualifies as Lamarckian, because any "environmental adaptations" of the ligand acquired during the local search will be inherited by its offspring.

In the Lamarckian genetic algorithm, genotypic mutation plays a somewhat different role than it does in traditional genetic algorithms. Traditionally, mutation plays the role of a local search operator, allowing small, refining moves that are not efficiently made by crossover and selection alone. With the explicit local search operator, however, this role becomes unnecessary, and is needed only for its role in replacing alleles that might have disappeared through selection. In LGA, mutation can take on a more exploratory role.

Scoring function

The approach implemented in AUTODOCK 3.0.5 to calculate the relationship between molecular structure and binding free energy (structure–based scoring function), uses the thermodynamic cycle of Wesson and Eisenberg Fig. 3.6 [147].

The form of the free energy function includes five terms:

$$\begin{aligned}\Delta G = W_{vdw}\sum_{i,j}(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}) - W_{hbound}\sum_{i,j}E(\phi)(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbound}) + \\ + W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} + W_{conf}N_{tor} + W_{sol}\sum_{i,j}S_{ij}V_j\exp(-r_{ij}^2/2\sigma^2)\end{aligned} \tag{3.19}$$

Where coefficients W were empirically determined using linear regression analysis [144]. The summations are performed over all pairs of ligand atoms, i, and protein atoms, j, in addition to all pairs of atoms in the ligand that are separated by three or more bonds.

The in vacuo contributions include three interaction energy terms, used in previous versions of AUTODOCK: a Lennard–Jones 12–6 dispersion/repulsion term;

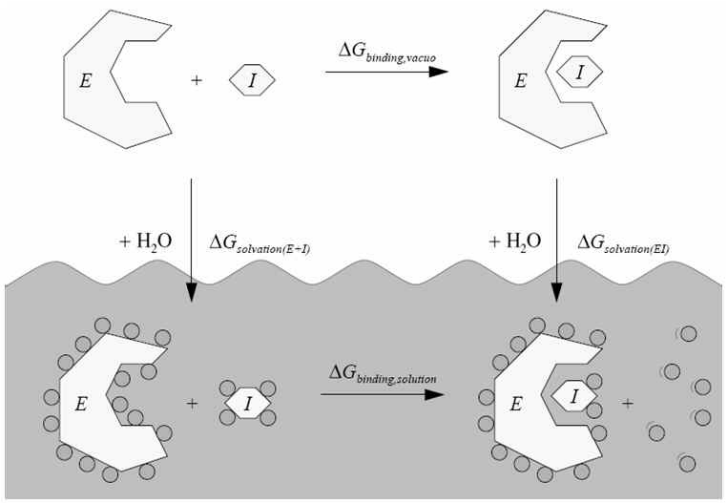


Figure 3.6: Thermodynamic cycle for the binding of an enzyme, E, and an inhibitor, I, in both the solvated phase and in vacuo. Note the solvent molecules are indicated by filled circles: they tend to be ordered around the larger molecules, but when E and I bind, several solvent molecules are liberated and become disordered. Taken from [148]

a directional 12–10 hydrogen bonding term, where $E(\phi)$ is a directional weight based on the angle ϕ, between the probe and the target atom and E_{hbound} is the estimated average energy of hydrogen bonding of water with a polar atom; and a screened Coulombic electrostatic potential. A measure of the unfavorable entropy of ligand binding due to the restriction of conformational degrees of freedom is added to the in vacuo function. This term is proportional to the number of sp3 bonds in the ligand, N_{tor}. Note that the internal or intramolecular interaction energy of the ligand is not included in the calculation of binding free energy.

The desolvation term is calculated using a volume–based approach [149]. For each atom in the ligand, fragmental volumes of surrounding protein atoms (V_j) are weighted by an exponential function and then summed, evaluating the percentage of volume around the ligand atom that is occupied by protein atoms. This percentage is then weighted by the atomic solvation parameter of the ligand atom (S_i) to give the desolvation energy.

Improvements

The new version of AUTODOCK (V. 4.0) used an improved version of the semiempirical free energy force field. It was calibrated using a large set of diverse protein–ligand complexes and includes two major advances [145].

1- An improved thermodynamic model of the binding process, which allows the inclusion of intramolecular terms in the estimated free energy. This gives also the

possibility to rotate protein side chains during the docking procedure.

The free energy of binding is estimated to be equal to the difference between (1) the energy of the ligand and the protein in separated states and (2) the energy of the ligand–protein complex.

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{bound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad (3.20)$$

Where the first two parentheses are intramolecular energies for the bound and unbound states of the ligand and the protein respectively and the third parentheses is the intermolecular energy between the bound and unbound state plus the conformational entropy lost upon binding.

$$\Delta S_{conf} = W_{conf} N_{tor} \quad (3.21)$$

2- An improve pair-wise evaluation for the force field, that includes a full desolvation model that includes terms for more atom types, including the favorable energetics of desolvating carbon as well as the unfavorable energetics of desolvating polar charges groups.

$$\begin{aligned} V = {} & W_{vdw} \sum_{i,j} (\frac{A_{ij}}{r_{ij}^{12}} \natural \frac{B_{ij}}{r_{ij}^{6}}) - W_{hbound} \sum_{i,j} E(\phi) (\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbound}) + \\ & + W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} (S_i V_j + S_j V_i \exp(-r_{ij}^2 / 2\sigma^2) \end{aligned} \quad (3.22)$$

The desolvation term is now calculated using the same approach as before [147] but now the atomic solvation parameter (S_i) is calculated as a function of the partial atomic charge (q_i):

$$S_i = (ASP_i + QASP \times |q_i|) \quad (3.23)$$

Where ASP_i is the atomic solvation parameter of atom i and $QASP$ is the charged-based atomic solvation parameter ($QASP = 0.01097$, $stderror = 0.0063$)

Another approach is implemented in the GOLD program [150, 151] but it has not been used in thesis. This approach uses a multi-population genetic algorithm and different types of scoring functions. It is presented in the Appendix B.

3.2 Molecular Dynamics

Molecular dyanmics (MD) simulations are based on the following assumptions:

1. The nuclei can be treated as classical particles
2. The Born-Oppenheimer approximation holds
3. The electronic degrees of freedom can be integrated out.

Under these greatly simplifying assumptions, the dynamics of the system can be described by the Newton second law:

$$\vec{F_i}(t) = \frac{d^2\vec{r_i}(t)}{dt^2} \cdot m_i = -\frac{dE_{tot}(t)}{d\vec{r_i}} \tag{3.24}$$

where is $\vec{r_i}$ the position of the i^{th} atom and E_{tot} is the total energy of the system of N particles. So, knowing the initial structure (by experiments or by computer modeling) and providing a velocity distribution consistent with the temperature simulated, one can provide the time-evolution of the system. MD average values of several properties can be evaluated from the resulting trajectory.

The MD run consist on several steps.

1. Defining of the initial conditions, potential interaction as a function of atom positions; position $\vec{r}$ and velocities $\vec{\nu}$ of all atoms in the system
2. Compute the force acting on each atom.
3. Solve the Newton's equations of motion and update configuration of the system.
4. Write the new positions, velocities, energies,..., etc
5. Back to point 2

In the followings, the form of the equations of mations used for proteins and some details of MD algorithms are given.

3.2.1 Equation of motions

Molecular dynamics simulation consists of the numerical, step-by-step, solution of the classical equations of motion (eq. 3.24).

Where $\vec{F_i}(t)$ is the total force acting on the $i-th$ atom, m_i is its mass, $\frac{d^2\vec{r_i}(t)}{dt^2}$ is its acceleration, $\vec{r_i}(t)$ is its position and E_{tot} is the total energy action on the atom. The integration time step is set to Δt, it is chosen at the begging of the simulation and remain unchanged during the all run.

The most common used algorithms for the integration of the equations of motion are the so called Verlet and Leap-Frog algorithm [152]. In both algorithms the positions of each atom are expressed by Taylor expansions. The lack of explicit velocities in the Verlet algorithm is remedied by the leap-frog algorithm [153] . Positions at times $t+\Delta t$ and t are given by the Taylor expansions around $t+\Delta t/2$.

$$\begin{aligned} \vec{v}_i(t+\Delta t/2) &= \vec{v}_i(t+\Delta t/2) + \frac{\vec{F}_i(t)}{m_i} \cdot \Delta t \\ \vec{r}_i(t+\Delta t) &= \vec{r}_i(t) + \vec{v}_i(t+\Delta t/2) \cdot \Delta t \end{aligned} \tag{3.25}$$

which make use of the positions and forces at time t and velocities at time $t+\Delta t/2$, alternating the update of the position and velocities.

First, the velocities are calculated at half time step, then these are used to calculate the positions at one time step. In this way, positions and velocities leap each other.

The advantage of this algorithm is that velocities are explicitly calculated, and the disadvantage is that they are not calculated at the same time as the positions. As a consequence, kinetic and potential energy cannot directly compute the total energy at time t, but energy evaluation is possible using the following approximate value of velocities at time t:

$$\vec{v}_i(t) = \frac{\vec{v}_i(t-\Delta t/2) + \vec{v}_i(t+\Delta t/2)}{2} \tag{3.26}$$

3.2.2 Force field

In order to solve numerically the equations of motion, we need to be able to calculate the forces acting on the atoms, and these are usually derived from a potential energy represents the complete set of 3N atomic coordinates.

The potential energy function, also called force field, used in most of the standard molecular simulation programs for biological systems consist of:

$$E_{tot} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{vdw} + E_{elect} \tag{3.27}$$

It takes form of the summation of different additive terms that correspond to bond distances stretching (E_{bonds}), bond angles bending (E_{angles}), bond dihedral or torsion angle ($E_{dihedrals}$), van der Waals potential (E_{vdw}) and electrostatic potential (E_{elect}). The first three terms are considered to be the intramolecular bonding interactions, each term involves a multiplet of atoms connected by chemical bonds. The other two terms represent the non-bonded interactions between atoms.

The specific form of each of these terms is shown below:

$$
\begin{aligned}
E_{bonds} &= \sum_{bonds} \frac{1}{2} k_r (r - r_{eq})^2 \\
E_{angles} &= \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_{eq})^2 \\
E_{dihedrals} &= \sum_{dihedrals} \frac{1}{2} V_r (1 + \cos(n\varphi) - \gamma) \\
E_{vdw} &= \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \\
E_{elec} &= \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}
\end{aligned}
\tag{3.28}
$$

In the bonding interactions, the bond stretching and angle bending terms are represented as harmonic energy functions where r_{eq} and θ_{eq} refer to equilibrium bond lengths and angles, k_r and k_θ are the vibrational constants. The equilibrium values of the bond and angle parameters are usually derived from structural databases, while the force constants are derived from infrared spectroscopy. In the third term, V_n is the torsional barrier corresponding to the n^{th} barrier of a given torsional angle with phase γ: dihedral parameters are calibrated on small model compounds, comparing the energies with those obtained by quantum chemical calculations. Improper dihedral angles are added to take into account quantum effects that are not present in E_{tot} as, for example, to preserve planarity in aromatic rings.

In the non-bonded interactions, the van der Waals potential is described by a Lennard-Jones potential, containing an attractive and a repulsive term, the parameters are defined so as to reproduce chemical-physical properties (e.g., densities, enthalpies of vaporization, solvation free-energies) in organic liquids. The electrostatic energy is evaluated by assuming the dielectric constant ε equal to 1, and using the restrained electrostatic potential model [141] to define partial atomic point charges: in this model, charges are assigned to the atom-centered points so as to fit the electrostatic potential derived from quantum chemistry calculations for a set of small representative molecules. Van der Waals (vdW) and electrostatic potentials are calculated between atoms belonging to different molecules or for atoms in the same molecule separated by at least three bonds. In principle, the non-bonded interaction is suppose to be calculated aver all pairs of atoms in the system, and they are the most expensive part of a MD calculation. In practical applications, however, the number of interactions is limited by a predefined cutoff distance, so the non-bonded interaction is calculated only between atoms separated by a distance not larger than the cutoff. For the vdW potential, this

truncation introduce only a small error in the energy but it is not the case for the electrostatic potential because the Coulomb interaction between charges q_i and q_j decays much less rapidly with distance. Hence it can not be truncated, but when periodic boundary conditions are used, it could be treated with schemes such as Particle Mesh Edwald, which approximate the exact result to an acceptable error similar to the error in the vdW potential.

3.2.3 Constrains

The integration time-step Δt is limited in the integration of the equations of motion to the fastest motions of the system. Bond stretching, in particular those involving hydrogen atoms, has the highest frequency and relatively low amplitude, and this restricts the time-step to 1 fs. Since these motions are of little interest in most cases, they are constrained to the equilibrium bond length and these allow an increase of Δt by a factor of 1.5-2 the time-step, with small affects on the accuracy of the simulation.

In the NAMD program [154], the algorithm used is called SHAKE algorithm. In this algorithm, the constrains are introduced through the Lagrangian [155] or Hamiltonian [156] formalism.

3.2.4 Periodic boundary conditions

MD simulations are usually performed under periodic boundary conditions (PBC), to minimize boundary effects and to mimic the presence of the "bulk" environment. In this approach, the system is surrounded with replicas of itself in all directions, to yield an infinite periodic lattice of identical cells. When a particle moves in the central cell, its periodic image in every another cell moves accordingly (Fig 3.7). As one molecule leaves the central cell, its image enters from the opposite side without "filling" the cell boundary.

The PBCs are taken into account only in the calculating of non-bonded interactions between atoms belonging to different molecules, and if the potential range is not too long (cutoff radius used to truncate non-bonded interactions does not exceed half the shortest box vector), the minimum image convention could be adopted. This means that each atom interacts only with the nearest atom or image in the periodic array.

When a macromolecule, such as a protein, is studied in solution, this restriction does not suffice. In principle a single molecule should not be able to "see" both sides of the macromolecule. This means that the length of each box vector must exceed the length of the macromolecule in the direction of that edge plus two time cutoff radius. But it is common to compromise in this respect, and make the solvent layer somewhat smaller in order to reduce the computational cost.

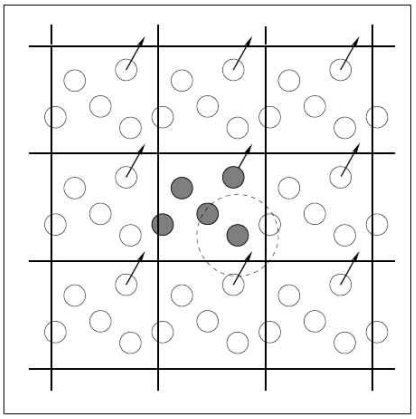


Figure 3.7: Periodic boundary conditions. As a particle moves out of the simulation box, and image particle moves in to replace it. In calculating particle interactions within the cutoff range, both real and image neighbors are included.

3.2.5 Neighbors list

Computing the non-bonded contribution to the interatomic forces in an MD simulation involves, in principle, a large number of pairwise calculations. Let us assume that the interaction potentials are of short range, equal to 0 if $r_{ij} > r_{cut}$. In this case, the program skips the force calculation, avoiding expensive calculations, and considers the next candidate j. Nonetheless, the time to examine all pair separations is proportional to the number of distinct pairs and this still consumes a lot of time.

Some economies result from the use of lists of nearby pairs of atoms. The potential cutoff sphere, of radius rcut, around a particular atom is surrounded by a 'skin', to give a larger sphere of radius rlist as shown in Fig. 3.8. At the first step in a simulation, a list is constructed of all the neighbours of each atom, for which the pair separation is within rlist. Over the next few MD time steps, only pairs appearing in the list are checked in the force routine. From time to time the list is reconstructed: it is important to do this before any unlisted pairs have crossed the safety zone and come within interaction range.

3.2.6 Long range forces

A long-range force is often defined as one in which the spatial interaction falls off no faster than r^{-d} where d id the dimensionality of the system. In this category is the charge-charge interactions between ions ($V(r) \approx r^{-1}$). These interactions are serious problem for the simulations, since their range is greater then half the box length.

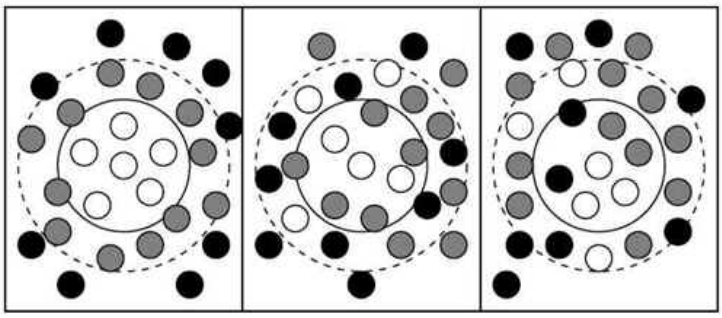


Figure 3.8: The potential cutoff range (solid circle), and the list range (dashed circle), are indicated. The list must be reconstructed before particles originally outside the list range (black) have penetrated the potential cutoff sphere.

Electrostatics

The Edwald sum is a technique for efficiently summing the interactions between ions and all its periodic images (see Appendix C for details). The potential energy can be written as

$$V = \frac{1}{2} \sum_{n}{}' \left(\frac{1}{4\pi\varepsilon_0} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{r_{ij}, \mathrm{n}} \right) \tag{3.29}$$

Where q_i and q_j are the charges, n is the box index, r_{ij} is the distance between the charges and the prime indicates that we omit $i = j$ fon $\mathrm{n} = 0$. For long-range potential, this sum is conditionally convergent, and thus extremely slow. Therefore, a very large number of images is required to achieve a reliable estimate of V. The idea behind the Ewald method is to surround every point charge by a charge distribution of equal magnitude and opposite sign ϱ_-, which spreads out radially from the charge. This distribution is conveniently taken to be Gaussian

$$\varrho_{\mathrm{i}}^{\mathrm{G}}(r) = q_i \left(\frac{\alpha}{\pi} \right)^{\frac{3}{2}} \exp\Big(-\alpha \left| r_i + nL \right|^2 \Big) \tag{3.30}$$

Where α is an arbitrary parameter which does not determine the final result, but can be adjusted to optimized the convergence rate (it is related to the width of the gaussian distribution). In this way an efficient screening is realized, so that interactions rapidly go to 0 and direct summation is possible. This extra distribution acts like an ionic atmosphere, to screen the interaction between neighboring charges. The screened interactions are now short-ranged, and the total screened potential is calculated summing over all molecules. A charge distribution of the same sign as the original charge, and the same shape as the distribution $\varrho_i^G(r)$ is

also added. This canceling distribution reduces the overall potential to that due to the original set of charges. In order to exclude self-interactions the contributions of these three charge densities should not be evaluated in r_i. However, it is convenient to keep self-interactions due to the cancelling charge distribution ϱ_+, since ϱ_+ is in this way periodic and can be represented as a rapidly converging Fourier sum. The spurious self-interaction can be easily subtracted separately. The canceling distribution is summed in the reciprocal space. In other words, the Fourier transforms of this distribution are added, and the total transformed back into real space. Thus, the total charge distribution of the system $\varrho(r)$ may be rewritten as:

$$\begin{aligned} \rho(r) &= \sum_i q_i \delta\left(r - r_i + nL\right) \\ \rho(r) &= \sum_i \left(q_i \delta\left(r - r_i + nL\right) - \rho_i^G(r)\right) + \sum_i \rho_i^G(r) \end{aligned} \tag{3.31}$$

Where the first sum (which only produce short ranged potentials) is calculated in the direct space, while the second is calculated in the reciprocal space:

$$\begin{aligned} V &= (\varepsilon_{dir} - \varepsilon_{self} + \varepsilon_{rec}) \\ \varepsilon_{dir} &= \frac{1}{2} \sum_n{}' \sum_{ij}^N \frac{q_i q_j}{|r_{ij} + nL|} erfc\left(\sqrt{\alpha}\,|r_{ij} + nL|\right) \\ \varepsilon_{self} &= \sqrt{\frac{\alpha}{\pi}} \sum_{i=1}^N q_1^2 \\ \varepsilon_{rec} &= \frac{2\pi}{V} \sum_{k\neq 0} \sum_{ij} \frac{\exp\left(\frac{k^2}{4\alpha}\right)}{k^2} q_i q_j \exp\left(\, -ik \cdot (r - r_j)\right) \end{aligned} \tag{3.32}$$

Where $erfc = \pi^{-\frac{1}{2}} \int_r^\infty \exp\left(\, -r^2\right) dx$ is the complementary error function.

ε_{dir} is very similar to eq. 3.29, although the long ranged $\frac{1}{r}$ function is here substituted by the short ranged $\frac{erfc(r)}{r}$: as a result, the interaction vanishes above a cutoff roughly equal to $\alpha^{-\frac{1}{2}}$, and for every i and j the interaction can be approximated by only one periodic image term. Typically, α is chosen in order to keep the truncation error in the order of $10^{-5}/10^{-6}$ of ε_{dir}.

ε_{self} is the self interaction of the gaussian charge distributions: it must be subtracted from the total, as the reciprocal space term ε_{rec} contains it, albeit it is a constant number, not depending on the atomic configuration.

ε_{rec} is a sum over an infinite number of terms, but the factor $\frac{k^2}{4\alpha}$ ensures a fast convergence in the reciprocal space, and normally no more than $5/10$ wave vectors in each direction are required. Its calculation is however the most consuming part in the Ewald scheme.

3.2.7 Temperature control

In MD simulations it is possible to realize different types of thermodynamics ensembles which are characterized by the control of certain thermodynamic quantities, the microcanonical ensemble (NVE) is obtained from the derivation of the equation of motion. However the microcanonical ensemble does not correspond to the conditions under which most experiments are carried out. The ensembles in which most of the experiments are carried out are the canonical (NVT) or the isothermal-isobaric ensembles. If one is interested in the behavior of the system in such ensembles, a thermostat is required to control the temperature during the simulation. Among the several methods to control the temperature usually available in MD software packages, the weak coupling method [157] and the coupling to an external bath algorithm [158, 159] are used in this work.

Berendsen thermostat

The instantaneous value of the temperature $T(t)$ of a system with N_{df} degrees of freedom is related to the kinetic energy E_{kin} via the particles velocities as follows:

$$\begin{aligned} E_{kin}(t) &= \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2(t) = \frac{1}{2} N_{df} k_B T(t) \\ T(t) &= \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{N_{df} k_B} \end{aligned} \tag{3.33}$$

An obvious way to alter the temperature is to scale the velocities by a factor λ . So:

$$\begin{aligned} \Delta T &= \sum_{i=1}^{N} \frac{m_i (\lambda v_i)^2}{N_{df} k_B} - \frac{m_i v_i^2}{N_{df} k_B} \\ \Delta T &= (\lambda^2 - 1) T(t) \end{aligned} \tag{3.34}$$

If T_0 is the desired temperature and $T(t)$ is the current temperature and the variation of the current temperature along the time is defined as:

$$\begin{aligned} \frac{dT(t)}{dt} &= \frac{T_0 - T(t)}{\tau_T} \\ \Delta T &= \frac{\partial t}{\tau_T} (T_0 - T) \end{aligned} \tag{3.35}$$

It is easy to demonstrate that:

$$\lambda(t) = \sqrt{1 + \frac{\Delta T}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)} \tag{3.36}$$

The factor λ is used to scale the velocities at each time-step, in order to relax the temperature toward the desired temperature value T_0. The relaxation rate is controlled by the time coupling constant τ_t, which should be small to achieve the required temperature but large enough to avoid disturbance of the physical properties of the system by coupling to the bath. Even though the Berendsen weak coupling algorithm is efficient for relaxing the system to the reference temperature, it does not allow for a proper sampling of the canonical ensemble. Canonical ensemble simulation is enabled by employing the Nosé-Hoover thermostat.

Nosé -Hoover temperature coupling

The Berendsen thermostat is extremely efficient for relaxing a system to a target temperature, but once the system has reached the equilibrium, it might be more important to probe a correct canonical ensemble. The extended system method was originally introduced by Nosé and subsequently developed by Hoover. So, in the Nosé -Hoover approach [157, 158], the idea is to consider the heat bath as an integral part of the system by addition of an artificial variable $\tilde{s}$ associated with a "mass parameter' $Q > 0$ as well as velocity .The magnitude of Q determines the coupling between the bath and the real system an so influences the temperature fluctuations. The artificial variable $\tilde{s}$ plays the role of a time-scaling parameter. The Nose Hamiltonian (H_N) for the extended system is chosen to be:

$$H_N = H_0(q, p/s) + gk_BT \ln s + p_s^2/2Q \tag{3.37}$$

Where H_0 is the usual Hamiltonian for a classical many body system, except that everywhere one would normally expect to see a momentum p appearing, it is replaced by p/s thus:

$$H_0(q, p/s) = \sum_i (p_i/s)^2/2m_i - U(q) \tag{3.38}$$

The Nosé equations of motions are:

$$\begin{aligned}
\dot{r}_i &= \frac{\partial H_N}{\partial p} = \frac{p_i}{ms^2} \\
\dot{p}_i &= \frac{\partial H_N}{\partial r} = F_i(r) \\
\dot{s} &= \frac{\partial H_N}{\partial p_s} = \frac{p_s}{Q} \\
\dot{p}_s &= \frac{\partial H_N}{\partial s} = \sum \frac{p_i^2}{ms^3} - \frac{N_{df} k_B T}{s}
\end{aligned} \tag{3.39}$$

These coupled first order equations take a simpler form if the time scale is reduced by s, so that $dt_{old} \equiv s dt_{new}$. All of the rates given before can then be expressed as derivatives with respect to t_{new}.

$$\begin{aligned}
\dot{r}_i &= \frac{\partial H_N}{\partial p} = \frac{p_i}{ms} \\
\dot{p}_i &= \frac{\partial H_N}{\partial r} = F_i s \\
\dot{s} &= \frac{\partial H_N}{\partial p_s} = \frac{p_s s}{Q} \\
\dot{p}_s &= \frac{\partial H_N}{\partial s} = \sum \frac{p_i^2}{ms^2} - N_{df} k_B T
\end{aligned} \tag{3.40}$$

And

$$\ddot{r}_i = \frac{d\dot{r}_i}{dt} = \frac{\dot{p}_i}{ms} - \frac{p_i}{m}\frac{\dot{s}}{s^2} = \frac{\vec{F}_i}{m_i} - \xi \dot{r}_i \tag{3.41}$$

where $\xi = \frac{p_s}{Q}$ is the thermodynamic friction coefficient that evolves in time according:

$$\dot{\xi} = \frac{\sum m\dot{r}^2 - N_{df} k_B T}{Q} = \frac{T(t) - T_0}{Q} \tag{3.42}$$

The time derivative is calculated from the difference between the current temperature $T(t)$ and the reference temperature T_0, and Q determines the strength of the coupling to the heat bath.

Langevin

In the Langevin thermostat [160], an additional random force term and a constant frictional coefficient are introduced to the system.

$$\ddot{r}_i(t) = \frac{F_i}{m_i} - \beta \dot{r}_i(t) + \eta_i(t) \tag{3.43}$$

where $n_i(t)$ is randomly extracted from a gaussian distribution having the following properties:

$$\begin{aligned} \langle \eta_i(t) \rangle &= 0 \\ \langle \eta_i(t) \cdot \eta_i(0) \rangle &= 2 k_B T_0 \beta_i \delta(t) \end{aligned} \tag{3.44}$$

with $\delta(t)$ indicating the Dirac delta. In many cases random numbers for $\eta_i(t)$ are extracted individually for each atom.

3.2.8 Pressure control

Berendsen barostat

The Berendsen algorithm, to control the pressure of the system, is similar to the Berendsen thermostat. In this case, the algorithm scales the coordinates and the box vectors. So the effect is a first-order kinetic relaxation of the instantaneous pressure $P(t)$ toward a reference pressure P_0, with a time constant τ_P that can be specified as an input parameter. The equations for the Berendsen algorithm barostat are:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \tag{3.45}$$

The pressure is given by

$$P = \frac{2}{3V}\big(E_k - \Theta\big), \tag{3.46}$$

where Θ is the internal virial for pair-additive potentials:

$$\begin{aligned} \Theta &= -\frac{1}{2} \sum_{i<j} r_{ij} F_{ij} \\ r_{ij} &= r_i - r_j \end{aligned} \tag{3.47}$$

and F_{ij} is the force on particle i due to particle j. Since intramolecular contributions to the pressure vanish, in molecular systems Eqs(2) and (3) can be evaluated using only the center of mass coordinates and velocities and forces acting on the centers of mass. A pressure change can be acomplished by changing the virial through scaling the interparticle distances. A simple proportional coordinate scaling, concomitant with volume scaling, minimizes local disturbances. So, an extra term in the equation of $\dot{x} = v$ is added, proportional to x ($\dot{x} = v + \alpha x$), while the volume changes accordingly:

$$\begin{aligned} \dot{V} &= 3\alpha V \\ \alpha &= -\frac{P_0 - P}{\tau_P}\frac{\beta}{3} \end{aligned} \tag{3.48}$$

The modified equation of motion is

$$\dot{x} = v - \frac{P_0 - P}{\tau_P}\frac{\beta}{3} \cdot x \tag{3.49}$$

This proportional scaling of coordinates and box length l (assuming an isotropic system in a cubic box) per time step from x to μx and l to μl with the scaling factor μ:

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_P}\big(P_0 - P\big) \tag{3.50}$$

Te compressibility, that may not be accurately known, occurs in the same expression for the scaling factor μ. Since the accuracy in β only influences the accuracy of the noncritical time constant τ_P, the imprecision of β is of no consequence for the dynamics. The equation 3.50 apply to an isotropic system. However, it is possible to modify the equations for anisotropic systems.

Nosé -Hoover / Langevin piston

One of the pressure control schemes used in this thesis is Nosè-Hoover / Langevin piston introduced in NAMD [154]. Such barostat is a hybrid method, combining two pressure control schemes, the Nosè-Hoover like barostat [161], and the Langevin–piston barostat [159]. Both barostats consider the unit cell vectors as virtual particles, and both require that a temperature control method is used in conjunction for the degrees of freedom of the extended set (atom coordinates plus unit cells).

Specifically, the barostat referred as Nosè-Hoover was introduced in order to remove the unphysical dependence of the trajectory on unit cell vectors [161], particularly in markedly anisotropic systems. The degrees of freedom of unit cell vectors are embedded in the temperature control scheme, using a formalism that leads to the correct isothermal/isobaric partition function [161].

On the other hand, the Langevin-piston barostat [159], while also explicitly considering the unit cell vectors as additional degrees of freedom, assigns them to a Langevin thermostat rather than a Nosè-Hoover one.

The equations of motion of each atom are modified by the introduction of an additional degree of freedom, the piston σ, which undergoes a Langevin like dynamics, in analogy with the motion of each particle (also in this case, the absolute coordinates are used rather than the scaled ones):

$$\begin{aligned} m_i \frac{d^2\vec{r}_i(t)}{d^2t} &= \vec{f}_i(t) - \beta_i \frac{d\vec{r}_i(t)}{dt} + \vec{\eta}_i(t) + \frac{d^2U(t)}{d^2t} U^{-1}\vec{r}_i(t) \\ \frac{d^2U(t)}{d^2t} &= \frac{V(U)}{W}(P - P_{ref}) - \beta_U \frac{dU(t)}{dt} + \eta_U(t) \end{aligned} \tag{3.51}$$

Where β_i and the vector $\vec{\eta}_i(t)$ are the Langevin friction coefficient and noise term for the simulated atoms, β_U and the matrix $\eta_U(t)$ are those for the unit cell vectors. Therefore, this barostat also includes explicitly the Langevin thermostat for the atomic degrees of freedom.

3.2.9 Analysis of molecular dynamics trajectories

The data obtained from MD trajectories can be used to calculate several structural and dynamical properties.

Root Mean Square Displacement (RMSD)

The displacement of the atoms or group of atoms along the trajectory could be estimated calculating the Root Mean Square Displacement (RMSD). The RMSD of a set of N atoms at time t, with respect to a reference conformation (ex: initial conformation), reads:

$$RMSD(t) = \sqrt{\frac{\sum_{i=0}^{N} \left|\vec{r}_i(t) - \vec{r}_i^0\right|^2}{N}} \tag{3.52}$$

Where $\left|\vec{r}_i(t) - \vec{r}_i^0\right|$ is the displacement of the i^{th} atom at time t from the reference position r_i^0. An increase of the RMSD indicates that the protein moves to a conformation different from the initial structure and thus suggests an incomplete sampling or a conformational change.

Root mean square fluctuation (RMSF)

Similar to the previous analysis is the calculation of the fluctuations of an atom or a groups of atoms, this calculated according:

$$RMSF_i = \sqrt{\left\langle (\vec{r}_i - \langle \vec{r}_i \rangle)^2 \right\rangle} \tag{3.53}$$

where $\vec{r}_i$ is the position vector of the i^{th} atom and the brackets stand for a temporal average. This analysis give clues about the most mobile regions of the protein.

Electrostatic Calculations

Electrostatic potentials have been calculated with the Poisson-Boltzmann equation [162]. In this approach, the solvent is represented as continuum and the electrostatic potential $\Psi(r)$ can be determined by solving the following equation:

$$\vec{\nabla} \cdot \left(\varepsilon(\vec{r}) \vec{\nabla} \Psi(\vec{r}) \right) = -4\pi\rho^f(\vec{r}) - 4\pi \sum_i c_i^\infty z_i q \exp\left(\frac{-z_i q \Psi(\vec{r})}{k_B T} \right) \lambda(\vec{r}) \tag{3.54}$$

Where $\rho^f(\vec{r})$ is the fixed charge distribution of the molecule, $\varepsilon(\vec{r})$ is the position dependent dielectric constant, c_i^∞ is the concentration of ion i at an infinite distance from the molecule, z_i is its valency, q is the proton charge, k_B is the Boltzmann constant, T is the temperature and $\lambda(\vec{r})$ describes the accessibility to ions at point $\vec{r}$.

An analytical solution of the equation can be obtained only for very simple geometries and charge distributions. For proteins, the equation can be solved numerically using the finite difference methods on a grid. As a first step, the molecular surface is generated using a spherical probe of 1.4 Å , which is the water radius. A grid is generated and, according to the surface, a high dielectric constant value is then assigned to the grid points that lie on the external side of the macromolecule, which are supposed to belong to the solvent, and a low dielectric constant value is assigned to the remaining grid points. Each atom charge is split to the nearest grid points of the mesh and a smoothing algorithms are used to improve the results. Finally, the electrostatic potential is numerically estimated for each grid point and

the energy of the charged grid is calculated. Since every atom charge is split to different grid points, different parts of the charge of the same atom interact each other. The self-grid energy is removed in the difference between energies obtained for the same protein conformation with the same grid in different conditions.

Here, the Poisson-Boltzmann equation is numerically solved using the APBS program [163].

Chapter 4

Computational approach to the investigation of alpha-synuclein/dopamine interactions

4.1 Introduction

The interplay between dopamine and α-synuclein (AS) plays a central role in Parkinson's disease. Unraveling structural determinants of dopamine binding may shed light on the mechanisms by which this molecule modulates AS fibrillization and toxicity,thus providing new clues for therapeutics intervention in PD and related diseases. In this chapter, to clarify the structural determinants of the interactions between dopamine and AS, a series of computational studies to identify the exact residues and the nature of the interactions mediating the dopamine binding to AS was performed. The molecular modeling results were later corroborated by biochemical experiments performed by a group in collaboration. Our modeling studies involve dopamine (DOP) and several of its derivatives (Fig. 4.2) because all of these species could be present under physiological conditions [164] and interacts with AS [100].

The ^{125}YEMPS^{129} region was predicted to form non-specific hydrophobic contacts with the ligands' aromatic ring. The ligands were also stabilized by significant long-range electrostatic interactions with E83 in the NAC region. By forming these interactions with AS, dopamine may affect AS fibrillation properties. The *in vitro* the effects of dopamine on the aggregation of mutants designed to alter the properties of these moieties were later investigated by the Group of Prof.

Gustincich. They found that point mutations in the ^{125}YEMPS^{129} region do not largely affect fibrillization, consistently with the fact that dopamine interacts non-specifically with that region. In contrast, as suggested by modeling, the E83A mutation abolishes the ability of dopamine to inhibit AS aggregation.

Together, these findings demonstrate how the use of structural models of even highly heterogeneous states of proteins determined by NMR spectroscopy in combination with molecular dynamics simulations can provide novel insight into the structural determinants of ligand binding. This strategy which enables ligands, that mimic the interactions between dopamine and AS, to be devised, and it will help to develop therapeutics for PD and related synucleinopathies.

4.2 Methodologies Used

4.2.1 Cluster analysis

The modeling is based on the 4,000 NMR conformations of AS [14]. First, the 938 conformations among them which exhibit 5 or more amino acids (not Gly) in the forbidden region of the Ramachandran plot were discarted, as defined by the "what_check" algorithm [165]. Then, we selected a set of 87 representative conformations out of the remaining 3,062 ones using the cluster analysis method proposed by Micheletti et al. [130]. This is a method for clustering conformational ensembles into suitable similarity classes, and it is based on the RMS pair distance distribution. A RMSD cutoff of 19 Å was used for the selection of the representative conformers. To crosscheck the results, we carried out a second analysis, based on a smaller number of starting structures (1,000 randomly selected out of the ensemble of 3,062 NMR structures above) using a different clustering algorithm [131, 134].

4.2.2 Docking.

The ligands in Fig. 4.2 were docked onto the first 6 representatives conformations of AS using AUTODOCK 3.0 [144]. The parameters for the molecules were calculated using its standard parameterization procedure (Morris G.M. et al., 1998). The Lamarckian Genetic Algorithm [144] was applied as a search method for the different docking results. Therefore, 54 complexes were considered. The potential grid map for each atom type was calculated using a cubic box of 252 grid points in each direction, with a distance of 0.5 Å between grid points. For each complex, 100 docking runs were performed resulting in a total of 5,400 calculations. The ligand location was then identified using cutoff distance criterion of 5 Å between the Cα of each amino acid and the center of mass of each ligand. The most probable region of interaction for each complex was identified using a cluster analysis

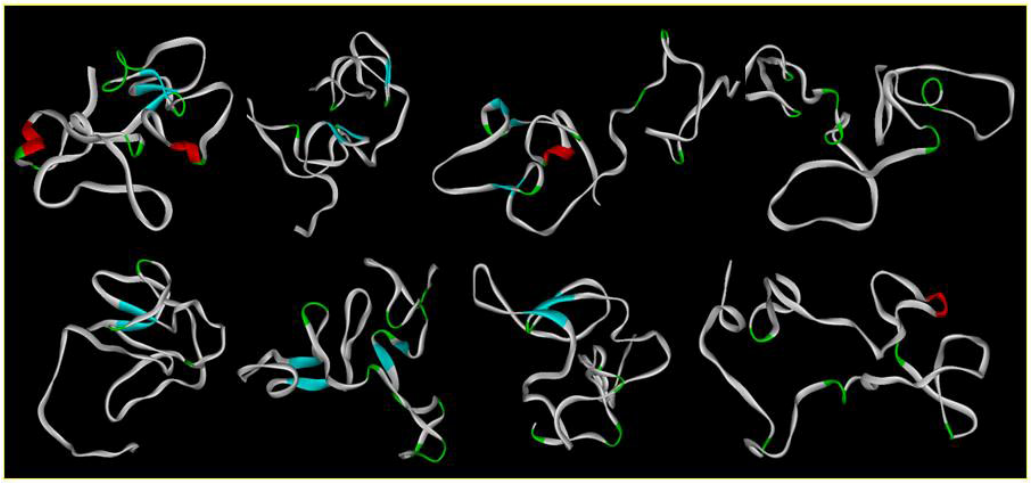


Figure 4.1: α-synuclein conformations chosen randomly from the 4,000 NMR conformations

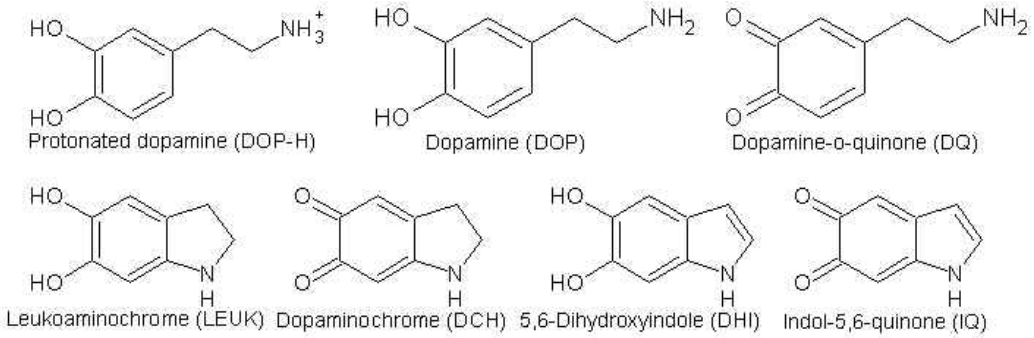


Figure 4.2: Dopamine forms (Adapted from [100]).

based on the RMSD distance between the ligand on each run (RMSD cutoff of 2 Å).

4.2.3 Molecular dynamics simulations.

Each of the 6 representative conformations, with each of the 7 ligands (Fig. 4.2) in the most populated positions obtained with AUTODOCK,were inserted into a box of water of 80 x 100 x 100 $Å^3$ edges. Periodic boundary conditions were applied by ensuring that the minimum distance between AS and its images was larger than 12 Å. The overall charge of the system studied was neutralized by the addition of 8 Na+ ions. The AMBER99 [166] and TIP3P [167] force fields were used for biomolecules with counter ions and water respectively. As for the ligands, RESP atomic charges [141] were calculated at the HF/6-31G* level of theory,

using the GAUSSIAN98 suite of programs [168] (Tab. D.1). The atom types and parameters (i.e.van der Waals, bond lengths, valence angles and dihedral angles) were assigned according the AMBER99 force field. Electrostatic interactions were calculated using the particle mesh Ewald method , with 64x64x64 grid points. A cutoff distance of 10 Å for the real part of the electrostatic and van der Waals interactions was used. The time-step was set to 2 fs. The SHAKE algorithm [169] was applied to fix all bond lengths. The simulations were performed at T=300K and P=1.013 bar, by coupling the systems with a Langevin thermostat and Nose-Hoover Langevin barostat [159]. The coupling coefficient of 5 ps-1 for the Langevin thermostat was used for all atoms in the complexes studied. For the Nose-Hoover Langevin barostat, an oscillation period of 200 fs and the damping timescale of 100 fs were used. The stability of the complexes studied on the timescale considered (6 ns) was estimated based on the RMSD and radius of gyration of the protein along the trajectory. MD calculations were performed using NAMD program [154] while the obtained results were analyzed using Gromacs and VMD programs. The Tanimoto coefficients for the electrostatic potential (Te) and shape (Ts) of the ligands [170] were calculated using EON code for chemical similarity analysis. Electrostatic interaction energies between residues of the NAC region and the ligands was calculated using the Amber force field as in the work by Guidoni et al [171]. We considered only the residues which form such interactions for at least 80% of the dynamics. This is a very approximate methodology and it was used here only to provide qualitative results.

4.2.4 Additional experiments

In order to prove the molecular modeling results, several experiments were done by the group of Prof. Gustincich

Human WT AS cDNA was cloned into the bacterial and the mutants were obtained using site-directed mutagenesis. Later, fibrillization studies were performed on the WT protein and all the mutants in absence or presence of an equimolar quantity of dopamine. Fibril formation was monitored by the ThT assay, circular dichroism (CD)spectroscopy and gel electrophoresis (SDS-PAGE). The final fibrils were then observev by Transmission Electron Microscopy (TEM).

4.3 Results

4.3.1 Docking of dopamine and its derivatives on AS.

A set of 6 structures representing about 75% of the total number of conformers were selected from the NMR ensemble of 4,000 structures using a cluster methodology

[130]. The adducts with dopamine (DOP) and its derivatives (Fig. 4.2) were constructed by molecular docking (4,200 complexes). Most ligands bind mostly to the C-terminal region that includes the residues ^{125}YEMPS^{129} , (Fig. 4.3) and Tab. 4.1).

This result is consistent with previous *in vitro* biochemical and biophysical studies implicating these residues in dopamine-induced inhibition of AS fibrillation *in vitro* [17, 16]. However, a significant number of hits are also found at the N-terminal region of AS for LEUK, DQ and DCH (Fig. 4.3).

Region	DCH	DHI	DOP	DOP-H	DQ	IQ	LEUK
100-140	40.7 %	39.2 %	62.2 %	71.3 %	74.7 %	42.7 %	38.3 %
110-140	32.3 %	27.3 %	43.4 %	51.9 %	55.9 %	30.0 %	27.7 %
120-140	21.1 %	18.3 %	32.3 %	37.0 %	38.2 %	19.1 %	17.6 %
125-129	7.6 %	5.4 %	10.4 %	9.5 %	5.9 %	5.1 %	4.6 %

Table 4.1: Relative contribution for the binding of the C-terminal regions, calculated as percentages of the total number of ligand-protein contacts.

4.3.2 Molecular dynamics simulations of the docked structures.

Because of the lack of solvent in the docking procedure, we next investigated the structural stability of the docked structures in aqueous solution by molecular dynamics simulations (MD). We considered the most populated complexes between the first 6 representatives and each ligand as obtained by the docking procedure (overall 42 starting complexes).

In more than 60% of the cases, the ligands form stable contacts with AS. In as much as 73% of the ensemble of the stable adducts, the binding region includes the $^{125}YEMPS^{129}$ residues in the C-terminal region (Fig. 4.4 and Tab. 4.2); specifically, it involves hydrophobic interactions between the aromatic ring of the ligand and hydrophobic side chains of AS, (Tab. 4.2 shows the interactions. In addition, the O and N groups of the ligands form, in some cases, hydrogen bonds (H-bonds) to polar groups in the same region of the AS protein. The similar binding mode is paralleled by the remarkable structural and electrostatic similarity between the ligands, which is pointed out by the calculated Tanimoto coefficients for the electrostatic potential (Te) and shape (Ts) [170] between each of them (values ranging between 0.7±0.3 and 0.8±0.1 for Te and Ts, respectively, see Tab. 4.3). We conclude that the ligands share a similar binding mode in which the most frequent contacts involve the ligands' aromatic ring moieties.

An analysis of the structural properties of the protein suggests that the C-terminal binding region assumes a specific conformation accommodating the ligand (Fig.

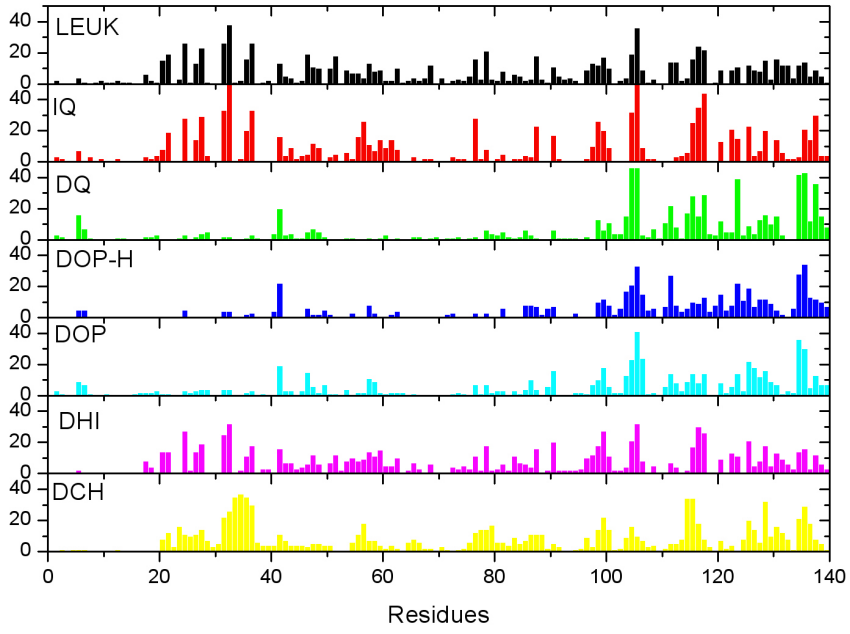


Figure 4.3: Number of hits between α-synuclein (AS) and the seven ligands as obtained by 4,200 docking runs. A hit is defined as a contact within a cutoff of 5 Å between the Cα of each amino acid and the center of mass of each ligand.

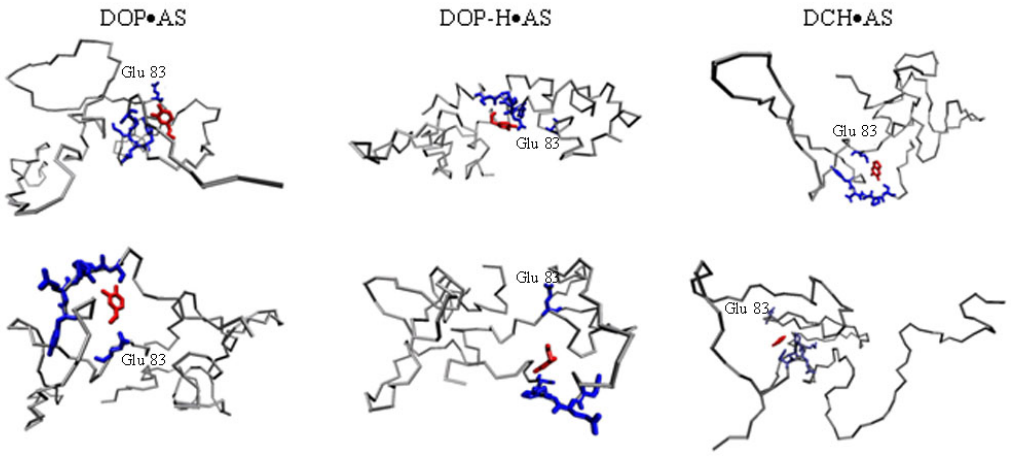


Figure 4.4: α-synuclein conformations of the 11 conformations out of 18, from the MD simulations, where the ligands bind to the ^{125}YEMPS^{129} region. Here we show the six structures (among those) which form an additional (water-mediated) contact with E83. The 125-129 residues and E83 are colored in blue, the ligand is colored in red

Cluster	Ligand	Hydrogen Bonds	Hydrophobic contacts
1 (39 %)	DCH	Met127(O)-DCH(N1), (D: 3.7±0.8 Å) Ser129(O)-DCH(O1), (D: 4.2±0.5 Å)	Tyr125, (D: 4.7±0.6 Å) Tyr136, (D: 6.4±0.6 Å)
	DHI	Glu137(OE2)-DHI(N1), (D: 2.8±0.2 Å)	Lys80, (D: 4.0±0.3 Å) Asp135, (D: 7.0±0.3 Å) Tyr136, (D: 9.9±0.5 Å)
	DOP		Lys96, (D: 6.2±0.7 Å) Val118, (D: 6.0±09 Å) Pro120, (D: 5.4±0.5 Å)
	DOP-H	Glu123(OE2)-DOP-H(O1), (D: 3.5±1.2 Å) Glu123(OE2)-DOP-H(O2), (D: 3.0±0.6 Å)	
	DQ	Asp135(N)-DQ(O2), (D: 3.3±0.4 Å) Gly111(O)-DQ(N1), (D: 3.9±1.0 Å)	Ile112, (D: 6.6±0.7 Å) Asp135, (D: 5.4±1.0 Å)
	IQ		Thr81, (D: 4.0±0.2 Å)
	LEUK		Glu131, (D: 7.4±2.2 Å) Gly132, (D: 5.8±2.1 Å)
2 (15 %)	DOP-H	Ser129(O)-DOP-H(N1), (D: 3.7±0.9 Å) Glu131(O)-DOP-H(N1), (D: 2.8±0.2 Å)	Ser129, (D: 5.4±0.7 Å)
3 (7 %)	DOP		Met127, (D: 7.8±2.4 Å)
	DQ	Thr81(N)-DQ(O2), (D: 4.6±0.5 Å)	Lys34, (D: 6.9±0.7 Å)
	IQ		Lys34, (D: 5.9±0.8 Å)
4 (6 %)	DCH	Ala90(N)-DCH(O1), (D: 3.3±0.7 Å) Lys97(NZ)-DCH(O2), (D: 3.0±0.5 Å)	Phe94, (D: 4.5±0.3 Å) Val118, (D: 5.7±0.6 Å) Tyr136 (D: 6.3±0.5 Å)
	DHI		Ala90, (D: 4.8±0.3 Å) Phe94, (D: 6.5±0.5 Å) Lys97, (D: 4.9±0.3 Å)
	DOP	Gly68(N)-DOP(O1), (D: 3.4±0.4 Å)	Gly67, (D: 4.8±0.6 Å) His50, (D: 5.0±0.5 Å) Val66, (D: 6.2±0.5 Å)
	DOP-H	Thr92(O)-DOP-H(N1), (D: 3.5±0.8 Å)	Tyr125, (D: 5.5±0.7 Å)
	DQ	Gln134(NE2)-DQ(O2), (D: 3.5±0.5 Å)	Tyr39, (D: 4.7±0.3 Å) Val49, (D: 6.3±0.5 Å)
	IQ	Glu123(OE2)-IQ(N1), (D: 3.2±0.8 Å)	Phe94, (D: 5.4±0.6 Å) Met116, (D: 5.5±0.7 Å)
5 (4 %)	DCH	Glu105(OE2)-DCH(N1), (D: 4.0±0.7 Å) Met116(O)-DCH(N1), (D: 3.5±0.4 Å)	Asp115, (D: 4.8±0.4 Å) Pro117, (D: 6.6±0.5 Å)
	DOP-H	Glu105(OE2)-DOP-H(O1), (D: 2.5±0.1 Å)	Val118, (D: 6.0±0.6 Å)
	IQ		Tyr125, (D: 6.2±1.9 Å)
6 (4 %)	DCH		Ser129, (D: 6.3±1.2 Å) Tyr133, (D: 5.7±1.4 Å)
	DHI		Gly41, (D: 6.4±0.8 Å) Pro128, (D: 4.8±0.6 Å)
	DOP	Ala89(O)-DOP(O1), (D: 3.5±0.9 Å)	Leu113, (D: 6.0±0.7 Å) Asp135, (D: 6.3±0.4 Å)
	DOP-H	Asp98(OD1)-DOP-H(N1), (D: 2.9±0.4 Å)	Asn65, (D: 5.4±0.6 Å) Glu61, (D: 5.5±0.2 Å)
	DQ	Ala89(O)-DQ(N1), (D: 3.3±0.5 Å)	Ala90, (D: 5.8±0.5 Å) Ser129, (D: 6.3±1.4 Å) Tyr133, (D: 6.1±1.3 Å)
	IQ		Ala56, (D: 5.5±1.0 Å) Glu57, (D: 5.5±0.9 Å

Table 4.2: H-bonds and hydrophobic contacts in Dopamine-AS 'stable' adducts. Several of then shown interactions with the C-terminal region, including the $^{125}YEMPS^{129}$ region. The distance (D) in the hydrogen bond column was measured between the heavy atoms. The hydrophobic contacts were measured as the distance (D) between the center of mass of the ligand and the specific amino acid.

Ts, Te	DCH	DHI	DOP	DQ	IQ	LEUK
DCH	1, 1	0.99, 0.38	0.85, 0.38	0.88, 0.82	0.99, 0.96	0.99, 0.46
DHI		1, 1	0.85, 0.54	0.86, 0.37	0.99, 0.38	0.99, 0.91
DOP			1, 1	0.99, 0.15	0.85, 0.02	0.86, 0.51
DQ				1, 1	0.86, 0.72	0.86, 0.43
IQ					1, 1	0.99, 0.51
LEUK						1, 1

Table 4.3: Similarity between the ligands in Fig. 4.2 The Tanimoto coefficients [170] for the shape (Ts) and the electrostatic potential (Te) between each ligand and all the others are shown. DOP-H is not shown due to its +1 charge.

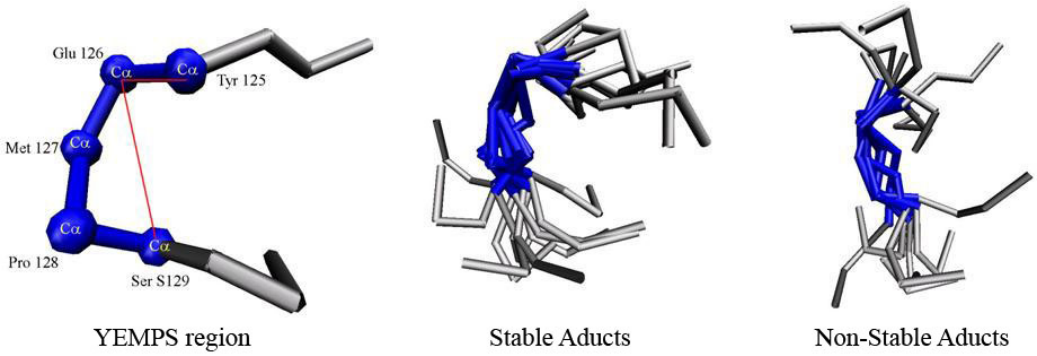


Figure 4.5: Close up on the ^{125}YEMPS^{129} Cα carbons (YEMPS region) after the MD simulation. If the ligands bind to this region (Stable Aducts), the angle between Cα 125-126-129 shows a relatively small spread (79°±11°). The remaining 7 conformations is also shown (Non-Stable Aducts) (105°±20°).

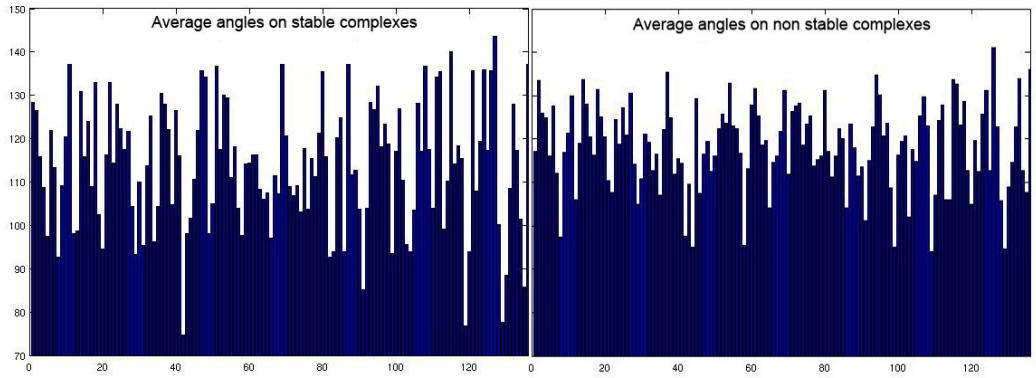


Figure 4.6: Average values of angles formed by Cα (n – n+1 – n+4) on stable (left) and non stable (right) adducts.

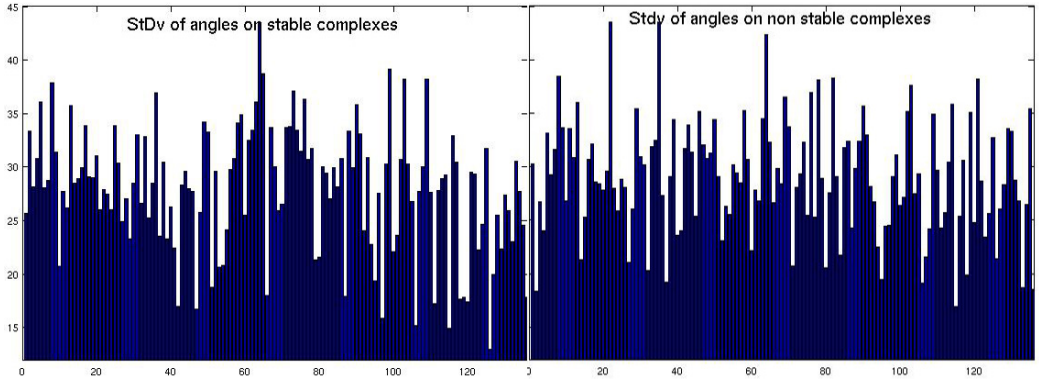
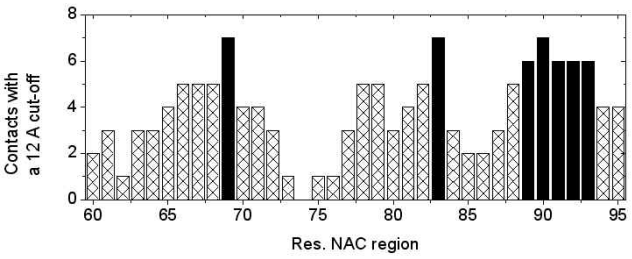


Figure 4.7: standard deviation of the angles formed by Cα (n – n+1 – n+4) on stable (left) and non stable (right) adducts (the average is 30° for the stable adducts and 28 for the non stable adducts).

4.5). To quantify this property, we calculated the standard deviation (SD) of the n, n+1, n+4 angles among the Cα's of the protein for the 26 stable and 16 unstable complexes. In the 26 stable complexes the smallest values are those of the C-terminal binding region (SD of the angle 125-126-129 is 13°; to be compared to the value of 28° (averaged over all the angles' SD) (Fig. 4.6 and 4.7). In addition, these deviations are smaller than any of the correspondent angles in the 16 unstable complexes (average 29°) (Fig. 4.6 and 4.7).

Taken together, our calculations suggest that almost all the ligands bind mostly to the C-terminal part of the protein that includes the ^{125}YEMPS^{129} region, fully consistent with previous *in vitro* experiments [17, 16] and in all cases, AS always assumes a similar, kinked conformation in its binding region (Fig. 4.6).

The discussion of these 'stable' adducts so far has been focusing on the C-terminal region. As for the rest of the protein, we notice that the ligands spend a significant



Res	Av	Std
69	1.0	0.3
83	**1.8**	**0.7**
89	1.5	1.3
90	1.0	1.1
91	0.9	1.1
92	1.2	1.1
93	1.2	0.8

Figure 4.8: Electrostatic analysis. Left: Number of times that NAC amino acids are found within a 12 Å from the ligands of Fig. 4.2.The residues selected for the electrostatic analysis are marked in black. Right. Averaged energies values, for the selected residues (Res), expressed in units of the total average (Av) electrostatic interaction energy of -2.7 Kcal/mol and total standard deviation (Std) of 2.2 Kcal/mol.

amount of the simulated time in proximity of the NAC residues (Fig. 4.8), although they never form a direct contact with them. According to a simple and very approximate electrostatic calculation, E83 is the NAC residue forming the largest long-range electrostatic stabilization of the ligands (at least 30% larger than any other residue in the region and with the lowest standard deviation, Fig. 4.8). Test calculations suggest that nullifying E83 charge results in a large decrease of such interactions ($\sim$ 60%). Within the limitations of our analysis, we suggest that additional stabilization of dopamine in its adducts with AS may arise from electrostatic interactions with E83.

In all the remaining simulations, the ligands do not bind in a stable manner with the protein. For example, in one case with DOP, the ligand moves from its starting binding region (residues 92, 93 and 103) towards residues 99, 105 and 106 after 3 ns, and does not form stable interactions with the protein during the time scale of the MD simulation.

4.3.3 Testing the domain of applicability of our predictions

We have repeated the entire computational procedure using a different clustering algorithm [131, 134] on a different number of initial conformers (1000 conformers randomly chosen among the 3,062 initial ones). The results, summarized in the Appendix D, are very similar to those of the other analysis, suggesting that our findings are robust and do not depend significantly on the chosen clustering analysis and/or the number of conformations chosen (Figs. D.1, D.2 and Tabs. D.2, D.3, D.4).

4.3.4 Experimental validation

Our calculations suggest that all the ligands bind mostly to the C-terminal part of the protein that includes the ^{125}YEMPS^{129} region, consistently with the experimental observation that dopamine binds *in vitro* to this region [17, 16]. The ligands form nonspecific hydrophobic interactions with all of the five residues in the ^{125}YEMPS^{129} region (Tab. 4.2) and form H-bond to E126 and S129 in some cases. In all cases, AS always assumes a similar, kinked conformation in its binding region (Fig. 4.5). In addition, the ligands are significantly stabilized by electrostatic interactions with E83. To test the validity of these conclusions, the aggregation properties of four alanine mutants of AS *in vitro*, which involve the ^{125}YEMPS^{129} residues in the C-terminal region were investigated by the group of Prof. Gustincich. The investigated mutants were E83A, E126A, S129A and E83A/E126A/S19A in the presence and absence of dopamine. Because our calculations suggest that ligand-AS interactions at the C-terminus are dominated by nonspecific hydrophobic interactions, we predict that the E126A and S129A mutations would not significantly alter ligand-AS interaction. In addition, in the case of S129A, the Ser to Ala mutation might not affect H-bonding with the ligand as it involves the backbone. In contrast, the E83A mutation is expected to affect dopamine affinity for AS (and therefore it might affect the fibrillation process).

The aggregation properties of the five proteins were determined in the absence or presence of equimolar dopamine. All the mutants show increased aggregation relative to the WT (Fig. 4.9.A). Interestingly, the two AS variants containing the E83A mutation retain the ability to form fibrils in the presence of an equimolar quantity of dopamine (Fig. 4.9.A), whereas fibril formation by the WT, E126A and S129A variants is abolished in the presence of dopamine. This result clearly suggests that the nature of dopamine-AS interactions in the C-terminal region is distinct from that of the interactions in the NAC region, consistent with our predictions.

A consistent picture is obtained by monitoring the loss of soluble protein during the fibrillization reaction by SDS-PAGE. We see a decrease of soluble content that is proportional to the degree of aggregation by each protein (Fig. 4.9.B). In the presence of dopamine, the majority of WT, S129A, and E126A protein remain in solution, whereas very little soluble protein remains in the case of the two E83A-containing mutants (Fig. 4.9, red rectangles).

CD spectra show in the absence of dopamine and after incubation for 72 hrs, the majority of WT, E126A, and S129A precipitates out of solution and the CD spectra of the remaining material exhibits a predominantly random coil structure, except for S129A which exhibits a spectra consistent with species (soluble aggregates) rich in β-sheet structure (Fig. 4.9.C). Co-incubation with dopamine prevents the transition from random coil to β-sheet in the case of WT, S129A, E126A, but not in the case of E83A or E83A/E126A/S129A, further confirming that dopamine is

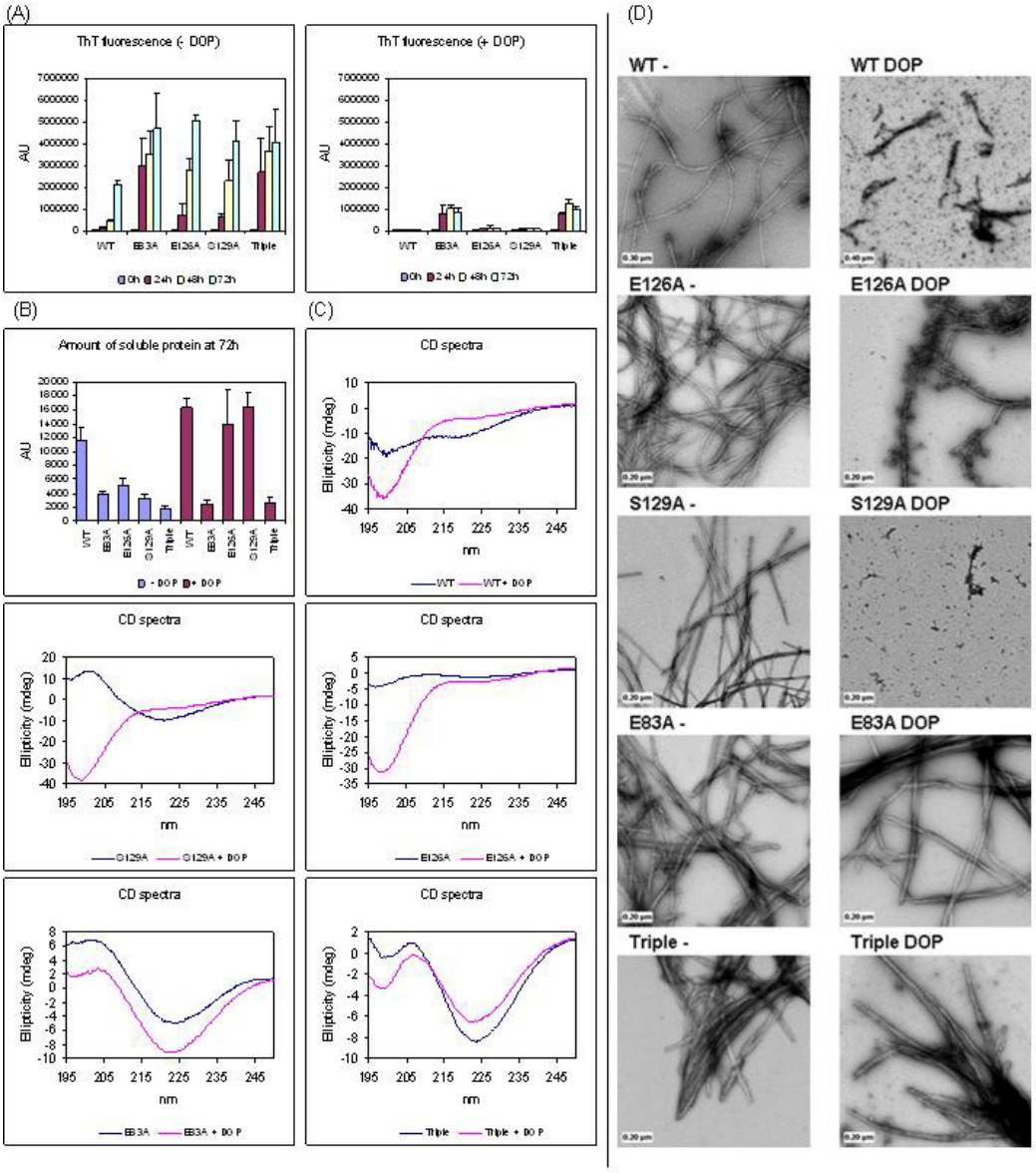


Figure 4.9: *In vitro* experiments of WT and mutant (E83A, E126A, S129A and Triple) AS in absence or presence dopamine.(A) Kinetics of fibrillization as monitored by the enhancement of Thioflavin-T (ThT) fluorescence intensity over time. (B) Amount of soluble protein remaining after 72 hrs, monitored by means of SDS-PAGE. (C) CD spectra of WT and mutant AS in absence (blue line) or presence (red line) of DOP. (D) TEM analysis.

able to prevent fibrillization of the WT, E126A and S129A mutants but not the E83A E83A/E126A/S129A mutants (Fig. 4.9.C).

The EM studies of AS in the absence of dopamine, after 48-72 hrs, the mutant proteins showed abundant fibrils resembling those formed by WT AS (Fig. 4.9.D left panels). In the presence of dopamine, WT AS, E126A and S129A formed predominantly soluble aggregates and no mature fibrils could be detected in these samples, consistent with previously reported data on the WT protein. On the contrary, addition of dopamine did not inhibit fibril formation or change the structure of the fibrils formed by both E83A containing mutants (Fig. 4.9.D right panels).

4.4 Concluding Remarks

In this chapter, the interaction between AS and dopamine was investigated by docking the latter and six of its derivatives onto selected representatives of a large ensemble of AS structures and by performing further molecular dynamics simulations on the most representative structures. Our calculations suggested that the C-terminal region including residues $^{125}YEMPS^{129}$ bind non-covalently with dopamine. This may be important for the inhibition of fibrillation and it is therefore consistent with *in vitro* experiments [17, 16]. The dopamine-binding region assumes a similar structure in all the cases that we investigated (Fig. 4.5). Thus, although the docking does not cover the entire set of conformers representing the initial configurations (it covers 75% of it), we expect that the structural determinants of the binding region may be similar for most of the non-covalent complexes. The fingerprint of dopamine-AS non-covalent complexes is the formation of nonspecific hydrophobic contacts between the ligands' aromatic ring and the $^{125}YEMPS^{129}$ region in the C-terminal (Tab. 4.2) helped by the particular conformation adopted by the binding region, accommodating the ligand (Fig. 4.5). These interactions are also complemented by nonspecific H-bonding interactions. In addition, dopamine and its derivatives are stabilized by significant electrostatic interactions with E83 in the NAC region (Fig. 4.8).

Based on our calculations, we hypothesized that the $^{125}YEMPS^{129}$-DOP, nonspecific hydrophobic interactions may affect the AS-DOP binding and hence DOP ability to modulate AS fibrillation [65]. In addition, replacement of E83 with alanine in the NAC region, affects the favorable long range electrostatic interactions with dopamine in its stable adducts. To test this hypothesis, several experiments, related to dopamine binding [79, 17, 99, 16]) were performed. Together, our findings suggest that hydrophobic interactions with the $^{125}YEMPS^{129}$ C-terminal region play a critical role in mediating the inhibition of AS fibril formation by dopamine. Instead, the E83A mutation in the NAC strongly impairs the ability of dopamine to inhibit AS aggregation. This may be consistent with our conclusion

that dopamine affinity may be stabilized by E83 long range electrostatic interactions and/or with the fact that such mutation may alter the properties of the NAC region.

These findings demonstrate how the use of structural models of even highly heterogeneous states of proteins determined by NMR spectroscopy in combination with molecular dynamics simulations can provide novel insight into the structural determinants of ligand binding.

Chapter 5

Design of dopamine mimics

5.1 Introduction

My previous study has suggested that dopamine and its derivatives form nonspecific hydrophobic contacts with the AS C-terminal as well as long-range electrostatic interactions with NAC region. Among all the residues in the NAC region, E83 was suggested to be the most important one for the interaction and consequently inhibition of the fibrillation. Here we attempt at identifying other chatechols molecules structurally and electrostatically similar to DOP and some of its derivatives [100], as these molecules might be able to inhibit AS fibrillation with higher affinity to the protein.

5.2 Methodology and Results

Dopamine mimic screening. Dopamine (DOP) and its derivatives in Fig. 4.2 were used as a template molecule for a virtual screening, based on well-established shape and electrostatic similarity based scoring functions developed by "OpenEye scientific software, Inc" [170, 139, 137].

The 1,159,274 molecules of the "ligand.info" meta-database [172] (most of them commercially available), were screened against each of the 7 template molecules (Fig. 4.2) (See Methodology 3.1.3 for details).

For each template, the top 10 molecules featuring the largest shape scoring-based scoring function and the largest electrostatic potential-based scoring functions were

Label	Name (IUPAC)	CID
1K	1H-indol-5-ol	16054
2K	1H-indole-5,6-dione	440728
3K	4-propylbenzene-1,2-diol	97638
4K	2-amino-4-propan-2-ylphenol	662440
5K	6-methylbenzene-1,2,4-triol	13026
6K	4-(2-methylaminoethyl)phenol	9727
7K	5-(2-aminoethyl)benzene-1,2,4-triol	4624
8K	2-amino-4-tert-butylphenol	70982
9K	2-methyl-3,4-dihydro-1H-isoquinoline-6,7-diol	37764
10K	1-methyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol	54456
11K	1H-indol-6-amine	256096
12K	4-(2-methylaminoethyl)benzene-1,2-diol	4382
13K	4-(2-aminoethyl)phenol	5610
14K	2,3,4-trimethyl-1H-indole-5,6-diol	266595
15K	5,6-dihydroxy-1H-indole-2-carboxylic acid	119405
16K	4-(2-aminoethyl)-2-methoxyphenol	1669
17K	2,4-diaminophenol	7266
18K	5-(2-aminoethyl)benzene-1,2,3-triol	114772
19K	2-(3,4-dihydroxyphenyl)ethyl-trimethylazanium	165581
20K	3-hydroxy-1-methyl-2,3-dihydroindole-5,6-dione	5898

Table 5.1: Names and CIDs of the selected molecules from the ligand.info meta-database. The CID is the PubChem compound ID.

The resulting 70 molecules were docked onto AS representatives (following exactly the same procedure as in the chapter 4*), resulting in 420 ligand/AS adducts. The molecules were given again a scoring, this time based on the number of contacts with the C-terminal region (i.e. the YEMPS region, as in Section 4.3.2). The procedure was identical to that of Chapter 4.

The 20 molecules with the largest score were selected (Tab. 5.1 and Fig. E.1 in Appendix E), all these molecules are commercially available.

These 20 molecules turn out to dock onto AS in a similar way as dopamine and its derivatives, as they bind at the C-term (and, at times, also at the N-term) and they form long range electrostatic interactions with residues in the NAC region (data not shown). Fig. 5.1 shows models labeled 1K, 4K, 8K and 10K as an example; all the other molecules bind in a similar way.

Dopamine mimic design. These 20 molecule were used to generate about 100 other molecules by changing or adding substituents E.1). For such procedure, the drug-like filter [143] was used (see Methodology 3.1.3 for details).

*Indeed, the representatives conformations of AS were also the same as in the previous chapter (the 6 representatives from the cluster of 3,062 NMR conformations of AS in solution).

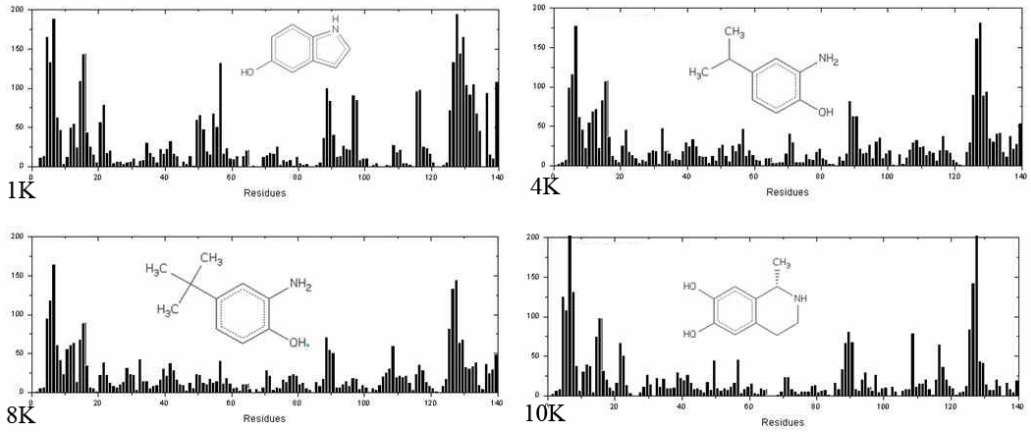


Figure 5.1: Average docking results for some of the known small compounds onto representatives conformations of AS.

By using exactly the same protocol as that used before, the new molecules were first screened against the 7 template molecules (Fig. 4.2). The two top molecules for each template were retained. Here only 2 molecules were selected because these new molecules have to be synthesized (unlike the others, which are commercially available) and so we prefer to focus only on the best ones according to our procedure. They were named here as the original template plus a digit (for instance, if a molecule has DOP as a template, its name will be DOP1, then the next DOP2 and so on).

The resulting 14 molecules underwent molecular docking as in the previous section (Fig. 5.2)

Among the 14 molecules, only the molecules named DCH1, DHI2, DOP1 and DOP2 were able to interact at the same time and in all the representatives conformations of AS with both the C-terminal and the N-terminal or NAC region (Fig. 5.3, Fig. 5.4 and Tab. 5.3), possibly because of their structural similarity: they are made by two condensed aromatic rings (which interact unspecifically with residues in the C-terminal region) and a hydrophilic tail (forming unspecific hydrogen bonds with the C-terminal and/or N-terminal regions) (see Fig. 5.4).

In vitro experiments are now called upon to test or disprove the predicted inhibition of fibrillation of the screened (20) and designed (4) molecules (Tab. 5.1, Figs. E.1 and E.2).

Label	Name (IUPAC)
DCH1	6-sulfanylidene-2,3-dihydro-1H-inden-5-one
DCH2	2-prop-1-en-2-yl-2,4-dihydro-1-benzofuran-5,6-dione
DHI1	5-bromo-1H-indol-6-amine
DHI2	5-fluoro-1H-indol-6-amine
DOP1	2-(2,3-dihydro-1-benzofuran-6-yl)ethanol
DOP2	N-ethyl-2,3-dihydro-1-benzofuran-6-amine
DOPH1	2-(3-fluoro-4-methylphenyl)ethylazanium
DOPH2	2-(3-chloro-4-methylphenyl)ethylazanium
DQ1	3-(2-aminoethyl)-6-sulfanylidenecyclohexa-2,4-dien-1-one
DQ2	4-(2-aminoethyl)cyclohexa-3,5-diene-1,2-dione
IQ1	5,6-difluoro-1H-indole
IQ2	5-chloro-6-methyl-2,3-dihydro-1H-indene
LEUK1	5-fluoro-6-methyl-2,3-dihydro-1H-indene
LEUK1	6-imino-2-prop-1-en-2-yl-3,7-dihydro-2H-inden-5-one

Table 5.2: IUPAC names of the designed new molecules.

AS Conf. Num.	N-terminal	NAC	C-terminal
1	10-14	85-90	136
2	50-58		125, 134-137
3			125-133
4		88-95	96-116
5	4-6, 14-15		
6	4-8		125-131

Table 5.3: Regions on the AS sequence where the 4 selected new molecules bind each of the 6 analyzed representative conformations of AS.

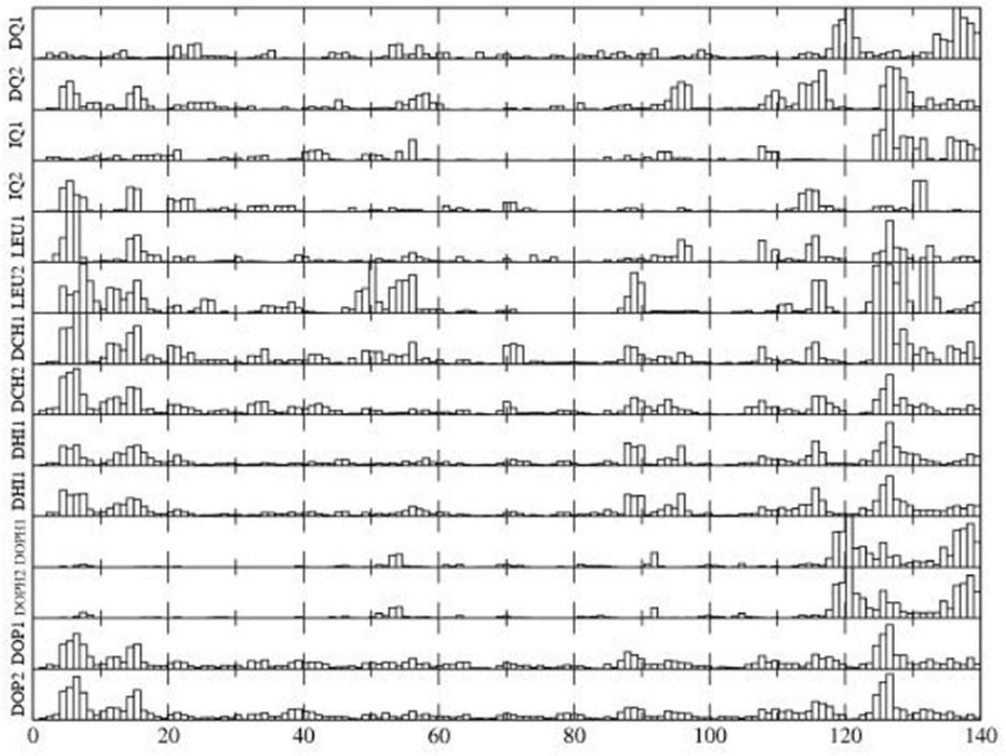


Figure 5.2: Average docking results for the first 14 compounds (two for each template) on AS's representatives.

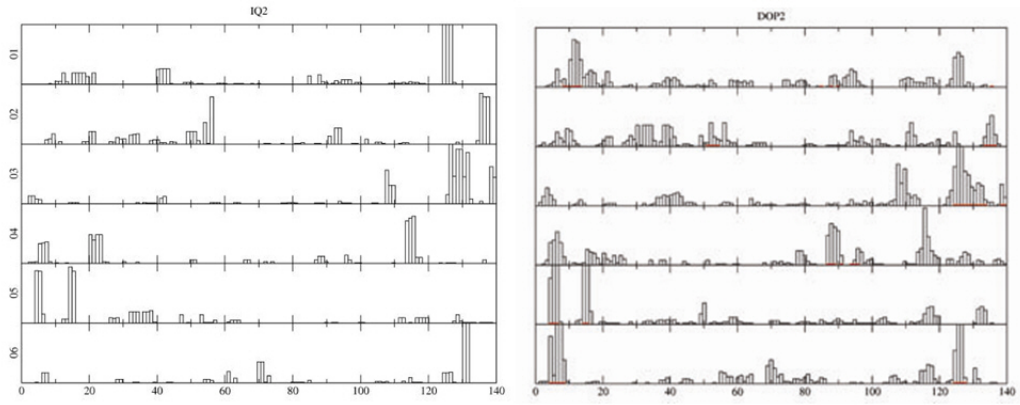


Figure 5.3: Docking results for two specific compounds (IQ2 and DOP2 in Tab. 5.2 and Fig. E.2) for each particular representative AS conformation

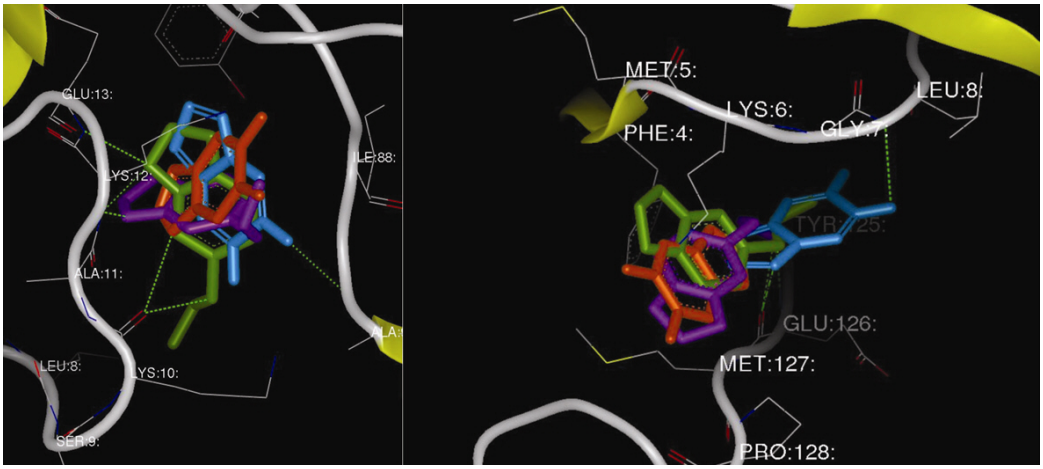


Figure 5.4: Superposition of DCH1, DHI2 DOP1 and DOP2 compounds (See Tab. 5.2 and Fig. E.2) after they have been docked onto two representatives of AS

Chapter 6

Insights into DJ-1 mutants by MD

6.1 Introduction

Mutations in the DJ-1 protein are present in patients suffering from familial forms of Parkinson's disease [110, 27].

Here we use computational methods and biological assays to investigate the relationship between DJ-1 missense mutations and the protein oligomeric state. Molecular dynamics (MD) calculations suggest that: (i) the structure of DJ-1 wild-type (WT) in aqueous solution, in both oxidized and reduced forms, is similar to the crystal structure of the reduced form [116, 102]; (ii) the PD-causing M26I variant is structurally similar to the WT, consistent with the experimental evidence showing this mutant to be a dimer as the WT [121, 115, 125] and in this work [33]; (iii) R98Q is structurally similar to the WT, consistent with the fact that this is a physiological variant [173]; (iv) the L166P monomer rapidly evolves towards a conformation significantly different than WT protein, suggesting a change in its ability to oligomerize [116, 102].

Molecular dynamics (MD) simulations based on the DJ-1 WT x-ray structure are carried out on homodimeric and monomeric forms of the WT. An extensive MD study of DJ-1 in the oxidized state has been performed (Fig. 6.1) as well as an analysis of the PD-causing mutations L166P and M26I.

Comparison is also made with MD simulations of a physiological variant (R98Q), which is expected to alter neither the DJ-1 fold nor the oligomeric state [173].

Our calculations show that: (i) the reduced and oxidized WT dimers in aqueous solution are very similar to the reduced dimer in the solid state; (ii) M26I in the

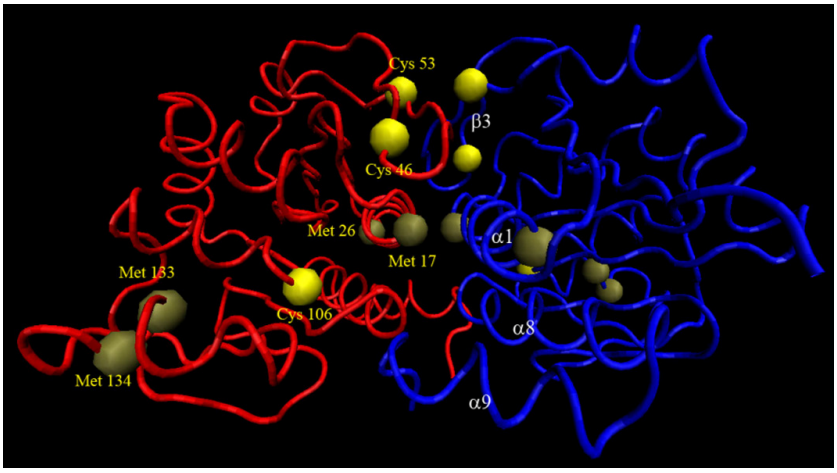


Figure 6.1: X-ray structure of the DJ-1 protein in the reduced state (PDB code 1UCF) [116]. The cysteine and methionine Cα atoms are shown as spheres. Labeling of selected secondary structure elements at the subunit-subunit interface (α1, α8, α9 helices and β3 β-sheet) as in Ref. [116].

monomeric and dimeric states maintains completely the fold of the WT; (iii) R98Q is also similar to the WT, fully consistent with the fact that this is a physiological variant; (iv) L166P causes local distortions at α8 helix (Fig. 6.1 and 6.2) and at the surface-forming contacts in the dimeric structure. We conclude that this mutation might affect the stability of the dimer by altering the structure of its local environment. Unfortunately, at present, the stability differences (i.e. the free energy differences) between the WT and these mutants cannot be firmly established by calculations alone. However, our conclusions based on the calculations are corroborated by previous biological assays *in vitro* on the WT and its mutants [125, 122, 111]. Furthermore, experiments carried out in this work confirm that L166P tends to form multimeric aggregates more than WT does [121, 119, 111].

Finally, our combined computational and experimental methodology is used to engineer a new mutation that, in contrast to L166P, does destabilize the subunit-subunit interactions without affecting the secondary structure elements of the protein. This mutation (R28A) is located at the subunit-subunit interface and it causes the disruption of salt bridges and hydrophobic contacts at the interface.

6.2 Methodologies Used

6.2.1 Molecular Dynamics simulations

Dimeric and monomeric structural models of reduced and oxidized DJ-1 WT protein together with M26I, L166P , R98Q and R28A were constructed based on the

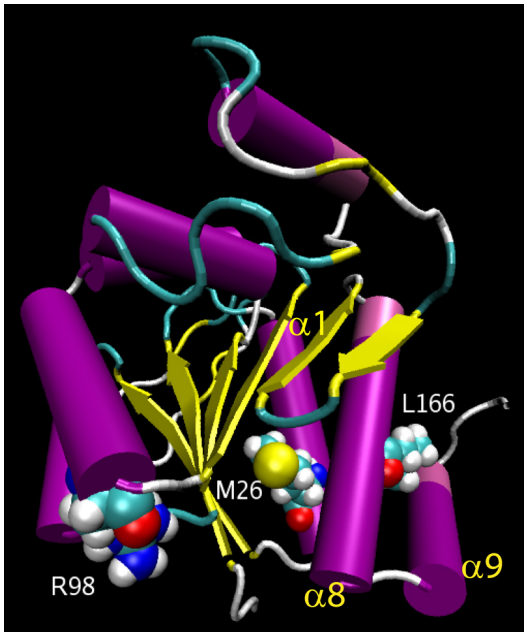


Figure 6.2: Structure of the DJ-1 monomer (PDB code 1P5F) [102], and location of PD-causing mutations (M26I, L166P) together with R98, which is mutated to Q in a physiological variant.

DJ-1 WT homodimer X-ray structure at a resolution of 1.95 Å [116].

In the oxidized form Cys and/or Met residues were assumed to be oxidized. We noticed that the shortest Cys-Cys distance (d[S-Cys53(A) and S-Cys53(B)]=3.2 Å) is too large for the formation of disulfide bridges. Thus, the three cysteines present in the protein (C46, C53 and C106) are replaced by cysteine sulphonic acid [117] (Fig. 6.1). The four methionines (M17, M26, M133 and M134, Fig. 6.1) are oxidized to methionine sulfoxide [29, 109] or methionine sulfone [109]. We further noticed that neither the cysteine residues nor the methiones interact directly with residues involved in the PD mutations. Several models were considered, based on the suggestion of Refs. [29, 109, 117]: (i) the protein with all the Cys residues oxidized and all the Met reduced (WT-OX_1); (ii) the protein with all the Cys oxidized and all the Met oxidized to methionine sulfoxide (WT-OX_2); (iii) the protein with all the Cys oxidized, two methionines (M26 and M134) oxidized to methionine sulfoxide and the remaining two (M17 and M133) oxidized to methionine sulfone (WT-OX_3).

The point mutations were obtained by simple residue substitution, taking care that the substituted residue would not clash with the rest of the protein and that the $\chi 1$ and $\chi 2$ torsion angles would fall in the most energetically favorable regions [174, 175]. The mutations in the dimeric structures were carried out for both subunits in order to obtain homodimers.

The proteins were immersed in a water box of edges ca. 69.0 Å x 66.0 Å x 69.0 Å and 81.0 Å x 76.0 Å x 94.0 Å for the monomeric and dimeric forms respectively. The solvent molecules were not included if the distance between any solvent atom and any protein atom was lower than the sum of their respective van der Waals' radii. Two or three sodium counterions were added to neutralize the systems. They were located in the regions of minimum electrostatic potential energy as calculated with the xleap program of the AMBER8 package [176, 177], namely close to Glu 59 and Asp 189. Periodic boundary conditions (PBC) were applied.

The AMBER99 [178, 166] and TIP3P [167] force fields were used for biomolecules with counterions and water respectively. The particle mesh Ewald method (PME) was applied to evaluate the long-range electrostatic interactions [179, 180, 181]. A cutoff of 8 Å was used for both the real part of the electrostatic interaction and the van der Waals non-bonded interaction evaluation. A timestep of 2 fs was applied to propagate nuclear degrees of freedom. The SHAKE algorithm [169] was used to fix all bond lengths. The investigated models first underwent 10,000 steps of energy minimization. Then, the systems were equilibrated at constant temperature and pressure for at least 0.5 ns. Subsequently, our models underwent MD simulations with constant temperature and volume for at least 10.0 ns. Constant temperature conditions were obtained by using a Langevin thermostat [160] at a target temperature of 300K with a coupling coefficient of 5 ps^{-1}. Constant pressure conditions were achieved with a Nosé-Hoover Langevin barostat [182, 159]. For the barostat, an oscillation period of 200 fs was used and the damping timescale used was 100 fs.

All calculations were performed using the NAMD program [154]. The obtained results have been further analyzed using the Gromacs [183, 184] and VMD [185] packages.

6.2.2 Computational Alanine Scanning

The mutation causing the largest destabilization of the dimer (R28A) was identified using the alanine scanning procedure (Methodology 3.1.2). This mutation was selected for subsequent MD simulations according to the protocol described in the previous subsection.

6.2.3 Additional experiments

The molecular modeling results were proved by experiments. The experiments were performed in Prof. Gustincich's laboratory. The WT DJ-1 protein and its mutants were analyzed by cross-linking assays and Co-immunoprecipitation experiments. A densitometric analysis was also performed to quantify ratios between monomeric and dimeric DJ-1.

6.3 Results

6.3.1 MD of the DJ-1 WT protein

Molecular dynamics simulations over an 11 ns timescale are used here to investigate the structure of the reduced and oxidized forms of the DJ-1 protein in aqueous solution. The simulations are performed for both the dimeric and monomeric forms of the reduced state and for the dimeric form of the oxidized state. In the latter state, either the Cys or the Met residues are assumed to be oxidized [109]. The obtained results are used as a reference for comparison with the corresponding mutants.

RMSD [Å]	Dimer			Monomer		
	Average	**SD**	**Final value**	**Average**	**SD**	**Final value**
WT	1.8	0.1	1.9	1.8	0.2	2.1
WT-OX_1	1.6	0.1	1.7	—	—	—
WT-OX_2	1.4	0.1	1.5	—	—	—
WT-OX_3	1.6	0.1	1.6	—	—	—-
L166P	—	—	—	2.0	0.2	1.9
M26I	1.7	0.1	1.7	1.5	0.1	1.8
R98Q	1.4	0.1	1.5	1.9	0.2	1.7
R28A	–	—	2.5	1.8	0.1	1.9

Table 6.1: Selected MD-averaged RMSD values of the DJ-1 WT and its mutants investigated in this study. For R28A, the RMSD increases during the dynamics and therefore only the final value is given.

In the reduced state, the Root Mean Square Deviation (RMSD) values fluctuate for the last 5 ns of the MD run around a value of 1.8 ± 0.1 Å and 1.8 ± 0.2 Å for the dimer and the monomer, respectively (Fig. 6.3 and Tab. 6.1. The standard deviation (SD) value suggests that the monomer undergoes larger structural fluctuations than the dimer, pointing to a key stabilizing role of the subunit-subunit association. In particular, the association stabilizes the flexible N- and C-terminal fragments. The distance values between Cα of the residues 1 and 189 fluctuate between 23 Å and 32 Å for the dimer while for the monomeric form they are between 20 Å and 34 Å (Fig. 6.3.ii). The average distance is 27.5 Å (Fig. 6.3.ii) for both forms. The secondary structure found in the X-ray experiments is conserved in both cases. Most of the subunit-subunit hydrogen bonds, along with the hydrophobic interactions and salt bridges are conserved during the entire MD run (Tabs. 6.2 and 6.3). However, a salt bridge between Asp 24 and Arg 48, not detected in the X-ray structure, is formed after a few hundred ps of dynamics. The average distance between the centers of mass of the DJ-1 subunits in the model structure is 28.1 ± 0.3 Å while it is 27.9 Å in the X-ray structure of the DJ-1 protein (Fig. 6.3.iii).

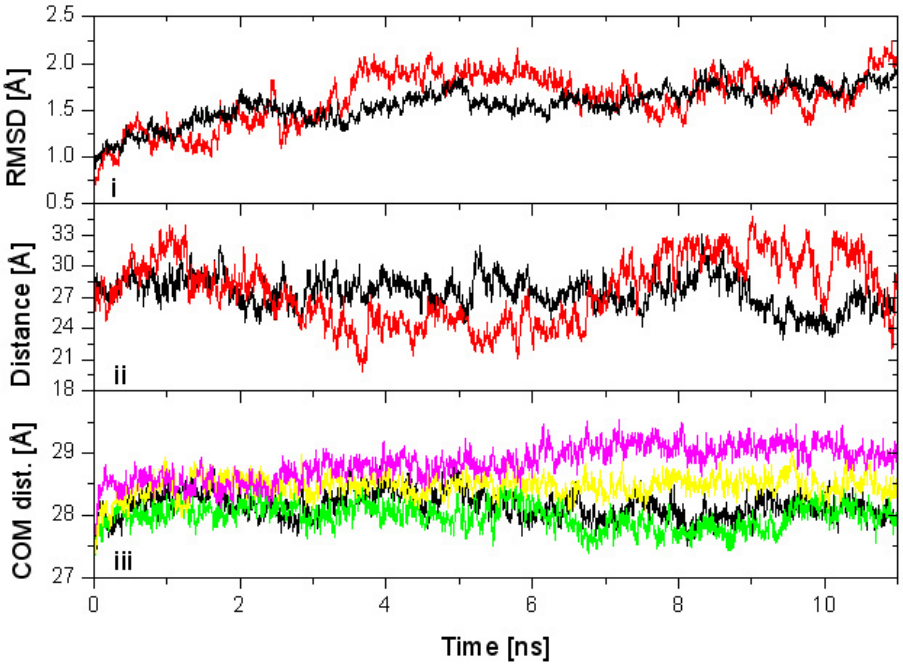


Figure 6.3: Molecular dynamics of the DJ-1 protein in aqueous solution. i. Root mean square deviation (RMSD) from the corresponding initial structure of backbone atoms of the dimeric (black) and monomeric (red) forms plotted as a function of simulation time. ii. Distance between the N-terminal and the C-terminal C-α of the dimer (black) and of the monomer (red). iii. Distances between the centers of mass (COM) of subunits in the reduced (black) and oxidized states (green (WT-OX_1), yellow (WT-OX_2) and magenta (WT-OX_3)).

Several oxidized forms of the dimeric protein are considered, following previous work [29, 109, 117]: (i) a form in which the cysteines are oxidized to cysteine sulphonic acid (WT-OX_1), (ii) one in which the cysteines are oxidized to cysteine sulphonic acid and the methionines oxidized to methionine sulfoxide (WT-OX_2) and (iii) one in which the cysteines are oxidized to cysteine sulphonic acid, two methionines (M26 and M134) oxidized to methionine sulfoxide and the remaining two (M17 and M133) oxidized to methionine sulfone (WT-OX_3) [109].

Our MD simulations show that such forms are fairly similar to the reduced form: (i) the RMSD ranges between 1.4 Å and 1.7 Å (Tab. 6.1), to be compared with 1.8 $\pm$ 0.1 Å of the WT in the reduced state; (ii) the contacts at the subunit-subunit interface are rather similar to those of the DJ-1 reduced form in aqueous solution (Tabs. 6.3 and 6.3); (iii) the MD-averaged distance between the centers of mass range between 28.0 Å and 29.1 Å (Fig. 6.3 similar to the value of the reduced state (28.1 $\pm$ 0.3 Å); (iv) the solvent accessible surface area (SASA) [186, 187] for each subunit is similar to that of the reduced form (MD-averaged values ranging from

Hydrogen Bonds

Contacts	Dist. Donor - Acceptor [Å]									Dist. Proton ... Acceptor [Å]							
	X-Ray	WT		WT-OX_1		WT-OX_2		WT-OX_3		WT		WT-OX_1		WT-OX_2		WT-OX_3	
	Dist.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
OD2, ASP24 (A) – NH1, ARG48(B)	5.1	2.9	0.4	3.0	0.5	2.8	0.2	2.9	0.3	2.0	0.4	2.1	0.5	1.9	0.2	2.0	0.3
NH1, ARG27(A) – O, ARG48(B)	2.8	2.8	0.1	2.8	0.1	2.9	0.2	3.0	0.2	1.9	0.1	1.9	0.1	2.0	0.2	2.0	0.2
NH1, ARG48(A) – OD2, ASP24(B)	5.1	2.9	0.5	2.8	0.1	3.0	0.6	2.9	0.3	2.0	0.5	1.9	0.1	2.1	0.7	2.0	0.3
O, ARG48(A) – NH1, ARG27(B)	2.8	2.8	0.1	2.8	0.1	3.0	0.4	3.0	0.2	1.9	0.1	1.9	0.1	2.1	0.4	2.0	0.3
O, VAL51(A) – N, CYS53(B)	4.3	2.9	0.2	3.0	0.2	3.0	0.2	3.1	0.2	2.0	0.2	2.1	0.2	2.1	0.2	2.1	0.2
N, CYS53(A) – O, VAL51(B)	4.3	2.9	0.4	3.2	0.3	3.0	0.2	3.0	0.2	2.0	0.4	2.3	0.3	2.1	0.2	2.1	0.2
NH1, ARG145(A) – O, VAL186(B)	3.1	3.2	0.9	3.2	0.7	3.1	0.3	3.1	0.3	2.4	0.9	2.3	0.7	2.4	0.4	2.2	0.4
N, GLY159(A) – O, LEU185(B)	2.6	2.8	0.1	2.9	0.1	2.8	0.1	2.9	0.1	2.0	0.1	2.0	0.1	1.9	0.1	2.0	0.2
O, LEU185(A) – N, GLY159(B)	2.6	2.8	0.1	2.8	0.1	2.9	0.2	4.0	0.9	1.9	0.1	2.0	0.1	2.1	0.2	3.5	0.9
O, VAL186(A) – NH2, ARG145(B)	3.1	3.2	0.5	3.4	1.3	3.2	0.3	3.2	0.4	2.4	0.5	2.6	1.3	2.4	0.5	2.4	0.5

Contacts	Angle Donor-Proton ... Acceptor [Å]							
	WT		WT-OX_1		WT-OX_2		WT-OX_3	
	Av.	SD	Av.	SD	Av.	SD	Av.	SD
OD2, ASP24 (A) – NH1, ARG48(B)	154	16	151	12	159	9	160	9
NH1, ARG27(A) – O, ARG48(B)	156	11	158	11	156	13	158	12
NH1, ARG48(A) – OD2, ASP24(B)	157	12	154	10	156	13	161	10
O, ARG48(A) – NH1, ARG27(B)	157	10	158	10	155	14	154	13
O, VAL51(A) – N, CYS53(B)	156	12	153	11	164	9	155	12
N, CYS53(A) – O, VAL51(B)	158	10	141	17	156	11	164	8.6
NH1, ARG145(A) – O, VAL186(B)	141	19	145	16	137	17	144	18
N, GLY159(A) – O, LEU185(B)	142	13	142	13	153	11	144	13
O, LEU185(A) – N, GLY159(B)	152	11	142	12	142	15	120	30
O, VAL186(A) – NH2, ARG145(B)	137	20	139	25	133	18	126	20

Table 6.2: Hydrogen bonds in the WT DJ-1 in the reduced and oxidized states as obtained by MD simulations and in the X-ray structure. MD-averaged (Av.) values along with their standard deviations (SD) are given.

Hydrophobic Contacts

	Distance [Å]								
	X-Ray	WT		WT-OX_1		WT-OX_2		WT-OX_3	
Residues	Dist.	Av.	SD	Av.	SD	Av.	SD	Av.	SD
(MET17)' – (VAL20, ILE21)"	6.0	6.2	0.3	6.0	0.5	5.1	0.2	6.1	0.3
(VAL20, ILE21)' – (MET17)"	6.0	6.2	0.4	6.0	0.3	5.6	0.3	6.2	0.5
(VAL51, ILE52)' – (VAL51, ILE52)"	5.7	5.9	0.3	6.1	0.2	5.9	0.2	6.0	0.2
(HIS126)' – (PRO184, LEU185, VAL186)"	8.5	8.5	0.4	8.5	0.5	9.2	0.3	9.1	0.6
(PHE162)' – (PHE162)"	11.1	11.9	0.4	12.0	0.4	11.2	1.2	10.5	0.6
(PRO184, LEU185, VAL186)' – (HIS126)"	8.5	8.2	0.4	8.3	0.4	10.2	0.8	9.0	0.8

Table 6.3: Hydrophobic contacts in the WT DJ-1 in the reduced and oxidized states as obtained by MD simulations and in the X-ray structure. The distances between the centers of mass of groups of residues involved in the interactions are presented. The chain A and B are identified by the ' and " respectively.

9,854 Å^2 to 10,150 Å^2, to be compared with 10,160 Å^2 for the reduced state).

We conclude that the oxidized and reduced forms of the DJ-1 protein in aqueous solution are very similar to the reduced form in the solid state.

6.3.2 MD of PD-causing DJ-1 mutations

Here we focus on the most studied PD-linked mutations: M26I and L166P [121, 115, 125, 50, 122, 111, 28]. Because the estimation of the free energy difference between the WT and these mutants in the dimeric and oligomeric states by MD is not possible at the present stage, we adopt here a simple approach which can provide some indirect hints: first, we compare MD results on the mutant monomers to the monomeric WT in the reduced state in aqueous solution. Within the limitation of the timescale investigated (∼10 ns), we make the plausible assumption that the way the monomers assemble in the dimeric structure may be similar to the WT if the structural determinants of mutant monomers are similar to those of the WT in the monomeric state.

In that case, we use the WT dimeric X-ray structure to construct the structure of the dimer mutants and investigate the stability of the dimer by MD simulations.

The RMSD of the M26I monomer turns out to be fairly similar to that of the WT in the monomeric state: it fluctuates around 1.9 Å over the last 5 ns, and the value at the end of the dynamics ranges between 1.6 Å - 2.5 Å (Tab. 6.1 and Fig. 6.4.A). The secondary structure elements are fully maintained during the dynamics and the overall fold is similar to the WT (Fig. 6.3).

Next, we compare structural determinants of the protein surface involved in the dimerization with those of the WT: (i) the distance (D) between the centers of

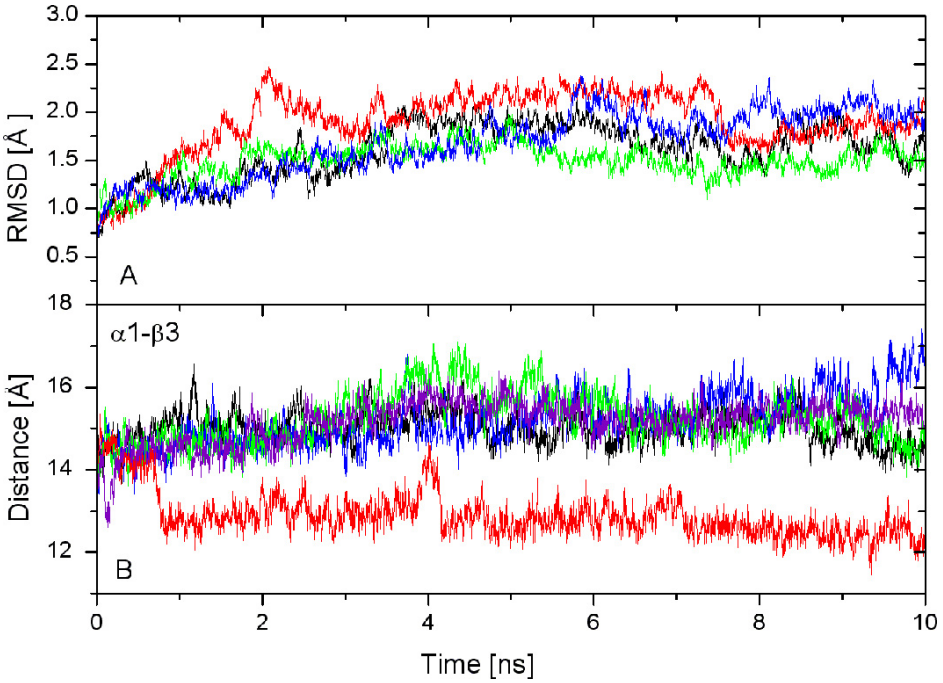


Figure 6.4: Selected properties of mutants in the monomeric state plotted as a function of time. (A) RMSD with respect to the initial structure. (B) Distance between the secondary structure elements at the subunit-subunit interface (defined in the text). Color coding is as follows: WT: black, M26I: green, L166P: red, R98Q: blue, and R28A violet.

	Monomers				**Dimer**			
Distance from Cα	**WT**		**M26I**		**WT**		**M26I**	
of residue 26	Dist.	SD	Dist.	SD	Dist.	SD	Dist.	SD
CE2, ILE31	4.5	0.2	4.8	0.3	4.7	0.3	5.2	0.2
CE2, ASP68	6.7	0.4	6.3	0.3	6.4	0.3	6.6	0.4
VAL, VAL169	6.1	0.4	5.9	0.3	6.1	0.4	5.9	0.3
Distance from Cα	**WT**		**L166P**					
of residue 166	Dist.	SD	Dist.	SD				
CE2, PHE162	7.5	2.1	10.4	0.9				
CG2, VAL169	5.8	0.3	4.5	0.6				
CG2, VAL181	5.9	0.6	4.7	0.7				
CD2, LEU185	8.0	1.2	9.0	1.7				

Table 6.4: Local interactions of residues in 26 and 166 positions in WT, M26I and L166P. The MD-averaged distances between the centers of mass of these residues and those of other residues in contact with them are given in Å.

mass of the secondary structure elements located at the interface (α1 helix and β3 β-sheet in Fig. 6.1) is similar to that of the WT (D = 15.0 ± 0.6 Å vs 14.8 ± 0.5 Å respectively, see Fig. 6.4.B); (ii) the Root Mean Square Fluctuation (RMSF) per residue for M26I is very similar to that of the WT (Fig. 6.6). The similarity between the WT and M26I can be rationalized observing that the mutation does not affect the intramolecular interactions (Tab. 6.4). The interactions formed by M26 in the WT are the same as those formed by I26 in the mutant, since the methionine and isoleucine have similar volume and shape [187, 188] (Tab. 6.5).

We conclude that the interface of the M26I mutant has a similar conformation to the WT and subsequently, has a similar tendency to form dimers.

We then proceed to investigate the dimeric structure of M26I. The RMSD value of M26I in the dimeric state is similar to that of the WT (Tab. 6.1 and Fig. 6.5.A) oscillating around 1.7 Å. The secondary structural elements are fully conserved during the molecular dynamics simulations. The non-bonded interactions at the subunit-subunit interface are also conserved during the MD run, analogously to the WT (Tab. 6.5). In addition, the centers of mass distance analysis showed that the distance between subunits in the mutant is similar to the distance between subunits obtained for the dimeric form of the WT (Fig. 6.5.B). The local interactions around residue 26 are also fully conserved (Tab. 6.4). Finally, the SASA for each subunit is similar to the WT (average value: 10,203 Å^2 and 10,160 Å^2 respectively).

Thus, both the monomeric and dimeric forms of M26I are similar to the WT.

L166P monomer shows a different behaviour than the M26I monomer. Its RMSD exhibits larger fluctuations than in other mutants (Fig. 6.4.A and Tab. 6.1). In addition, properties (i)-(ii) are different from that obtained for the WT: (i) the distance (D) between α1 helix and β3 β-sheet is smaller (D =12.5 ± 0.4 Å and 14.8 ± 0.5 Å, Fig. 6.4.B); (ii) the RMSF of Asp 49 (located on the subunit surface

Hydrogen Bonds

Contacts	Dist. Donor-Accep. X-Ray Dist	M26I Av.	M26I SD	R98Q Av.	R98Q SD	Dist. Proton...Accep. M26I Av.	M26I SD	R98Q Av.	R98Q SD	Angle Donor-H...Accep. M26I Av.	M26I SD	R98Q Av.	R98Q SD
OD2, ASP24 (A) – NH1, ARG48(B)	5.1	2.8	0.2	2.9	0.2	1.9	0.2	2.0	0.2	153	10	141	12
NH1, ARG27(A) – O, ARG48(B)	2.8	2.8	0.1	2.9	0.1	1.9	0.1	1.9	0.2	153	12	159	11
NH1, ARG48(A) – OD2, ASP24(B)	5.1	3.5	0.9	2.8	0.1	2.6	0.8	1.9	0.1	139	26	155	10
O, ARG48(A) – NH1, ARG27(B)	2.8	2.8	0.1	2.8	0.1	1.9	0.1	1.9	0.1	157	11	155	11
O, VAL51(A) – N, CYS53(B)	4.3	3.0	0.2	3.0	0.2	2.1	0.2	2.1	0.2	158	12	160	10
N, CYS53(A) – O, VAL51(B)	4.3	3.1	0.2	3.0	0.2	2.1	0.2	2.1	0.2	161	11	162	10
NH1, ARG145(A) – O, VAL186(B)	3.1	4.1	1.0	3.1	0.3	3.4	0.9	2.2	0.3	128	19	140	20
N, GLY159(A) – O, LEU185(B)	2.6	2.8	0.1	2.8	0.1	1.9	0.1	1.9	0.1	148	12	151	12
O, LEU185(A) – N, GLY159(B)	2.6	2.9	0.1	3.0	0.9	2.0	0.1	2.3	0.9	142	13	129	32
O, VAL186(A) – NH2, ARG145(B)	3.1	4.0	1.6	3.3	0.9	3.2	1.2	2.4	0.9	137	30	144	19

Hydrophobic contacts

Residues	Distances X-Ray Dist.	M26I Av.	M26I SD	R98Q Av.	R98Q SD
(MET 17)(A) – (VAL 20, ILE 21)(B)	6.0	6.3	0.4	7.3	0.7
(VAL 20, ILE 21)(A) – (MET 17)(B)	6.0	6.0	0.4	6.0	0.6
(VAL 51, ILE 52)(A) – (VAL 51, ILE 52)(B)	5.7	6.0	0.2	6.0	0.3
(HIS 126)(A) – (PRO 184, LEU 185, VAL 186)	8.5	8.7	0.4	9.3	0.8
(PHE 162)(A) – (PHE 162)(B)	11.1	11.8	0.6	11.8	0.5
(PRO 184, LEU 185, VAL 186)(A) – (HIS 126)(B)	8.5	6.3	0.4	10.1	1.2

Table 6.5: Selected subunit-subunit contacts in M26I and R98Q obtained from MD simulations. The same structural parameters as those in Tabs. 6.2 and 6.3, are presented.

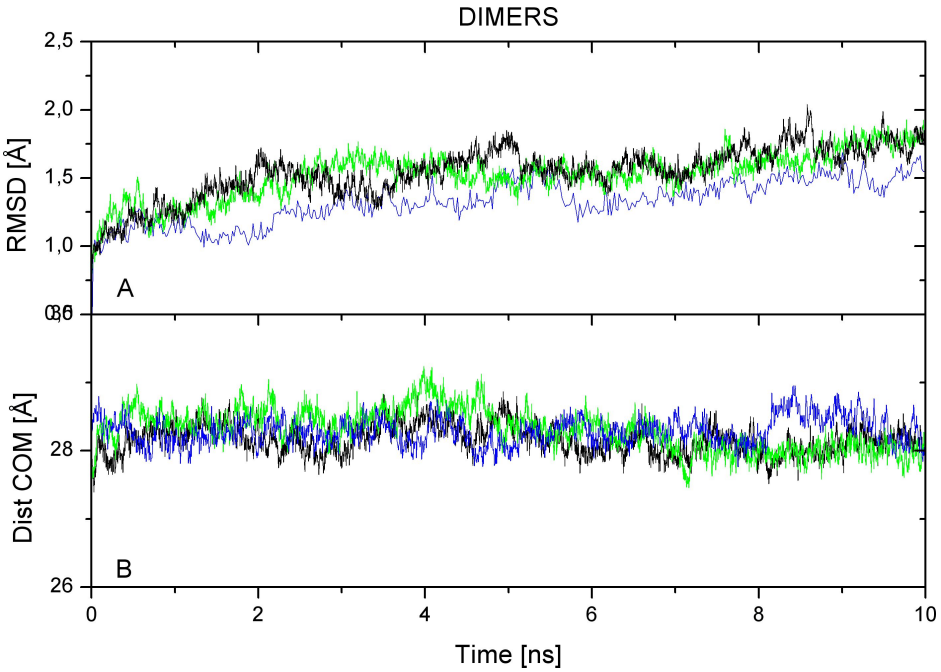


Figure 6.5: MD simulations of DJ-1 WT and its mutants in the dimeric state. (A) Root mean square deviation (RMSD) of the WT (black), M26I (green), R98Q (blue) from the initial structure. (B) Distance between the centers of mass (COM) of each subunit. Same color coding as (A).

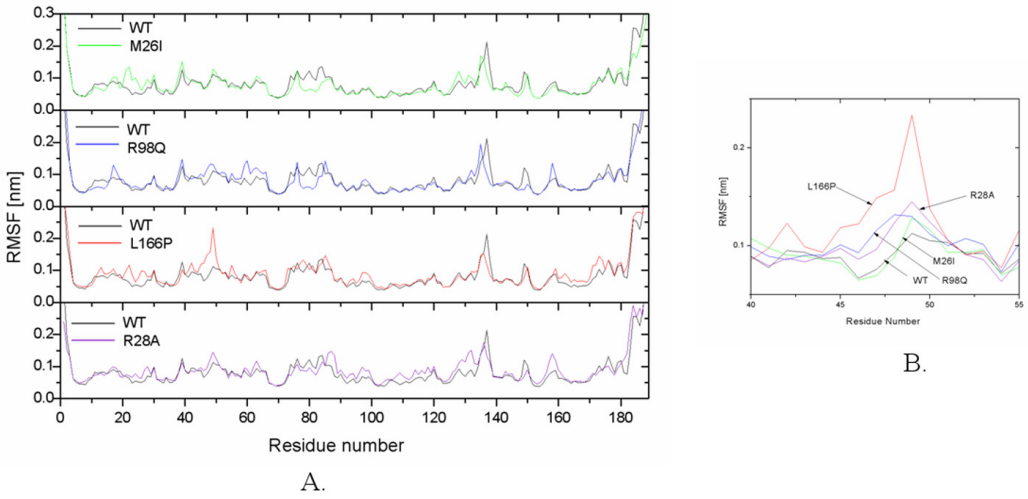


Figure 6.6: Fluctuations of mutant monomers. (A) RMSF of WT, M26I, R98Q, and L166P. (B) Close up of the region (residues 40 to 55) exhibiting the most significant residues fluctuations. Color coding as in Fig. 6.4

involved in the dimerization process) is larger (Fig. 6.6). The discrepancies may be caused, as already suggested [102, 28], by the fact that the replacement of Leu with Pro disrupts $\alpha 8$ helix and therefore the local interactions of residue 166 are different (Tab. 6.4).

Thus, our results suggest the lowered dimerization efficiency of the L166P mutant may be due to structural differences in the monomer. These results are valid within the limits of the computational power of our analysis.

6.3.3 MD of R98Q physiological variant

Since it is a physiological variant of the DJ-1, R98Q is expected to alter none of DJ-1's biochemical functions [173]. From a structural point of view, the replacement of R98, which is indeed exposed to the solvent, with a polar residue such as Q should not dramatically affect the thermodynamic stability of the protein. To investigate the effect of such a mutation, we follow the same computational protocol as for the PD-causing mutations.

The mutant monomer shows a similar behaviour to the WT form. Its RMSD does not show any large fluctuation compared to the WT (Fig. 6.4.A and Tab. 6.1). In addition, properties (i)-(ii) are similar to those of the WT: (i) the distance (D) between $\alpha 1$ helix and $\beta 3$ β-sheet is comparable to that of the WT (14.9 $\pm$ 0.5 Å and 14.8 $\pm$ 0.5 Å respectively, Fig. 6.4.B); (ii) the RMSF shows a similar behaviour to those of the WT and M26I mutation (Fig. 6.6). Furthermore, the

RMSD value of the dimeric form is comparable to that of the WT dimer (Tab. 6.1 and Fig. 6.5.A) oscillating around 1.4 Å with small fluctuations after 5 ns. The behaviour of this physiological variant is similar to that of the dimeric forms of the DJ-1 WT and of the M26I mutant (Tab. 6.5). The analysis of centers of mass distance shows that the DJ-1 variant behaves similarly to the DJ-1 WT (Fig. 6.5.B). Finally, the SASA for each subunit is similar to that of the WT (average value: 10,040 $Å^2$ and 10,160 $Å^2$, respectively).

Within the limitations of our approach, we conclude that R98Q variation, like the M26I, affects neither the dimeric nor the monomeric structures of the WT.

6.3.4 Engineered mutant

Our calculations suggest that L166P is the only mutation among those considered that may alter dimer stability by modifying the fold of the single subunits. These in turn might assemble differently than the WT to form HMW structures.

Here we attempt to identify a mutation that does affect stability by disrupting none of the secondary structure elements. We perform Baker's Computational Alanine Scanning procedure [136, 135], which may quickly (albeit approximately) estimate changes in the interaction free energy ($\Delta\Delta G$) between the two subunits upon mutation of each interface residue with Ala. The largest $\Delta\Delta G$ value turns out to be associated with the R28A mutation (Tab. 6.6). This value is indeed twice as large than any other Ala substitutions presented in Tab. 6.6.

Residue	**ΔΔG (bind) kcal/mol**
M17	1.3
V20	1.3
R27	1.6
R28	**2.9**
V50	1.4
I52	1.1
G159	1.5

Table 6.6: Computational Alanine Scanning - based [135, 136] free energies obtained for the DJ-1 dimeric structure taken from X-ray experiments [116].

Following the MD protocol adopted for the other mutations, we have investigated by MD simulation the structural properties of the R28A monomeric form and we compared them to those of the WT monomer MD structure. The R28A monomer turns out to be similar to the WT: the RMSD value oscillates around 1.8 Å after 5 ns (Tab. 6.1 and Fig. 6.7A). In addition, properties (i)-(ii) are similar to those of the WT: (i) the distance (D) between the α1 helix and the β3 β-sheet is comparable to that of the WT (14.9 ± 0.5 Å and 14.8 ± 0.5 Å for the R28A mutant and the

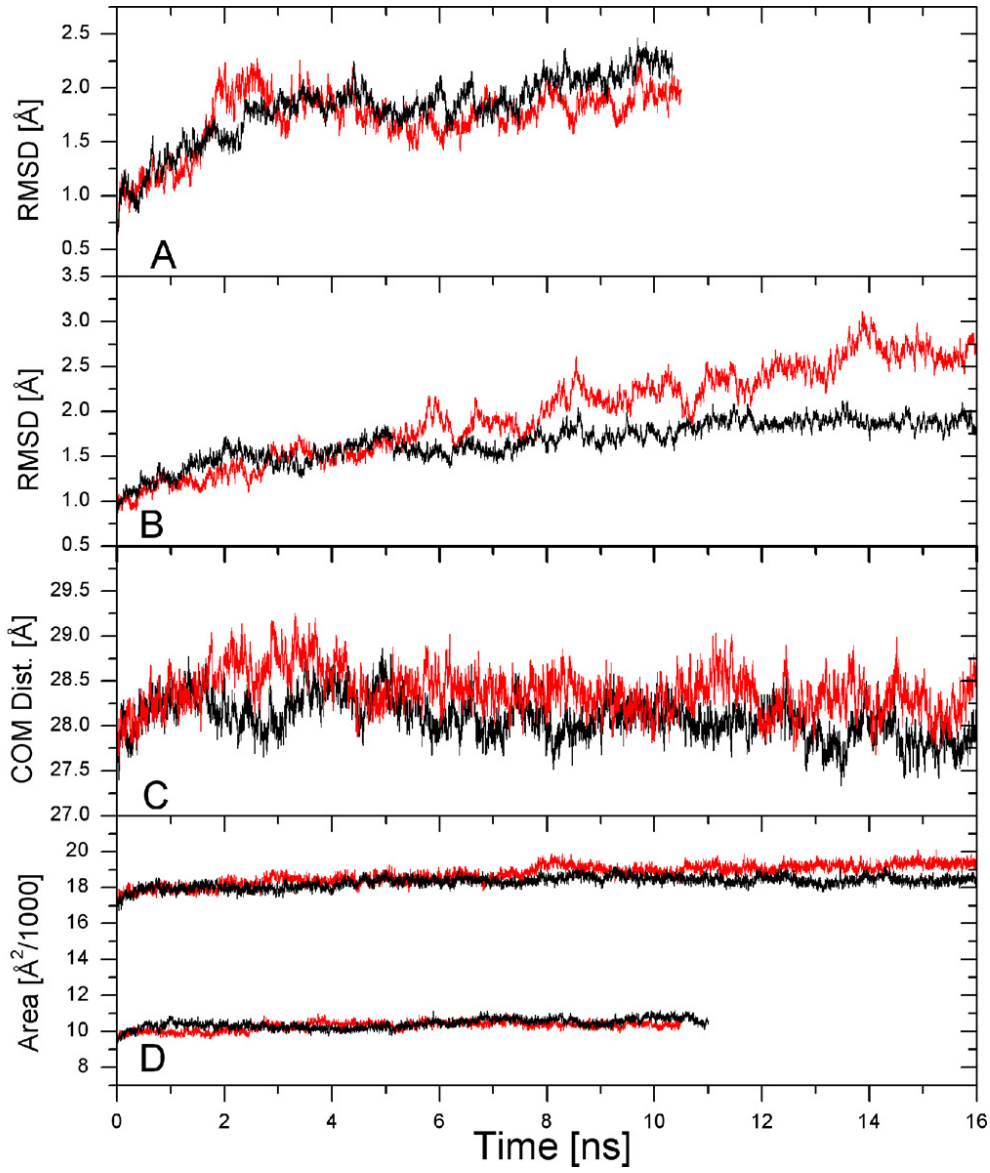


Figure 6.7: Comparison between selected properties of WT and R28A, plotted as a function of time. RMSD of backbone atoms of the WT (black) and R28A mutant (red) for: (A) monomers and (B) dimers. (C) Distance between the centers of mass of subunits of the WT and R28A dimers. (D) Solvent accessible surface area for monomers (lower values) and dimers (larger values) of WT and R28A.

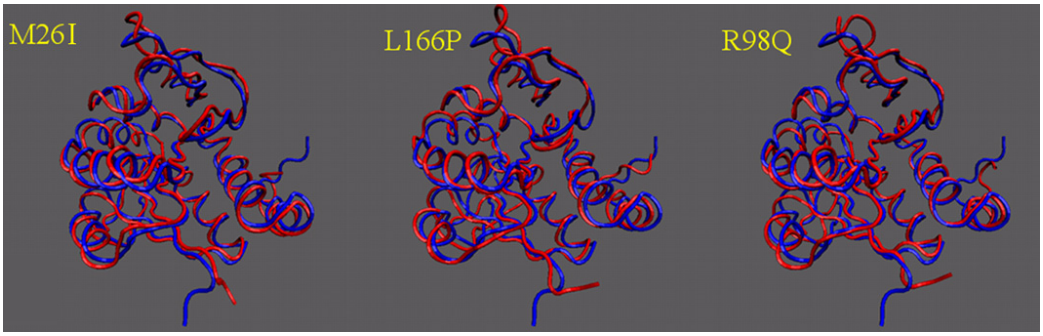


Figure 6.8: Structures obtained after 11 ns of MD simulations of M26I and L166P monomers (red line) superimposed on the corresponding structure of the WT monomer (blue line).

WT DJ-1, see Fig. 6.4.B); (ii) the RMSF is similar to those of the WT and the M26I proteins (Fig. 6.6). The proteins fully maintain the secondary structure elements (Fig. 6.8). This is expected as the mutation is on the surface of the monomer.

We proceed then to investigate the dimeric form of R28A by MD simulation (Fig. 6.7.B). The mutation deeply destabilizes the subunit-subunit interactions and the mutant shows a completely different behaviour compared to that of the WT protein: (i) The distance between the subunit centers of mass along with the SASA increases during the dynamics: at the end of the simulation it is much larger than those of the WT (Fig. 6.7). This suggests that the structure is evolving towards another minimum, in which the subunit-subunit interaction is less strong than in the WT. Unfortunately, observing an eventual complete detachment of the two subunits is well beyond the present domain of applications of MD simulations; (ii) consistently, the RMSD value increases during the dynamics up to 3.0 Å (Fig. 6.7.B), which is significantly larger than the final value found for the WT (1.9 Å, Tab. 6.1 and Fig. 6.7.B). Thus, it is not correct to take the average values of the RMSD (the structure is still evolving). We then calculate for this mutant the average value of the RMSD and the standard deviation (SD) of each subunit during the MD run. The values are 2.0 ± 0.5 Å for both subunits in the mutant, to be compared with 1.6 ± 0.1 Å and 1.7 ± 0.1 Å for each subunit of the WT dimer in the reduced form. This suggests that the R28A DJ-1 mutant is more flexible than the WT.

6.3.5 Experimental validation

To compare quantitatively the ability of PD-causing mutations to form dimers *in vitro*, chemical cross-linking experiments were performed. Human HEK 293 cells were transfected with FLAG-tagged DJ-1 WT, M26I, L166P, and R28A (Fig.

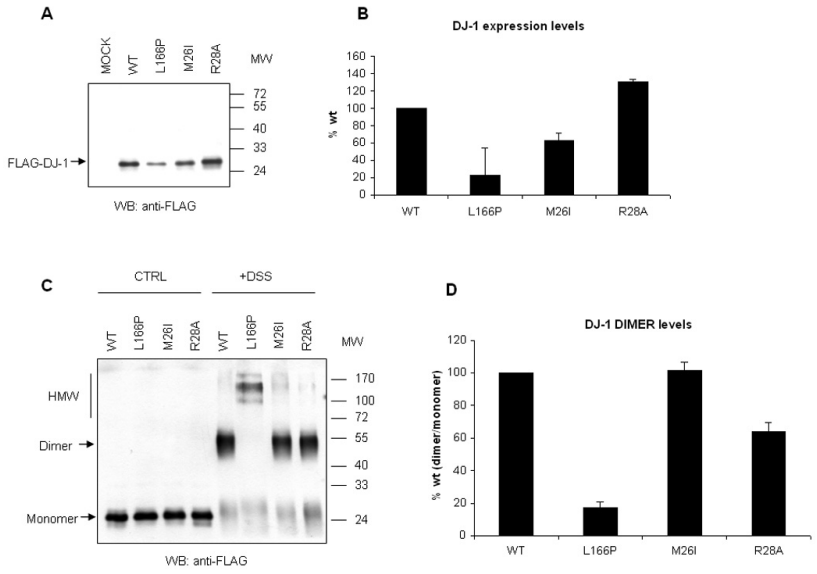


Figure 6.9: *In vitro* analysis of DJ-1 dimer formation. (A) HEK 293 cells were transiently transfected with FLAG-tagged DJ-1 contructs. Equal amount of total protein (5 micrograms) was loaded on gel and immunoblotted with anti-FLAG antibody. (B) Densitometric analysis of protein bands on two independent experiments. Relative expression was normalized to DJ-1wt protein. (C) HEK 293 cells were transfected as in A. Protein lysates were treated with chemical cross-linker DSS or with DMSO as indicated. Normalized quantities of transfected proteins were loaded on gel. Monomer, dimer and HMW forms of DJ-1 were visualized with anti-FLAG antibody. (D) Densitometric analysis of monomer and dimer bands on two independent experiments. Dimer to monomer ratio was calculated and normalized to DJ-1wt protein. The experiments were performed in Prof. Gusticich's laboratory.

6.9). The amount of L166P mutant was 20-40% of the DJ-1 WT, and the levels of R28A were higher than those of WT (Fig. 6.9 A and B). To obtain a measure of the dimerization efficiency, densitometry analysis were performed and the ratio between dimer to monomer was calculated and compared with the various mutants and the WT Fig. 6.9. It was confirmed that the L166P mutant has an impaired capacity to form a homodimer, while displaying a tendency to exist as HMW complexes. The efficacy in L166P dimer formation was calculated to be 20% of the WT DJ-1 or even less. The M26I mutant behaved as shown by Moore et al. [28], since dimer formation capability is equivalent to that of the WT DJ-1. Under these conditions, we observed that in the R28A (computationally selected mutant) the ability to form dimers is impaired and reduced to 50% of DJ-1 WT.

6.4 Concluding Remarks

Several mutations of the DJ-1 gene have been associated with familiar cases of PD [27, 115, 50]. Some of them clearly result in a loss of function since no DJ-1 protein is synthesized due to large deletions and/or splicing errors. The molecular basis of neurodegeneration induced by a group of DJ-1 missense mutations is less clear. A working hypothesis in the field is that DJ-1 carries out its function as a dimer and that missense mutations impair the quantity of the dimeric DJ-1 in the cell [27, 116, 120, 111, 189]. In addition, since oxidized DJ-1 has been found in post mortem brains of PD patients [109], it is of interest to determine the structural determinants of such modifications.

Here we have used MD simulations to assess the effects on DJ-1 protein structure of oxidative stress and of two missense mutations causing familiar PD (L166P, M26I) [53] along with a physiological variant (R98Q) [173].

Our MD simulations suggest that:

(i) WT in solution is very similar to the X-ray structure in both reduced and oxidized states and in both monomeric and dimeric forms.

(ii) The M26I monomer is structurally similar to the WT monomer, because the mutation does not disrupt intra-protein interactions of the WT (Tab. 6.4). Thus, this monomeric structure may assemble similarly to the WT to form dimers.

(iii) The local conformation around P166 in the L166P monomer evolved towards a conformation significantly different from that of the WT. This might be due to the disruption of $\alpha 8$ helix, which in turn affects the conformations of the secondary structure elements at the subunit-subunit interface ($\alpha 1$, $\alpha 9$ helices and $\beta 3$ β-sheet). Thus, although P166 is not located at the subunit surface involved in the dimerization, it changes its shape by modifying some of the intra-protein interactions. This is shown by a comparison of structural determinants of the protein surface of the mutant with the WT (Tab. 6.4). The different conformation of L166P monomeric units might affect the mutant's ability to assemble in dimeric and multimeric structures. Indeed, by gel filtration and other assays, L166P has been previously shown to be present mostly as an HMW complex that may contain either DJ-1 oligomers and/or aggregates with other proteins [121, 119, 111]. This led to the hypothesis that L166P may have additional, unidentified, dominant-negative effects. Parkin, CHIP and hsp70 were all able to interact with L166P as part of a large complex [120]. Furthermore, Parkin interaction provoked the sequestration of DJ-1 mutants into insoluble fractions [120]. By *in vitro* experiments we confirmed previous data that the ectopic expression of L166P is very low and that dimer formation is very negligible. Interestingly, we detected the previously identified HMW complex that may be involved in abnormal oligomerization and protein/protein interactions [121, 119, 111].

(iv) MD simulations show that the dimeric structures of M26I and R98Q DJ-1s are similar to that of the WT and they keep the subunit-subunit interactions intact. The dimeric structures turn out to be less flexible than the monomer, especially in the flexible N- and C-terminal fragments. This fact points to a stabilizing role of the subunit-subunit association.

These results for M26I are consistent with our *in vitro* assays, which show that M26I has the same ability as WT to form dimers. Therefore, our combined computational and experimental approach, along with previous experimental work [125, 121], provides a coherent picture in which M26I assumes mostly a dimeric structure. The results for R98Q are consistent with the fact that this mutant is a physiological variant.

We conclude that the PD-causing M26I mutation does not largely affect the stability of the dimer and does not destabilize it by changing the shape of the subunit-subunit contact surface.

To identify a mutation that decreases the affinity for the dimeric structure without changing the fold of the monomer, we next perform Baker's Computational Alanine Scanning procedure [136, 135] and MD calculations. We find that R28A is indeed able to cause destabilization at the interface. Furthermore, the MD structure of the R28A monomer turns out to be similar to that of the WT monomer, as shown by the comparison of selected structural determinants (Fig. 6.4.B and Fig. 6.6.AB). This may be caused by the fact that A28 faces the solvent. However, the MD simulations point to the destabilization of the dimer as already observed in the timescale investigated ($\sim$10 ns).

The SASA for water and the distances between the centers of mass increase on passing from the WT to R28A (Fig. 6.7.D). This is caused by the disruption of subunit-subunit interactions, rather than a change of shape of the subunit surface. The computational results are consistent with *in vitro* experiments that showed R28A has a reduced, although not abolished, capacity to form homodimers as compared to DJ-1 WT. This effect is not caused by a reduced stability of the R28A mutant since the protein is expressed at even higher levels than the DJ-1 WT (Fig. 6.9 A and B).

Chapter 7

Concluding remarks

I have presented a computational study on two important proteins involved in the derangement of familial Parkinson's Disease.

I have focused first on interactions with dopamine to alpha-synuclein. Dopamine is known to inhibit fribillation experimentally [16, 17, 79, 100, 99], although the protein is unstructured in aqueous solution. Thus, modeling dopamine/interactions is really challenging. Here we have used a rather elaborated computational protocol, involving structural clustering, docking and molecular dynamics techniques along with the experimental validation of the results. Our calculations suggest that the C-terminal region of the protein, including the ^{125}YEMPS^{129} region bind non-covalently with dopamine and several of its derivatives. Moreover, the C-terminal region binding region adopts a specific kinked conformation upon binding (Fig. 4.5). The ligands are stabilized into the binding region by unspecific hydrogen bond and hydrophobic interactions (Tab. 4.2). Furthermore, there are long range electrostatic interaction between the ligands and specific amino acids in the NAC region, which stabilize the ligand. The most important residue in the NAC region involved in such interactions is GLU83 (Fig. 4.8). The experimental work carried out in SISSA and in EPFL corroborates this result.

As next step I have attempted to design molecules structurally and electrostatically similar to dopamine and its derivatives studied here. *In vitro* experiments - carried out exactly as in the previous study, may be performed to test whether the molecules identified by study (reported in Fig. E.1) do affect AS fibrillation.

Our findings demonstrate how the use of structural models of even highly heterogeneous states of proteins determined by NMR spectroscopy in combination with molecular dynamics simulations can provide novel insight into the structural determinants of ligand binding as well as it can be used to design new ligands.

In my thesis work, I have then moved my attention to the role of PD-linked mutations for the oligomeric state of the DJ-1 protein. Whilst the WT protein is

a dimer [116], the mutant L166P - found in patients from familial cases of PD - is a monomer and may form high molecular weight complexes [28, 121, 111, 119]. In addition, M26I - found also in familial cases of PD - is still a dimer [121, 102] but it is a matter of debate whether dimer formation occurs at the same rate as WT or less efficiently. Our calculations suggest that the M26I mutation, which is known experimentally to reduce the thermal stability of DJ-1 in solution [190], affects neither the monomeric structure nor the dimeric one [33, 125], and it has a similar tendency to form dimers as that of the WT. This suggests that the pathological role of this mutant for PD may arise from factors other than structural determinants. Our calculations provide also a possible rationale on why L166P mutation favors the formation of multimers, as shown experimentally [121, 111, 119]. Indeed, this is is probably due to the structural changes of the monomeric subunit upon mutation [116]: although the mutated residue is located on a inner helix (the $\alpha 8$ helix), it produces changes in the interactions among secondary structure elements ($\alpha 1$ helix and $\beta 3$ beta sheet, Fig. 6.4), which in turn may affect subunit-subunit interactions in a sort of 'domino' effect [191]. Thus, this mutation may allow the protein to form *in vivo* oligo or monomeric structures. This proposal is supported experimentally here and in the previous work [28]. Our methodology is subsequently used to search for mutations that decrease the affinity for the dimer *without* affecting DJ-1 fold (as it occurs in L166P). This leads to identify a key residue on the subunit-subunit interface and to develop a new engineered mutant R28A. This mutant experimentally shows a reduce ability to form dimers but due to a different mechanism than the L166P mutant. This is consistent with with the fact that this mutant does not affect the structural properties of the monomer. Instead, it affects the subunit-subunit interactions (indeed, the mutation is located in subunit-subunit interface of the dimer).

In conclusion, my combined computational and experimental work may be a valuable strategy to address key issues related to the molecular medicine of PD. In particular, different computational methodologies along with experimental studies have helped addressing a difficult and medically relevant problem such as ligand binding to a completely unstructured protein. In spite of its important pharmacological implications, this issue has never been addressed before, to the best of my knowledge.

Chapter 8

Acknowledgments

I wish to thank my Supervisor Paolo Carloni for his help and support during my study in the last few years and for giving me the opportunity to come to Trieste for completion of my PhD research and thesis.

I gratefully acknowledge the supportive attitude of all the people of the Sector of Statistical and Biological Physics, who yielded invariably precious and helpful encouragements, which were crucial to afford my study.

A very special thanks goes out to all of our collaborators, without their help it would have been impossible to carried out this work.

I would also like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my wife, Mariela, without whose love, advice and assistance, I would not have finished this thesis.

I particularly would like also to mention and to acknowledge my father, who passed away last year while I was here in Italy, and it was very painful for me, for all the support, encouragement and the example of life he gave to me.

Thank to everyone.

Appendix A

The neurotransmitter dopamine

A.1 Dopaminergic neurons: synthesis of DOP

In the dopaminergic (DA) neurons, dopamine (DOP) is synthesized by tyrosine hydroxylase (TH) in the cytosol. Like all catecholamines, DOP is unstable and prone to autoxidation, with production of ROS and reactive DOP derivatives such as quinones (Q) and dopaminochrome (DCH) which are cytotoxic species able to cause oxidative damage to cellular proteins and DNA. This toxicity is avoided by rapid sequestration of DOP in synaptic vesicles by vesicular monoamine transporter 2 (VMAT2). Interestingly, mice that express low levels of VMAT2 display age-associated nigrostriatal dopamine dysfunction that ultimately results in neurodegeneration [192]. After release in the synaptic cleft, DOP is re-uptaken in the cytosol by the dopamine transporter (DAT) (Fig. A.1).

A.2 AS/Dopamine: covalent and non-covalent adducts

Covalent adducts: Conway *et al.*, in 2001 performed an screening of a commercially available library of 169 compounds to identify drug-like molecules able to block AS fibrillization *in vitro*, 14 of the 15 inhibitors found were found to be catecholamines related to dopamine. The inhibitory activity of dopamine depended on its oxidative ligation to AS and was selective for the protofibril-to-fibril conversion, causing accumulation of the AS protofibril [98].

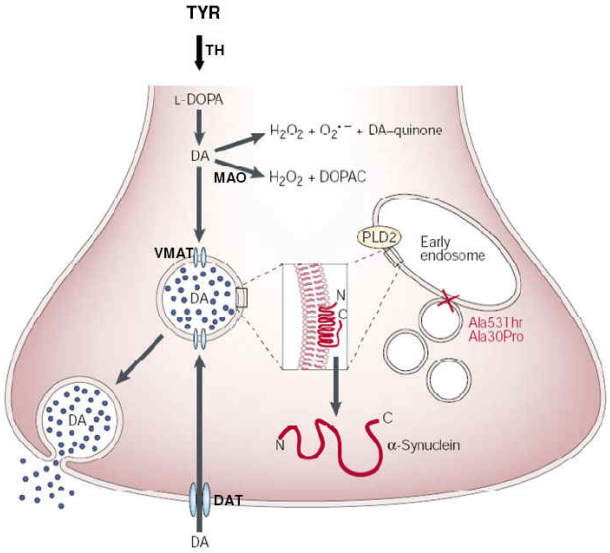


Figure A.1: Synthesis of dopamine in the presynaptic terminal of dopaminergic neurons. The potential function of AS in intracellular dopamine storage is also presented (adapted from [11]).

DOP modulates differently the stability of protofibrils and fibrils composed of wild-type or variants of AS (A30P and A53T) as shown by a high hydrostatic pressure (HHP) study. Although fibrillization of AS was inhibited by DOP, long incubation times (18 days) eventually led to the formation of fibrils with a decreased stability. As for the protofibrils, in the absence of DOP they all exhibited identical stability, while in its presence, the mutant ones acquired a greater stability, which can shed light onto why these mutations are so aggressive. Indeed the A30P protofibrils, when added to mesencephalic and cortical neurons in culture, decreased the number and length of neurites and increased the number of apoptotic cells. The toxic effects were abolished by HHP treatment, which is able to break the protofibrils into smaller aggregates, as seen by atomic force microscopy. This suggests that strategies aimed at breaking and/or clearing these aggregates are promising in alleviating the symptoms of PD. The exact mechanism of DOP/AS interaction has not been fully elucidated yet, and given its potential pharmacological and therapeutical relevance, is the object of many ongoing research efforts.

In another work, Li *et al.*, tested the inhibitory abilities of DOP and several analogs including chemically synthetic and natural polyphenols *in vitro*. The MS and NMR characterizations demonstrated that DOP and its analogs inhibit AS fibrillization by a mechanism where the oxidation products (quinones) of DOP analogs react with the amino groups of AS chain, generating AS/quinone adducts. NMR experiments suggest that at least three amino groups of lysine residues in a AS molecule are modified by DOP to generate imino groups that link to the aromatic

ring of quinone and the side chains of the protein. The covalently cross-linked AS adducts by DOP are primarily large molecular mass oligomers, while those by catechol and benzoquinone (or hydroquinone) are largely monomers or dimers, probably because DAQ reacts with the amino group of DOP molecule to form melanin that further covalently cross-links DOP/AS adducts to form more complicated DOP/AS oligomers. The DOP quinoprotein retains the same cytotoxicity as the intact AS, suggesting that the oligomeric intermediates are the major elements that are toxic to the neuronal cells [99].

Non-covalent adducts: There is strong debate about the covalent [79, 100, 99] or non-covalent [16, 17] nature of the DOP/AS complexes.

Norris *et al.* proved that dopamine inhibition of AS fibrillization generated exclusively spherical oligomers that depended on dopamine autoxidation but not AS oxidation, because mutagenesis of Met, His, and Tyr residues in AS did not abrogate this inhibition. However, truncation of AS at residue 125 restored the ability of AS to fibrillize in the presence of dopamine. Mutagenesis and competition studies with specific synthetic peptides identified AS residues 125 to 129 (*i.e.* YEMPS) as an important region in the dopamine- induced inhibition of AS fibrillization. They also show that the inhibition of the interaction is reversible if the system is treated with urea and renaturated by dialysis in PBS, demostrating that the DOP/AS interactions is non covalent. Thus, their data indicate that dopamine and also dopaminochrome (an autooxidation product of DOP) inhibits AS fibrillization, probably inducing structural changes in AS that can occur through the interaction of dopaminochrome with the ^{125}YEMPS^{129} motif of AS forming non covalent adducts [16].

On the other hand, Mazzulli *et al.*, using molecular and biochemical approaches were able to demostrated the important role of the C-terminal region of AS for the inhibition of its fibrillization. Interestingly, they have shown that the intracellular inhibition of AS aggregation requires the oxidation of catechols and the specific non-covalent interaction of the oxidized catechols with residues ^{125}YEMPS^{129} in the C-terminal region of the protein [17].

In addition, Bisaglia *et al.* carried out a kinetic and structural analysis of the early oxidation products of dopamine. Specifically, considering the potential high toxicity of dopaminochrome for both cells and mitochondria, they focused on its rearrangement to 5,6-dihydroxyindole. After the spectroscopic characterization of the products derived from the oxidation of dopamine, the structural information obtained was used to analyze the reactivity of quinones toward AS. The results suggest that all these molecules should be able to interacts with AS forming covalent or non-covalent adducts, but indole-5,6- quinone (IQ), is the most reactive species [100] (Fig. A.2).

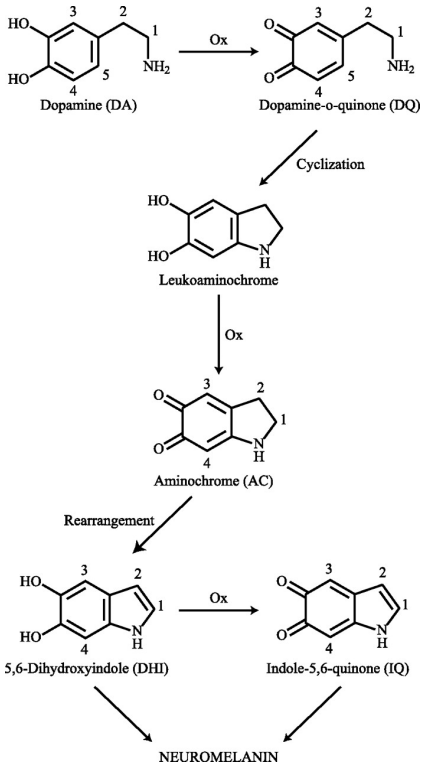


Figure A.2: Characterization of dopamine forms (Adapted from [100]).

A.3 Oxidation pathways of DOP

A summarizing scheme of the oxidative pathway DOP and of the proposed interactions with AS is shown in Fig. A.2. The initial reaction involves the oxidation of dopamine (DOP) to yield the corresponding dopamine-oquinone (DAQ). The two following steps are the cyclization of the quinone to give leukoaminochrome and its subsequent oxidation to dopaminochrome (DCH). Then, dopaminochrome rearranges to 5,6-dihydroxyindole (DHI) which can be oxidized to indole- 5,6-quinone (IQ) and polymerize to form neuromelanin [100].

Appendix B

Other docking algorithms

B.1 Multi-population GA

In the section 3.1.4 we have explained the structure of a single population genetic algorithm as it is implemented in the programs AUTODOCK 3 and 4 (Fig B.1)

In other programs, like the GOLD program [150, 151], it is applied instead a multi-population genetic algorithm. This program was not used in this thesis, but it will be used in the future work. With this program, better results are obtained by introducing multiple subpopulations. Every subpopulation evolves over a few generations isolated (like the single population genetic algorithm) before one or more individuals are exchanged between the subpopulation (migration). The multi-population evolutionary algorithm models the evolution of a species in a way more similar to nature than the single population genetic algorithm. Fig. B.2 shows the structure of such an extended multi-population genetic algorithm.

Goldscore and Chemscore scoring functions: The GOLD program is able to use different scoring functions, GoldScore, ChemScore and a User Defined Score.

The Goldscore fitness function is evaluated in six stages as follows. (1) A conformation of the ligand and protein active site is generated. (2) The ligand is placed within the active site using a least-squares fitting procedure. (3) A hydrogen bonding energy H-Bond-Energy, is obtained for the complex. (4) A pairwise energy, Complex-Energy, is obtained for the steric energy of interaction between the protein and the ligand. (5) Molecular mechanics expressions are used to generate the term Internal-Energy which is a measure of the internal energy of the ligand. (6) The energy terms are summed together to give a final fitness score [151].

The ChemScore fitness function [193, 194] was derived empirically from a set of 82 protein-ligand complexes for which measured binding affinities were available.

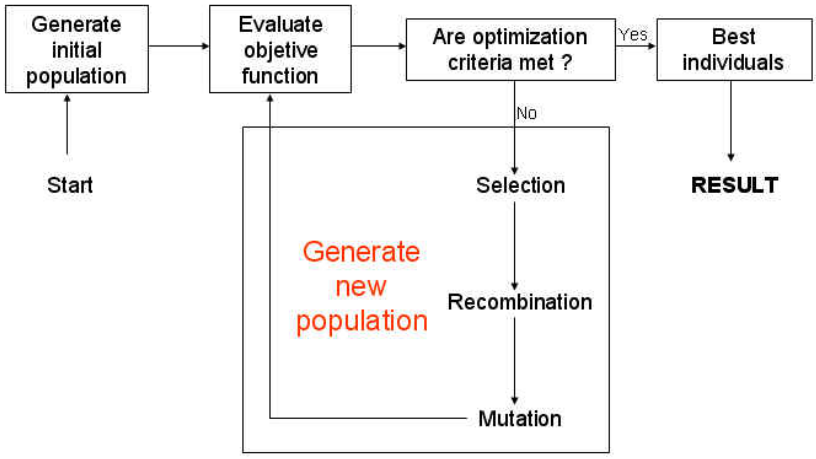


Figure B.1: Scheme of the genetic algorithm used in the AUTODOCK program

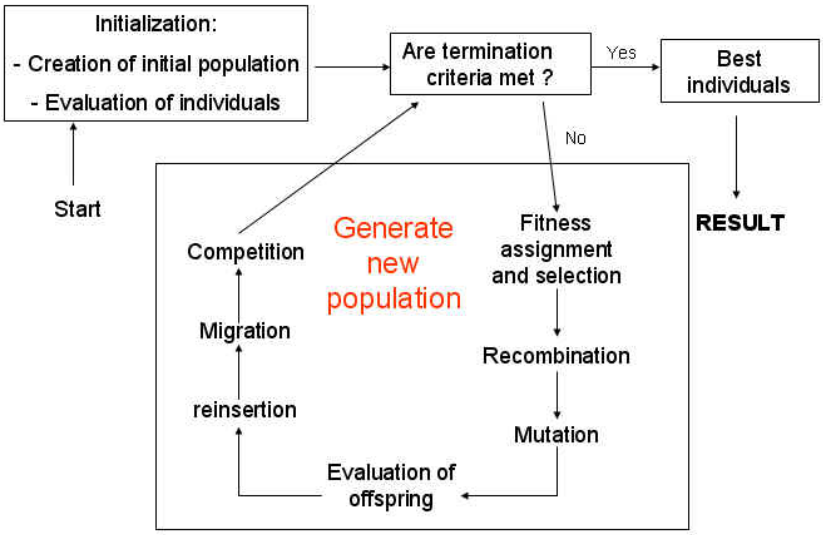


Figure B.2: Scheme of the genetic algorithm used in GOLD program

Unlike GoldScore, the ChemScore function was trained by regression against measured affinity data, although there is no clear indication that it is superior to GoldScore in predicting affinities. It estimates the total free energy change that occurs on ligand binding as:

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot} \tag{B.1}$$

Each component of this equation is the product of a term dependent on the magnitude of a particular physical contribution to free energy (e.g. hydrogen bonding) and a scale factor determined by regression.

The final ChemScore value is obtained by adding in a clash penalty and internal torsion terms, which militate against close contacts in docking and poor internal conformations. Covalent and constraint scores may also be included.

$$ChemScore = \Delta G_{binding} + P_{clash} + c_{internal} P_{internal} + (c_{covalent} P_{covalent} + P_{constrains}) \tag{B.2}$$

Appendix C

Ewald Sum

The Ewald sum is a technique for efficiently summing the interaction between an ion and all its periodic images. It was originally developed in the study of ionic crystals [195, 196]. The electrostatic potential energy U^{Coul} for a neutral system composed of N positively and negatively particles of charge q_i, placed at $\mathbf{r}_i$ and enclosed in a volume $V = L^3$, takes the form:

$$U^{Coul} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i) = \frac{1}{2}\sum_{\mathbf{n}=0}^{\prime}\sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r}_{\mathrm{ij}} + \mathbf{n}\mathrm{L}|} \tag{C.1}$$

where $\mathbf{n}$ are the direct lattice vectors and each particle interacts with all the other $N-1$ particles into the simulation box and with all the N particle images in an infinite 3D array of periodic cells. The prime on the summation indicates that the sum is over all the images $\mathbf{n}$, and $i \neq j$ if $\mathbf{n} = 0$. The convergence of this series is conditional and requires a very large number of images to achieve a reliable estimate of U^{Coul}. The idea behind the Ewald method is to surround every charge q_i with a density cloud ϱ_- of opposite sign and total charge $-q_i$. In this way an efficient screening is realized, so that interactions rapidly go to 0 and direct summation is possible. Then, a *cancellation* charge density ϱ_+ of total charge q_i centered at $\mathbf{r}_i$ is also added to recover the original distribution. In order to exclude self-interactions the contributions of these three charge densities should be not evaluated in $\mathbf{r}_i$. However, it is convenient to keep self-interactions due to the canceling charge distribution ϱ_+, since ϱ_+ is in this way periodic and can be represented as a rapidly converging Fourier sum. The spurious self-interaction can be easily subtracted separately. In the original formulation Ewald choose two Gaussian charge distributions, $\varrho_-^G(\mathbf{r})$ and $\varrho_+^G(\mathbf{r})$ of width $\sqrt{2/\alpha}$ and total charge $-q_i$ and q_i, respectively. Thus, the total charge distribution of the system $\varrho(\mathbf{r})$ may be rewritten as:

$$\begin{aligned}\varrho(\mathbf{r}) &= \varrho(\mathbf{r}) + \varrho_-^G(\mathbf{r}) + \varrho_+^G(\mathbf{r} \\ &= \sum_{\mathbf{n}} \left[\left(\sum_{i=1}^{N} q_i \delta(\mathbf{r} - (\mathbf{r}_\mathrm{i} + \mathbf{n}\mathrm{L})) + \sum_{i=1}^{N} (-q_i) \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha|\mathbf{r}-(\mathbf{r}_\mathrm{i}+\mathbf{n}\mathrm{L})|^2} \right) \right. \\ &\left. + \sum_{i=1}^{N} q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha|\mathbf{r}-(\mathbf{r}_\mathrm{i}+\mathbf{n}\mathrm{L})|^2} \right]\end{aligned} \tag{C.2}$$

At this point, $\varrho + \varrho_-^G$ and ϱ_+^G contributions to the total charge distribution can be considered separately. From the first term one gets a potential energy U^{dir} rapidly going to 0 at large r, due only to the fraction of q_i unscreened by ϱ_-^G. From ϱ_+^G the long-range interaction energy U^{rec} comes, that can be evaluated in the reciprocal space. To this aim, we derive the electrostatic potential in this space through Poisson's equation $\nabla^2\phi(\mathbf{r}) = -4\pi\varrho(\mathbf{r})$. Fourier transforming the members of this equation gives:

$$\nabla_{\mathbf{r}}^2\phi(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}} \tilde{\phi}(\mathbf{k})\nabla_{\mathbf{r}}^2 e^{i\mathbf{k}\cdot\mathbf{r}} = -\frac{1}{V}\sum_{\mathbf{k}} k^2\tilde{\phi}(\mathbf{k})e^{i\,\mathbf{k}\cdot\mathbf{r}} \tag{C.3}$$

$$\varrho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{k}} \tilde{\varrho}(\mathbf{k})e^{i\,\mathbf{k}\cdot\mathbf{r}} \tag{C.4}$$

where

$$\tilde{f}(\mathbf{k}) = \int_V d\mathbf{r} f(\mathbf{r}) e^{-i\,\mathbf{k}\cdot\mathbf{r}} \tag{C.5}$$

and $\mathbf{k} = (2\pi/L)\mathbf{l}$, $\mathbf{l}$ being reciprocal lattice vectors. The Poisson's equation takes the particularly simple form:

$$k^2\tilde{\phi}(\mathbf{k}) = 4\pi\tilde{\varrho}(\mathbf{k}) \tag{C.6}$$

Substituting the expression for $\varrho_+^G(\mathbf{r})$ in eq. C.5 we have:

$$\begin{aligned}\tilde{\varrho}_+^G(\mathbf{k}) &= \int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{n}}\sum_{j} \left(\frac{\alpha}{\pi}\right)^{3/2} q_j e^{-\alpha|\mathbf{r}-(\mathbf{r}_j+\mathbf{n}L)|^2} \\ &= \sum_{j} \left(\frac{\alpha}{\pi}\right)^{3/2} q_j \int_V d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{n}} e^{-\alpha|\mathbf{r}-(\mathbf{r}_j+\mathbf{n}L)|^2} \\ &= \int_{allspace} d\mathbf{r} \sum_{j} \left(\frac{\alpha}{\pi}\right)^{3/2} q_j e^{-i\mathbf{k}\cdot\mathbf{r}} e^{-\alpha|\mathbf{r}-\mathbf{r}_j|^2}\end{aligned} \tag{C.7}$$

where the relation $\sum_n \int_{-a}^{a} f(x+na) = \int_{-\infty}^{\infty} f(x)$ for periodic functions has been used. With the change of variables $\mathbf{r} \to \mathbf{r}' = \mathbf{r} - \mathbf{r}_j$ we obtain finally the expression for $\tilde{\varrho}_+^G$ in the reciprocal space:

$$\begin{aligned}\tilde{\varrho}_+^G(\mathbf{k}) &= \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} \left[\left(\frac{\alpha}{\pi}\right)^{3/2} \int_{allspace} d\mathbf{r}' e^{-i\mathbf{k}\cdot\mathbf{r}'} e^{-\alpha r'^2} \right] \\ &= \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j} e^{-k^2/4\alpha} = \tilde{\varrho}_P(\mathbf{k})\tilde{\gamma}(\mathbf{k})\end{aligned} \tag{C.8}$$

where $\tilde{\varrho}_P(\mathbf{k}) = \sum_j q_j e^{-i\mathbf{k}\cdot\mathbf{r}_j}$ and $\tilde{\gamma}(\mathbf{k}) = e^{-k^2/4\alpha}$ are the Fourier transform of a point charges set and of the smearing Gaussian, respectively.

Inserting in Poisson's equation the potential takes the form:

$$\tilde{\phi}_+^G(\mathbf{k}) = \tilde{g}(\mathbf{k})\tilde{\varrho}_P(\mathbf{k})\tilde{\gamma}(\mathbf{k}) \tag{C.9}$$

where $\tilde{g}(\mathbf{k}) = \frac{4\pi}{k^2}$ is the Green's function.

Antitrasforming the potential gives the electrostatic energy:

$$\begin{aligned}U^{rec} &= \frac{1}{2}\sum_i q_i \phi_+^G(\mathbf{r}_i) = \frac{1}{2}\sum_i q_i \left[\frac{1}{V} \sum_{\mathbf{k}\neq 0} \tilde{\phi}_+^G(\mathbf{k}) \right] \\ &= \frac{2\pi}{V} \sum_{\mathbf{k}\neq 0} \sum_{ij} \frac{e^{-k^2/4\alpha}}{k^2} q_i q_j e^{-i\mathbf{k}\cdot(\mathbf{r}_i - \mathbf{r}_j)} = \frac{1}{2V} \sum_{\mathbf{k}\neq 0} \tilde{g}(\mathbf{k}) \left|\tilde{\varrho}_+^G(\mathbf{k})\right|^2 e^{-k^2/4\alpha}\end{aligned} \tag{C.10}$$

s mentioned above, this term contains the self-interaction of each charge q_i with a Gaussian distribution $\varrho_+^{self} = q_i(\alpha/\pi)^2 e^{-\alpha|\mathbf{r}-\mathbf{r}_i|^2}$. The potential due to this cloud is obtained solving the Poisson's equation in spherical coordinates ($x \equiv |\mathbf{r} - \mathbf{r}_i|$):

$$\frac{1}{x} = \frac{\partial x \phi_+^{self}}{\partial x^2} = 4\pi \varrho_+^{self} \to \phi_+^{self} = \frac{q_i}{x}\left(\frac{2}{\sqrt{\pi}}\right) \int_0^{\sqrt{\alpha}x} e^{-x'^2} dx' = \frac{q_i}{x} \mathrm{erf}(\sqrt{\alpha}x) \tag{C.11}$$

where $\mathrm{erf}(x) = (2/\sqrt(\pi)) \int_0^x \exp(-u^2) du$. Finally, we get:

$$U^{self} = \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 \tag{C.12}$$

The advantage of this scheme is now clear: the self-contribution energy is simply a constant, independent on system configuration, and does not enter in the calculation of the forces. Using the result for ϕ^{self} it is easy to calculate the short-range interaction energy U^{dir}. Indeed, this energy arises from a potential generated by charge densities $q_i - \varrho_-^G(\mathbf{r})$ centered at $\mathbf{r}_i$, which thus is given at any point $\mathbf{r}$ by:

$$\phi^{dir}(\mathbf{r}) = \sum_{\mathbf{n}} \sum_j \frac{q_j}{|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|} \mathrm{erfc}\left(\sqrt{\alpha}\,|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|\right) \tag{C.13}$$

The total electrostatic interaction energy can be finally expressed as:

$$\begin{aligned} U_E^{Coul} w &= \frac{1}{2}\sum_i q_i \phi^{tot}(\mathbf{r}_i) \\ &= U^{dir} - U^{self} + U^{rec} = \frac{1}{2}\sum_i q_i \left[\phi^{dir}(\mathbf{r}_i) + \phi^{rec}(\mathbf{r}_i) - \phi^{self}(\mathbf{r}_i)\right] \\ &= \frac{1}{2}\sum_{\mathbf{n}}{}' \sum_{ij}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \mathrm{erfc}\left(\sqrt{\alpha}\,|\mathbf{r}_{ij} + \mathbf{n}L|\right) \\ &+ \frac{2\pi}{V} \sum_{\mathbf{k}\neq 0} \sum_{ij} \frac{e^{-k^2/4\alpha}}{k^2} q_i q_j e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_j)} \\ &- \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 \end{aligned} \tag{C.14}$$

The accuracy of the method, defined as the difference $\epsilon = U_{Ew}^{Coul} - U^{Coul}$, is related to the optimal choice of the parameters α, width of the Gaussian distribution, and of the two cutoffs r_c and k_c in real and reciprocal space, respectively. It can be shown that ϵ for both the real and reciprocal space terms is proportional to $\exp(-s^2)/s^2$ [197], where $s \equiv r_c\alpha$ and $s \equiv \pi n_c/\alpha L$, respectively. Here n_c is a positive integer used to define the cutoff in the reciprocal space vectors as $k_c = 2\pi/Ln_c$, so that the number of Fourier components within this cutoff is $(4\pi/3)n_c^3$. If one requires the same accuracy on the evaluation of U^{dir} and U^{rec} the value of s is fixed, and the choice of α determines both the number of Fourier components required to calculate U^{rec} and the cutoff r_c. The optimal choice of α is obtained minimizing the total computational time

$$\tau = \tau^{dir} N^{dir} + \tau^{rec} N^{rec} \tag{C.15}$$

where τ^{dir} is the time needed to evaluate the interaction between a couple of particles, τ^{rec} that needed to evaluate the interaction per k vector and per particle,

N^{dir} and N^{rec} are the number of cycles per MD step needed for the evaluation of U^{dir} and U^{rec}, respectively. For a homogeneous system of density $D = N/L^3$ we have

$$\begin{aligned} N^{dir} &= \frac{4}{3}\pi \left(\frac{r_c}{L}\right)^3 N^2 = \frac{4}{3}\pi \left(\frac{s}{\alpha L}\right)^3 N^2 \\ N^{rec} &= \frac{4}{3}\pi n_c^3 N = \frac{4}{3}\pi \left(\frac{s\alpha L}{\pi}\right)^3 N \end{aligned} \tag{C.16}$$

which gives for α and τ the values:

$$\begin{aligned} \alpha &= \left(\frac{\tau^{dir}\pi^3 N}{\tau^{rec} L^6}\right)^{1/6} \sim N^{-1/6} \\ \tau &= \frac{8\sqrt{\tau^{dir}\tau^{rec}}N^{3/2}s^3}{3\sqrt{\pi}} = \mathcal{O}(N^{3/2}) \end{aligned} \tag{C.17}$$

C.1 Particle-Mesh Approaches

The computational cost for solving eq. C.14 thus scales as $\mathcal{O}(N^{3/2})$. In general, r_c can be safely set to a value $< L/2$, restricting the calculation of U^{dir} to the original box only ($|\mathbf{n}| = 0$). However, for convenience r_c is sometimes set equal to the cut-off for short-range interactions, and in this case α does not need to have its optimum value. This implies that, although the real part of the Ewald sum goes as $\mathcal{O}(N)$, the Fourier part goes as $\mathcal{O}(N^2)$. To see this, simply substitute $s = \alpha r_c$ in eq. C.16, and recall that $N = DL^{-3}$ and $N_c \sim Dr_c^{-3}$, where N_c is the number of particles within the sphere of radius r_c. Either or noth α takes its optimum value, the computational cost of the Ewald sum becomes prohibitive for large ($N > 10^4$). To cope with this problem the so-called Particle-Mesh Approaches have been developed [197], in which the point charges are spread on a space grid. This leads to a Poisson equation expressed through Discrete Fourier Transform (DFT), which is solved using the Fast Fourier Transforms (FFT) [198], scaling as $\mathcal{O}(M \log_2 M)$, where M is the number of grid points or Discrete Fourier Components (in practice $\sim N$?). Since the FFT is a grid transformation, there are obvious discretization problems to be solved (*in primis* those of charge assignment and calculation of forces from grid points back to atoms) and corresponding discretization errors to be minimized. In this sense an improvement in the accuracy has been made by further splitting interactions in the reciprocal space into a short and long range term, and evaluating the first directly as particle-particle and the second through the use of a mesh (Particle-Particle Particle-Mesh, PPPM or P^3M). What should

be kept in mind is that Particle-Mesh approaches have been introduced as the allow to implement Ewald summation using the FFT, not for reasons of accuracy. Without their advent, however, computer simulations would have never found spread use in biological problems.

Appendix D

Computational approach to the investigation of alpha-synuclein/dopamine interactions

D.1 Resp charges on ligands

	DCH	DHI	DOP	DOP-H	DQ	IQ	LEUK
C1	0.185	0.070	-0.072	-0.033	-0.029	0.228	0.093
H1	0.186	0.179	0.469	0.477	0.158	0.194	0.177
C2	-0.446	-0.328	-0.327	-0.293	-0.281	-0.397	-0.378
H2	0.212	0.186	0.461	0.471	0.189	0.202	0.220
C3	0.627	0.226	0.329	0.300	0.512	0.534	0.197
O1	-0.498	-0.566	-0.626	-0.594	-0.485	-0.482	-0.663
H4	0.028	0.208	0.178	0.196	0.071	0.191	0.018
C4	0.494	0.284	0.268	0.281	0.588	0.550	0.314
O2	-0.515	-0.557	-0.658	-0.625	-0.493	-0.501	-0.610
H5	0.034	0.318	0.162	0.179	0.071	0.322	0.027
C5	-0.496	-0.385	-0.331	-0.301	-0.291	-0.449	-0.353
H3	0.028	0.162	0.217	0.205	0.169	0.197	0.018
C6	0.040	0.037	-0.175	-0.207	-0.180	0.083	-0.020
C7	0.054	-0.221	-0.031	-0.064	-0.046	-0.350	0.031
H6	0.034	0.426	0.059	0.073	-0.008		0.027
H7	0.302	0.424	0.059	0.073	-0.008		0.294
C8	0.058	-0.252	0.296	0.034	0.342	-0.053	0.050
H8			0.004	0.098	0.395		0.470
H9			0.004	0.098	0.395		0.467
N1	-0.325	-0.209	-1.047	-0.278	-1.070	-0.269	-0.378
H10			0.381	0.303			
H11			0.381	0.303			
H12				0.303			
Net Charge	0.0	0.0	0.0	1.0	0.0	0.0	0.0

Table D.1: RESP atomic charges of the ligands in Fig. 4.2.

D.2 Tables of distances

The table D.2 shown the results obtained from the MD simulations of the first six representatives conformation of AS with all the analyzed ligands.

D.3 Clustering of 1,000 conformations of AS in solution

The following figures and tables are related to the clustering of 1,000 conformations of AS in solution.

Representative from cluster 1 (distances in [Å] between the ligands and the C-terminal residues)

Res.	DCH		DHI		DOP		DOP-H		DQ		IQ		LEUK	
Num.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
83	18.9	0.9	11.6	1.0	19.8	2.2	21.5	2.3	7.8	1.1	8.3	1.6	33.7	2.5
110	11.2	1.0	9.4	0.7	13.6	1.3	16.0	2.3	7.1	0.9	6.9	0.5	23.1	2.3
111	11.0	0.9	7.6	1.2	14.9	1.6	16.5	2.2	4.4	0.5	11.1	0.7	20.0	2.1
112	12.3	0.8	10.2	0.8	12.3	2.9	19.4	2.2	6.6	0.7	12.3	1.3	22.1	2.1
113	7.0	0.9	12.6	0.9	11.3	1.5	16.2	2.5	8.2	0.8	12.8	0.7	18.3	2.4
114	10.4	0.8	16.6	0.8	14.3	1.6	18.1	1.9	12.5	0.8	16.6	1.1	20.1	2.6
115	11.2	0.7	19.8	1.2	14.1	1.3	15.7	2.0	14.9	0.7	18.3	1.1	23.1	3.1
116	9.0	1.3	20.6	1.1	10.5	1.1	11.9	1.9	15.3	0.8	16.4	1.4	25.0	3.5
117	6.6	0.7	18.4	1.3	10.4	1.0	11.5	2.0	14.4	2.4	12.6	1.2	22.6	3.2
118	8.3	0.7	18.7	1.2	6.0	0.9	11.7	3.8	18.1	2.5	12.0	0.8	25.4	2.7
119	8.5	0.8	15.4	0.9	6.7	1.0	10.3	3.5	17.8	4.2	8.7	0.8	24.4	2.5
120	12.4	0.8	16.2	1.5	5.4	0.5	9.3	3.3	21.0	4.5	9.9	0.6	28.1	2.6
121	13.2	0.8	16.8	0.8	9.6	1.0	9.3	2.6	20.8	4.5	9.6	1.4	26.0	2.8
122	9.3	0.8	16.3	0.6	9.4	1.1	8.7	1.7	17.2	3.3	11.1	2.1	21.5	2.9
123	12.7	0.8	21.2	0.7	11.8	1.8	6.2	1.4	20.0	2.3	15.0	2.1	22.3	3.5
124	8.9	0.9	19.3	1.0	11.9	2.1	9.9	1.4	18.9	2.3	15.7	2.7	18.3	3.4
125	4.7	0.6	14.5	1.3	13.1	1.6	13.5	2.5	16.0	4.0	15.0	2.0	13.3	3.2
126	8.0	1.1	18.6	1.2	16.3	2.4	13.6	1.7	20.2	2.9	18.4	2.1	13.6	3.2
127	8.0	1.0	14.6	1.3	18.4	1.7	16.3	2.1	18.8	3.0	20.4	1.7	12.6	2.5
128	7.4	0.8	16.9	1.7	18.7	1.7	19.9	1.7	18.0	2.9	20.7	1.1	8.4	2.6
129	5.3	0.8	14.4	1.9	18.0	1.0	20.5	1.8	14.4	2.7	18.4	1.0	8.9	1.8
130	9.6	0.8	17.3	2.8	21.6	0.8	25.1	1.8	15.9	2.6	22.2	1.1	10.4	2.0
131	9.8	0.6	19.2	1.8	22.3	1.0	26.8	1.8	14.7	1.2	24.0	0.9	7.4	2.2
132	6.8	0.7	16.6	0.8	18.9	1.1	23.2	1.8	12.3	0.8	20.7	0.9	5.8	2.1
133	6.6	1.3	14.3	1.4	13.8	1.4	18.1	2.0	12.3	0.8	16.3	1.2	8.1	3.1
134	7.8	0.4	11.5	0.7	16.3	0.9	21.0	2.3	7.2	0.4	15.0	0.8	10.0	1.9
135	7.2	0.8	7.0	0.3	13.6	0.8	19.0	2.2	5.4	1.0	10.2	0.8	13.4	2.1
136	6.4	0.6	9.9	0.5	11.9	0.8	17.6	1.8	7.1	1.3	12.1	1.6	16.5	2.1
137	11.6	1.3	5.8	0.3	17.7	0.9	22.5	1.8	6.2	0.9	13.6	1.0	16.4	1.5
138	13.3	0.9	8.4	0.6	20.1	0.9	26.1	2.1	9.2	0.6	18.1	1.0	14.8	1.3
139	15.3	1.4	8.3	1.0	21.5	1.7	27.4	2.7	10.0	0.7	19.9	1.0	14.1	1.5
140	18.2	1.1	7.9	1.9	25.2	1.4	30.6	2.3	9.3	1.2	20.8	1.6	16.7	1.7

Representative from cluster 2 (distances in [Å] between the ligands and the C-terminal residues)

Res.	DCH		DHI		DOP		DOP-H		DQ		IQ		LEUK	
Num.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
83	34.1	4.8	18.2	1.7	28.7	6.9	37.4	3.8	29.2	5.2	18.3	2.1	11.2	1.7
110	56.9	21.7	27.6	5.0	39.2	7.2	29.0	6.2	36.6	7.1	33.2	2.0	31.0	2.7
111	57.8	21.9	27.2	5.4	37.3	7.3	26.3	7.0	34.6	7.9	35.0	2.7	32.0	2.5
112	55.0	22.3	23.4	5.9	33.2	7.3	26.1	6.6	31.3	7.4	32.6	3.8	29.1	2.9
113	55.2	24.4	21.2	4.8	33.5	7.2	22.7	5.4	30.6	8.8	31.3	3.6	25.6	2.2
114	58.9	23.7	25.7	4.4	37.0	7.3	22.9	5.6	33.9	10.6	36.1	3.7	28.9	2.0
115	59.8	25.5	25.1	3.8	33.2	7.4	18.6	5.5	32.2	11.2	34.7	3.2	27.6	1.5
116	58.0	27.8	21.2	3.6	30.0	7.1	20.3	4.3	31.2	9.7	31.4	3.9	22.9	1.5
117	60.2	28.5	21.7	3.4	28.9	7.1	17.7	3.6	30.3	10.6	30.0	2.9	24.2	1.7
118	61.2	29.3	24.0	2.6	32.1	7.1	20.2	3.4	34.0	10.1	31.6	2.6	23.2	1.7
119	61.6	31.9	21.1	3.5	30.2	6.8	20.7	2.8	35.0	9.3	28.1	3.0	24.1	2.3
120	63.9	31.1	24.1	3.9	30.7	5.4	24.1	2.8	38.2	9.1	28.4	2.8	23.6	2.8
121	63.1	32.9	23.1	3.1	29.1	4.8	23.9	2.9	36.8	8.5	29.1	2.8	23.0	2.1
122	63.9	34.3	20.6	3.8	28.8	4.6	21.6	2.8	36.0	7.7	25.9	3.5	23.1	2.6
123	60.8	35.1	20.9	2.0	24.6	4.6	17.4	2.9	33.0	8.7	24.2	2.3	20.7	2.4
124	59.0	34.6	18.0	1.7	22.3	4.6	18.1	2.7	30.0	8.6	22.3	3.7	17.3	2.5
125	56.7	33.9	20.0	1.9	20.4	5.8	13.9	2.6	29.8	10.1	18.9	2.4	19.1	2.1
126	54.3	35.5	18.1	3.5	17.5	4.9	18.0	2.5	26.8	10.4	18.0	4.8	15.0	2.8
127	52.8	36.1	19.8	3.3	16.8	5.5	15.1	2.8	27.6	12.7	17.6	4.2	15.7	3.0
128	50.5	34.6	17.4	2.9	14.2	6.5	13.0	2.7	25.9	13.5	17.1	2.2	16.6	2.6
129	51.9	33.2	17.9	2.3	17.4	7.1	5.4	0.7	25.4	13.4	19.3	2.1	20.2	2.8
130	50.8	34.7	19.4	2.4	16.1	6.4	10.9	3.0	24.6	14.3	21.7	3.0	22.8	3.2
131	50.7	33.1	23.3	2.4	17.8	6.4	10.3	4.4	22.4	13.6	20.4	3.4	21.1	3.2
132	50.9	30.8	23.0	2.9	19.6	7.3	9.4	3.6	22.7	12.5	16.9	2.9	18.3	3.1
133	50.8	30.7	20.4	2.4	22.0	7.9	12.3	2.9	22.3	10.4	14.3	2.3	17.8	3.3
134	52.2	27.1	23.5	2.3	23.9	7.3	8.6	1.9	23.2	10.2	18.4	2.3	23.6	2.9
135	50.9	25.5	26.2	3.3	27.7	7.8	6.8	1.8	23.1	8.5	16.6	2.6	25.9	3.4
136	52.1	23.5	27.4	3.6	28.6	7.3	8.8	1.5	26.1	6.4	21.5	2.4	26.7	4.0
137	50.4	22.3	30.2	3.7	33.8	7.8	5.8	1.2	26.1	7.5	21.6	3.1	30.8	4.8
138	51.6	20.0	32.7	3.7	36.4	7.7	7.7	1.8	28.4	8.8	24.3	2.8	32.5	5.7
139	51.8	20.9	33.3	3.6	39.7	7.1	11.5	2.5	27.0	10.0	27.6	3.4	35.6	5.1
140	51.2	18.9	34.6	3.9	41.8	7.8	12.8	2.7	29.2	10.4	25.8	3.4	38.3	5.7

Representative from cluster 3 (distances in [Å] between the ligands and the C-terminal residues)

Res.	DCH		DHI		DOP		DOP-H		DQ		IQ		LEUK	
Num.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
83	21.8	3.1	24.8	4.2	25.4	6.0	17.6	2.6	11.6	2.8	21.3	4.1	26.3	6.7
110	37.5	2.8	34.0	5.2	26.5	2.4	24.7	6.2	31.9	2.8	42.6	4.2	12.6	6.9
111	39.1	3.9	30.9	5.2	25.3	2.4	25.7	5.7	34.0	3.0	44.8	3.8	12.2	4.8
112	38.4	4.2	27.9	5.5	21.6	2.6	23.6	5.9	36.3	3.1	46.6	3.9	11.7	3.2
113	35.5	5.6	25.2	4.8	23.0	2.5	23.0	6.1	33.1	3.2	41.9	3.8	13.8	2.9
114	33.3	5.1	22.5	5.7	18.1	3.1	18.0	6.1	35.8	3.1	43.8	4.9	12.4	4.2
115	31.5	6.2	20.8	4.6	21.3	3.0	17.9	7.5	35.5	4.6	44.0	4.6	14.6	3.8
116	35.2	6.6	22.4	4.2	21.9	2.5	14.9	7.2	37.1	4.4	44.5	3.5	16.1	6.3
117	35.1	7.2	20.2	4.2	19.7	2.7	12.6	6.4	35.4	4.8	41.6	3.4	14.3	7.0
118	38.3	5.9	22.0	4.8	16.4	2.8	14.4	4.5	36.4	3.3	41.4	2.2	16.5	8.4
119	38.2	4.2	22.4	5.0	16.4	2.1	12.4	3.2	36.7	4.2	40.3	2.0	16.2	8.9
120	39.0	4.5	21.3	4.8	13.3	2.7	13.2	2.9	36.2	4.2	41.0	3.2	17.5	9.2
121	41.1	3.8	24.6	5.4	12.8	2.7	12.0	2.6	35.7	3.8	42.0	3.3	16.2	8.5
122	43.4	4.0	26.0	5.4	16.1	2.6	11.5	4.0	37.2	4.5	44.8	3.1	19.2	8.3
123	42.0	3.7	27.0	6.4	14.4	2.5	9.8	2.5	34.6	4.9	41.9	3.0	19.9	7.5
124	38.3	3.8	23.3	6.3	13.6	2.8	8.2	2.9	32.3	6.2	44.1	3.3	20.2	8.6
125	34.1	3.3	23.1	6.6	12.5	3.1	8.1	2.3	28.9	5.8	40.0	3.3	25.4	9.2
126	37.5	3.4	19.4	5.3	11.0	2.5	10.1	2.7	34.0	6.4	44.7	4.0	25.8	9.2
127	37.1	2.9	20.2	6.5	7.8	2.4	12.9	2.5	34.9	5.4	41.8	4.3	25.8	9.1
128	34.0	2.8	18.7	6.9	7.5	2.2	14.0	3.5	33.6	4.7	38.7	4.6	28.9	9.0
129	36.3	3.7	17.9	6.0	10.5	1.7	14.8	3.8	31.2	4.9	35.5	3.7	27.1	9.4
130	35.5	4.7	18.4	6.1	9.3	2.4	18.1	4.6	29.7	4.4	33.0	4.9	29.6	8.9
131	37.5	5.1	21.4	5.9	10.3	2.3	19.5	3.8	31.2	4.8	35.1	6.0	27.6	8.9
132	40.2	4.4	22.2	5.5	11.5	2.2	17.4	3.6	32.8	4.1	38.0	5.0	24.7	9.3
133	39.9	3.3	22.7	5.6	10.6	2.4	15.4	3.6	35.4	4.6	39.3	3.8	20.8	9.6
134	42.2	4.1	23.5	6.7	14.8	2.6	20.2	4.0	37.1	3.8	34.9	4.7	23.1	8.4
135	44.6	4.0	26.9	5.5	14.7	3.2	21.0	3.5	39.1	4.1	36.6	3.8	20.0	8.2
136	45.7	3.7	26.6	6.1	15.2	2.5	19.4	3.0	38.1	4.4	34.2	3.5	17.1	8.8
137	48.4	3.4	29.8	6.6	18.2	2.7	24.0	3.2	38.2	4.9	38.4	3.2	18.9	7.5
138	50.8	3.8	32.0	6.4	21.3	2.7	24.8	3.2	40.5	5.0	40.7	2.7	17.6	5.6
139	50.8	4.1	32.6	7.8	22.2	3.7	26.7	3.6	42.2	4.5	44.3	3.2	21.0	5.7
140	53.1	4.5	35.2	7.9	25.3	3.4	30.3	3.6	43.9	5.4	46.4	2.2	22.8	4.5

Representative from cluster 4 (distances in [Å] between the ligands and the C-terminal residues)

Res.	DCH		DHI		DOP		DOP-H		DQ		IQ		LEUK	
Num.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
83	15.4	0.9	11.8	1.5	24.0	1.6	29.3	1.6	23.3	1.2	17.5	1.0	12.1	1.2
110	17.8	1.6	15.3	1.4	22.2	1.0	30.4	1.3	18.5	0.7	19.4	1.1	12.2	3.1
111	15.5	0.8	15.3	0.9	19.2	0.8	27.5	1.4	14.5	0.8	18.6	1.3	14.6	3.7
112	12.9	1.3	13.3	0.8	14.7	0.8	24.6	1.3	10.3	0.7	15.5	1.4	15.7	3.8
113	11.0	0.7	10.2	1.0	14.4	1.1	21.8	1.3	14.3	0.8	12.7	1.2	10.7	4.7
114	6.3	0.9	6.6	0.9	10.0	0.9	17.4	1.3	12.4	0.6	8.5	1.5	12.9	4.1
115	9.4	0.9	7.7	0.8	9.7	2.1	18.5	1.2	15.4	0.9	8.4	1.1	12.0	5.1
116	5.8	0.8	5.3	1.3	9.1	1.3	15.5	1.3	16.8	0.7	5.5	0.7	13.2	3.1
117	8.4	0.8	8.8	1.5	10.1	2.0	17.5	1.2	19.3	1.0	8.8	1.3	11.8	2.2
118	5.7	0.6	8.6	2.3	14.4	1.9	12.9	1.2	23.8	0.9	9.4	2.0	9.4	1.2
119	9.9	0.9	11.1	1.2	16.9	1.4	14.1	0.9	26.1	0.9	9.3	3.0	12.6	1.7
120	10.0	1.9	13.7	1.5	19.6	1.8	13.9	1.6	28.9	1.0	12.6	2.8	14.3	2.4
121	13.6	1.1	14.4	0.8	21.3	1.8	13.0	1.2	31.5	0.9	11.8	2.8	17.3	2.0
122	13.2	1.3	12.4	0.7	18.0	2.6	10.7	1.1	29.7	1.1	11.7	2.6	16.1	1.9
123	9.9	1.2	7.9	0.2	15.4	1.9	6.7	0.7	26.0	1.0	7.4	2.2	19.2	1.0
124	10.8	0.7	11.3	0.8	16.0	1.3	8.2	1.0	25.0	1.2	10.5	1.4	19.2	1.8
125	6.4	0.7	8.0	0.6	14.1	1.7	5.5	0.7	22.2	1.1	8.1	1.7	22.4	1.9
126	11.0	0.9	13.5	1.1	12.5	1.1	8.7	1.3	19.5	1.2	11.4	0.9	22.7	2.8
127	10.0	1.5	14.2	1.2	13.1	0.9	11.1	1.7	18.8	1.6	13.5	1.4	26.3	2.7
128	9.5	0.8	12.6	1.4	8.6	0.8	13.1	1.6	14.5	1.2	10.1	1.6	23.7	3.0
129	14.1	0.7	13.8	1.3	8.3	1.0	16.5	1.5	11.7	1.1	12.1	1.7	24.8	3.8
130	15.1	0.9	17.4	1.4	11.0	0.6	20.4	1.9	9.4	1.2	16.0	1.8	26.3	4.4
131	16.4	0.9	16.7	1.1	9.0	0.5	20.9	1.6	5.6	1.0	13.9	1.7	23.1	4.7
132	13.1	0.9	12.9	1.0	6.2	0.4	19.1	1.4	7.5	0.8	10.6	1.6	20.4	4.3
133	10.0	0.7	9.5	1.4	6.3	0.4	15.9	1.3	10.9	0.9	7.5	1.6	18.7	3.1
134	9.6	1.1	11.3	1.3	11.9	0.6	20.5	1.4	8.4	0.7	12.4	1.6	18.6	3.6
135	11.1	0.7	10.9	1.5	13.3	0.5	17.1	1.4	13.6	0.8	11.0	1.6	18.5	3.0
136	6.3	0.5	9.6	1.5	13.4	1.3	11.6	1.3	15.5	1.0	7.2	1.5	19.6	3.0
137	11.2	0.8	12.3	1.1	16.4	1.0	11.6	1.1	15.8	1.3	9.8	1.6	23.7	3.1
138	12.7	1.5	12.6	1.6	17.2	1.7	7.5	1.1	17.0	1.5	10.3	1.9	25.2	3.4
139	16.3	0.9	16.9	1.5	21.3	1.5	8.4	1.4	13.6	1.8	14.8	1.9	28.0	2.9
140	18.4	1.8	18.8	1.8	22.9	1.5	6.7	1.3	17.4	2.3	16.9	1.8	30.3	3.9

Representative from cluster 5 (distances in [Å] between the ligands and the C-terminal residues)

Res.	DCH		DHI		DOP		DOP-H		DQ		IQ		LEUK	
Num.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
83	16.5	2.7	35.0	8.6	12.0	2.3	26.6	2.6	50.6	7.0	23.2	3.4	21.7	3.4
110	14.9	0.7	22.6	11.2	29.8	1.9	11.6	0.7	24.7	12.7	11.7	2.1	13.2	1.5
111	11.9	0.6	19.7	10.9	29.2	1.9	10.3	0.6	24.8	13.3	12.3	2.3	14.0	2.0
112	8.8	0.7	16.7	10.9	25.3	1.9	6.6	0.6	23.8	15.0	9.9	2.4	11.2	2.1
113	8.3	0.6	20.2	11.8	26.8	2.7	9.2	0.9	24.3	14.2	13.0	3.9	14.3	3.5
114	6.4	0.5	17.5	11.8	22.9	3.5	7.1	1.2	21.3	13.7	11.3	5.1	12.0	4.5
115	4.8	0.4	20.1	10.9	23.6	3.6	9.1	1.9	23.4	15.3	15.1	4.7	14.1	4.4
116	7.5	0.4	21.7	12.0	24.0	2.6	8.3	1.2	24.8	15.6	13.3	3.4	12.7	3.4
117	6.6	0.5	24.6	12.8	20.4	2.4	8.2	1.5	25.2	15.7	15.0	3.7	11.0	3.5
118	8.6	0.5	25.4	13.0	19.0	2.1	6.0	0.6	22.6	15.6	12.1	2.7	7.5	2.5
119	12.8	0.5	29.2	13.2	18.3	2.6	10.2	1.1	26.0	14.9	14.4	2.0	8.8	2.0
120	14.3	0.8	31.8	12.7	15.7	2.7	11.9	1.1	26.7	15.0	16.4	2.2	7.3	3.0
121	17.1	0.6	33.3	13.2	18.9	4.0	13.9	1.2	27.9	14.3	16.6	2.6	10.9	2.7
122	13.8	0.6	29.2	13.3	16.7	4.1	9.9	1.5	24.0	14.0	12.6	2.7	11.2	2.1
123	18.1	0.9	32.5	12.5	15.9	6.2	13.4	1.8	25.1	12.8	13.7	4.8	12.8	2.1
124	17.1	1.8	30.9	13.2	13.2	4.4	9.9	2.2	22.4	12.6	6.3	3.0	12.0	2.4
125	15.8	1.0	29.0	12.8	15.3	4.2	10.5	0.9	22.6	12.5	6.2	1.9	15.7	2.1
126	17.4	0.7	30.3	13.3	18.2	5.2	12.5	1.2	25.7	13.2	7.6	2.5	14.7	1.5
127	19.1	1.4	28.0	13.1	20.3	3.9	12.5	0.9	28.6	13.6	12.6	2.6	16.1	2.4
128	16.4	1.4	30.3	14.6	21.6	4.7	13.8	0.8	29.6	12.9	14.3	2.6	13.2	2.8
129	20.4	1.5	32.3	13.3	24.8	5.6	18.5	0.7	32.4	12.8	17.2	2.4	15.2	2.9
130	21.3	2.2	34.3	14.0	28.5	5.4	20.1	2.6	34.7	12.6	19.8	2.7	15.1	4.1
131	24.0	1.9	35.7	12.4	30.1	6.4	22.9	2.4	38.0	12.2	23.1	2.5	18.1	4.4
132	21.1	1.4	34.2	11.1	26.4	6.3	21.8	1.7	38.8	12.8	22.5	2.3	19.6	3.8
133	25.1	1.7	32.9	8.9	30.0	6.5	25.6	2.5	41.3	11.8	25.2	3.1	23.3	2.7
134	25.3	1.1	32.5	10.3	26.7	6.5	23.5	2.1	37.8	11.7	25.9	3.1	23.7	4.1
135	29.0	0.9	31.5	8.6	30.3	6.1	27.6	2.0	37.7	11.5	28.4	3.7	26.7	4.1
136	27.5	1.7	31.2	9.3	26.5	6.1	23.3	2.3	33.8	11.5	29.8	4.0	27.2	5.5
137	23.3	1.1	31.8	8.0	25.2	5.9	21.6	2.3	34.2	11.8	31.6	3.3	29.4	4.7
138	23.5	2.1	33.1	6.0	27.4	6.2	24.4	2.6	34.8	11.0	31.9	4.2	30.7	4.2
139	20.5	2.7	34.1	6.4	28.0	6.1	22.2	3.5	39.3	11.2	35.7	4.2	33.3	5.4
140	20.3	2.2	33.8	8.3	28.6	5.0	21.2	2.6	39.7	11.4	36.6	5.6	35.0	5.3

Representative from cluster 6 (distances in [Å] between the ligands and the C-terminal residues)

Res.	DCH		DHI		DOP		DOP-H		DQ		IQ		LEUK	
Num.	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD	Av.	SD
83	30.3	5.8	11.9	1.4	19.0	1.5	13.1	0.6	21.3	3.4	28.5	1.8	20.1	10.2
110	27.6	5.3	20.9	2.4	15.3	0.7	20.4	2.9	18.7	2.4	16.3	1.1	26.3	13.8
111	23.8	5.6	17.2	2.3	12.4	0.7	20.7	1.3	14.8	2.2	19.5	1.0	25.1	13.0
112	23.0	4.9	14.8	3.3	11.6	0.5	18.3	1.3	13.8	2.1	15.8	1.0	23.0	12.9
113	18.0	4.3	12.2	1.1	6.0	0.7	20.8	1.4	11.7	2.9	19.7	1.3	19.8	13.4
114	19.2	3.5	14.0	1.4	10.2	0.5	14.9	1.5	11.5	2.1	17.8	1.6	18.2	13.0
115	15.9	2.3	10.4	0.7	7.2	0.8	19.4	1.5	8.6	2.4	21.4	1.9	19.3	11.5
116	18.3	1.3	14.2	1.1	11.7	0.9	20.7	1.6	12.9	2.2	19.2	2.5	17.8	12.9
117	14.7	1.4	12.7	1.1	10.6	1.3	21.3	1.4	11.8	2.0	23.9	2.4	19.0	10.5
118	16.1	2.9	16.8	1.5	13.9	2.0	26.0	1.3	13.2	2.1	25.2	3.1	18.7	10.8
119	19.8	1.8	18.2	1.1	16.0	1.7	25.6	2.2	15.7	2.5	22.5	3.0	17.9	12.8
120	21.7	1.1	18.1	1.8	16.4	1.4	23.6	2.9	16.8	2.5	19.3	2.5	17.4	14.3
121	22.8	1.2	19.5	1.6	19.3	1.0	25.6	3.2	20.5	2.3	23.3	2.1	19.9	13.2
122	18.1	1.8	15.9	1.0	15.7	1.1	24.3	1.9	18.8	3.0	25.4	2.3	20.0	10.8
123	18.8	1.7	16.1	1.4	18.1	1.1	22.5	2.4	19.6	2.1	27.3	1.5	22.7	10.0
124	15.0	1.5	12.5	1.0	14.5	1.2	23.6	1.9	15.8	2.1	25.6	1.3	22.0	9.5
125	15.0	2.1	11.1	1.0	12.8	1.1	19.7	2.3	15.6	1.7	29.8	1.3	24.1	7.8
126	11.4	1.8	6.2	0.8	8.0	1.5	23.5	1.5	10.0	1.7	28.7	1.8	24.0	8.2
127	11.2	1.3	6.7	0.4	11.4	2.5	25.4	2.5	11.5	1.5	27.9	1.4	26.5	8.7
128	7.3	1.3	4.8	0.6	9.4	3.1	25.1	1.8	8.5	1.1	30.3	1.2	27.9	8.6
129	6.3	1.2	8.2	0.4	10.1	1.9	29.2	1.8	6.3	1.4	33.1	1.2	30.8	7.4
130	6.7	1.1	11.5	1.0	12.7	1.8	29.7	1.7	9.0	1.4	36.9	1.2	34.3	6.8
131	10.4	1.0	13.7	0.8	13.1	0.8	34.3	1.6	10.9	1.6	40.4	1.2	35.4	6.4
132	8.9	1.4	10.8	0.5	9.6	1.0	32.7	2.2	8.4	2.0	40.1	1.6	33.1	7.1
133	5.7	1.4	6.6	0.5	5.6	1.0	27.2	2.4	6.1	1.3	35.1	1.8	29.1	7.9
134	7.6	0.7	6.9	0.5	7.0	0.8	30.2	2.5	10.4	1.3	37.5	1.4	31.7	8.0
135	6.0	0.6	5.5	0.4	6.3	0.4	26.1	2.4	8.4	1.4	34.5	2.1	29.8	8.4
136	11.1	0.7	10.5	0.5	11.9	0.8	30.1	2.5	6.9	2.3	37.1	2.0	31.4	9.4
137	10.3	2.0	7.1	1.0	7.5	0.7	25.8	2.1	9.2	1.7	31.0	1.8	32.8	9.4
138	12.5	1.8	10.6	0.8	11.7	0.7	27.9	2.2	12.6	1.5	30.8	1.4	32.5	9.0
139	14.0	2.7	9.8	1.6	12.6	0.6	24.6	2.5	13.4	1.8	29.0	1.5	32.4	10.1
140	12.8	3.7	11.6	2.4	12.7	1.0	22.2	4.1	16.6	1.7	32.7	1.6	34.3	9.9

Region	DOP	DOP-H	DCH
125-129	9.1 %	12.5 %	7.4 %
130-135	14.8 %	16.7 %	9.5 %
125-135	23.9 %	29.2 %	16.9 %

Table D.2: Docking: Relative contribution of the binding to selected regions in the C-terminal region, calculated as percentage of the total number of ligand-protein contacts.

Cluster		**Hydrogen Bonds**	**Hydrophobic contacts**
1 (5 %)			
2 (5 %)	DOPH	Glu130(OE2)-DOP-H(O1), (D: 2.9±0.5 Å)	Met127 (D: 4.4±0.4 Å)
	DCH		Ile112 (D: 6.7±1.1 Å) Glu126 (D: 7.7±0.5 Å) Tyr136 (D: 4.3±0.5 Å)
3 (4 %)	DOP	Ser129(O)-DOP(O1), (D: 2.7±0.1 Å)	Gln134 (D: 4.4±0.4 Å) Lys43 (D: 5.2±0.5 Å)
	DCH		Glu126 (D: 5.9±1.5 Å) Gln134 (D: 6.6±1.3 Å)
4 (4 %)	DOP	Glu123(OE2)-DOP(O1), (D: 2.6±0.3 Å)	Gln134 (D: 4.7±0.7 Å)
	DOPH	Pro128(O)-DOP-H(O2), (D: 2.8±0.2 Å)	
	DCH		Glu123 (D: 4.3±0.4 Å)
5 (4 %)	DOP	Glu131(OE2)-DOP(O2), (D: 3.0±0.9 Å)	Tyr125 (D: 7.3±1.8 Å)
	DCH		Thr81 (D: 5.7±0.6 Å) Pro128 (D: 4.4±0.6 Å)
6 (3 %)	DOP-H	Glu126(O)-DOP-H(N1), (D: 2.8±0.1 Å) Glu126(OE2)-DOP-H(O2), (D: 2.7±0.4 Å)	
	DCH		Val66 (D: 5.2±1.1 Å) Glu126 (D: 4.0±0. 5 Å)

Table D.3: Hydrogen bonds and hydrophobic contacts for 11 out of the 18 analyzed complexes forming interactions with the protein. Several of then shown interactions with the C-terminal region, including the ^{125}YEMPS^{129} region. The distance (D) in the hydrogen bond column was measured between the heavy atoms. The hydrophobic contacts were measured as the distance (D) between the center of mass of the ligand and the specific amino acid.

Representative from cluster 1 (Ligs. - C-term residues distances [Å])							Representative from cluster 2 (Ligs. - C-term residues distances [Å])						
Res.	**DCH**		**DOP**		**DOP-H**		**Res.**	**DCH**		**DOP**		**DOP-H**	
Num.	Av.	SD	Av.	SD	Av.	SD	**Num**	Av.	SD	Av.	SD	Av.	SD
83	39.9	2.6	27.0	1.9	37.0	4.3	**83**	34.1	1.8	29.4	2.5	23.4	1.8
110	12.6	2.2	10.0	2.6	14.1	3.2	**110**	12.6	0.9	6.5	1.1	13.4	0.9
111	13.3	2.3	13.8	2.6	11.7	3.6	**111**	8.3	1.0	9.7	0.8	14.2	0.6
112	11.2	2.6	15.4	2.3	8.2	4.1	**112**	6.7	1.1	10.2	0.8	17.8	0.7
113	11.8	2.7	10.7	2.1	9.9	4.6	**113**	10.1	1.5	6.3	1.1	18.0	0.9
114	14.7	2.6	12.6	1.1	13.2	2.8	**114**	13.1	1.1	7.0	0.9	16.3	0.5
115	18.5	2.7	12.0	1.1	15.9	3.8	**115**	16.9	1.0	5.9	1.0	16.3	0.6
116	19.0	3.1	11.9	0.9	16.9	3.1	**116**	19.3	1.0	9.8	1.1	15.7	0.7
117	22.5	3.0	10.8	1.3	17.8	3.5	**117**	19.5	1.4	11.6	1.1	16.7	0.7
118	20.6	3.3	6.9	1.6	16.0	3.4	**118**	17.4	0.9	11.9	1.7	12.2	0.7
119	24.6	3.2	6.9	2.6	17.5	3.8	**119**	17.8	0.9	14.9	2.2	13.5	0.7
120	27.9	3.2	10.6	2.6	20.8	4.1	**120**	14.0	1.1	13.4	2.8	13.9	0.8
121	28.5	2.9	9.0	2.6	18.1	4.4	**121**	12.9	0.8	15.7	3.7	8.9	0.8
122	26.9	2.6	7.8	2.2	18.3	4.7	**122**	13.8	1.2	19.4	3.7	9.5	0.9
123	27.0	3.6	6.9	1.8	17.5	4.7	**123**	12.6	0.7	18.8	4.4	6.2	0.7
124	28.4	2.5	10.1	1.6	21.5	4.7	**124**	10.8	0.5	13.7	4.5	6.4	0.5
125	32.7	2.4	15.1	1.7	23.1	4.9	**125**	7.2	0.6	12.4	4.3	10.8	0.5
126	32.0	2.9	14.0	1.8	25.3	5.3	**126**	7.7	0.5	18.5	4.6	9.8	0.5
127	36.3	2.9	17.6	1.9	25.7	5.4	**127**	12.8	0.6	20.4	4.9	4.4	0.4
128	37.5	3.4	21.1	1.8	28.1	4.7	**128**	11.6	1.1	15.7	5.3	7.8	0.4
129	35.2	3.9	22.1	1.5	31.3	5.3	**129**	15.9	0.8	16.8	5.3	7.1	0.6
130	37.6	3.6	24.3	1.6	33.5	4.9	**130**	17.1	1.7	17.0	6.2	4.2	0.4
131	39.7	2.8	28.2	1.5	37.5	5.0	**131**	17.8	1.1	20.8	6.3	10.4	0.6
132	37.9	3.1	26.8	1.6	37.0	5.9	**132**	14.6	0.8	20.3	5.7	9.1	0.5
133	35.0	3.4	24.2	1.7	34.8	6.6	**133**	12.7	1.0	21.9	5.6	8.6	0.5
134	33.6	3.4	20.9	1.5	32.8	5.2	**134**	8.7	0.8	16.7	5.4	11.6	0.4
135	35.9	2.8	23.7	1.6	37.2	5.1	**135**	8.6	0.6	21.2	4.9	14.8	1.0
136	35.2	3.3	19.9	1.7	34.1	4.9	**136**	4.3	0.5	20.6	5.6	15.6	0.8
137	40.4	3.5	26.4	1.7	39.4	4.8	**137**	8.2	0.7	23.6	4.8	19.5	0.6
138	42.9	3.0	28.3	1.8	41.7	4.7	**138**	6.6	1.0	21.6	4.7	18.1	1.1
139	44.9	3.4	31.4	1.7	44.9	4.7	**139**	10.7	1.3	23.6	4.2	22.1	1.2
140	43.8	3.8	32.6	1.7	44.1	4.6	**140**	11.7	1.7	23.0	3.7	24.4	0.9

Representative from cluster 3 (Ligs. - C-term residues distances [Å])							Representative from cluster 4 (Ligs. - C-term residues distances [Å])						
Res.	**DCH**		**DOP**		**DOP-H**		**Res.**	**DCH**		**DOP**		**DOP-H**	
Num.	Av.	SD	Av.	SD	Av.	SD	**Num**	Av.	SD	Av.	SD	Av.	SD
83	28.2	3.1	31.7	3.4	32.9	3.4	**83**	8.2	0.8	6.0	1.2	9.2	0.9
110	14.0	1.7	20.2	2.7	13.7	1.6	**110**	30.6	1.0	30.2	2.0	26.9	4.0
111	15.1	1.6	18.3	2.2	17.7	1.6	**111**	27.8	1.3	27.0	1.9	26.0	2.5
112	16.4	1.4	17.6	2.1	18.5	1.7	**112**	28.4	1.1	25.7	2.1	24.3	3.1
113	12.8	1.5	14.6	1.3	14.0	1.5	**113**	24.2	1.4	21.1	2.2	21.0	2.2
114	17.3	1.2	16.2	1.2	18.5	1.6	**114**	20.9	1.1	21.4	1.3	24.3	1.1
115	13.5	1.4	12.7	1.2	14.6	1.7	**115**	17.6	0.9	18.2	1.5	19.9	1.1
116	17.6	2.0	16.6	1.6	14.4	1.5	**116**	16.0	0.8	15.0	1.4	17.5	1.9
117	14.8	2.7	13.2	1.3	10.3	1.6	**117**	11.5	0.8	11.4	1.4	14.4	0.9
118	18.4	2.1	16.8	1.8	14.4	1.5	**118**	11.1	1.0	11.1	0.8	12.7	0.6
119	16.8	2.0	15.3	1.5	13.7	1.3	**119**	6.9	1.0	7.1	0.6	7.7	0.6
120	16.7	2.5	15.4	2.0	12.4	1.6	**120**	8.9	1.1	9.5	0.6	10.5	0.6
121	14.2	1.7	12.7	1.3	14.3	1.5	**121**	6.3	0.8	7.0	1.5	6.6	0.9
122	10.7	1.3	11.6	1.3	12.0	1.5	**122**	9.5	0.5	9.3	0.5	9.4	1.0
123	11.8	1.1	11.5	1.6	8.9	1.1	**123**	4.3	0.4	4.6	0.6	6.2	1.4
124	7.1	1.2	7.2	1.6	8.9	2.0	**124**	5.6	0.7	5.4	1.2	4.6	0.6
125	9.3	0.9	9.8	1.6	8.6	1.6	**125**	9.6	0.7	9.1	0.9	9.7	1.6
126	5.9	1.5	10.2	1.5	5.4	1.7	**126**	8..6	0.8	10.6	1.0	10.0	0.9
127	6.8	0.9	8.7	1.2	6.6	0.6	**127**	6.2	0.7	9.3	1.7	4.2	1.6
128	6.4	1.0	6.4	1.3	9.2	1.4	**128**	4.4	0.6	10.2	1.7	2.9	0.7
129	5.8	0.9	4.9	0.8	9.7	1.6	**129**	7.8	0.6	8.7	2.0	4.9	0.9
130	8.4	1.9	5.9	0.7	9.3	1.8	**130**	10.4	0.7	12.4	1.1	9.2	0.6
131	9.7	2.2	8.2	0.8	9.7	1.8	**131**	14.9	0.7	13.2	1.3	13.8	0.7
132	9.0	1.1	9.1	0.8	9.1	1.2	**132**	11.3	0.8	9.1	1.5	9.6	0.8
133	6.3	1.4	7.1	0.8	6.3	1.2	**133**	7.6	0.6	5.8	0.9	5.9	0.6
134	6.6	1.3	4.3	0.4	5.4	1.1	**134**	6.9	0.6	4.7	0.7	6.1	0.6
135	7.0	1.8	5.5	1.0	6.9	1.3	**135**	12.5	0.7	10.2	0.9	11.0	0.7
136	10.4	1.3	7.8	1.5	9.2	1.5	**136**	10.7	0.7	9.7	0.9	10.9	0.8
137	10.8	2.8	9.9	1.1	11.0	1.6	**137**	16.3	0.9	14.3	1.4	15.8	0.6
138	12.4	4.2	9.4	2.8	12.0	2.0	**138**	19.1	0.8	17.2	0.9	18.8	0.8
139	11.7	5.6	12.5	3.3	9.9	3.1	**139**	21.4	0.8	19.9	1.4	19.5	1.3
140	12.3	4.3	12.4	2.6	14.0	2.7	**140**	25.4	0.8	18.9	2.0	17.0	1.9

Representative from cluster 5 (Ligs. - C-term residues distances [Å])							Representative from cluster 6 (Ligs. - C-term residues distances [Å])						
Res.	**DCH**		**DOP**		**DOP-H**		**Res.**	**DCH**		**DOP**		**DOP-H**	
Num.	Av.	SD	Av.	SD	Av.	SD	**Num**	Av.	SD	Av.	SD	Av.	SD
83	8.3	0.9	10.1	1.9	21.6	2.7	**83**	23.5	0.8	16.8	2.9	19.5	1.9
110	20.8	2.9	23.3	3.9	10.0	2.8	**110**	34.8	0.7	19.8	5.4	32.8	1.2
111	22.1	2.4	21.3	3.1	9.4	3.9	**111**	29.7	0.6	15.8	4.8	28.0	1.3
112	19.7	2.0	24.7	2.7	13.1	3.3	**112**	28.3	0.9	15.2	4.4	26.8	1.1
113	24.0	2.5	24.0	3.0	11.2	2.4	**113**	24.2	0.6	11.4	4.6	24.0	2.4
114	23.3	2.5	21.7	3.5	9.7	2.4	**114**	26.9	0.8	12.8	4.8	25.3	1.8
115	19.6	2.1	17.6	3.3	8.2	1.7	**115**	27.8	0.6	15.9	3.9	25.4	1.2
116	14.9	2.2	18.2	3.0	6.0	0.9	**116**	26.5	0.7	16.7	3.4	24.4	0.7
117	15.5	1.6	14.3	2.9	5.9	0.9	**117**	24.5	0.6	18.4	2.7	23.1	0.8
118	13.1	2.0	16.1	2.7	9.5	1.7	**118**	22.7	0.7	16.8	3.2	21.1	0.9
119	12.8	2.2	13.6	2.6	12.9	1.3	**119**	18.7	0.5	13.0	3.0	17.8	0.6
120	13.4	1.8	14.2	2.4	15.2	1.7	**120**	19.3	0.7	12.0	4.2	17.6	0.8
121	11.5	2.3	10.7	2.3	18.2	1.8	**121**	15.3	0.5	12.4	4.7	13.1	0.7
122	15.0	2.6	12.8	2.4	22.9	1.9	**122**	14.6	0.4	15.1	3.8	12.5	1.3
123	12.6	2.1	10.0	1.5	23.9	1.8	**123**	10.5	0.4	15.7	4.1	8.8	2.0
124	11.6	2.0	9.2	1.5	22.0	2.1	**124**	11.3	0.5	12.0	4.0	9.3	0.4
125	6.3	1.5	7.3	1.8	23.1	2.7	**125**	10.8	0.6	13.0	2.5	9.8	0.4
126	10.3	1.5	11.7	1.6	24.1	3.0	**126**	4.0	0.5	16.5	3.4	3.9	0.2
127	10.1	0.7	12.4	1.7	27.6	3.3	**127**	6.5	0.5	16.5	4.6	4.9	0.3
128	4.4	0.6	11.4	1.8	28.3	4.1	**128**	7.9	0.4	15.4	3.9	6.8	0.2
129	8.8	1.3	9.6	1.1	29.4	3.0	**129**	7.1	0.3	19.8	4.1	7.1	0.3
130	9.9	1.8	7.3	0.9	32.2	2.7	**130**	11.3	0.4	19.1	4.2	11.2	0.5
131	13.0	1.6	5.0	0.4	31.3	2.2	**131**	11.7	0.5	17.2	4.0	11.7	0.4
132	11.6	1.3	8.0	0.8	28.2	2.2	**132**	11.3	0.4	21.6	4.1	11.3	0.6
133	8.9	1.2	5.6	1.0	25.8	2.2	**133**	11.4	0.5	24.5	4.3	11.5	1.0
134	13.4	1.6	6.4	1.3	24.5	1.6	**134**	16.2	0.5	22.3	3.2	15.7	1.0
135	11.8	1.4	9.2	2.4	20.2	1.4	**135**	17.0	0.4	23.8	4.1	16.2	0.8
136	17.4	1.5	12.7	1.7	24.3	2.2	**136**	14.1	0.4	24.2	4.6	14.0	0.8
137	17.7	1.8	13.9	2.2	20.7	1.8	**137**	11.6	0.6	26.0	4.1	11.1	1.0
138	21.3	1.7	17.0	1.9	22.4	3.4	**138**	16.0	0.6	27.6	4.0	15.5	1.1
139	23.1	2.2	16.5	2.7	22.8	2.0	**139**	17.2	0.5	31.5	4.0	17.1	1.0
140	26.6	2.0	19.7	2.3	25.1	1.8	**140**	18.9	0.8	32.4	4.1	17.9	1.1

Table D.4: Molecular dynamics simulations of the all dopamine forms/AS adducts studied with the conformations from the second clustering algorithm (18 complexes). The average values (Av.) with their standard deviations (SD) of the distance between the center of mass of E83 along with the C-Terminal residues (from 110 to 140) and dopamine, along the trajectory, are reported here

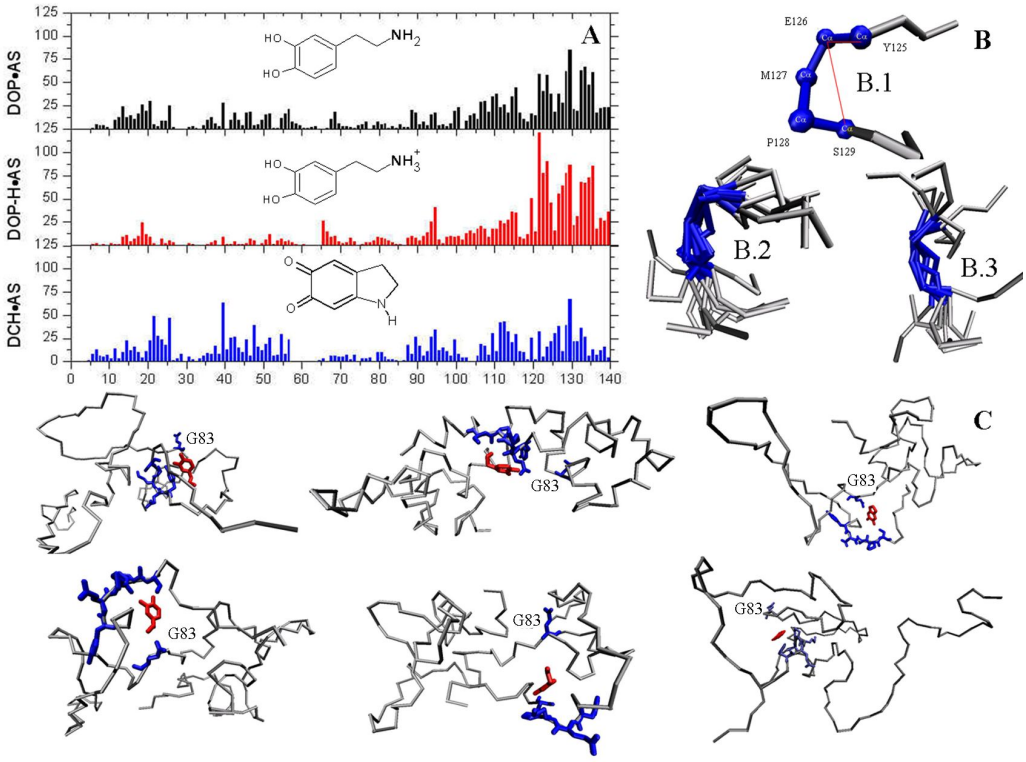
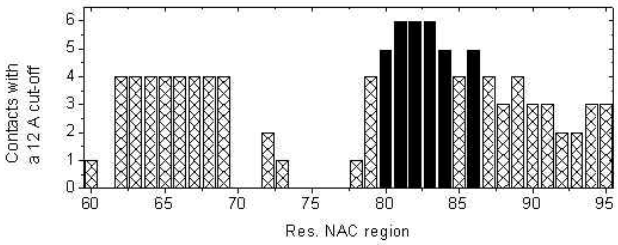


Figure D.1: Molecular docking and MD simulations of dopamine and its derivatives onto α-synuclein (AS): A) Number of hits between AS and dopamine (DOP), protonated dopamine (DOP-H) and dopaminochrome (DCH) as obtained in the 5,400 docking. B) Close up on the ^{125}YEMPS^{129} Cα carbons (B.1) after the MD simulation. If the ligands bind to this region (B.2), the angle between Cα 125-126-129 shows a relatively small spread ($79^\circ \pm 11^\circ$). The remaining 7 conformations is also shown (B.3) ($105^\circ \pm 20^\circ$). C) In 11 simulations out of 18, the ligands bind to the ^{125}YEMPS^{129} region. Here we show the six structures (among those) which the ligand close to E83. The 125-129 residues and E83 are colored in blue, the ligand is colored in red.



Res	Av	Std
80	2,3	4,4
81	2,7	4,2
82	0,2	0,7
83	**4,2**	**1,5**
84	0,1	0,8
86	0,06	0,74

Figure D.2: TElectrostatic analysis. Left:Number of times that NAC amino acids are found within a 12 Åfrom the analyzed ligands.The residues selected for the electrostatic analysis are marked in black. Right. Averaged energies values, for the selected residues (Res), expressed in units of the total average (AV) electrostatic interaction energy of -1.4 Kcal/mol and total standard deviation (Std) of 1.3 Kcal/mol.

Appendix E

Design of dopamine mimics

E.1 Database screening

The results from the screening of the database of known small compounds is shown in the Fig. E.1

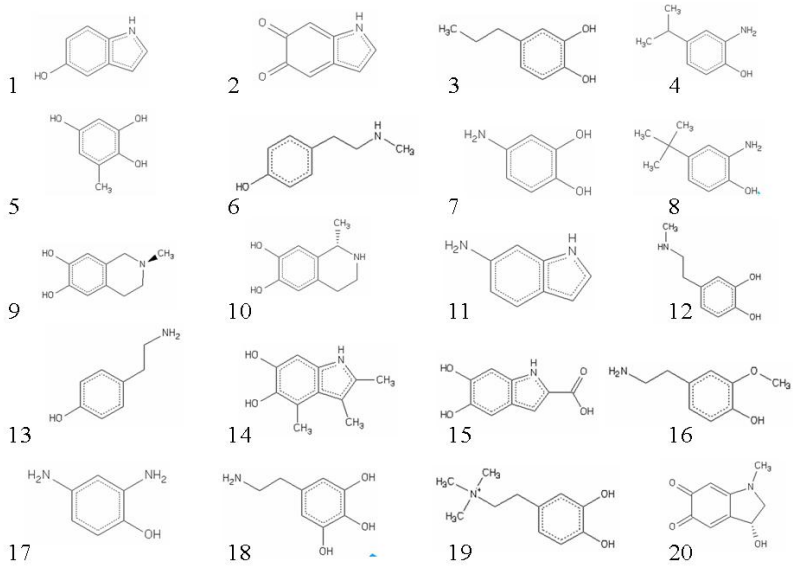


Figure E.1: Known molecules after the screening protocol, the 20 candidates from the set of 70 molecules are shown.

E.2 New compounds

The designed new molecules are shown in the Fig. E.2.

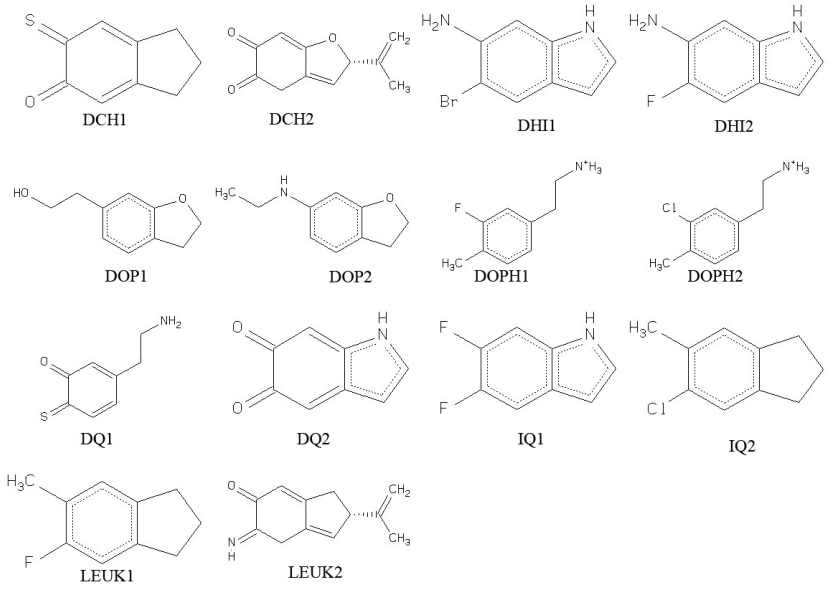


Figure E.2: Design of new molecules: the 14 molecules from the design protocol selected for the docking analysis are shown.

Bibliography

[1] Lang A.E. and Lozano A.M. Parkinson's disease. first of two parts. *New Engl. J. Med.*, 339:1044–1053, 1998.

[2] Lang A.E. and Lozano A.M. Parkinson's disease. second of two parts. *New Engl. J. Med.*, 339:1130–1143, 1998.

[3] Shadrina M., Nikopensius T., Slominsky P., Illarioshkin S., Bagyeva G., Markova E., Ivanova-Smolenskaia I., Kurg A., Limborska S., and Metspalu A. Association study of sporadic Parkinson's disease genetic risk factors in patients from russia by apex technology. *Neurosci. Lett.*, 405:212–216, 2006.

[4] Polymeropoulos M.H., Lavedan C., Leroy E., Dehejia A., Dutra A., Pike B., Root H., Rubenstein J., Boyer R., Stenroos E.S., Chandrasekharappa S., Athanassiadou A., Papapetropoulos T., Johnson W.G., Lazzarini A.M., Duvoisin R.C., Di Iorio G., Golbe L.I., and Nussbaum R.L. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*, 276:2045–2047, 1997.

[5] Abeliovich A, Schmitz Y, Farinas I, Choi-Lundberg D, Ho WH, Castillo PE, Shinsky N, Verdugo JM, Armanini M, Ryan A, Hynes M, Phillips H, Sulzer D, and Rosenthal A. Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system. *Neurom.*, 25:239–252, 2000.

[6] Barcelo-Coblijn G., Golovko M.Y., Weinhofer I., Berger J., and Murphy E.J. Brain neutral lipids mass is increased in alpha-synuclein gene-ablated mice. *J. Neurochem.*, 101:132–141, 2007.

[7] Willinghan S., Outeiro T.F., DeVit M.J., Lindquist S.L., and Muchowski P.J. Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein. *Science*, 302:1769–1772, 2003.

[8] George JM, Jin H, Woods WS, and Clayton DF. Characterization of a novel protein regulated during the critical period for song learning in the zebra finch. *Neurom.*, 15:361–372, 1995.

[9] Fountaine T.M. and Wade-Martins R. Interference-mediated knockdown of alpha-synuclein protects human dopaminergic neuroblastoma cells from+ toxicity and reduces dopamine transport. *J. Neurosci. Res.*, 85:351–363, 2007.

[10] Sharon R., Goldberg M.S., Bar-Josef I., Betensky R.A., Shen J., and Selkoe D.J. Alpha-synuclein occurs in lipid-rich high molecular weight complexes, binds fatty acids, and shows homology to the fatty acid-binding proteins. *Proc. Natl. Acad. Sci. USA.*, 98:9110–9115, 2001.

[11] Lotharius J. and Brundin P. Pathogenesis of Parkinson's Disease: dopamine, vesicles and alpha-synuclein. *Nat. Rev. Neurosci.*, 3:932–942, 2002.

[12] Souza J.M., Giasson B.I., Lee V.M., and Ischiropoulos H. Chaperone-like activity of synucleins. *FEBS Lett.*, 474:116–119, 2000.

[13] Chandra S., Gallardo G., Fernandez-Chacon R., Schluter O.M., and Sudhof T.C. Alpha-synuclein cooperates with CSP alpha in preventing neurodegeneration. *Cell*, 123:383–396, 2005.

[14] Dedmon M.M., Lindorff-Larsen K., Christodoulou J., Vendruscolo M., and Dobson C.M. Mapping long-range interactions in alpha-synuclein using spin-label and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.*, 127:476–477, 2005.

[15] Bertoncini C.W., Jung Y.S., Fernandez C.O., Hoyer W., Griesinger C., Jovin T.M., and Zweckstetter M. Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc. Natl. Acad. Sci. USA.*, 102:1430–1435, 2005.

[16] Norris E.H., Giasson B.I., Hodara R., Xu S., Trojanowski J.Q., Ischiropoulos H., and Lee V.M. Reversible inhibition of alpha-synuclein fibrillization by dopaminochrome-mediated conformational alterations. *J. Biol. Chem.*, 280:21212–21219, 2005.

[17] Mazzulli J.R., Armakola M., Dumoulin M., Parastatidis I., and Ischiropoulos H. Cellular oligomerization of alpha-synuclein is determined by the interaction of oxidized catechols with a c-terminal sequence. *J. Biol. Chem.*, 4282:31621–31630, 2007.

[18] Sickmeier M., Hamilton J.A., LeGall T., Vacic V., Cortese M.S., Tantos A., Szabo B., Tompa P., Chen J., Uversky V.N., Obradovic Z., and Dunker A.K. Disprot: the database of disordered proteins. *Nucleic Acids Res.*, 35:D786–93, 2007.

[19] Dunker A.K., Brown C.J., Lawson J.D., Iakoucheva L.M., and Obradovic Z. Intrinsic disorder and protein function. *Biochemistry*, 41:65736582, 2002.

[20] Uversky V.N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.*, 11:739756, 2002.

[21] Dyson H.J. and Wright P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, 6:197–208, 2005.

[22] Mo H., Moore R.C., Cohen F.E., Westaway D., Prusiner S.B., Wright P.E., and Dyson H.J. Two different neurodegenerative diseases caused by proteins with similar structures. *Proc. Natl. Acad. Sci. USA.*, 98:2352–2357, 2001.

[23] Orosz F., Kovacs G.G., Olah J., Lehotzky A., Vincze O., Zotter A., and Ovadi J. Tppp/p25: the role of a new intrinsically unstructured protein in neurodegeneration. *FEBS J.*, 272:S1, 2005.

[24] Schlessinger A., Punta M., and Rost B. Natively unstructured regions in proteins identified from contact predictions. *Bioinformatics*, 23:2376–2384, 2007.

[25] Sugase K., Dyson H.J., and Wright P.E:. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447:1021–1027, 2007.

[26] Fuxreiter M., Simon I., and Tompa P. Recognition of intrinsically unstructured proteins: the facile-binding model. *FEBS J.*, 271:G2–042P, 2005.

[27] Lev N., Roncevich D., Ickowicz D., Melamed E., and Offen D. Role of DJ-1 in Parkinson's disease. *J. Mol. Neurosci.*, 29:215–225, 2006.

[28] Moore D.J., Zhang L., Dawson T.M., and Dawson V.L. A missense mutation (L166P) in DJ-1, linked to familial Parkinson's disease, confers reduced protein stability and impairs homo-oligomerization. *J. Neurochem.*, 87:1558–1567, 2003.

[29] Zhou W., Zhu M., Wilson M.A., Petsko G.A., and Fink A.L. The oxidation state of DJ-1 regulates its chaperone activity toward alpha-synuclein. *J. Mol. Biol.*, 356:1036–1048, 2006.

[30] Taira T., Marta S., Iguchi-Ariga M., and Ariga H. Co-localization with DJ-1 is essential for the androgen receptor to exert its transcription activity that has been impaired by androgen antagonists. *Biol. Pharm. Bull.*, 27:574, 2004.

[31] Mitsumoto A. and Nakagawa Y. DJ-1 is an indicator for endogenous reactive oxygen species elicited by endotoxin. *Free Radic. Res.*, 35:885–893, 2001.

[32] Canet-Aviles R.M., Wilson M.A., Miller D.W., Ahmad R., McLendon C., Bandyopadhyay S., Baptista M.J., Ringe D., Petsko G.A., and Cookson M.R. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc. Natl. Acad. Sci. USA*, 101:9103–9108, 2004.

[33] Herrera F.E., Zucchelli S., Jezierska A., Lavina Z.S., Gustincich S., and Carloni P. On the oligomeric state of DJ-1 protein and its mutants associated with Parkinson Disease. a combined computational and in vitro study. *J. Biol. Chem.*, 282:24905–24914, 2007.

[34] Wood-Kaczmar A., Gandhi S., and Wood N.W. Understanding the molecular causes of Parkinson's disease. *Trends Mol. Med.*, 12:521–528, 2006.

[35] Shagam JY. Unlocking the secrets of Parkinson's disease. *Radiol Technol.*, 79:227–239, 2008.

[36] Perier C., Tieu K., Guegan C., Caspersen C., Jackson-Lewis V., Carelli V., Martinuzzi A., Hirano M., Przedborski S., and Vila M. Complex i deficiency primes bax-dependent neuronal apoptosis through mitochondrial oxidative damage. *Proc. Natl. Acad. Sci. U.S.A.*, 102:19126–19131, 2005.

[37] Braak H., Del Tredici K., Rub U., de Vos R.A., Jansen Steur E.M, and Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging*, 24:197–211, 2003.

[38] Forno L.S. Neuropathology of Parkinson's disease. *J. Neuropathol. Exp. Neurol.*, 55:259–272, 1996.

[39] Spillantini M.G., Schmidt M.L., Lee V.M., Trojanowski J.Q., Jakes R., and Goedert M. Alpha-synuclein in lewy bodies. *Nature*, 388:839–840, 1997.

[40] Langston J.W., Ballard P., Tetrud J.W., and Irwin I. Chronic parkinsonism in humans due to a product of meperidine analog synthesis. *Science*, 219:979–980, 1983.

[41] Dawson T.M. and Dawson V.L. Molecular pathways of neurodegeneration in Parkinson's disease. *Science*, 302:819–822, 2003.

[42] Hicks A.A., Ptursson H., Jonsson T., Stefansson H., Johannsdottir H.S., Sainz J., Frigge M.L., Kong A., Gulcher J.R., Stefansson K., and Sveinbjornsdottir S. A susceptibility gene for late-onset idiopathic Parkinson's disease. *Ann. Neurol.*, 52:549–555, 2002.

[43] Marder K., Tang M.X., Mejia H., Alfaro B., Cote L., Louis E., Groves J., and Mayeux R. Risk of Parkinson's disease among first-degree relatives: A community-based study. *Neurology*, 47:155–160, 1996.

[44] Sveinbjornsdottir S., Hicks A.A., Jonsson T., Petursson H., Gugmundsson G., Frigge M.L., Kong A., Gulcher J.R., and Stefansson K. Familial aggregation of Parkinson's disease in Iceland. *N. Engl. J. Med.*, 343:1765–1770, 2000.

[45] Singleton A.B., Farrer M., Johnson J., Singleton A., Hague S.M., Kachergus J., Hulihan M., Peuralinna T., Dutra A., Nussbaum R, Lincoln S., Crawley A., Hanson M., Maraganore D., Adler C., Cookson M.R., Muenter M., Baptista M., Miller D, Blancato J., Hardy J., and Gwinn-Hardy K. Alpha-synuclein locus triplication causes Parkinson's disease. *Science*, 302:841, 2003.

[46] Kitada T., Asakawa S., Hattori N., Matsumine H., Yamamura Y., Minoshima S., Yokochi M., Mizuno Y., and Shimizu N. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*, 392:605–608, 1998.

[47] Gasser T., Mller-Myhsok B., Wszolek Z.K., Oehlmann R., Calne D.B., Bonifati V., Bereznai B., Fabrizio E., Vieregge P., and Horstmann R.D. A susceptibility locus for Parkinson's disease maps to chromosome 2p13. *Nat. Genet.*, 18:262–265, 1998.

[48] Leroy E., Boyer R., Auburger G., Leube B., Ulm G., Mezey E., Harta G., Brownstein M.J., Jonnalagada S., Chernova T., Dehejia A., Lavedan C., Gasser T., Steinbach P.J., Wilkinson K.D., and Polymeropoulos MH. The ubiquitin pathway in Parkinson's disease. *Nature*, 395:451–452, 1998.

[49] Valente E.M., Abou-Sleiman P.M., Caputo V., Muqit M.M., Harvey K., Gispert S., Ali Z., Del Turco D., Bentivoglio A.R., Healy D.G., Albanese A., Nussbaum R., Gonzlez-Maldonado R., Deller T., Salvi S., Cortelli P., Gilks W.P., Latchman D.S., Harvey R.J., Dallapiccola B., Auburger G., and Wood N.W. Hereditary early-onset Parkinson's disease caused by mutations in pink1. *Science*, 304:1158–1160, 2004.

[50] Bonifati V., Rizzu P., van Baren M.J., Schaap O., Breedveld G.J., Krieger E., Dekker C.J., Squitieri F., Ibanez P., Joose M., van Dongen J.W., Vanacore N., van Switen J.C., Brice A., Meco G., van Duijin C.M., Oostra B.A., and Heutink P. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science*, 299:256–259, 2003.

[51] Funayama M., Hasegawa K., Kowa H., Saito M., Tsuji S., and Obata F. A new locus for Parkinsons disease (PARK8) maps to chromosome 12p11.2q13.1. *Ann. Neurol.*, 51:296–301, 2002.

[52] Pankratz N., Nichols W.C., Uniacke S.K., Halter C., Rudolph A, Shults C., Conneally P.M., Foroud T., and Parkinson Study Group. Significant linkage of Parkinson disease to chromosome 2q3637. *Am. J. Hum. Genet.*, 72:1053–1057, 2003.

[53] Cookson M.R. The biochemistry of Parkinson's disease. *Ann. Rev. Biochem.*, 74:29–52, 2005.

[54] Wong P.C., Cai H., Borchelt D.R., and Price D.L. Genetically engineered mouse models of neurodegenerative diseases. *Nat. Neurosci.*, 5:633–639, 2002.

[55] Jarret J.T. and Lansbury P.T. Jr. Seeding "one dimensional crystallization" of amyloid: a pathogenic mechanism in alzheimer's disease and scrapie? *Cell*, 73:1055–1058, 1993.

[56] Caughey B. and Lansbury P.T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Ann. Rev. Neurosci.*, 26:267–298, 2003.

[57] Zhang L., Shimoji M., Thomas B., Moore D.J., Yu S.W., Marupudi N.I., Torp R., Torgner I.A., Ottersen O.P., Dawson T.M., and Dawson V.L. Mitochondrial localization of the Parkinson's disease related protein DJ-1: implications for pathogenesis. *Hum. Mol. Genet.*, 14:2063–2073, 2005.

[58] Xu J., Zhong N., Wang H., Elias J.E., Kim C.Y., Woldman I., Pifl C., Gygi S.P., Geula C., and Yankner B.A. The Parkinson's disease-associated DJ-1 protein is a transcriptional co-activator that protects against neuronal apoptosis. *Hum. Mol. Genet.*, 14:1231–1241, 2005.

[59] Mythri R.B., Jagatha B., Pradhan N., Andersen J., and Bharath M.M. Mitochondrial complex i inhibition in Parkinson's disease: how can curcumin protect mitochondria? *Antioxid. Redox Signal.*, 9:399–408, 2007.

[60] Yokota T., Sugawara K., Ito K., Takahashi R., Ariga H., and Mizusawa H. Down regulation of DJ-1 enhances cell death by oxidative stress, er stress, and proteasome inhibition. *Biochem. Biophys. Res. Commun.*, 312:1342–1348, 2003.

[61] Takeda A., Hasegawa T., Matsuzaki-Kobayashi M, Sugeno N., Kikuchi A., Itoyama Y., and Furukawa K. Mechanisms of neuronal death in synucleinopathy. *J. Biomed. Biotechnol.*, 2006:19365, 2006.

[62] Uversky V.N. A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders. *J. Biomol. Struc. Dyn.*, 21:211–234, 2003.

[63] Hashimoto M., Rockenstein E., Mante M., Mallory M., and Masliah E. beta-synuclein inhibits alpha-synuclein aggregation: a possible role as an anti-parkinsonian factor. *Neuron*, 32:213–223, 2001.

[64] Fan Y., Limprasert P., Murray I.V., Smith A.C., Lee V.M., Trojanowski J.Q., Sopher B.L., and La Spada A.R. Beta-synuclein modulates alpha-synuclein neurotoxicity by reducing alpha-synuclein protein expression. *Hum. Mol. Genet.*, 15:3002–3011, 2006.

[65] Fink A.L. The aggregation and fibrillation of alpha-synuclein. *Acc. Chem. Res.*, 39:628–634, 2006.

[66] Ulmer T.S. and Bax A. Comparison of structure and dynamics of micelle-bound human alpha-synuclein and Parkinson disease variants. *J. Biol. Chem.*, 280:43179–43187, 2005.

[67] Bennett M.C. The role of alpha-synuclein in neurodegenerative diseases. *Pharmacol. ther.*, 105:311–331, 2005.

[68] Gandhi S. and Wood N.W. Molecular pathogenesis of Parkinson's disease. *Hum. Mol. Genet.*, 14:2749–2755, 2005.

[69] Fujiwara H., Hasegawa M., Dohmae N., Kawashima A., Masliah E., Goldberg M.S., Shen J., Takio K., and Iwatsubo T. alpha-synuclein is phosphorylated in synucleinopathy lesions. *Nat. Cell Biol.*, 4:160–164, 2002.

[70] Chen L., Cagniard B., Mathews T., Jones S., Koh H.C., Ding Y., Carvey P.M., Ling Z., Kang U.J., and Zhuang X. Age-dependent motor deficits and dopaminergic dysfunction in-1 null mice. *J. Biol. Chem.*, 280:21418–21426, 2005.

[71] Bisaglia M., Trolio A., Bellanda M., Bergantino E., Bubacco L., and Mammi S. Structure and topology of the non-amyloid-beta component fragment of human alpha-synuclein bound to micelles: implications for the aggregation process. *Protein Sci.*, 15:1408–1416, 2006.

[72] Ueda K., Fukushima H., Masliah E., Xia Y., Iwai A., Yoshimoto M., Otero D.A.C., Kondo J., Ihara Y., and Saitoh T. Molecular cloning of cdna encoding an unrecognised component of amyloid in alzheimers disease. *Proc. Natl. Acad. Sci. USA*, 90:1128211286., 1993.

[73] Li H.T., Du H.N., Tang L., Hu J., and Hu H.Y. Structural transformation and aggregation of human alpha-synuclein in trifluoroethanol: Non-amyloid component sequence is essential and beta-sheet formation is prerequisite to aggregation. *Biopolymers*, 64:221–226, 2002.

[74] Polymeropoulos M.H., Higgins J.J., Golbe L.I., Johnson W.G., Ide S.E., Di Iorio G., Sanges G., Stenroos E.S., Pho L.T., Schaffer A.A., Lazzarini A.M., Nussbaum R.L., and Duvoisin R.C. Mapping of a gene for Parkinsons disease to chromosome 4q21q23. *Science*, 274:11971199, 1996.

[75] Kruger R., Kuhn W., Muller T., Woitalla D., Graeber M., Kosel S., Przuntek H., Epplen J.T., Schols L., and Riess O. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nat. Genet.*, 18:106–108, 1998.

[76] Zarranz J.J., Alegre J., Gomez-Esteban J.C., Lezcano E., Ros R., Ampuero I., Vidal L., Hoenicka J., Rodriguez O., Atares B., Llorens V., Gomez T.E., del S.T., Munoz D.G., and de Yebenes J.G. The new mutation, e46k, of alpha-synuclein causes Parkinson and lewy body dementia. *Ann. Neurol.*, 55:64–173, 2004.

[77] Pals P., , Lincoln S., Manning J., Heckman M., Skipper L., Hulihan M., Van den Broeck M., De Pooter T., Cras P., Crook J., Van Broeckhoven C., and Farrer M.J. Alpha-Synuclein promoter confers susceptibility to Parkinson's disease. *Ann. Neurol.*, 56:591–595, 2004.

[78] Tsigelny I.F., Bar-On P., Sharikov Y., Crews L., Hashimoto M., Miller M.A., Keller S.H., Platoshyn O., Yuan J.X., and Masliah E. Dynamics of alpha-synuclein aggregation and inhibition of pore-like oligomer development by beta-synuclein. *FEBS J.*, 274:1862–1877, 2007.

[79] Conway K.A., Lee S.J., Rochet J.C., Ding T.T., Williamson R.E., and Lansbury P.T. Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. USA.*, 97:571–576, 2000.

[80] Ding T.T., Lee S.J., Rochet J.C., and Lansbury P.T.Jr. Annular alpha-synuclein protofibrils are produced when spherical protofibrils are incubated in solution or bound to brain-derived membranes. *Biochemistry*, 41:10209–10217, 2002.

[81] Conway K.A., Harper J.D., and Lansbury P.T. Accelerated in vitro fibril formation by a mutant alpha-synuclein linked to earlyonset Parkinson disease. *Nat. Med.*, 4:1318–1320, 1998.

[82] Li J., Uversky V.N., and Fink A.L. Effect of familial Parkinsons disease point mutations a30p and a53t on the structural properties, aggregation, and fibrillation of human alpha-synuclein. *Biochemistry*, 40:604–613, 2001.

[83] Lansbury PT Jr. Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl. Acad. Sci. USA*, 96:33423344, 1999.

[84] Goldberg M.S. and Lansbury P.T. Is there a cause-and-effect relationship between alpha-synuclein fibrillization and Parkinson's disease? *Nat. Cell Biol.*, 2:115–119, 2000.

[85] Volles M.J., Lee S.J., Rochet J.C., Shtilerman M.D., Ding T.T., Kessler J.C., and Lansbury P.T. Jr. Vesicle permeabilization by protofibrillar alpha-synuclein: implications for the pathogenesis and treatment of Parkinson's disease. *Biochemistry*, 40:7812–7819, 2001.

[86] Volles M.J. and Lansbury P.T. Jr. Vesicle permeabilization by protofibrillar alpha-synuclein is sensitive to Parkinsons diseaselinked mutations and occurs by a pore-like mechanism. *Biochemistry*, 41:45954602, 2002.

[87] Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.*, 4:49–60, 2003.

[88] Li J., Zhu M., Manning-Bong A.B., Di Monte D.A., and Fink A.L. Dopamine and-dopa disaggregate amyloid fibrils: implications for Parkinson's and Alzheimer's disease. *FEBS J.*, 18:962–964, 2004.

[89] Rasia R.M., Bertoncini C.W., Marsh D., Hoyer W., Cherny D., Zweckstetter M., Griesinger C., Jovin T.M., and Fernandez C.O. Structural characterization of copper) binding to alpha-synuclein: Insights into the bioinorganic chemistry of Parkinson's disease. *Proc. Natl. Acad. Sci. USA.*, 102:4294–4299, 2005.

[90] Cole N.B., Murphy D.D., Lebowitz J., Di Noto L., Levine R.L., and Nussbaum R.L. Metal-catalyzed oxidation of alpha-synuclein: helping to define the relationship between oligomers, protofibrils, and filaments. *J. Biol. Chem.*, 280:9678–9690, 2005.

[91] Ischiropoulos H. Oxidative modifications of alpha-synuclein. *Ann. N. Y. Acad. Sci.*, 991:93–100, 2003.

[92] Xu J., Kao S.Y., Lee F.J., Song W., Jin L.W., and Yankner B.A. Dopamine-dependent neurotoxicity of alpha-synuclein: a mechanism for selective neurodegeneration in Parkinson disease. *Nat. Med.*, 8:600–606, 2002.

[93] Lee F.J., Liu F., Pristupa Z.B., and Niznik H.B. Direct binding and functional coupling of alpha-synuclein to the dopamine transporters accelerate dopamine-induced apoptosis. *FEBS J.*, 15:916–926, 2001.

[94] Perez R.G., Waymire J.C., Lin E., Liu J.J., Guo F., and Zigmond M.J. The role for alpha-synuclein in the regulation of dopamine biosynthesis. *J. Neurosci.*, 22:3090–3099, 2002.

[95] Jenco J.M., Rawlingson A., Daniels B., and Morris A.J. Regulation of phospholipase d2: selective inhibition of mammalian phospholipase isoenzymes by alpha- and beta-synucleins. *Biochemistry*, 37:4901–4909, 1998.

[96] Lotharius J, Barg S., Wiekop P., Lundberg C., Raymon H.K., and Brundin P. Effect of mutant alpha-synuclein on dopamine homeostasis in a new human mesencephalic cell line. *J. Biol. Chem.*, 277:38884–38894, 2002.

[97] Mazzulli J.R., Mishizen A.J., Giasson B.I., Lynch D.R., Thomas S.A., Nakashima A., Nagatsu T., Ota A., and Ischiropoulos H. Cytosolic catechols inhibit alpha-synuclein aggregation and facilitate the formation of

intracellular soluble oligomeric intermediates. *J. Neurosci.*, 26:10068–10078, 2006.

[98] Conway K.A., Rochet J.C., Bieganski R.M., and Lansbury P.T.Jr. Kinetic stabilization of the alpha-synuclein protofibril by a dopamine-alpha-synuclein adduct. *Science*, 294:1346–1349, 2001.

[99] Li H.T., Lin D.H., Luo X.Y., Zhang F., Ji L.N., Du H.N., Song G.Q., Hu J., Zhou J.W., and Hu H.Y. Inhibition of alpha-synuclein fibrillization by dopamine analogs via reaction with the amino groups of alpha-synuclein. implication for dopaminergic neurodegeneration. *FEBS J.*, 272:3661–3672, 2005.

[100] Bisaglia M., Mammi S., and Bubacco L. Kinetic and structural analysis of the early oxidation products of dopamine: Analysis of the interactions with alpha-synuclein. *J. Biol. Chem.*, 282:15597–15605, 2007.

[101] Rochet J.C., Outeiro T.F., Conway K.A., Ding T.T., Volles M.J., Lashuel H.A., Bieganski R.M., Lindquist S.L., and Lansbury P.T. Interactions among alpha-synuclein, dopamine, and biomembranes: some clues for understanding neurodegeneration in Parkinson's disease. *J. Mol. Neurosci.*, 23:23–34, 2004.

[102] Wilson M.A., Collins J.L., Hod Y., Ringe D., and Petsko G.A. The 1.1 å resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease. *Proc. Natl. Acad. Sci. USA.*, 100:9256–9261, 2003.

[103] Le Naour F., Misek D.E., Krause M.C., Deneux L., Giordano T.J., Scholl S., and Hanson M. Proteomics-based identification of DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin. Cancer Res.*, 7:3328–3335, 2001.

[104] Allard L., Burkhard P.R., Lescuyer P., Burgess J.A., Walter N., Hochstrasser D.F., and Sanchez J.C. Park7 and nucleoside diphosphate kinase a as plasma markers for the early diagnosis of stroke. *Proteomics and Protein Markers*, 51:2043–2051, 2005.

[105] Inden M., Taira T., Kitamura Y., Yanagida T., Tsuchiya D., Takata K., Yanagisawa D., Nishimura K., Tanaguchi T., Kiso Y., Yoshimoto K., Agatsuma T., Koide-Yoshida S., Iguchi-Ariga S.M., and Shimohama S.Ariga H. Park7-1 protects against degeneration of nigral dopaminergic neurons in Parkinson's disease rat model. *Neurobiol. Dis.*, 24:144–158, 2006.

[106] Waragai M., Wei J., Fujita M., Nakai M., Ho G.J., Masliah E., Akatsu H., Yamada T., and Hashimoto M. Increased level of DJ-1 in the cerebrospinal fluids of sporadic Parkinson's disease. *Biochem. Biophys. Res. Commun.*, 345:967–972, 2006.

[107] Taira T., Saito Y., Niki E., Iguchi-Ariga S.M., Takahashi K., and Ariga H. DJ-1 has a role in antioxidative stress to prevent cell death. *EMBO Rep*, 5:213–218, 2007.

[108] Bandopadhyay R., Kingsbury A.E., Cookson M.R., Reid A.R., Evans I.M., Hope A.D., Pittman A.M., Lashley T., Canet-Aviles R., Miller D.W., McLendon C., Strand C., Leonard A.J., Abou-Sleiman P.M., Healy D.G., Ariga H., Wood Nicholas W., de Silva R., Revesz T., Hardy J.A., and Lees A.J. The expression of-1 (park7) in normal human cns and idiopathic Parkinson's disease. *Brain*, 127:420–430, 2004.

[109] Choi J., Sullards M.C., Olzmann J.A., Rees H.D., Weintraub S.T., Bosrwick D.E., Gearing M., Levey A.I., Chin L.S., and Li L. Oxidative damage of DJ-1 is linked to sporadic Parkinson and Alzheimer diseases. *J. Biol. Chem.*, 281:10816–10824, 2006.

[110] Moore D.J., West A.B., and Dawson T.M. Molecular pathophysiology of Parkinson's disease. *Ann. Rev. Neurosci.*, 28:57–87, 2005.

[111] Olzmann J.A., Brown K., Wilkinson K.D., Rees H.D., Huai Q., Ke H., Levey A.I., Li L., and Chin L.S. Familial Parkinson's disease-associated L166P mutation disrupts DJ-1 protein folding and function. *J. Biol. Chem.*, 279:8506–8515, 2004.

[112] Neumann M., Muller V., Gorner K., Kretzschmar H.A., Haass C., and Kahle P.J. Pathological properties of the Parkinson's disease-associated protein DJ-1 in alpha-synucleinopathies and tauopathies: relevance for multiple system atrophy and pick's disease. *Acta Neuropathol.*, 107:489–486, 2004.

[113] Rizzu P., Hinkle D.A., Zhukareva V., Bonifati V., Severijnen L.A., Martinez D., Ravid R., Kamphorst W., Eberwine J.H., Lee V.M., Trojanowski J.Q., and Heutink P. DJ-1 colocalizes with tau inclusions: a link between parkinsonism and dementia. *Ann. Neurol.*, 55:113–118, 2004.

[114] Hague S.M., Rogaeva E., Hernandez D., Gulick C., Singleton A., Hanson M., Johnson J., Weiser R., Gallardo M., Ravina B., Gwinn-Hardy K., Crawley A., St George-Hyslop P.H., Lang A.E., Heutink P., Bonifati V., Hardy J.A., and Singleton A. Early-onset Parkinson's disease caused by a compound heterozygous DJ-1 mutation. *Ann. Neurol.*, 54:271–274, 2003.

[115] Abou-Sleiman P.M., Healy D.G., Quinn N., Lees A.J., and Wood N.W. The role of pathogenic DJ-1 mutations in Parkinson's disease. *Ann. Neurol.*, 54:283–286, 2003.

[116] Honbou K., Suzuki N.N., Horiuhci M., Niki T., Taira T., Ariga H., and Inagaki F. The crystal structure of DJ-1, a protein related to male fertility and Parkinson's Disease. *J. Biol. Chem.*, 278:31380–31384, 2003.

[117] Kinumi T., Kimata J., Taira T., Ariga H., and Niki E. Cysteine-106 of DJ-1 is the most sensitive cysteine residue to hydrogen peroxide-mediated oxidation in vivo in human umbilical vein endothelial cells. *Biochem. Biophys. Res. Commun.*, 317:722–728, 2004.

[118] Canet-Aviles R.M., Wilson M.A., Miller D.W., Ahmad R., McLendon C., Bandyopadhyay S., Baptista M., Ringe D., Petsko G.A., and Cookson M.R. The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization. *Proc. Natl. Acad. Sci. USA.*, 101:9103–9108, 2004.

[119] Macedo M.G., Anar B., Bronner I.F., Cannella M., Squitieri F., Bonifati V., Hoogeveen A., Heutink P., and Rizzu P. The DJ-1 L166P mutant protein associated with early onset Parkinson's disease is unstable and forms higher-order protein complexes. *Hum. Mol. Genet.*, 12:2807–2816, 2003.

[120] Moore D.J., Zhang L., Troncoso J., Lee M.K., Hattori N., Mizuno Y., Dawson T.M., and Dawson V.L. Association of DJ-1 and parkin mediated by pathogenic DJ-1 mutations and oxidative stress. *Hum. Mol. Genet.*, 14:71–84, 2005.

[121] Baulac S., LaVole M.J., Strahle J., Schiossmacher M.G., and Xia W. Dimerization of Parkinson's disease-causing DJ-1 and formation of high molecular weight complexes in human brain. *Mol. Cell. Neurosci.*, 27:236–246, 2004.

[122] Gorner K., Holtorf E., Waak J., Pham T.T., Vogt-Weisenhorn D.M., Wurst W., Haass C., and Kahle P.J. Structural determinants of the terminal Helix-Kink-Helix motif essential for protein stability and survival-promoting activity of DJ-1. *J. Biol. Chem.*, Epub, 2007.

[123] Gorner K., Holtorf E., Odoy S., Nuscher B., Yamamoto A., Regula J.T., Beyer K., Haass C., and Kahle P.J. Differential effects of Parkinson's disease-associated mutations on stability and folding of DJ-1. *J. Biol. Chem.*, 279:6943–6951, 2004.

[124] Takahashi K., Niki T., Taira T., Iguchi-Ariga S.M., and Ariga H. Reduced anti-oxidative stress activities of DJ-1 mutants found in Parkinson's disease patients. *Biochem. Biophys. Res. Commun.*, 320:389–397, 2004.

[125] Blackinton J., Ahmad R., Miller D.W., van der Brug M.P., Canet-Aviles R.M., Hague S.M., Kaleem M., and Cookson M.R. Effects of DJ-1 mutations and polymorphisms on protein stability and subcellular localization. *Mol. Brain Res.*, 134:76–83, 2005.

[126] Bonifati V., Oostra B.A., and Heutink P. Linking DJ-1 to neurodegeneration offers novel insights for understanding the pathogenesis of Parkinson's disease. *J. Mol. Med.*, 82:163–174, 2004.

[127] Miller D.W., Ahmad R., Hague S., Baptista M., Canet-Aviles R., McLendon C., Carter D.M., Stadler J., Chandran J., Klinefelter G.R., Blackstone C., and Cookson M.R. L166P mutant DJ-1, causative for recessive Parkinson's disease, is degraded through the ubiquitin-proteasome system. *J. Biol. Chem.*, 278:36588–36595, 2003.

[128] MacQueen J.B. *Some Methods for classification and Analysis of Multivariate Observations.* University of California Press, Berkeley, 1967.

[129] Daura X., Gademann K., Jaun B., Seebach D., van Gunsteren W.F., and Mark A.E. Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.*, 38:236–240, 1999.

[130] Micheletti C., Seno F., and Maritan A. Recurrent oligomers in proteins - an optimal scheme reconciling accurate and concise backbone representations in automated folding and design studies. *Proteins: Str. Func. and Genet.*, 40:662–674, 2000.

[131] Kelley L.A., Gardner S.P., and Sutcliffe M.J. An automated approach for clustering an ensemble of nmr-derived protein structures into conformationally related subfamilies. *Protein Eng.*, 9:1063–1065, 1996.

[132] Bezdek J.C. *Pattern Recognition with Fuzzy Objective Function Algoritms.* Plenum Press, New York, 1981.

[133] Zaiane O.R. Data clustering. In *Principles of Knowledge Discovery in Databases.*, 1999.

[134] Sutcliffe M.J. Representing an ensemble of-derived protein structures by a single structure. *Protein Sci.*, 2:936–944, 1993.

[135] Kortemme T. and Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA.*, 99:14116–14121, 2002.

[136] Kortemme T., Kim D.E., and Baker D. Computational alanine scanning of protein-protein interfaces. *Sci.STKE*, 219:1–8, 2004.

[137] OpenEye Scientific Software, Inc. *ROCS manual, version 2.2.*

[138] Grant J.A., Gallardo M.A., and Pickup B.T. A fast method of molecular shape comparison: A simple application of a gaussian description of molecular shape. *J. Comp. Chem.*, 17:16531666, 1996.

[139] OpenEye Scientific Software, Inc. *EON manual, version 1.1.*

[140] Jakalian A., Jack D.B., and Bayly C.I. Fast, efficient generation of high-quality atomic charges. am1bcc model:. parametrization and validation. *J. Comp. Chem.*, 23:1623–1641, 2002.

[141] Bayly C.I., Cieplak P., Cornell W.D., and Kollman P.A. Well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: The model. *J. Phys. Chem.*, 97:10269, 1993.

[142] Bayly C.I., Cieplak P., Cornell W.D., and Kollman P.A. Application of charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.*, 115:9620, 1993.

[143] OpenEye Scientific Software, Inc. *FILTER. Database preparation and Property calculation. Manual, version 2.0.1.*

[144] Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., and Olson A.J. Automated docking using a lamarckian genetic algorithm and and empirical binding free energy function. *J. Comput. Chem.*, 19:1639–1662, 1998.

[145] Huey R., Morris G.M., Olson A.J., and Goodsell D.S. A semiempirical free energy force field with charge-based desolvation. *J. Comp. Chem.*, 28:1145–1152, 2007.

[146] Solis F.J. and Wets R.J.-B. Minimization by random search techniques. *Math. Oper. Res.*, 6:19–30, 1981.

[147] Wesson L. and Eisemberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.*, 1:227–235, 1992.

[148] Garret M., Goodsell D.S., Huey R., Hart W.E., Halliday S., Belew R., and Olson A.J. *Autodock 3.0.5 user's guide: automated docking of flexible ligands to receptors.*

[149] Stouten P.F.W., Frommel C., Nakamura H., and Sander C. An effective solvation term based on atomic accupancies for use in protein simulations. *Mol. Simul.*, 10:97, 1993.

[150] Jones G., Willet P., and Glen R.C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 245:43–53, 1995.

[151] Jones G., Willet P., Glen R.C., Leach A.R., and Taylor R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267:727–748, 1997.

[152] Feyman R.P., Leighton R.B., and Sands M. *The Feyman Lectures on Physics.* Addison-Wesley, 1963.

[153] van Gunsterren W.F. and Berendsen H.J.C.A. The leap-frog algorithm for stochastic dynamics. *Mol. Simul.*, 1:173–185, 1998.

[154] Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., and Schulten K. Scalable molecular dynamics with namd. *J. Comput. Chem.*, 26:1781–1802, 2005.

[155] Goldstein H. *Classical Mechanics*, volume second. Addison Wesley, Reading, Massachusetts, 1980.

[156] Deleeuw S.W. and Perram J.W. Hamilton equations for constrained dynamics systems. *J. Stat. Phys*, 61:1203–1222, 1990.

[157] Nose S. molecular dynamics method for simulation in the canonical ensemble. *Mol. Physics*, 52:255–268, 1984.

[158] Evans D.J. and Holian B.L. The nose-hoover thermostat. *J. Chem. Phys.*, 83:4069–4074, 1985.

[159] Feller S.E., Zhang Y., Pastor R.W., and Brooks B.R. Constant pressure molecular dynamics simulation: the langevin piston method. *J. Chem. Phys.*, 103:4613–4621, 1995.

[160] Adelman S.A and Doll J.D. Generalized langevin equation approach for atom-solid-surface scattering - general formulation for classical scattering off harmonic solids. *J. Chem. Phys.*, 64:2375–2388, 1976.

[161] Martyna G., Tobias D., and Klein M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101:4977–4981, 1994.

[162] Fogolari F, Brigo A., and Molinaru H. The poisson-boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.*, 15:377392, 2002.

[163] Baker N.A., Sept D., Joseph S., Holst M.J., and McCammon J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA.*, 98:10037–10041, 2001.

[164] Corona-Avendano S., Alarcon-Angeles G., Rosquete-Pina G.A., Rojas-Hernandez A., Gutierrez A., Ramirez-Silva M.T., Romero-Romo M., and Palomar-Pardave M. New insights on the nature of chemical species involved during the process of dopamine deprotonation in aqueous solution: theoretical and experimental study. *J. Phys. Chem. B*, 111:1640–1647, 2007.

[165] Hooft R.W.W., Vriend G., Sander C., and Abola E.E. Errors in protein structures. *Nature*, 381:272, 1996.

[166] Ponder J.W. and Case D.A. Force field for protein simulations. *Ann. Rev. Biochem.*, 66:27–85, 2003.

[167] Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., and Klein M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[168] Frisch M.J, Trucks G.W., Schlegel H.B., Scuseria G.E., Robb M.A., Cheeseman J.R., Montgomery Jr. J.A., Vreven T., Kudin K.N., Burant J.C., Millam J.M., Iyengar S.S., Tomasi J., Barone V., Mennucci B., Cossi M., Scalmani G., Rega N., Petersson G.A., Nakatsuji H., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Klene M., Li X., Knox J.E., Hratchian H.P., Cross J.B., Bakken V., Adamo C., Jaramillo J., Gomperts R., Stratmann R.E., Yazyev O., Austin A.J., Cammi R., Pomelli C., Ochterski J.W., Ayala P.Y., Morokuma K., Voth G.A., Salvador P., Dannenberg J.J., Zakrzewski V.G., Dapprich S., Daniels A.D., Strain M.C., Farkas O., Malick D.K., Rabuck A.D., Raghavachari K., Foresman J.B., Ortiz J.V., Cui Q., Baboul A.G., Clifford S., Cioslowski J., Stefanov B.B., Liu G., Liashenko A., Piskorz P., Komaromi I., Martin R.L., Fox D.J., Keith T., Al-Laham M. A., Peng C.Y., Nanayakkara A., Challacombe M., Gill P.M.W., Johnson B., Chen W., Wong M.W., Gonzalez C., and Pople J. A. *Gaussian 98*. Gaussian, Inc., Pittsburgh P.A., 2001.

[169] Ryckaert J.P., Ciccotti G., and Berendsen H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23:327–341, 1977.

[170] Willett P. Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, 38:983–996, 1998.

[171] Guidoni L., Torre V., and Carloni P. Water potassium dynamics inside the Kcsa K+ channel. *FEBS Letters*, 477:32–42, 2000.

[172] von Grotthuss M., Koczyk G., Pas J., Wyrwicz L.S., and Rychlewski L. Ligand.info small-molecule meta-database. *Comb. Chem. High Throughput Screen.*, 7:757–761, 2004.

[173] Hedrich K., Schafer N., Hering R., Hagenah J., Lanthaler A.J., Schwinger E., Kramer P.L., Ozellus L.J., Bressman S.B., Abbruzzese G., Martinelli P., Kostic V., Pramstaller P.P., Vieregge P., Taanman J.W., Riess O., and Klein C. The R98Q variation in DJ-1 represents a rare polymorphism. *Ann. Neurol.*, 55:145–146, 2004.

[174] Laskowski R.A., MacArthur M.W., Moss D.S., and Thornton J.M. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26:283–291, 1993.

[175] Morris A.L., MacArthur M.W., Hutchinson E.G., and Thornton J.M. Stereochemical quality of protein structure coordinates. *Proteins*, 12:345–364, 2007.

[176] Case D.A., Pearlman D.A., Caldwell J.W., Cheatham III T.E., Wang J., Ross W.S., Simmerling L., Darden T.A., Merz K.M., Cheng A.L., Vincent J.J., Crow ley M., Tsui V., Gohlke H., Radmer R.J., Duan Y., Pitera J., Massova I., Seibel G.L., Singh U.C., Weiner P.K., and Kollman P.A. *8*. University of California, San Francisco, 2002.

[177] Case D.A., Cheatham III T.E., Darden T., Gohlke H., Luo R., Merz K.M., Onufriev A., Simmerling C., Wang B., and Woods R.J. The biomolecular simulation programs. *J. Comput. Chem.*, 26:1668–1688, 2005.

[178] Wang J., Cieplak P., and Kollman P.A. How well does a restrained electrostatic potential) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.*, 21:1049–1074, 2000.

[179] Darden T., York D., and Pedersen L.G. Particle mesh ewald: an n.log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[180] Essman U., Perela L., Berkowitz M.L., Darden T., Lee H., and Pedersen L.G. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8592, 1995.

[181] Sagui C. and Darden T. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.*, 28:155–179, 1999.

[182] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant-pressure molecular-dynamics algorithms. *J. Chem. Phys.*, 101:4177–4189, 1994.

[183] Berendsen H.J.C., van-derSpoel D., and van-Drunen R. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.

[184] Lindhal E., Hess B., and van der Spoel D. Gromacs 3.0: package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7:306–317, 2001.

[185] Humphrey W., Dalke A., and Schulten K. Vmd - visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.

[186] Connolly M.L. Analytical molecular surface calculation. *J. Appl. Crystallogr.*, 16:548–558, 1983.

[187] Janin J. Surface and inside volumes in globular proteins. *Nature*, 277:491–492, 1979.

[188] Wolfenden R., Andersson P., Cullis P., and Southgate C. Affinities of amino acid side chains for solvent water. *Biochemistry*, 157:105–132, 1982.

[189] Ariga H., Honbou K., Suzuki N.N., Horiuchi M., Niki T., Taira T., Ariga H., and Inagaki F. The crystal structure of DJ-1, a protein related to male fertility and Parkinson's Disease. *J. Biol. Chem.*, 278:31380–31384, 2003.

[190] Lakshminarasimhan M., Maldonado M.T., Zhou W., Fink A.L., and Wilson M.A. Structural impact of three Parkinsonism-associated missense mutations on human DJ-1. *Biochemistry*, 47:1381–1392, 2008.

[191] Pantano S., Zaccolo M., and Carloni P. Molecular basis of the allosteric mechanism of cAMP in the regulatory PKA subunit. *FEBS Letters*, 579:2679–2685, 2005.

[192] Caudle W.M., Richardson J.R., Wang M.Z., Taylor T.N., Guillot T.S., McCormack A.L., Colebrooke R.E., Di Monte D.A., Emson P.C., and Miller G.W. Reduced vesicular storage of dopamine causes progressive nigrostriatal neurodegeneration. *J. Neurosci.*, 27:8138–8148, 2007.

[193] Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., and Mee R.P. Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.*, 11:425–445, 1997.

[194] Baxter C.A., Murray C.W., Clark D.E., Westhead D.R., and Eldridge M.D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins*, 33:367–382, 1998.

[195] Ewald P. Die berechnug optischer und elektrostatischer gitterpotentiale. *Ann. Phys.*, 64:253–287, 1921.

[196] Madelung E. Das elektrische feld in systemen von regelmassig angeordneten punktladungen. *Phys. Z.*, 19:524–532, 1918.

[197] Frenkel D. and Smit B. *Understanding Molecular Simulations: From Algorithms to Applications.* Academic Press, San Diego, 2002.

[198] Press W.H., Flannery B.P., Teukolsky S.A., and Vetterling W.T. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 1986.