Scuola Internazionale Superiore di Studi Avanzati - Trieste

Condensed Matter Theory Sector

Nuclear Magnetic Resonance Study of Complex Molecular Crystals From First Principles: Case of Cholesterol

Thesis submitted for the degree of Doctor Philosophiæ
Academic Year 2010/2011

CANDIDATE
Emine Kucukbenli

SUPERVISOR
Prof. Stefano de Gironcoli

October 2011

SISSA - Via Bonomea 265 - 34136 TRIESTE - ITALY

International School for Advanced Studies

Condensed Matter Theory Sector

Nuclear Magnetic Resonance Study of Complex Molecular Crystals From First Principles: Case of Cholesterol

Thesis submitted for the degree of Doctor Philosophiæ
Academic Year 2010/2011

CANDIDATE
Emine Kucukbenli

SUPERVISOR
Prof. Stefano de Gironcoli

October 2011

SISSA - Via Bonomea 265, 34136 Trieste - ITALY

Contents

Chapter 1

Introduction

Nuclear Magnetic Resonance (NMR) is a powerful technique in structure determination and it is widely used in several fields ranging from solid state physics to molecular biology. Earlier NMR experiments were mostly focused on spectral features of molecules in solution, i.e. solution-state NMR spectroscopy. Recently, thanks to the developments in many experimental aspects, from hardware [1] to pulse sequences [2], solid-state NMR (SSNMR) is also commonly used. This method became an important tool in the field of crystallography, alongside other spectral techniques such as X-ray diffraction, infrared spectroscopy and scanning electron microscopy. The advantage of SSNMR over these other methods becomes apparent in certain fields such as medical NMR: X-ray diffraction analysis requires fine crystals which are often not available for biological samples, infrared spectroscopy may not be able to detect all components of a complex biological material in a single experiment without ambiguity. SSNMR instead is rapid, non-invasive and highly reproducible and does not require fine crystals, therefore is highly applicable to organic systems and in particular to diagnostic medicine.

Our study is motivated by one such application: a recent SSNMR experiment [3] reports that different pathologies of gallbladder result in gallstones with distinct NMR features. If the structures giving rise to these spectra could be identified using this relationship, important information on the growth of gallstones and pathogenesis of gallbladder diseases could be revealed.

The gallstones investigated in Ref. [3] are mostly made of cholesterol crystals. Cholesterol crystals, in general, are of great interest in biomedical studies as they play an important role in human diseases such as atherosclerosis [4] and, as mentioned, gallstone formation [5]. Crystallographically, cholesterol crystals are challenging systems: In crystal forms, much like in membranes, cholesterol molecules form a bilayer structure, where the molecules are ar-

ranged in a head-to-head fashion. These crystals also tend to have several molecules in the asymmetric unit cell, increasing their complexity, which is the reason why X-ray crystallographers have spent decades for an unambiguous structure determination of known cholesterol crystal phases [6, 7, 8, 9].

In terms of establishing a relationship between an NMR spectrum and the underlying structure, there is no generally valid 'recipe'. Although for small systems one can make reference to empirical databases to interpret the measured spectra, for large systems such as cholesterol crystals more reliable and predictive tools are required for an unambiguous mapping of spectra to the correct structure.

In this direction, several methods have been developed to determine the NMR spectrum from first principles. Among them, quantum chemistry methods have been extensively used for the interpretation of solution-state NMR spectra, as they are limited to studies of molecules and finite systems. One can also study solids by quantum chemistry methods with extrapolations from clusters, however this approach may not represent the long-range interactions accurately enough for NMR calculations and is limited to small molecules. For periodic systems, a solid-state physics approach, Gauge-Including Projector-Augmented Wave (GIPAW) method, was proposed in 2001 [10] and has been widely employed since. It has proved to successfully reproduce experimental solid-state NMR chemical shifts in a wide variety of systems [11, 12, 13].

The GIPAW method enables calculation of electronic response to a finite magnetic field within the framework of Density Functional Perturbation Theory (DFPT) [14]. It is based on Blochl's projector augmented wave (PAW) [15] formalism to correctly reconstruct the all-electron valance electron density within the core region of pseudopotentials. It was originally implemented using norm-conserving pseudopotentials [10] and later extended to Vanderbilt's ultrasoft [16] pseudopotentials [17]. For this study, we extended the implementation of GIPAW method to PAW pseudopotentials in the Quantum Espresso simulation package (QE) [18]. By doing so, we keep the level of theory consistent with all-electron accuracy throughout our calculations with minimal computational cost. This new implementation enables us to study systems of on the order of 1000 atoms, such as cholesterol crystals which, to our knowledge, are the largest systems examined with GIPAW method to date.

In this work we obtain the NMR chemical shifts of the three known phases of cholesterol crystals from first principles and compare our results with the experimental findings. Furthermore we combine spectral editing experiments with our theoretical predictions to distinguish the source nuclei of each signal in cholesterol spectra. For high accuracy, we account for the van der

Waals interactions using a recent implementation of van der Waals-aware functionals in QE.

The layout of this thesis is as follows: In Chapter 2 we present the theoretical background relevant to our study, namely NMR, Density Functional Theory (DFT), PAW and GIPAW. We also report the performed computational tests that validate our implementation. In Chapter 3 we first give the experimentally determined structural properties of cholesterol crystals for all known phases and compare our theoretical results with these reports. In this chapter we also examine the adequacy of classical force field calculations in structure determination through comparisons with both *ab initio* results and experiment. In Chapter 4 we report our results for NMR spectrum of all known phases and comment on the accuracy of our calculations. Factors affecting the *ab initio* NMR calculations are investigated in detail, such as the impact of structural optimization and exchange-correlation functionals. A systematic error observed in GIPAW calculations is also discussed and a possible correction is proposed. We further test the validity of the introduced correction using the results of spectral editing experiments. Finally using all the information gathered, we perform peak assignment for the observed NMR spectra. We also compare the NMR spectrum obtained from molecules and crystals to draw conclusions on the intermolecular interactions present in cholesterol crystals.

Chapter 2

ab initio Solid State NMR

In this chapter, we will briefly review the theoretical background underlying the computational methods used in the present work. First a general summary on some theoretical and experimental Nuclear Magnetic Resonance (NMR) concepts are given. Later we give some fundamental information on the theories our work is based on, namely Density Functional Theory (DFT) and Projector-Augmented Wave (PAW) formalism.

In section 2.4 we outline the main characteristics of Gauge-Including Projector-Augmented Wave (GIPAW) method, which enables us to obtain NMR parameters for a periodic system. Then we present a series of computational tests validating our implementation of GIPAW in Quantum Espresso *(QE) package.*

2.1 Nuclear Magnetic Resonance

In this section we give a basic introduction to chemical shifts within the framework of Nuclear Magnetic Resonance (NMR). NMR theory is vast, from relativistic interactions to the effects of experimental techniques on spin dynamics. In this introduction we will limit ourselves only to the concepts relevant to our studies:

Nucleus in a static uniform field

The Hamiltonian of an isolated spin in a static uniform magnetic field $\mathbf{B}_0$, without considering any other interaction is simply the Zeeman Hamiltonian:

$$H = \boldsymbol{\mu} \cdot \mathbf{B}_0 \tag{2.1}$$

where $\boldsymbol{\mu}$ is the nuclear magnetic moment operator:

$$\boldsymbol{\mu} = \gamma \hbar \mathbf{I} \tag{2.2}$$

with γ being the nuclear gyromagnetic ratio and $\mathbf{I}$ the nuclear spin operator. The direction of the field, which can be chosen as $\hat{z}$, gives the quantization axis. Therefore the Hamiltonian H is proportional to the operator I_z and the eigenfunctions of I_z, which can be labeled as $|I, m\rangle$, are also the eigenfunctions of the Hamiltonian:

$$H\,|I, m\rangle = E_{I,m}\,|I, m\rangle \tag{2.3}$$

with eigenvalues $E_{I,m} = -\gamma\hbar B_0 m$ using $I_z\,|I, m\rangle = m\,|I, m\rangle$.

Trivially, for a spin with $I = \frac{1}{2}$, $m = \pm\frac{1}{2}$, there are two possible eigenstates with energies $E_{\pm\frac{1}{2}} = \pm\frac{1}{2}\gamma\hbar B_0$, which are commonly called Zeeman states.

These two eigenstates have different energies so that any superposition state non-parallel to the eigenstates will have a net time evolution, which results in a precession with the frequency proportional to the energy difference between the eigenstates:

$$\Delta E = \gamma\hbar B_0 = \hbar\omega_0 \tag{2.4}$$

This phenomenon is referred to as Larmor precession and the frequency ω_0 is called Larmor frequency. At a given magnetic field B_0, Larmor frequency can be used to distinguish different nuclei from each other as it depends on nuclear mass and charge through the gyromagnetic ratio γ.

In a macroscopic sample, the spin states are populated according to Boltzmann distribution, where population of a state ψ with energy E_ψ, at temperature T is:

$$p_\psi = \frac{e^{-E_\psi/kT}}{\sum_{\psi'} e^{-E_{\psi'}/kT}} \tag{2.5}$$

Therefore, in NMR experiments, only the ensemble averages are measured for any observable.

The effect of radio-frequency pulses

In an NMR experiment, a radio-frequency (rf) pulse introduces an oscillating magnetic field, $\mathbf{B}_1(t)$, therefore the Hamiltonian and its eigenstates become time dependent. For a field oscillating with frequency ω_{rf} along $\hat{x}$, the Hamiltonian now reads:

$$\begin{aligned} H &= -\gamma\hbar(I_z B_0 + I_x B_1 \cos(\omega_{rf} t)) \\ &= -\frac{\gamma\hbar}{2}(I_z B_0 + B_1 e^{-i\omega_{rf} t I_z} I_x e^{i\omega_{rf} t I_z}) + \text{c.c.} \end{aligned} \tag{2.6}$$

In equation 2.6, we have used the property that an alternating field can be seen as two fields rotating in opposite directions and, near resonance, only one of them has a significant effect on the spin dynamics.

To find a static solution for the above Hamiltonian, we move to a rotating frame, about $\mathbf{B}_0$ with angular frequency ω_{rf}. In the rotating frame the Hamiltonian becomes

$$H' = -\hbar((\gamma B_0 - \omega_{rf})I_z + \gamma B_1 I_x) \tag{2.7}$$

with eigenstates

$$\psi' = e^{-i\omega_{rf}tI_z}\psi \tag{2.8}$$

Through the above rotation, the rf field mixes the Zeeman states such that in the rotating frame, the moment acts as though it experiences an effective static field in the (x-z) plane and precesses. There is no net absorption of energy from this precession, as all the energy taken to tilt $\boldsymbol{\mu}$ from B_0 is returned in a complete cycle. This is why alternating fields suggest a simple way of observing the magnetic resonance in NMR experiments: For example, if we initially magnetize a sample along B_0 and then turn on the alternating field for a short time, we can expect that during precession the moment periodically will be lined up normal to B_0. Following the turn-off of the alternating field at that moment, the magnetic moment would remain at rest in the rotating frame, normal to B_0. The interaction of spins with their environment would then cause a decay (free induction decay) in the excess magnetization, through a precession with a frequency $\gamma\hbar B_0$ which can be measured as the resonance frequency.

Chemical shifts

So far we have only considered the interactions with external magnetic fields, and neglected other sources of magnetic fields internal to the sample. In NMR experiments the applied static field is, in general, orders of magnitude larger than the internal local fields, therefore its direction should still be taken as the axis of quantization, however local fields in the direction of the static field, or precessing in the plane perpendicular to the static field near the resonance frequency significantly affect the spin states. Some of the important sources of changes in spin states and effective field can be listed as chemical shielding effect, nuclear spin dipole-dipole coupling through space and electric quadrupole moment coupling. Now we will focus on the chemical shielding interaction.

If only the bare nucleus were to be considered, all carbons with the same nuclear mass would result in very similar nuclear magnetic resonance frequencies. However the electrons surrounding each nucleus react to the magnetic field as well and produce a secondary field, which changes the resonance frequency of each nucleus. Due to the response of the electrons, we observe different resonance frequencies for carbons in different chemical environments.

This shift in resonance frequency is called the 'chemical shift' as it provides information on the chemical environment of the nuclei.

The local field contributions from shielding interactions can be divided in two types. *i*) A diamagnetic contribution: primary induced current of electrons that produce a magnetic field opposing the applied field. *ii*) A paramagnetic contribution: The external magnetic field can mix the excited states with paramagnetic properties with the ground state, inducing paramagnetism, which would generate a field supporting the applied field.

Since the electron distribution around a current is not spherically symmetric, the shielding associated to a nucleus is given by a second rank tensor, denoted by $\boldsymbol{\sigma}$, and the perturbative Hamiltonian due to this effect can be written as

$$H_{cs} = -\gamma\hbar \mathbf{I} \cdot \boldsymbol{\sigma} \cdot \mathbf{B}_0 \tag{2.9}$$

where $\mathbf{I}$ is the spin of the nucleus of interest. For B_0 in $\hat{z}$ direction, the first order contribution to the energy of a spin level $|I, m\rangle$ from chemical shielding, E_{cs} is:

$$\begin{aligned} E_{cs} = &- \gamma\hbar\sigma_{zz} B_0 \langle I, m|\mathbf{I}_z|I, m\rangle \\ &- \gamma\hbar\sigma_{zz} B_0 m \end{aligned} \tag{2.10}$$

Note also that it is useful to symmetrize the shielding tensor as the symmetric component is the only one with a significant effect on the chemical shifts.

While describing $\boldsymbol{\sigma}$, it is possible to choose an axis frame different from the laboratory frame, such as the one where $\boldsymbol{\sigma}$ becomes diagonal. This axis frame is called the principal axis frame (PAF). The diagonal elements of $\boldsymbol{\sigma}$ in PAF are often expressed in terms of the isotropic value σ_{iso}, the anisotropy Δ and the asymmetry η, defined as follows:

$$\begin{aligned} \sigma_{iso} &= \frac{1}{3}(\sigma_{xx}^{PAF} + \sigma_{yy}^{PAF} + \sigma_{zz}^{PAF}) \\ \Delta &= \sigma_{zz}^{PAF} - \sigma_{iso} \\ \eta &= (\sigma_{xx}^{PAF} - \sigma_{yy}^{PAF})/\sigma_{zz}^{PAF} \end{aligned} \tag{2.11}$$

Using these definitions, for an axially symmetric shielding tensor and external field in $\hat{z}$, we can write the chemical shift frequency as a function of the polar angles (θ, ϕ) that define the orientation of the external magnetic field B_0 in the PAF:

$$\omega_{cs}(\theta, \phi) = -\omega_0\sigma_{iso} - \frac{1}{2}\omega_0\Delta\left[3\cos^2\theta - 1 + \eta\sin^2\theta\cos 2\phi\right] \tag{2.12}$$

where $-\omega_0\sigma_{iso}$ is the chemical shift frequency relative to a bare nucleus with Larmor frequency ω_0. In powder samples, the lineshape can be deduced from the chemical shift parameters and vice versa.

In NMR experiments, instead of the absolute frequency $\omega_0 + \omega_{cs}$, the frequency relative to a specific signal of a reference sample is measured. This relative shift is called the chemical shift, $\boldsymbol{\delta}$, and the isotropic chemical shift δ_{iso} is given as follows:

$$\delta_{iso} = \sigma_{iso}(ref) - \sigma_{iso} \tag{2.13}$$

Often in experiments the PAF is not known and the principal values of chemical shift tensor are labeled by convention as $\delta_{11}^{PAF} \geq \delta_{22}^{PAF} \geq \delta_{33}^{PAF}$. The chemical shift anisotropy Δ_{cs} and asymmetry η_{cs} are defined as:

$$\Delta_{cs} = \delta_{11}^{PAF} - \delta_{iso}\eta_{cs} = \frac{\delta_{33}^{PAF} - \delta_{22}^{PAF}}{\delta_{11}^{PAF}} \tag{2.14}$$

and the chemical shift that is measured in experiments can be written as

$$\delta = \delta_{iso} + \frac{1}{2}\Delta_{cs}\left[3\cos^2\theta - 1 + \eta_{cs}\sin^2\theta\cos 2\phi\right] \tag{2.15}$$

So far we have given a basic introduction to the chemical shielding interaction concept. A theoretical scheme to calculate the NMR chemical shielding tensor by perturbation theory is given in section 2.4.

Magic-angle spinning

In solution state NMR experiments, the effects due to chemical shift anisotropy and dipolar coupling are rarely observed. The angular dependencies of the quantities, i.e. $(3\cos^2\theta - 1)$ in equation 2.15 are averaged to zero due to the rapid tumbling of the molecules in solution and sharp peaks are observed. Magic-angle spinning (MAS) is a method that is devised to play the role of tumbling for solid samples. By spinning the sample around an axis at an angle $\theta_{MAS} = 54.74^\circ$ with respect to the external magnetic field at a rate higher than the anisotropy frequency, the measured shift corresponds to an average of the θ dependence such that $\langle 3\cos^2\theta - 1\rangle$ is zero and sharp peaks are, again, observed.

In addition to the angular dependence of the shielding tensor MAS experiments allow to average out the dipole-dipole magnetic interactions in the sample that would broaden the peaks. If the resonating nucleus has a spin larger than 1/2 additional couplings are present. The most important is quadrupolar coupling that depends on the electric quadrupole at the nucleus position and induces characteristic line shapes in powder MAS experiments in systems with large spin such as ^{27}Al (I=5/2) which will be discussed in Appendix B.

Cross polarization and spectral editing

Observing the resonance signals from isotopes with low abundance, such as ^{13}C is challenging. Due to their low abundance, signal to noise ratio is

generally poor and due to the lack of homonuclear dipolar couplings, relaxation times tend to be very long. Long relaxation times require long recovery time between scans of a sample, often in the order of minutes. To increase the already low signal to noise ratio one generally needs several thousand scans, and the time for the overall experiment becomes very long.

A technique to overcome this problem is cross polarization method. With a specific pulse sequence, it transfers the magnetization from a nearby abundant spin to the dilute spin. The experimental results in this study which were generously provided by the research group of Dr. Sinha in Centre of Centre of Biomedical Magnetic Resonance, Lucknow, India, uses cross polarization between ^{13}C and ^{1}H spins. The theory of cross polarization is rather lengthy to discuss here however it can be outlined as the following:

One can excite an abundant nucleus type which is in the vicinity of the low-abundance isotope, such as ^{1}H for ^{13}C in organic molecules. Using special pulse sequences, the spin state levels for the two types of nucleus can be aligned such that magnetization transfer from the ^{1}H to ^{13}C occurs. The interaction that is exploited to allow the magnetization transfer between these nuclei is the dipole-dipole coupling interaction.

The efficiency of the magnetization transfer from the abundant to the rare nuclei depends on a number of experimental details such as pulse sequence, length of the contact time interval and strength of the dipole-dipole interaction which depends on the chemical environment. Ingenious experimental setups can be devised so as to enhance/modify the recorded signals depending on the local environment. Such sophisticated experimental techniques go under the name of spectral editing and may provide valuable additional information as will be discussed in chapter 4.

2.2 Density Functional Theory

The study of condensed matter systems at the nano-scale requires both electrons and nuclei to be treated quantum mechanically as a part of a many body wavefunction of both electronic and ionic coordinates. The Born-Oppenheimer (also called adiabatic) approximation, introduced in 1927, allows to disentangle the ionic degrees of freedom from the electronic ones by considering that although the momenta of these particles are comparable, the ratio of the electronic mass to nuclei mass is very small ($m_e/M_I \leq 10^{-3}$). Therefore the time scale of ionic dynamics is much longer than the electronic one, which allows us to neglect the ionic kinetic energy when studying the electronic motion, so that we can solve the problem of N interacting electrons, considering fixed ions acting as an external potential V_{ext}.

Within the adiabatic approximation, the many-body Schrödinger equation that describes the quantum mechanical properties of a system of N_e electrons in the external potential of the nuclei reduces to the following:

$$H\psi_{\mathbf{R}}(\mathbf{r}) = E(\mathbf{R})\psi_{\mathbf{R}}(\mathbf{r}) \tag{2.16}$$

with the following Hamiltonian

$$H = -\sum_{i=1,N_e} \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I}\frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} \tag{2.17}$$

where $\mathbf{R}_I$ and $\mathbf{r}_i$ are coordinates of nuclei and electrons respectively, $Z_I e$ is the charge of the I-th nucleus and electron charge and mass are $-e$ and m_e respectively. $\psi_{\mathbf{R}}(\mathbf{r})$ is the many-body wavefunction of electrons at fixed nuclear positions $\mathbf{R}$. The first term describes the electronic kinetic energy, T_e, the second term is for electron-electron Coulomb interaction, U_{ee}, and the last term defines the external potential of nuclei V_{ext} and parametrically depends on nuclei positions. To this the direct ion-ion Coulomb interaction must also be added, which is just a constant term with respect to the electronic problem, and will be neglected in the following discussion.

The electronic wavefunction $\psi_{\mathbf{R}}(\mathbf{r})$ is a function of all $3N_e$ electronic degrees of freedom supplemented by the spin degrees of freedom. Therefore the complexity of equation 2.16 increases exponentially with the number of electrons and solving it to obtain the ground state wavefunction becomes an intractable problem even for nano-scale systems.

A numerically efficient and often successful approach to solve this problem is density-functional theory (DFT), which is also adopted throughout this thesis. DFT was introduced in 1964 by Hohenberg and Kohn (HK) [19], who proved a one-to-one correspondence between an external potential V_{ext} and a non-degenerate ground-state electronic charge density $n_0(\mathbf{r})$:

$$n_0(\mathbf{r}) = \langle\psi_0(\mathbf{r}_1, ..., \mathbf{r}_N)|\sum_{i=1}^{N}\delta(\mathbf{r} - \mathbf{r}_i)\,|\psi_0(\mathbf{r}_1, ..., \mathbf{r}_N)\rangle \tag{2.18}$$

where ψ_0 is the non-degenerate ground state wavefunction. In other words, HK theory suggests not only that there is one possible ground-state density $n_0(\mathbf{r})$ for each external potential V_{ext}, but also that there is only one external potential that has $n_0(\mathbf{r})$ as its ground-state density. Since the external potential also determines the many-body wavefunction of the ground state ψ_0, every observable of the system in its ground state, can be expressed as a functional of the ground state electron density $n_0(\mathbf{r})$.

Let us write the ground state total energy as a functional of the electronic charge density:

$$E[n_0(\mathbf{r})] = \langle\psi_0| H |\psi_0\rangle = F[n_0(\mathbf{r})] + \int_V d^3\mathbf{r} V_{ext}(\mathbf{r}) n_0(\mathbf{r}) \tag{2.19}$$

where the integral is over the volume of the system, V. The functional $F[n_0(\mathbf{r})]$ is the universal functional which is the integral of kinetic and interaction terms evaluated at the ground state. Due to HK theorem, the minimum of total energy functional $E[n(\mathbf{r})]$ occurs at the ground state charge density $n_0(\mathbf{r})$. Therefore, minimizing the total energy functional in 2.19 with respect to density $n(\mathbf{r})$, for a given number of electrons N, yields the ground state total energy and electronic density.

However the exact form of the kinetic and interaction density functional $F[n(\mathbf{r})]$ is not known. In 1965 Kohn and Sham [20] introduced a mapping of this many-body functional into one of an auxiliary non-interacting electron system with the same ground state density and defined it in three separate contributions:

$$F[n] = T_s[n] + E_H[n] + E_{xc}[n] \tag{2.20}$$

The kinetic energy of the non-interacting system, $T_s[n]$, is given by

$$T_s[n] = \sum_i f_i \langle\psi_i| - \frac{\hbar^2}{2m_e}\nabla^2 |\psi_i\rangle \tag{2.21}$$

where f_i and ψ_i are the occupation and wavefunction of state i respectively. The classical Hartree electrostatic energy of a system with charge density n is defined as:

$$E_H[n] = \frac{1}{2}\int d^3\mathbf{r} d^3\mathbf{r}' n(\mathbf{r}) \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} n(\mathbf{r}') \tag{2.22}$$

The remaining term, the so called exchange-correlation functional, E_{xc}, is defined as the remainder of equation 2.20. This is the term which holds all the many-body details of the system, after mapping from interacting to non-interacting system is performed. The energy functional associated to the external potential can also be written as:

$$E_{ext}[n] = \sum_i f_i \langle\psi_i| V_{ext} |\psi_i\rangle \tag{2.23}$$

And now we can rewrite equation 2.19 as:

$$E[n] = T_s + E_H + E_{xc} + E_{ext} \tag{2.24}$$

and minimize this functional with respect to the orthonormal single particle wavefunctions ψ_i using Lagrange multipliers. The solution of this minimization leads to Kohn-Sham equations

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + V_{KS}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}) \tag{2.25}$$

where the effective Kohn-Sham potential V_{KS} is defined as:

$$V_{KS} = V_{ext} + V_H + V_{xc} \tag{2.26}$$

where $V_H(r) = \int d^3\mathbf{r}' n(\mathbf{r}')e^2/|\mathbf{r} - \mathbf{r}'|$ is the Hartree potential and the exchange-correlation potential is defined as the functional derivative of the exchange-correlation energy term:

$$V_{xc}(\mathbf{r}) = \left.\frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}\right|_{n(\mathbf{r})=n_0(\mathbf{r})} \tag{2.27}$$

Therefore from the solution of KS equations one can obtain the auxiliary one electron wavefunctions, ψ_i, and construct the ground state density through

$$n_0(\mathbf{r}) = \sum_i f_i|\psi_i(\mathbf{r})|^2 \tag{2.28}$$

Based on these single particle equations, one can interpret the KS scheme as a mean-field approach where an electron is affected by the potential of the surrounding nuclei and interacts with other electrons through the charge density. Therefore the ground state charge density depends on the single particle solutions, which in turn, depend on the density. The solution can be achieved via a self-consistent iterative procedure or via global minimization strategies such as conjugate gradient [21]. In this study we use a self-consistent iterative procedure.

The DFT theory explained so far is exact within Born-Oppenheimer approximation, but not useful since the analytical expression for the exchange correlation density functional, E_{xc}, is not known. A critical step in DFT is to replace this unknown functional with an approximation. The first and simplest approach to be proposed was the Local Density Approximation (LDA) where the exchange correlation energy density of an homogeneous gas of interacting electrons with constant charge density, $\varepsilon_{xc}^{hom}(n)$, was used as an approximation for the non-homogeneous system at hand:

$$E_{xc}[n(\mathbf{r})] = \int_V d^3\mathbf{r}\varepsilon_{xc}^{hom}(n(\mathbf{r}))n(\mathbf{r}) \tag{2.29}$$

where $\varepsilon_{xc}^{hom}(n)$ is a well-defined function that can been calculated using Monte Carlo simulations [22].

Since the birth of DFT, many functional forms have been proposed for XC energy density. The Local Density Approximation (LDA), as proposed by Kohn and Sham, has been used in several early works. Different interpolation schemes have been proposed for ε_{xc}^{hom}, such as the one of Perdew and Zunger [23], and are still in use. A significant step that improved the performance of DFT was the proposal of Perdew and Wang in 1991 [24]. They introduced a new functional form which depends not only on the local density but also on its gradient, which is better known today as Generalized Gradient Approximation (GGA). GGA improved successfully the shortcomings of LDA especially for systems with strong inhomogeneities in the charge density. Many different reformulations and extensions of GGA have been proposed and tested over the years. One of the most commonly used GGA functionals in the condensed matter community is the PBE functional, proposed by Perdew, Burke, and Ernzerhof [25]. A revision of PBE was suggested by Zhang and Yang and demonstrated improved atomic energies for several elements[26].

Both PBE and revPBE do not reproduce the van der Waals interactions correctly, which is a common problem of any local and semi-local density functional. A recent solution to this problem has been proposed by Dion et *al.* [27] with a non-local functional called vdW-DF. However, being in its original formulation computationally very expensive, this non-local functional approximation did not find much area of use. Only recently, a new development [28], has made its use only marginally more computationally expensive than a regular GGA calculation. The development of non-local functionals is still ongoing, some very recent studies propose several new functionals such vdW-DF2 [29] and c09x [30], which do not include any drastic changes but are mostly aiming to tune the exchange functionals that can be used with the nonlocal correlation part of vdW-DF for better results.

In this study, we use PBE, revPBE and vdW-DF, which employs the revPBE gradient correction as exchange part in the XC functional. We compare the performance of these three functionals in the case of cholesterol crystals where long-range interactions are of significant importance.

2.3 Projector-Augmented Waves Method

The Projector-Augmented Waves (PAW) method, introduced by Blochl in 1991 [15], is an elegant way to combine the concepts of linearly-augmented plane wave method (LAPW) [31] with the one of ultrasoft pseudopotentials

[16]. Assuming familiarity with the basics of pseudopotential theory, here we provide a summary of the PAW method.

The PAW method is based on the assumption that all-electron single particle wavefunction solutions of KS equation can be constructed from a set of pseudo wavefunctions. In his work, Blochl defines a linear mapping operator, T,that transforms the pseudo wavefunctions into all-electron ones:

$$\begin{aligned} \left|\psi_\nu^{AE}\right\rangle &= T\left|\psi_\nu^{PS}\right\rangle \\ &= \left|\psi_\nu^{PS}\right\rangle + \sum_i (\left|\phi_i^{AE}\right\rangle - \left|\phi_i^{PS}\right\rangle)\left\langle p_i^{PS}|\psi_\nu^{PS}\right\rangle \end{aligned} \tag{2.30}$$

where $\left|\phi_i^{AE}\right\rangle$ and $\left|\phi_i^{PS}\right\rangle$ are the partial waves, which can be chosen to be the solutions of atomic calculations, for all-electron and pseudo atom respectively. The $\left|p_i^{PS}\right\rangle$ are projectors which are constructed to obey the following orthonormality conditions:

$$\left\langle p_i^{PS}|\phi_j^{PS}\right\rangle = \delta_{ij} \tag{2.31}$$

The projector functions are localized in a small 'core' region around the nuclei therefore the composite index i runs over all ionic sites and pseudization channels, i.e. angular momentum and different reference energies. Therefore PAW formalism allows one to reconstruct the all-electron wavefunction of band index ν, $\left|\psi_\nu^{AE}\right\rangle$, starting from the pseudo wavefunction $\left|\psi_\nu^{PS}\right\rangle$, using one-centered corrections based on partial waves in the core region.

In the same manner, the expectation value of any all-electron operator can also be reconstructed. The mapping of an all-electron operator O^{AE} to its pseudo counterpart O^{PS} can be obtained using the transformation operator T

$$\begin{aligned} \left\langle\psi^{AE}\right| O^{AE}\left|\psi^{AE}\right\rangle &= \left\langle\psi^{PS}T\right| O^{AE}\left|T\psi^{PS}\right\rangle \\ &= \left\langle\psi^{PS}\right| T^\dagger O^{AE} T\left|\psi^{PS}\right\rangle \end{aligned} \tag{2.32}$$

Therefore we can write O^{PS} such as

$$\begin{aligned} O^{PS} &= T^\dagger O^{AE} T \\ &= O^{AE} + \sum_{i,j}\left|p_i^{PS}\right\rangle\left(\left\langle\phi_i^{AE}\right| O^{AE}\left|\phi_j^{AE}\right\rangle - \left\langle\phi_i^{PS}\right| O^{AE}\left|\phi_j^{PS}\right\rangle\right)\left\langle p_j^{PS}\right| + \Delta O \end{aligned} \tag{2.33}$$

where ΔO is defined by the following expression:

$$\begin{aligned} \Delta O = \sum_i \Bigg[&\left|p_i^{PS}\right\rangle\left(\left\langle\phi_i^{AE}\right| - \left\langle\phi_i^{PS}\right|\right) O^{AE}\left(1 - \sum_j \left|\phi_j^{PS}\right\rangle\left\langle p_j^{PS}\right|\right) \\ &+ \left(1 - \sum_j \left|p_j^{PS}\right\rangle\left\langle\phi_j^{PS}\right|\right) O^{AE}\left(\left|\phi_i^{AE}\right\rangle - \left|\phi_i^{PS}\right\rangle\right)\left\langle p_j^{PS}\right| \Bigg] \end{aligned} \tag{2.34}$$

and can be shown to vanish for local or semi-local operators such as the charge density, the kinetic energy, etc.

A critical assumption that is made in order to demonstrate the vanishing of ΔO is that any pseudo wavefunction can be expanded as a sum of partial waves inside the augmentation region:

$$\left|\psi_\nu^{PS}\right\rangle = \sum_i \left|\phi_i^{PS}\right\rangle \left\langle p_i^{PS}|\psi_\nu^{PS}\right\rangle \tag{2.35}$$

The validity of this assumption in the range of interest should be tested upon pseudopotential generation.

Having now the recipe (eq. 2.33) to reconstruct the pseudo operators that would yield the all-electron expectation values, one could continue and write the descriptions for charge density and Hamiltonian operators. However the summary given so far is sufficient to let us discuss the Gauge-Including Projector Augmented Waves (GIPAW) method. Further information on the PAW method and its implementation in Quantum Espresso distribution can be found in the original works [15], [32], [33].

2.4 Gauge-Including Projector-Augmented Waves Method

2.4.1 Introduction and History

Historically, the development of computational methods to calculate NMR parameters parallels the development in experimental methods. In the last decade, solid-state NMR (SSNMR) spectroscopy has gone through several developments from hardware [1] to pulse sequences [2]. As these technical improvements rendered the SSNMR method effective and reliable, further theoretical methods to calculate NMR methods in extended systems became essential. Although the traditional quantum chemical techniques [34] have been successfully applied to assign the solution-state NMR spectra of molecular systems [35], in order to apply these techniques to SSNMR, it has been necessary to use cluster approaches. In this approach, one models the local environment around a site of interest through a finite cluster. Such an approach has proven to be useful in obtaining quantitative insight into relationships between NMR parameters and local structural features [36]. However, there are many circumstances where the cluster approximation can only provide limited information on realistic solids, such as solids with strong long range intermolecular or electrostatic interactions.

In 1996, Mauri, Pfrommer, and Louie [37] presented an *ab initio* approach to calculate NMR chemical shieldings for systems with periodic boundary conditions, i.e. calculation of the response to a periodic magnetic field with a finite wavelength q, subsequently extrapolated in the limit $q \to 0$. This approach, also known as MPL, was formulated for an all-electron Hamiltonian, and was applied in the context of calculations based on norm-conserving pseudopotentials, neglecting the contribution of the core electrons and valence electrons in the core region. The demonstration that the contribution of core electrons to the chemical shielding is nearly independent of the chemical environment [38] was a promising step for MPL method. Thus, the core contribution needs only to be evaluated once, typically from a simple atomic calculation. However neglecting the reconstruction of valance electrons in the core region limited the MPL approximation only to light elements in which the pseudization error was expected to be small.

In 2001, Mauri and Pickard extended the MPL method [10] by combining it with the PAW formalism [15]. The resulting approach is known as Gauge Including Projector Augmented Waves (GIPAW) method, since a significant improvement introduced with this method was the gauge-invariance, which was established by using a gauge-including transformation operator. This transformation operator maps the pseudo response to magnetic field to the all-electron one in a PAW-like fashion.

The GIPAW method was originally developed for norm-conserving pseudopotentials (NCPs) however many transition and rare-earth metals require very large numbers of plane waves to accurately describe the valence wave functions. In order to improve its efficiency but also to widen its range of applicability to the mentioned segment of the periodic table, GIPAW method has been recently extended [17] to allow the use of ultrasoft pseudopotentials [16].

The extension of GIPAW to allow PAW type pseudopotentials follows very closely the implementation for ultrasoft pseudopotential scheme, especially for codes that already allow the use of PAW formalism in conjunction with density functional perturbation theory [14]. The extension of GIPAW to allow PAW type pseudopotentials is not challenging, but has the benefit of consistency as GIPAW theory is based on the PAW formalism.

Below we give a summary of GIPAW method in the extended form that accommodates the PAW pseudopotentials, as we have implemented it in Quantum Espresso (QE) [18] distribution. The details of the method can be found in the original papers [10] and [17].

2.4.2 GIPAW with PAW pseudopotentials

When a diamagnetic insulator is placed in a uniform magnetic field, $\mathbf{B}_{\text{ext}}$ the induced current in the material gives rise to induced magnetic field, $\mathbf{B}_{\text{in}}^{(1)}(\mathbf{r})$. Due to this induced magnetic field, the resonance frequency observed in the NMR experiments varies slightly depending on the local environment of the nucleus of interest. This phenomenon is called chemical shielding effect. The chemical shielding tensor $\vec{\sigma}(\mathbf{r})$ is defined by the following relationship:

$$\mathbf{B}_{\text{in}}^{(1)}(\mathbf{r}) = -\vec{\sigma}(\mathbf{r})\mathbf{B}_{\text{ext}} = \frac{1}{c}\int d^3\mathbf{r}'\mathbf{j}^{(1)}(\mathbf{r}') \times \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^3} \tag{2.36}$$

where $\mathbf{j}^{(1)}$ is the first order induced current. As can be seen in equation 2.36, the induced magnetic field requires an accurate description of the induced current, especially close to the nuclear center. Therefore using the PAW formalism and reconstructing the all-electron properties in the core region is very important in obtaining an accurate chemical shielding tensor. However, the original PAW formalism is not built for systems subject to magnetic field so that the PAW transformation operator does not include any information on the effect of magnetic field. For example, in a uniform magnetic field, a rigid translation of all atoms by a vector $\mathbf{t}$ results in a field-dependent phase factor, which in a symmetric gauge such as the Coulomb one ($\mathbf{A}(\mathbf{r}) = 1/2\mathbf{B} \times \mathbf{r}$), can be shown to be:

$$\langle \mathbf{r}|\psi'\rangle = e^{(i/2c)\mathbf{r}\cdot\mathbf{t}\times\mathbf{B}} \langle \mathbf{r}-\mathbf{t}|\psi\rangle \tag{2.37}$$

To overcome this problem and still keep the benefits of the PAW formalism, Pickard and Mauri proposed a modified version of the PAW transformation operator, $T_{\mathbf{B}}$:

$$T_{\mathbf{B}} = 1 + \sum_{\mathbf{R},n} e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}}(\left|\phi_{\mathbf{R},n}^{AE}\right\rangle - \left|\phi_{\mathbf{R},n}^{PS}\right\rangle) \left\langle p_{\mathbf{R},n}^{PS}\right| e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} \tag{2.38}$$

where the sum runs over nuclei positions $\mathbf{R}$ and partial wave basis elements n. Comparing the GIPAW operator with the one of the PAW described in equation 2.30, one can see that the main difference lies in the exponential operators, which ensure that both pseudo, ψ^{PS}, and all electron wavefunctions, $\psi^{AE} = T_{\mathbf{B}}\psi^{PS}$, gain the same phase under translation described in equation 2.37.

Following the literature, in the rest of this summary we will use an overbar as in $\bar{\psi}$ for the pseudo wavefunctions and operators that are linked to the all electron ones (for which no symbols will be used) through GIPAW

transformation operator, as in $\psi = T_{\mathbf{B}}\bar{\psi}$; and for the ones that are linked by PAW operator we will use a tilde, as in $\psi = T\tilde{\psi}$

With the transformation operator at hand, for any all-electron local or semi-local operator O, we can write its GIPAW counterpart as follows:

$$\begin{aligned}\bar{O} = O + \sum_{\mathbf{R},n,m} & e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} \left|\tilde{p}_{\mathbf{R},n}\right\rangle \\ & \times \left(\left\langle\phi_{\mathbf{R},n}\right| e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} O e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} \left|\phi_{\mathbf{R},m}\right\rangle \right. \\ & \left. - \langle\tilde{\phi}_{\mathbf{R},n}|e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} O e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} |\tilde{\phi}_{\mathbf{R},m}\rangle\right)\langle\tilde{p}_{\mathbf{R},m}|e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}}\end{aligned} \tag{2.39}$$

Using the GIPAW transformation operator and treating the magnetic field as a perturbation, we will now derive each ingredient necessary to calculate the response of the system to the magnetic field, and in particular the induced current operator that is used to calculate the chemical shift tensor.

GIPAW Hamiltonian

Let us first build the perturbative GIPAW Hamiltonian to first order in the magnetic field. In a uniform magnetic field the all-electron Hamiltonian is

$$H = \frac{1}{2}\left[\mathbf{p} + \frac{1}{c}\mathbf{A}(\mathbf{r})\right]^2 + V(\mathbf{r}) \tag{2.40}$$

where $V(\mathbf{r})$ is the total potential felt by the electrons, which, in DFT, is the Kohn-Sham potential V_{KS} described in equation 2.26. Using the GIPAW transformation operator in equation 2.37 and the Coulomb gauge, we can write the GIPAW Hamiltonian as:

$$\begin{aligned}\bar{H} = & \frac{1}{2}\mathbf{p}^2 + V^{loc}(\mathbf{r}) + \sum_{\mathbf{R}} e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} V^{nl}_{\mathbf{R}} e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} \\ & + \frac{1}{2c}\mathbf{L}\cdot\mathbf{B} + \frac{1}{8c^2}(\mathbf{B}\times\mathbf{r})^2 \\ & + \sum_{\mathbf{R},n,m} e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} (\beta_1 + \beta_2) e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}}\end{aligned} \tag{2.41}$$

where $V^{loc}(\mathbf{r})$ is due to the identity operator in the transformation $T_{\mathbf{B}}$ and the $V^{nl}_{\mathbf{R}}$ results from the 'sandwiching' of partial waves in GIPAW transformation. In the same manner, the partial wave part of the transformation acting on $\mathbf{L}\cdot\mathbf{B}$ and $(\mathbf{B}\times\mathbf{r})^2$ are denoted as β_1, and β_2 respectively:

$$\beta_1 = \frac{1}{2c}\left|\tilde{p}_{\mathbf{R},n}\right\rangle \left[\langle\phi_{\mathbf{R},n}|(\mathbf{L}_{\mathbf{R}}\cdot\mathbf{B})|\phi_{\mathbf{R},m}\rangle - \langle\tilde{\phi}_{\mathbf{R},n}|(\mathbf{L}_{\mathbf{R}}\cdot\mathbf{B})|\tilde{\phi}_{\mathbf{R},m}\rangle\right]\langle\tilde{p}_{\mathbf{R},m}| \tag{2.42}$$

and

$$\beta_2 = \frac{1}{8c^2}\left|\tilde{p}_{\mathbf{R},n}\right\rangle \left[\langle\phi_{\mathbf{R},n}|[\mathbf{B}\times(\mathbf{r}-\mathbf{R})]^2|\phi_{\mathbf{R},m}\rangle - \langle\tilde{\phi}_{\mathbf{R},n}|[\mathbf{B}\times(\mathbf{r}-\mathbf{R})]^2|\tilde{\phi}_{\mathbf{R},m}\rangle\right]\langle\tilde{p}_{\mathbf{R},m}| \tag{2.43}$$

Considering a magnetic field in $\hat{z}$ direction, we rewrite the β_1 term as:

$$\begin{aligned}\beta_1 &= \frac{1}{2c} B_z m_z \left|\tilde{p}_{\mathbf{R},n}\right\rangle \left[\langle\phi_{\mathbf{R},n}|\phi_{\mathbf{R},m}\rangle - \langle\tilde{\phi}_{\mathbf{R},n}|\tilde{\phi}_{\mathbf{R},m}\rangle\right] \langle\tilde{p}_{\mathbf{R},m}| \\ &= \frac{1}{2c}(\mathbf{L}_\mathbf{R}\cdot\mathbf{B})Q_\mathbf{R}\end{aligned} \tag{2.44}$$

where $\mathbf{L}_\mathbf{R}$ is the angular momentum centered on atomic site $\mathbf{R}$ and the charge augmentation operator, which would be zero in the case of norm-conservation, can be defined as

$$\begin{aligned}Q_\mathbf{R} &= \left|\tilde{p}_{\mathbf{R},n}\right\rangle q_{\mathbf{R},n,m} \langle\tilde{p}_{\mathbf{R},m}| \\ &= \left|\tilde{p}_{\mathbf{R},n}\right\rangle \left[\langle\phi_{\mathbf{R},n}|\phi_{\mathbf{R},m}\rangle - \langle\tilde{\phi}_{\mathbf{R},n}|\tilde{\phi}_{\mathbf{R},m}\rangle\right] \langle\tilde{p}_{\mathbf{R},m}|\end{aligned} \tag{2.45}$$

Keeping only the first order terms in $\mathbf{B}$, we can now write the GIPAW perturbative Hamiltonian:

$$\bar{H}^{(1)} = \frac{1}{2c}\left(\mathbf{L} + \sum_\mathbf{R} \mathbf{R}\times\frac{1}{i}[\mathbf{r}, V^{nl}_\mathbf{R}] + \sum_\mathbf{R} \mathbf{L}_\mathbf{R} Q_\mathbf{R}\right)\cdot\mathbf{B} \tag{2.46}$$

With this Hamiltonian, we now have access to first order perturbation in wavefunctions and current as will be explained in the next sections.

Overlap operator

In the PAW formalism the norm of the pseudo-partial-waves in the augmentation region is different from that of the corresponding all-electron partial waves, giving rise to an overlap operator different from identity

$$S = 1 + \sum_\mathbf{R} Q_\mathbf{R} \tag{2.47}$$

With the GIPAW transformation, and expanding to the first order in B, we obtain the first order overlap operator

$$\begin{aligned}\bar{S} &= 1 + \sum_\mathbf{R} e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} Q_\mathbf{R} e^{-(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}} = S^{(0)} + \bar{S}^{(1)} + \dots \\ \bar{S}^{(1)} &= \frac{1}{2ic}\sum_\mathbf{R} \mathbf{R}\times[\mathbf{r}, Q_\mathbf{R}]\cdot\mathbf{B}\end{aligned} \tag{2.48}$$

We also introduce the augmented velocity operator to be used later at an unperturbed eigenvalue $\varepsilon^{(0)}_o$ of the occupied subspace o:

$$\begin{aligned}\mathbf{v}(\varepsilon^{(0)}_o) &= \frac{1}{i}[\mathbf{r}, (\bar{H}^{(0)} - \varepsilon^{(0)}_o \bar{S}^{(0)})] \\ &= \frac{1}{i}\nabla + \frac{1}{i}\sum_\mathbf{R}[\mathbf{r}, (V^{nl}_\mathbf{R} - \varepsilon^{(0)}_o Q_\mathbf{R})]\end{aligned} \tag{2.49}$$

With the Hamiltonian and overlap operator, we can write the perturbed wavefunction projected onto empty bands :

$$\left|\psi_o^{(1e)}\right\rangle = G(\varepsilon_o^{(0)})(\bar{H}^{(1)} - \varepsilon_o^{(0)}\bar{S}^{(1)})\left|\psi_o^{(0)}\right\rangle \tag{2.50}$$

with Green's function operator projected onto empty bands e given as :

$$G(\varepsilon) = \sum_e \frac{\left|\psi_e^{(0)}\right\rangle\left\langle\psi_e^{(0)}\right|}{\varepsilon - \varepsilon_e} \tag{2.51}$$

Current operator

Although algebraically cumbersome, calculating the induced current with GIPAW transformation is straightforward. Let us first write the current operator $\mathbf{J}(\mathbf{r}')$ in the all-electron case:

$$\mathbf{J}(\mathbf{r}') = \frac{1}{c}\mathbf{A}(\mathbf{r}')\left|\mathbf{r}'\right\rangle\left\langle\mathbf{r}'\right| - \frac{\mathbf{p}\left|\mathbf{r}'\right\rangle\left\langle\mathbf{r}'\right| + \left|\mathbf{r}'\right\rangle\left\langle\mathbf{r}'\right|\mathbf{p}}{2} \tag{2.52}$$

where the first term is the diamagnetic current $\mathbf{J}^d(\mathbf{r}')$ and the second term is the paramagnetic current $\mathbf{J}^p(\mathbf{r}')$. Using GIPAW transformation (equation 2.37), the GIPAW current operator can be written as a sum of four terms: i) $\mathbf{J}^d(\mathbf{r}')$ and $\mathbf{J}^p(\mathbf{r}')$ themselves due to the identity term in the transformation and their partial wave counterparts due to the partial wave 'sandwich' of the transformation operator. Since we are only interested in the components up to the first order in magnetic field, we can proceed with the following argument: $\mathbf{J}^d(\mathbf{r}')$ is already first order in the magnetic field. So the only components up to first order in the perturbation arising from this term, would be the ones where the exponential operators, $e^{(i/2c)\mathbf{r}\cdot\mathbf{R}\times\mathbf{B}}$, are expanded to zeroth order in B, namely the identity operator. For $\mathbf{J}^p(\mathbf{r}')$, since it is of order zero in B, the terms contributing to the current operator up to first order would also arise from expansion of the exponential operators to the first order in B. Let us now write the zeroth and first order GIPAW current operators:

$$\bar{\mathbf{J}}(\mathbf{r}') = \bar{\mathbf{J}}^{(0)}(\mathbf{r}') + \bar{\mathbf{J}}^{(1)}(\mathbf{r}') + O(B^2) \tag{2.53}$$

$$\bar{\mathbf{J}}^{(0)}(\mathbf{r}') = \mathbf{J}^p(\mathbf{r}') + \sum_{\mathbf{R}} \Delta\mathbf{J}^p_{\mathbf{R}}(\mathbf{r}') \tag{2.54}$$

$$\bar{\mathbf{J}}^{(1)}(\mathbf{r}') = -\frac{\mathbf{B}\times\mathbf{r}'}{2c}\left|\mathbf{r}'\right\rangle\left\langle\mathbf{r}'\right| + \sum_{\mathbf{R}}\left[\Delta\mathbf{J}^d_{\mathbf{R}}(\mathbf{r}') + \frac{1}{2ci}[\mathbf{B}\times\mathbf{R}\cdot\mathbf{r}, \Delta\mathbf{J}^p_{\mathbf{R}}(\mathbf{r}')]\right] \tag{2.55}$$

where the augmentation operators (indicated with a Δ) are defined by:

$$\Delta\mathbf{J}^p_{\mathbf{R}}(\mathbf{r}') = \sum_{n,m} |\tilde{p}_{\mathbf{R},n}\rangle \left[\langle\phi_{\mathbf{R},n}|\mathbf{J}^p(\mathbf{r}')|\phi_{\mathbf{R},m}\rangle - \langle\tilde{\phi}_{\mathbf{R},n}|\mathbf{J}^p(\mathbf{r}')|\tilde{\phi}_{\mathbf{R},m}\rangle\right] \langle\tilde{p}_{\mathbf{R},m}| \quad (2.56)$$

for the paramagnetic augmentation operator, and

$$\begin{aligned}\Delta\mathbf{J}^d_{\mathbf{R}}(\mathbf{r}') = -\frac{\mathbf{B}\times(\mathbf{r}'-\mathbf{R})}{2c} \sum_{n,m} |\tilde{p}_{\mathbf{R},n}\rangle \\ \left[\langle\phi_{\mathbf{R},n}|\mathbf{r}'\rangle\langle\mathbf{r}'|\phi_{\mathbf{R},m}\rangle - \langle\tilde{\phi}_{\mathbf{R},n}|\mathbf{r}'\rangle\langle\mathbf{r}'|\tilde{\phi}_{\mathbf{R},m}\rangle\right] \langle\tilde{p}_{\mathbf{R},m}|\end{aligned} \quad (2.57)$$

for the diamagnetic augmentation operator.

So now, we can write the induced current density using the unperturbed wavefunctions $|\bar{\psi}^{(0)}_n\rangle$ and its linear variation due to perturbation $|\bar{\psi}^{(1)}_n\rangle$:

$$\mathbf{j}^{(1)}(\mathbf{r}') = 2\sum_o [\langle\bar{\psi}^{(1)}_n| \bar{\mathbf{J}}^{(0)}(\mathbf{r}') |\bar{\psi}^{(0)}_n\rangle + \langle\bar{\psi}^{(0)}_n| \bar{\mathbf{J}}^{(0)}(\mathbf{r}') |\bar{\psi}^{(1)}_n\rangle + \langle\bar{\psi}^{(0)}_n| \bar{\mathbf{J}}^{(1)}(\mathbf{r}') |\bar{\psi}^{(0)}_n\rangle] \quad (2.58)$$

the summation runs over all occupied bands.

Current density

Now we have all the necessary ingredients to calculate the induced current density $\mathbf{j}^{(1)}(\mathbf{r}')$. Using equation 2.50 and 2.58, and after some algebra (see Appendix B of Ref. [17]):

$$\begin{aligned}\mathbf{j}^{(1)}(\mathbf{r}') =& 4\sum_o Re[\langle\bar{\psi}^{(0)}_o|\bar{\mathbf{J}}^{(0)}(\mathbf{r}')G(\varepsilon^{(0)}_o)(\bar{H}^{(1)} - \varepsilon^{(0)}_o\bar{S}^{(1)})|\bar{\psi}^{(0)}_o\rangle \\ &- 2\sum_{o,o'}\langle\bar{\psi}^{(0)}_o|\bar{\mathbf{J}}^{(0)}(\mathbf{r}')|\bar{\psi}^{(0)}_{o'}\rangle\langle\bar{\psi}^{(0)}_{o'}|\bar{S}^{(1)}|\bar{\psi}^{(0)}_o\rangle \\ &+ 2\sum_o\langle\bar{\psi}^{(0)}_o|\bar{\mathbf{J}}^{(1)}(\mathbf{r}')|\bar{\psi}^{(0)}_o\rangle\end{aligned} \quad (2.59)$$

Note that the second term arises from paramagnetic current operators, only in the case of non-norm-conserving pseudopotentials, due to possible mixing of occupied levels by this perturbative operator.

We can substitute equations 2.53, 2.46 and 2.48 in equation 2.59, and regroup the current density in three terms: *i)* one that does not involve any augmentation current as $\mathbf{j}^{(1)}_{bare}(\mathbf{r}')$, *ii)* one with paramagnetic augmentation current as $\mathbf{j}^{(1)}_{\Delta p}(\mathbf{r}')$ and *iii)* one with diamagnetic augmentation current as $\mathbf{j}^{(1)}_{\Delta d}(\mathbf{r}')$. Before giving the final forms of these terms, we use the generalized f-sum rule outlined in Appendix C of Ref. [17], so the induced current density becomes:

$$\mathbf{j}^{(1)}(\mathbf{r}') = \mathbf{j}^{(1)}_{bare}(\mathbf{r}') + \mathbf{j}^{(1)}_{\Delta p}(\mathbf{r}') + \mathbf{j}^{(1)}_{\Delta d}(\mathbf{r}') \tag{2.60}$$

where

$$\begin{aligned}\mathbf{j}^{(1)}_{bare}(\mathbf{r}') =& 4\sum_{o} Re[\langle\bar{\psi}^{(0)}_o|\mathbf{J}^p(\mathbf{r}')G(\varepsilon^{(0)}_o)(\bar{H}^{(1)} - \varepsilon^{(0)}_o\bar{S}^{(1)})|\bar{\psi}^{(0)}_o\rangle \\ &- \langle\bar{\psi}^{(0)}_o|\mathbf{J}^p(\mathbf{r}')G(\varepsilon^{(0)}_o)\sum_{\mathbf{R}}\frac{\mathbf{B}\times\mathbf{r}'}{2c}\cdot\mathbf{v}(\varepsilon^{(0)}_o)|\bar{\psi}^{(0)}_o\rangle] \\ &- \sum_{\mathbf{R},o,o'}\langle\bar{\psi}^{(0)}_o|\mathbf{J}^p(\mathbf{r}')|\bar{\psi}^{(0)}_{o'}\rangle\langle\bar{\psi}^{(0)}_{o'}|(\mathbf{R}-\mathbf{r}')\times\frac{1}{2ic}[\mathbf{r},Q_{\mathbf{R}}]\cdot\mathbf{B}|\bar{\psi}^{(0)}_o\rangle\end{aligned} \tag{2.61}$$

and

$$\begin{aligned}\mathbf{j}^{(1)}_{\Delta p}(\mathbf{r}') =& \sum_{\mathbf{R}',o} Re[\langle\bar{\psi}^{(0)}_o|\Delta\mathbf{J}^p_{\mathbf{R}'}(\mathbf{r}')G(\varepsilon^{(0)}_o)(\bar{H}^{(1)} - \varepsilon^{(0)}_o\bar{S}^{(1)})|\bar{\psi}^{(0)}_o\rangle \\ &- \langle\bar{\psi}^{(0)}_o|\Delta\mathbf{J}^p_{\mathbf{R}'}(\mathbf{r}')G(\varepsilon^{(0)}_o)\sum_{\mathbf{R}}\frac{\mathbf{B}\times\mathbf{R}'}{2c}\cdot\mathbf{v}(\varepsilon^{(0)}_o)|\bar{\psi}^{(0)}_o\rangle] \\ &- \sum_{\mathbf{R},\mathbf{R}',o,o'}\langle\bar{\psi}^{(0)}_o|\Delta\mathbf{J}^p_{\mathbf{R}'}(\mathbf{r}')|\bar{\psi}^{(0)}_{o'}\rangle\langle\bar{\psi}^{(0)}_{o'}|(\mathbf{R}-\mathbf{R}')\times\frac{1}{2ic}[\mathbf{r},Q_{\mathbf{R}}]\cdot\mathbf{B}|\bar{\psi}^{(0)}_o\rangle\end{aligned} \tag{2.62}$$

and

$$\mathbf{j}^{(1)}_{\Delta d}(\mathbf{r}') = 2\sum_{\mathbf{R},o}\langle\bar{\psi}^{(0)}_o|\Delta\mathbf{J}^d_{\mathbf{R}}(\mathbf{r}')|\bar{\psi}^{(0)}_o\rangle. \tag{2.63}$$

Using the above equations for the induced current density and Biot-Savart formula (see equation 2.36) one can obtain the chemical shielding tensor.

Infinitely periodic systems

The equations for the induced current density contain terms where the expectation value of the position operator should be calculated, which is typically ill-defined in periodic systems. Therefore it is convenient to define the reciprocal space correspondents of these equations. This motivated Pickard and Mauri to replace any position operator component in the previous equations with a form more suitable for a calculation in reciprocal space. To simply illustrate their approach let us use $f(r') = Re[\langle\phi|\,(r-r')\,|\phi\rangle]$, which is replaced by $f(r') = \lim_{q\to 0}\frac{1}{2q}[s(q,r') - s(-q,r')]$ where a small wave vector q is introduced in the auxiliary function $s(q,r') = Re[\langle\phi|\,e^{i\mathbf{q}\cdot(r-r')}\,|\phi\rangle]$. Expanding the exponential in the $q\to 0$ limit, $e^{i\mathbf{q}\cdot\mathbf{x}} = 1 + i\mathbf{q}\cdot\mathbf{x} + O((qx)^2)$, and choosing the appropriate Cartesian components, one can recover the original

$f(r')$ in a way which is suitable to work in reciprocal space with plane waves. For practical purposes $\mathbf{q}$ can be chosen small enough to accurately represent the limit but not so small that the numerical noise dominates.

This approach can be straightforwardly applied to equations 2.61 and 2.62. For demonstration purposes here we give the expression for $\mathbf{j}^{(1)}_{bare}(\mathbf{r}')$ in reciprocal space:

$$\begin{aligned}\mathbf{j}^{(1)}_{bare}(\mathbf{r}') =& \lim_{q\to 0}\frac{1}{2q}\left[\mathbf{S}_{bare}(\mathbf{r}',q)-\mathbf{S}_{bare}(\mathbf{r}',-q)\right]\\ &+\frac{2}{c}\sum_{\mathbf{R},o} Re[\langle\bar{\psi}^{(0)}_o|\mathbf{J}^p(\mathbf{r}')G(\varepsilon_o)\mathbf{B}\cdot\mathbf{L}_{\mathbf{R}}Q_{\mathbf{R}}|\bar{\psi}^{(0)}_o\rangle]\end{aligned} \tag{2.64}$$

with

$$\begin{aligned}\mathbf{S}_{bare}(\mathbf{r}',q) =& \frac{2}{cN_k}\sum_{i=x,y,z}\sum_{o,\mathbf{k}} Re\left[\frac{1}{i}\langle\bar{u}^{(0)}_{o,\mathbf{k}}|\mathbf{J}^p_{\mathbf{k},\mathbf{k}+\mathbf{q}_i}(\mathbf{r}')G_{\mathbf{k}+\mathbf{q}_i}(\varepsilon_{o,\mathbf{k}})\mathbf{B}\right.\\ &\times\hat{\mathbf{u}}_i\cdot\mathbf{v}_{\mathbf{k}+\mathbf{q}_i,\mathbf{k}}(\varepsilon_{o,\mathbf{k}})|\bar{u}^{(0)}_{o,\mathbf{k}}\rangle\\ &-\sum_{o'}\langle\bar{u}^{(0)}_{o,\mathbf{k}}|\mathbf{J}^p_{\mathbf{k},\mathbf{k}+\mathbf{q}_i}(\mathbf{r}')|\bar{u}^{(0)}_{o',\mathbf{k}+\mathbf{q}_i}\rangle\\ &\left.\times\langle\bar{u}^{(0)}_{o',\mathbf{k}+\mathbf{q}_i}|\mathbf{B}\times\hat{\mathbf{u}}_i\cdot\mathbf{s}_{\mathbf{k}+\mathbf{q}_i,\mathbf{k}}|\bar{u}^{(0)}_{o,\mathbf{k}}\rangle\right]\end{aligned} \tag{2.65}$$

where $\hat{\mathbf{u}}_i$ are unit vectors in Cartesian coordinates, N_k is the number of k points, $\bar{u}^{(0)}_{o,\mathbf{k}}$ are the cell periodic part of the Bloch functions and $\mathbf{s}_{\mathbf{k},\mathbf{k}'}$ is the commutator of the zeroth order overlap operator with position operator: $\mathbf{s}_{\mathbf{k},\mathbf{k}'} = -i[r, Q_{\mathbf{k},\mathbf{k}'}]$.

One can obtain, in a similar manner the paramagnetic and diamagnetic augmentation terms and obtain the induced magnetic field through Biot-Savart equation in reciprocal space.

Notes on practical implementation

It is important to note that in the implementation of GIPAW with PAW pseudopotentials, it is assumed that augmentation currents within an augmentation sphere, only have impact on the chemical shift of that particular site but not on the ones of other nuclei around. To realize this on-site assumption, we have generated pseudopotentials with rather small augmentation spheres.

As it can be seen by now, all augmentation-including terms greatly depend on how accurately the all-electron wavefunctions are reconstructed. Therefore one needs to pay attention to the accuracy of reconstruction while generating the pseudopotentials. Implementation of PAW pseudopotentials in GIPAW enables us to have consistency, in the sense that all contributions are calculated with the same set of projectors, from the unperturbed

Hamiltonian to induced augmentation currents. We note that the computational cost of GIPAW with PAW pseudopotentials is comparable to the one of NC pseudopotentials, at the same kinetic energy and charge density cutoff. However, by allowing the use of PAW pseudopotentials we greatly reduce the minimum cutoff that is required, which, in total, results to be very advantageous especially for large systems.

Although not mentioned in our summary, the Biot-Savart equation 2.36 when transformed in reciprocal space, has a singularity at zero reciprocal lattice vector $G = 0$ (a thorough discussion of this can be found in the original works [37][10][17]). This component is not a bulk property and is strongly related to the macroscopic shape of the sample. In the GIPAW implementation in QE, the possibility of using different macroscopic shapes is available. In our calculations involving molecular crystals the NMR parameters were found to be rather insensitive to the macroscopic shape and we have used a spherical sample shape correction.

2.5 Computational Tests

In this section, we perform tests to verify the accuracy of our implementation by comparing our results with the ones in literature, both for NC pseudopotentials and quantum chemical methods. For a better comparison with the published results [10], we use Perdew-Zunger LDA functional [23] and for small molecules, calculations are performed in a cubic cell of 6000 bohr^3 volume. PAW type pseudopotentials from the PSLibrary project [39] have been adapted in the calculations such that GIPAW and PAW reconstruction schemes use the same projectors.[1] A kinetic energy cutoff of 60 Ry and charge density cutoff of 350 Ry were used for all pseudopotentials for simplicity in comparison; note however that many elements could be pseudized more aggressively in the case of PAW.

2.5.1 Small molecules

Here we first compare the GIPAW with PAW results to the theoretical results of Ref. [10] where GIPAW method was used with norm-conserving pseudopotentials and of Ref. [40] where Gaussian-based quantum chemistry method IGAIM [41] was used. In Table 2.1, for each molecule considered we report the total isotropic chemical shielding calculated with the mentioned setting.

[1] The only modification to the pseudopotentials from PSLibrary project was the change of the pseudization scheme for Phosphorus element (from PSQ to BESSEL)

Molecule	σ_{iso} PAW	σ_{iso} NC	σ_{iso} IGAIM
^{13}C Signal			
CH_4	189.87	190.96	191.22
CH_3F	98.63	98.47	99.66
CH_3NH_2	145.94	149.77	150.44
C_6H_6	38.69	36.12	39.52
CO	-23.72	-22.93	-21.16
^{1}H Signal			
CH_4	30.47	30.87	30.99
CH_3F	25.61	26.13	26.50
C_6H_6	22.97	22.74	23.25
^{31}P Signal			
P_2	-366.24	-360.75	-375.45

Table 2.1: The isotropic chemical shielding for several small molecules compared to NC pseudopotentials with GIPAW [10] and to results obtained with IGAIM method [40]

Note that to ensure a proper comparison we have used the same cell size. All chemical shielding values are given in ppm.

Looking at the results we see that the agreement between PAW method and the literature is satisfactory. The most significant discrepancies are observed for P_2 and CH_3NH_2 molecules. Note that for P_2 dimer, the absolute difference between NC GIPAW and IGAIM methods is as large as 15 ppm. For this molecule PAW GIPAW results in a satisfactory value in between, considering the variation with different methods. Instead for CH_3NH_2 molecule, there is an agreement between NC GIPAW and IGAIM results and our PAW GIPAW calculations give a smaller result by 4 ppm.

Throughout our studies we will compare our results with experimental findings as well. A comparison between experiment and theory can be conducted through chemical shifts, instead of shieldings. Experimentally, ^{13}C are referenced with respect to tetramethylsilane (TMS) compound. Instead of performing calculations on TMS solution, a common procedure in computational NMR calculations is to use a secondary reference to obtain the isotropic chemical shielding of TMS σ_{TMS}. In our calculations for cholesterol crystals we use cyclobutane (cyclic C_4H_8) as our secondary reference as it is easy to calculate and like many carbon atoms in cholesterol, C-C bonds are saturated and it forms a ring.

To obtain the σ_{TMS}, we use the experimental $\delta_{cyclobutane}$:

$$\sigma_{TMS} = \delta^{exp}_{cyclobutane} + \sigma^{theory}_{cyclobutane} \tag{2.66}$$

which yields approximately 167.5 ppm, as PAW GIPAW results in 144.53 ppm for $\sigma^{theory}_{cyclobutane}$ and the reported experimental values for $\delta^{exp}_{cyclobutane}$ are 22.86 and 23.3 ppm [42].Now using this value for σ_{TMS} we can calculate the chemical shifts of cholesterol molecules as:

$$\delta^{theory}_{iso} = \sigma_{TMS} - \sigma_{theory} \tag{2.67}$$

While calculating this secondary reference through cyclobutane calculations we used PAW pseudopotentials that were used for the cholesterol crystal calculations and vdW-DF formalism for the sake of consistency. The details of the calculations are the same as for the cholesterol crystals, given later on in section 3.2.1. Cyclobutane calculations were performed in a large cubic simulation cell with volume 800Å^3, to prevent any interaction with the periodic images in the presence of vdW interactions.

We perform a quick test for the accuracy of referencing by comparing the experimental and calculated chemical shift of another hydrocarbon ring, benzene. The reported experimental chemical shift of benzene is 128.5 ppm [42] and our calculations yield 128.81 ppm. This agreement validates our secondary reference procedure. Further examination of the use of secondary reference scheme will be given in section 4.4.

2.5.2 Solids

We have also tested our method on crystalline systems, mainly crystalline glycine. Glycine has two co-existing crystalline phases at room temperature and ambient pressure, α and γ. Like cholesterol crystals, in glycine crystals van der Waals interactions are important and due to their small size, these crystals are suitable systems to test the performance of van der Waals functional in structure determination.

Computationally inexpensive calculations of glycine crystals are a good test for van der Waals contributions to structure. Before performing GIPAW calculations we fully relax the glycine structures, as in the case of cholesterol crystals, which will be given in more detail in section 3.2.1.

At variance with cholesterol, glycine contains a Nitrogen atom. For this element, we have generated a pseudopotential with d as local angular momentum channel, with cutoff radius of 1.1 a.u and a reference energy of 0.15 Ry. We have used two projectors for both s and p angular momentum channels. The cutoff radii were 1.30 and 1.40 a.u. for s and p projectors, respectively. In addition to the eigenvalues, 0.05 Ry was used as reference energy for both s and p projectors.

For Brilluoin sampling, we have used Monkhorst-Pack(MP) [43] sampling with $4 \times 2 \times 4$ and $3 \times 3 \times 3$ grids for α and γ phases, respectively. The grid

is shifted by half the interval size in all directions. The NMR spectrum

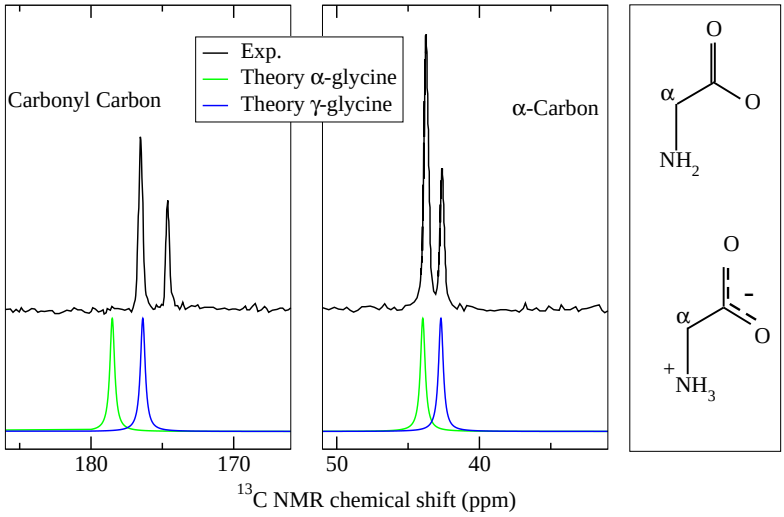


Figure 2.1: ^{13}C NMR chemical shift in ppm for α and γ glycine, comparison between experiment and theory. The experimental spectra is for a mixture of both phases. On the right, representations of neutral glycine molecule (top) and glycine molecule in zwitterion form (bottom) are given. The α carbon is indicated.The resonance signal at the low-shift region is from the α-carbon, the one at the high-shift region is from the carbonyl-carbon. In both phases, glycine takes a zwitterion form.

calculated at fully relaxed structure for α and γ glycine phases and the experimental spectrum of a sample that contains both phases, obtained from the research group of Dr. Sinha at Centre of Biomedical Magnetic Resonance, Lucknow, India, are compared in Fig. 2.1.

We see that the agreement between theory and experiment is good, especially in the case of α-carbon signal. For both high-shift and low shift-signals, the differences between phases, which are less than 2 ppm, are very well reproduced, validating our implementation. We note that the peak-position agreement for carbonyl carbon can be improved greatly with a constant shift of -2 ppm, as the difference between the two peak position of the two phases is already in good agreement with the experiment. We also note that from the point of chemical environment, the carbons from the reference compound cyclobutane is more similar to α-carbon than to carbonyl carbon, which may have a role in the lesser agreement of carbonyl carbon peak positions. A possible systematic correction using environment dependent referencing will be further examined in section 4.4.

Chapter 3

Structural Properties of Cholesterol Crystals

Cholesterol, the principal sterol synthesized by animals, is mainly found as a component in cell membranes and lipoproteins. It contains a hydrophilic C3 hydroxyl group which causes this otherwise hydrophobic molecule to occupy a position at polar-nonpolar interfaces. All known forms of cholesterol crystals form double layer structures with a head-for-head arrangement, similar to the one of membranes. Another common feature of these structures is the existence of several molecules in the asymmetric unit cell, which results in a complicated crystal structure. This complexity is the reason why X-ray crystallographers have spent decades for an unambiguous structure determination of cholesterol crystals.

In this chapter we first introduce the experimental data obtained so far on cholesterol crystals. After mentioning the theoretical details of our calculations, we report the resulting optimized structures for all phases and compare them with the experimental data. Our findings reveal that for these structures, room temperature X-ray experiments are useful in characterization of the phases but the atomic positions obtained with this method perform poorly for structure-sensitive calculations. Additionally we present the optimized structures using force fields and diagnose certain issues that may result in inadequate structure determination using this method.

3.1 Experimental Findings

3.1.1 Cholesterol Molecule

Cholesterol, principal sterol synthesized by animals, is an essential structural component of mammalian cell membranes and also takes part in the manufacturing of bile acids and steroid hormones. Its chemical formula is $C_{27}H_{46}O$. It is made up of one pentagonal and 3 hexagonal carbon rings and a flexible hydrocarbon tail. The standard numbering of carbon atoms is given in Fig.3.1.

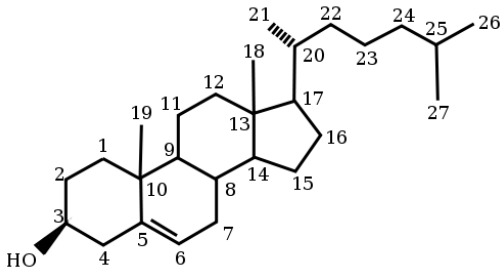


Figure 3.1: Cholesterol molecule representation with the standard numbering of carbon atoms.

3.1.2 Cholesterol Monohydrate (ChM)

The first complete structure analysis for monohydrate phase of cholesterol was reported by Craven [6]. The crystal is found to be in triclinic symmetry, thus in P1 space group, with reduced cell parameters a=12.39 Å, b=12.41 Å, c=34.36 Å, α=91.9°, β=98.1°, γ=100.8°. The unit cell is found to contain 8 cholesterol molecules and 8 water molecules as in $8(C_{27}H_{46}O) \cdot 8H_2O$. Large amplitudes of atomic thermal vibrations were observed through high values of isotropic Debye-Waller factor (also known as B factor) of 7.6 Å^2 in X-ray experiments. In this phase a remarkable twinning of molecules was observed, giving raise to pseudosymmetry. In ChM unit cell, there are two nearly identical but differently oriented subcells with translations $\vec{b}/2$ and $\vec{a}/2$. Therefore the molecules of the unit cell can be divided in two subcells as in Fig. 3.2.

Molecules such as A and B (as well as E and F, C and D, G and H) which are related by subcell translations have the same conformation. Molecules such as A and G (as well as B and H, E and C, F and D) which are related by approximate twofold local symmetry (pseudosymmetry) have very similar

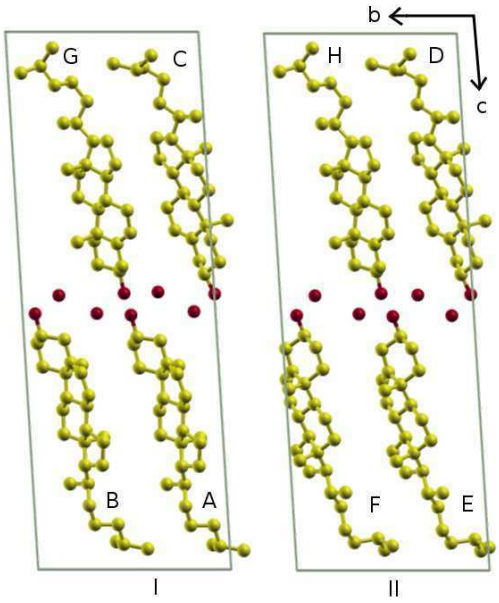


Figure 3.2: Cholesterol monohydrate unit cell, shown in two subcells (see text). Molecule labeling and atomic positions are taken from Ref. [6]. Hydrogen atoms are omitted for clarity.

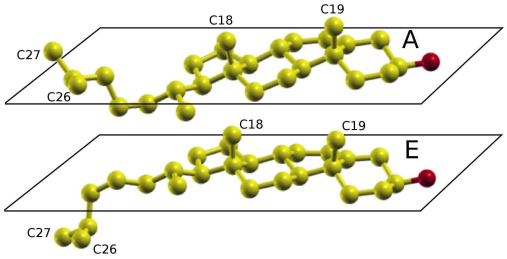


Figure 3.3: Side view of molecules A and E , placed on sterol ring plane. Molecule E is rotated from its crystal cell orientation for a better comparison with molecule A, to demonstrate the tail twist to different directions. Hydrogen atoms are omitted for clarity.

conformations. Molecules such as A and E in the same subcell have different conformations. The C17 chain is twisted towards the α side in molecule E and to β side in molecule A, so that the chains are approximately parallel in the crystal (see Fig. 3.3)

It is also found that the hydrocarbon tails in these molecules are not fully extended resulting in a condensed structure at the hydrophobic center of the bilayer. The ring systems of all cholesterol molecules are found to be parallel to (111). The C18 and C19 methyl groups are suggested to have important role in the packing of this phase, where these methyl groups from different pairs of molecules (such as in A and F) are close enough to cause significant van der Waals interactions. At the hydroxyl end of the bilayer, the C3 hydroxyl groups and water molecules form 2D hydrogen bond networks along $\hat{a}$ and $\hat{b}$ where each oxygen atom forms three hydrogen bonds, oxygens belonging to water molecules being H-donors in two of these and acceptor in one. Due to the high values for the mean square thermal amplitudes of vibration of the oxygen atoms ($0.11 < u^2 < 0.16$ Å^2), it has been suggested that the hydrogen bond network is weak and that the water molecules have an important space-filling role in stabilizing the structure during crystal growth. Migration of water molecules along tunnels parallel to $\hat{a}$ has been suggested as an explanation of the phenomenon of these crystals easily dehydrating on exposure to air.

3.1.3 Anhydrous Cholesterol - Low-Temperature Phase (ChAl)

The first structural analysis of low-temperature phase of anhydrous cholesterol was reported by Shieh, Hoard and Nordman [7]. This phase, like the monohydrate one, is found to form triclinic crystals in P1 space group with eight cholesterol molecules in the unit cell. The cell dimensions are reported to be a=14.172(7) Å, b=34.209(18) Å, c=10.481(5)Å, α=94.64(4)°, β=90.67(4)°, γ=96.32(4)°.

Unlike the ChM phase where water molecules were binding the two layers of the crystal with hydrogen bond networks, in ChAl phase there are two independent hydrogen-bond chains running parallel to c axis (Fig. 3.4). Unlike ChM, the steroid planes of ChAl show no parallel packing. Adjacent unit cells of ChAl demonstrate a local non-crystallographic twofold symmetry with symmetry axis at $x \approx 0.5; y \approx 1.0$. A 180° rotation maps the molecules A, B, C, D into close coincidence with the molecules E, F, G, H of the adjacent cell, respectively. This local pseudosymmetry, although not exactly established, is however reported to be obeyed to a remarkably high degree,

and claimed to leave four unique molecules in the cell (A, B, C, D) which are not held together by hydrogen bonds but only hydrophobic interactions.

The steroid ring skeletons are found to be similar in the eight molecules, with the greatest variation occurring in the torsional units of C4-C5-C10-C1 and C5-C10-C1-C2. This is proposed to be due to C5-C6 double bond introducing flexibility in the steroid region. The hydrocarbon tail is observed to vary more significantly, where two distinct side-chain conformations were reported: In molecules A, B and D the side chains are all-trans, while molecule C has a gauche-trans-gauche sequence at C20-C22-C23-C24.

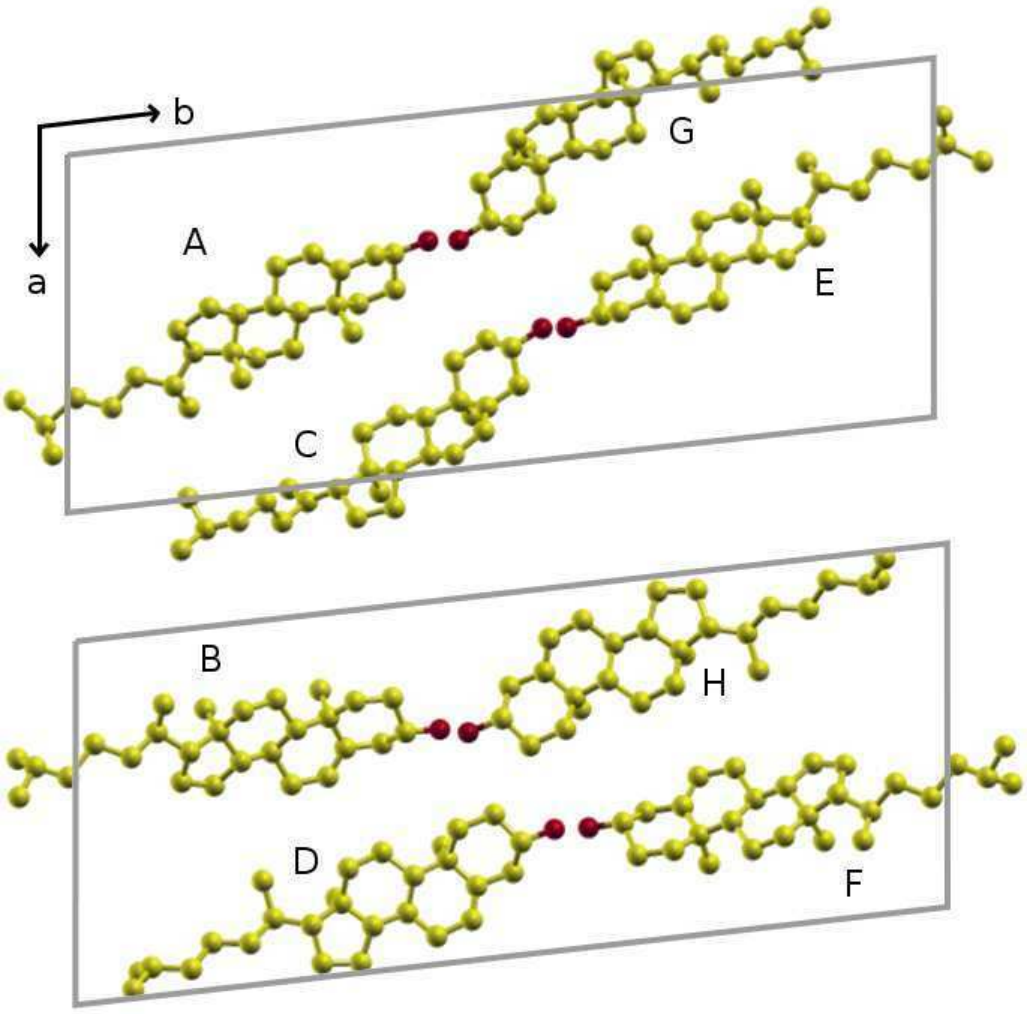


Figure 3.4: ChAl unit cell divided in two groups to improve visibility. Molecule labeling and atomic positions are taken from Ref. [7]. Hydrogen atoms are omitted for clarity.

3.1.4 Anhydrous Cholesterol - High-Temperature Phase (ChAh)

Single crystal anhydrous cholesterol goes through reversible phase transition at 31.8 ± 0.1°C. High temperature phase of anhydrous cholesterol belongs to triclinic P1 space group as do the other known phases [8]. Unlike the common trend of increased symmetry in high temperature phases, ChAh shows a larger unit cell with 16 independent molecules, twice the amount in ChAl phase. The ChAh unit cell was primarily modeled as two ChAl cells, one translated by $a/2$ relative to the other. Further refinement of this model showed that considerable rotation of molecules about their long axis is required for a better model. The reported cell dimensions are a=27.565(10) Å, b=38.624(16) Å, c=10.748(4) Å, α=93.49(3)°, β=90.90(3)°, γ=117.15(3)°.

ChAh is found to have the same bilayer structure of ChAl, with alternating hydrophobic and hydrophilic layers (Fig. 3.5). The hydrogen bond networks lie along c axis and follow the same pattern in both structures.

The differences in molecular conformations between high and low temperature phases can be summarized as follows: Molecules A1 and E2, related with pseudosymmetry in the low temperature phase, turn by 157° about their long axis, while A2 and E1, translationally related counterparts of A1 and E2 remain mostly the same. Similarly, B1 and F2 turn by 87°, while B2 and F1 remain the same. The rest of the molecules C, D, G, H do not display rotation of the steroid skeleton more than by 4°. However side chains of molecules C2 and G1 undergo an important conformational change from gauche-trans to the energetically more favorable trans-trans conformations. The terminal isopropyl groups including carbons C25-C26-C27 of these molecules exhibit extreme thermal motion, which is modeled as disorder, destroying the pseudosymmetry that is observed in ChAl. A rather recent X-ray study of this phase [9] reports precise carbon and also hydrogen atom positions, which will be used and evaluated in our calculations.

3.2 Computational Details

3.2.1 First Principles

Ab initio calculations are performed in the framework of Density Functional Theory using the Quantum Espresso distribution [18]. Quantum Espresso distribution is capable of electronic-structure simulations using periodic boundary conditions and a plane-wave basis set.

Throughout the thesis, three different exchange-correlation functionals

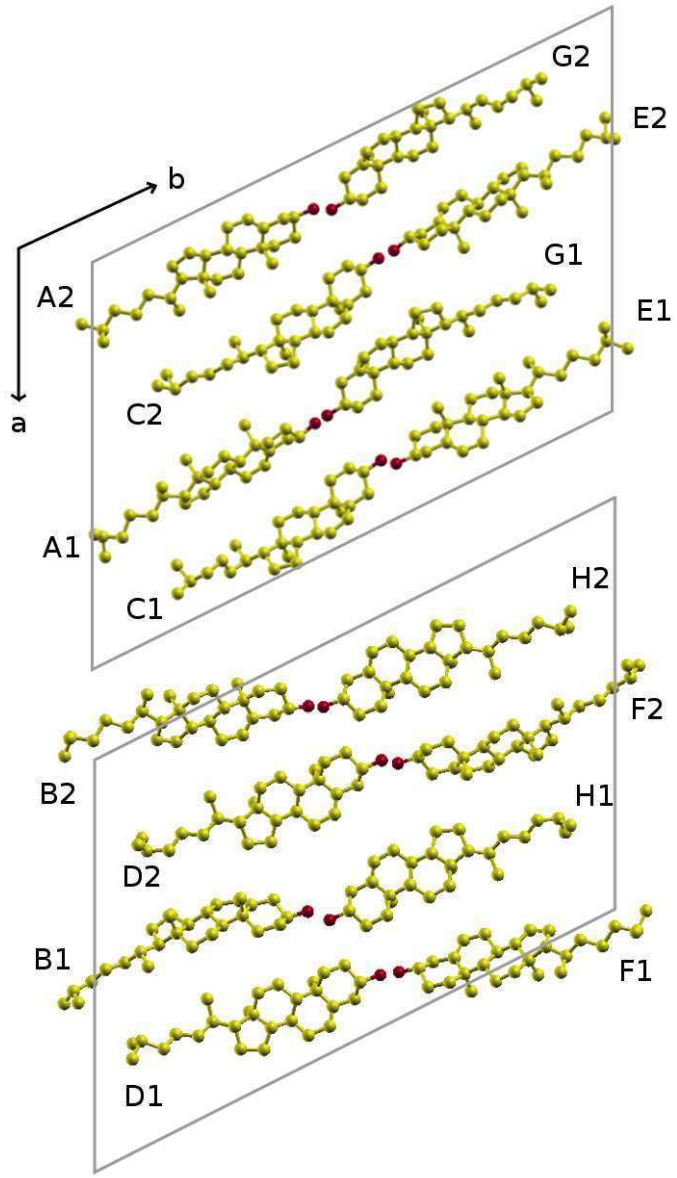


Figure 3.5: ChAh unit cell divided in two groups to improve visibility. Molecule labeling and atomic positions are taken from Ref. [9]. Hydrogen atoms are omitted for clarity.

were employed: the standard Perdew-Burke-Ernzerhof (PBE) [25], exchange part of this functional modified by Zhang-Yang, also known as revised-PBE (revPBE)[44], the van der Waals-aware vdW-DF correlation functional [27] which employs revPBE exchange, recently made computationally efficient by the modifications by Soler [28], and imported to QE package.

The majority of the calculations were done using the vdW-DF functional to account for the van der Waals interactions. PBE and revPBE were used to demonstrate the effect of chosen XC on the NMR chemical shifts later in section 4.3.3. The use of vdW-DF functional enables the relaxation of cell parameters as well as the atomic positions. The intermolecular interactions in cholesterol crystals are mainly due to van der Waals interactions [45]. Due to this property, PBE and revPBE calculations were performed using the experimentally determined cell parameters. This approach, using XC functionals that do not include van der Waals interactions in conjunction with experimental cells is a common practice in studies involving molecular crystals and its validity will be discussed.

We chose to model the atoms in the PAW formalism. As mentioned earlier PAW formalism aims at recovering the accuracy of all-electron calculations with a reasonable cost. Furthermore, GIPAW theory heavily relies on an accurate PAW reconstruction of the valence one-electron wavefunctions in the core region. We constructed our pseudopotentials to be as accurate as possible without being excessively computationally expensive. The cutoff radii (in atomic units) and reference energies (in Rydberg) where the eigenvalues were not used are given in table 3.1:

The projectors are generated following the Troullier-Martins recipe [46] in scalar-relativistic formalism, non-linear core correction was employed for carbon and oxygen pseudopotentials. The local channels were chosen to be l=2, 1, 2 respectively for C, H, O pseudopotentials.

A kinetic energy cutoff of 45 Ry is found sufficient for all calculations. The augmentation charges are expanded up to 220 Ry. Due to the molecular nature of the crystal and large size of the unit cell, calculations are done using Γ point in the Brillouin zone. With this set of parameters, total energy is converged within 1 mRy per atom, the NMR chemical shifts are converged within 0.5 ppm. During geometry optimizations, atomic positions were relaxed until the maximum Cartesian component of the force was less than 1.5 mRy/bohr for any atom. Both ionic and cell optimizations are done using Broyden-Fletcher-Goldfarb-Shanno method as implemented in QE.

Element	(n,l)	r_{cut} [a.u.]	energy [Ry]
C	2s	1.3	eigenvalue
	2s	1.3	0.05
	2p	1.45	eigenvalue
	2p	1.45	0.05
	3d	1.1	0.10
O	2s	1.35	eigenvalue
	2s	1.35	0.05
	2p	1.35	eigenvalue
	2p	1.35	0.05
	3d	1.25	0.15
H	1s	1.0	eigenvalue
	1s	1.0	0.05
	2p	0.75	0.05

Table 3.1: Core radii and reference energies used while generating the pseudopotentials used for cholesterol crystal calculations.

3.2.2 Classical Force Field

As we have seen so far, cholesterol systems can display very complex structures, large unit cells and may display important conformational changes at finite temperature. Although the focus of this work is on first principles calculations, further studies concerning these aspects may not be feasible with *ab initio* methods. Therefore we also performed classical force field calculations as a benchmark of accuracy of this method for future studies.

Classical molecular mechanics calculations were carried out with LAMMPS code [47] using an all-atom force field (FF) as parametrized in Ref. [45]. In the case of ChM, TIP3P water model as implemented in CHARMM [48] was used for the water molecules. Non-bonded interactions were cutoff at 12 Å [49] and calculated with a modified Ewald sum technique [50]. For geometry optimization we first performed molecular dynamics calculations with Langevin thermostat [51] where thermostat temperature was gradually decreased from 298 K to 10 K in 25 ns with a timestep of 0.5 fs. Then several minimization algorithms [52] were consecutively employed to ensure a force convergence of less than 1 mRy/bohr. The use of several minimization techniques is fundamental since minimizations with force fields can pathologically stop before convergence due to discontinuities introduced by interaction cutoff radii. In fact, we observed that this phenomenon occurs more frequently in the case of relaxation of cell parameters. Therefore, to ensure the consistency of precision for both *ab initio* and classical force fields, we compared

our results between these methods only for calculations where cell parameters were fixed at experimentally reported values and only the atomic positions were allowed to relax.

The accuracy of classical force field calculations depends strongly on the force field. It should be noted that the force field that is used in this work, generated by Cournia *et al.*, is fitted mainly for cholesterol molecule in gas phase using the Automated Frequency Matching Method [53]. Nevertheless, it is reported as successful in determining the crystal volume [45]. By repeating our calculations with the positions obtained from this force field, and comparing with experiment findings, we will be performing an important benchmark on the quality of this new FF as well.

3.2.3 Geometry Optimization

We perform our calculations using several levels of geometry optimization from experimental positions as given to fully relaxed structures. We will later study the adequacy of these different procedures in NMR calculations in section 4.3.1. The notation that is followed throughout the thesis is presented as follows:

For the high-temperature anhydrous form (ChAh), all atomic positions including the proton ones are known from X-ray diffraction experiments [9]. Calculations using these experimental positions without further optimization are labeled as OPT-0. As a starting point for the calculations on low temperature phases, ChAl and ChM, we considered the reported experimental structures [6, 7] for non-hydrogen atoms and added hydrogen atoms following the methods of Ref. [54] and Ref. [45]. We then performed several levels of *ab initio* geometry optimization for all three structures: *i*) allowing only the proton positions to relax (OPT-1) *ii*) allowing all atomic positions to relax (OPT-2) *iii*) allowing both atomic positions and unit cell parameters to relax (OPT-3). Results of the molecular mechanics calculations at experimental cell parameters described in the preceding section are labeled as cOPT-2.

3.3 Theoretical results

In this section we report our theoretical results on the structural properties of cholesterol crystals and compare them with available experimental results from the literature.

3.3.1 Cholesterol Monohydrate

Starting from experimentally published atomic positions we perform step by step geometry optimization as explained in subsection 3.2.3. We observe that using the experimental positions for non-hydrogen atoms and positions given by Ref.[54] for hydrogen atoms result in forces as large as 0.18 Ry/au on non-hydrogen atoms when evaluated from first principles. Even though the hydrogen atom positions were relaxed to their equilibrium positions by force field methods while keeping the other atoms fixed, we observe considerably large *ab initio* forces (0.02Ry/au) on hydrogen atoms as well. Nevertheless here we report the experimental (OPT-0) bond lengths and angles that define the cholesterol molecules [Fig. 3.6a and 3.7a].

Note that bond lengths show large variability both within the molecule and for each bond between different molecules, therefore neither the molecules that are linked with subcell translation, nor the ones linked with pseudosymmetry can be recognized using these atomic positions. This result clearly states that experimental positions are far from equilibrium for any calculation which depends strongly on structural properties. This is probably connected with the fact that the amplitudes of atomic thermal vibrations that were observed in X-ray experiments were large [6].

To make sure that the reason for this rather-crude atomic positions is not the inaccuracy in FF relaxations, we perform *ab initio* relaxation of H atoms (OPT-1) [Fig. 3.6b and 3.7b]. The C-H bond lengths change up to 2% however the differences between molecules are even more strongly pronounced verifying that the poor description of the crystal (lack of twinning due to pseudosymmetry or subcell translation that is strongly observed in experiments) originates from the poorly determined carbon positions, and no further relaxation of hydrogen atoms would change the quality of the representation. These observations point out that, for structure sensitive probes such as NMR, direct use of experimentally determined positions, even including a partial relaxation, is likely invalid.

Naturally, we continue allowing the carbon atom positions to relax as well (OPT-2). The profound changes in bond lengths and angles can be seen in the Fig. 3.6c and 3.7c.

In general, the bond lengths and angles gain homogeneity among molecules,

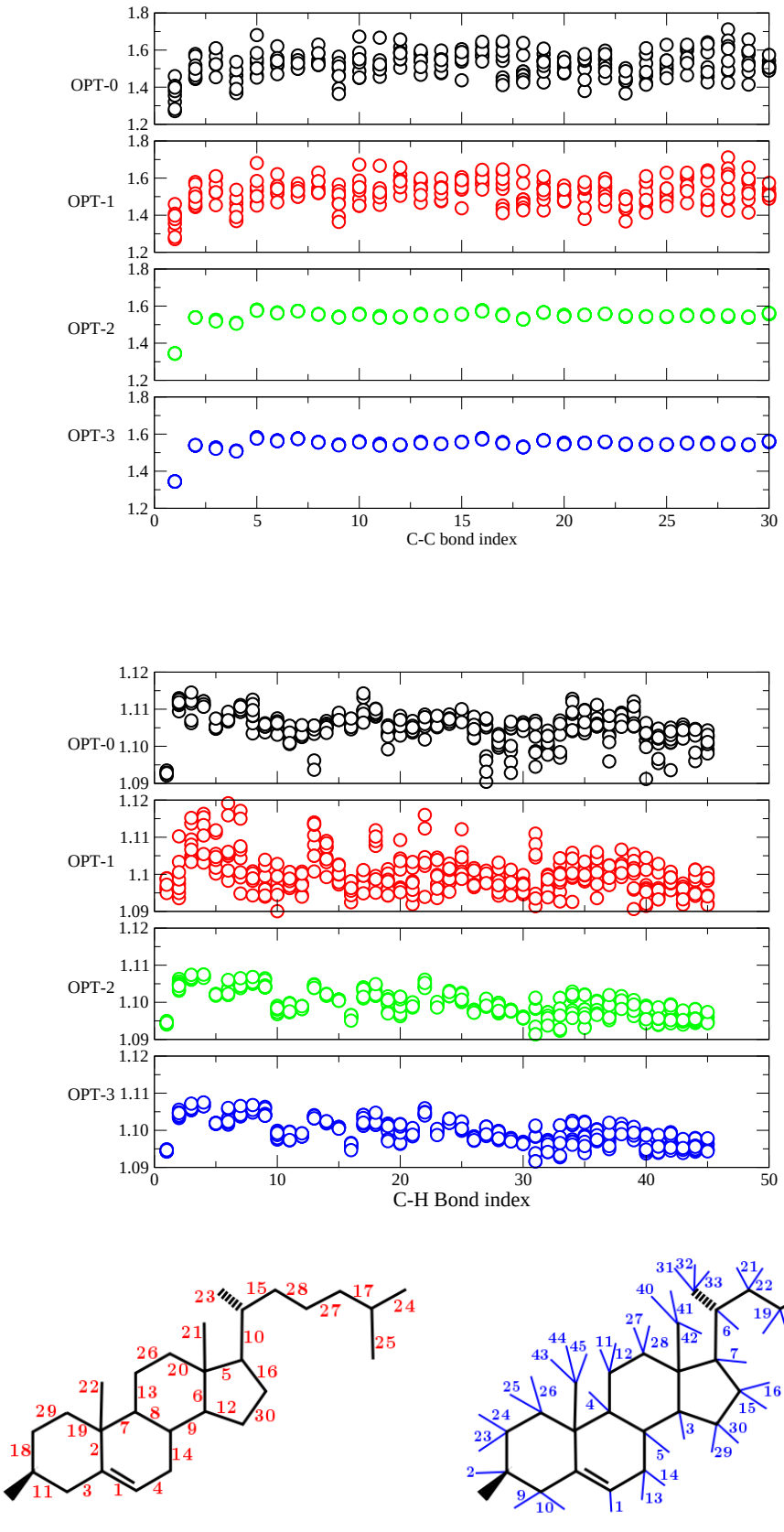


Figure 3.6: C-C (top) and C-H (bottom) bond lengths (in Å) for ChM phase for all levels of optimization considered. Bond lengths belonging to all 8 molecules are shown for a given bond index. In the lower panel the indexing of C-C and C-H bonds are given.

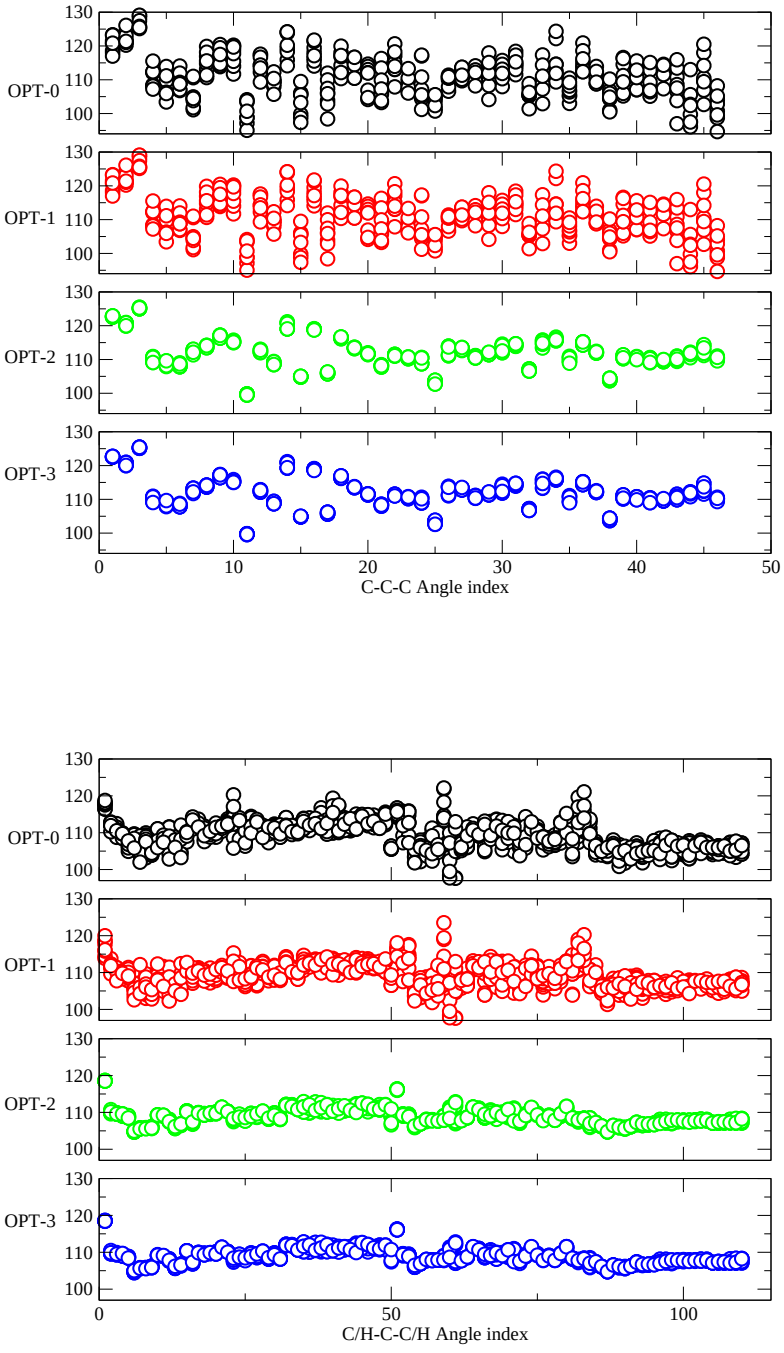


Figure 3.7: C-C-C(top) and C/H-C-C/H (bottom) angles (in degrees) for ChM phase for all levels of optimization considered. Angles belonging to all 8 molecules are shown for a angle index. Since these data are presented for qualitative comparison only, the long list of angle index is not given.

resulting in molecules with similar conformations. Furthermore, with this relaxation twinning among molecules that was observed experimentally is now seen. Molecules (A-B-G-H) and (C-D-E-F) establish two distinct groups for several bond lengths and angles in the molecule, which is more easily spotted based on C-H bond lengths and C-C-C angles. The resulting relaxation looks very satisfactory by reproducing the best known properties of ChM crystals and does this with bond lengths and angles well within the variability observed experimentally [Fig. 3.6a and 3.7a]. However it is obtained with the experimental unit cell. For further applications it is important that intermolecular interactions of these molecules are accurately described. To test for this, we allow the cell parameters to relax fully as well (OPT-3).

We see that relaxing cell parameters does not cause any significant change in the details of the molecular structure. Bond lengths vary less than 0.5% with respect to OPT-2 results remaining compatible with experiment. At this stage we also see that full relaxation changes the cell parameters insignificantly (a=12.37 Å, b=12.47 Å, c=34.50 Å, α=92.6°, β=98.8°, γ=101.7°) and the optimized volume agrees very well with the experimental one, with a very small deviation of +0.19%.

3.3.2 Anhydrous Cholesterol-Low Temperature Phase

A similar analysis of relaxation is performed for ChAl phase. This time the experimentally obtained positions for non-hydrogen atoms are combined with FF relaxation on hydrogen positions by another research group, based on another force field as explained in section 3.2.3. Our calculations show that this set of atomic positions also results in considerable *ab initio* forces, as large as 0.14 Ry/au, on non-hydrogen atoms. Furthermore, the force field relaxed hydrogen atoms also show non-negligible *ab initio* forces on them reaching up to 0.03 Ry/au on single Cartesian coordinate. Since we have seen the same behavior in ChM phase too, we can conclude that for these systems, experimental positions should be used only as guidelines to the structure and further *ab initio* relaxation should be performed to consistently obtain equilibrium positions, especially if one is interested in structure sensitive calculations such as NMR parameters or phonon dispersion.

Nevertheless we continue with the step by step relaxation procedure that have been established. The resulting bond lengths and angles are given in Fig. 3.8, 3.9.

It is obvious comparing Fig. 3.8a 3.9a to the other panels that in OPT-0 level, the structure is over-simplified by almost all C-H bonds sharing the same bond length and further relaxation with *ab initio* reveals a more complex profile. It should also be noted that C positions relax less for this phase

than in ChM, indicating that experimental positions for these atoms are more reliable than in the case of ChM. This can be understood considering the denser, more rigid structure of ChAl, leaving less uncertainty due to vibrational motion. Again, first principle relaxation tend to make C-C bonds and angles more uniform among different molecules with values within the experimental variability.

Still, the analysis of bond lengths and angles shows that the homogeneity of ChM structure is not seen in this phase. Although the range of bond lengths and angles are very similar to those of ChM, ChAl shows a greater variety between molecules. This finding is consistent with the fact that experimentally, a much weaker pseudosymmetry is observed for this phase. Judging from our bond length and angle analysis, we came to the conclusion that twinning occurs only partially between the molecules, and it does not separate the molecules in two groups of equal number of elements as in ChM, or as suggested experimentally (4-4, two groups each having 4 members), although we have observed groups in the form of 3-3-2 (three groups with 3, 3 and 2 members) or 6-2 (two groups with 6 and 2 members). However it should be noted that the effect of temperature is not included in our studies. As the head and tail regions of cholesterol molecules are very flexible, temperature averaging of anharmonic motion may reduce the multiplicity of molecules, giving rise to more similar groups.

For this phase too, a full relaxation of the cell (OPT-3) does not introduce significant changes. Optimized values for cell parameters are a=14.15 Å, b=34.24 Å, c=10.53 Å, α=94.66°, β=90.62°, γ=96.58° and the optimized volume agrees very well with the experimental one, with a small overestimation of +0.36%.

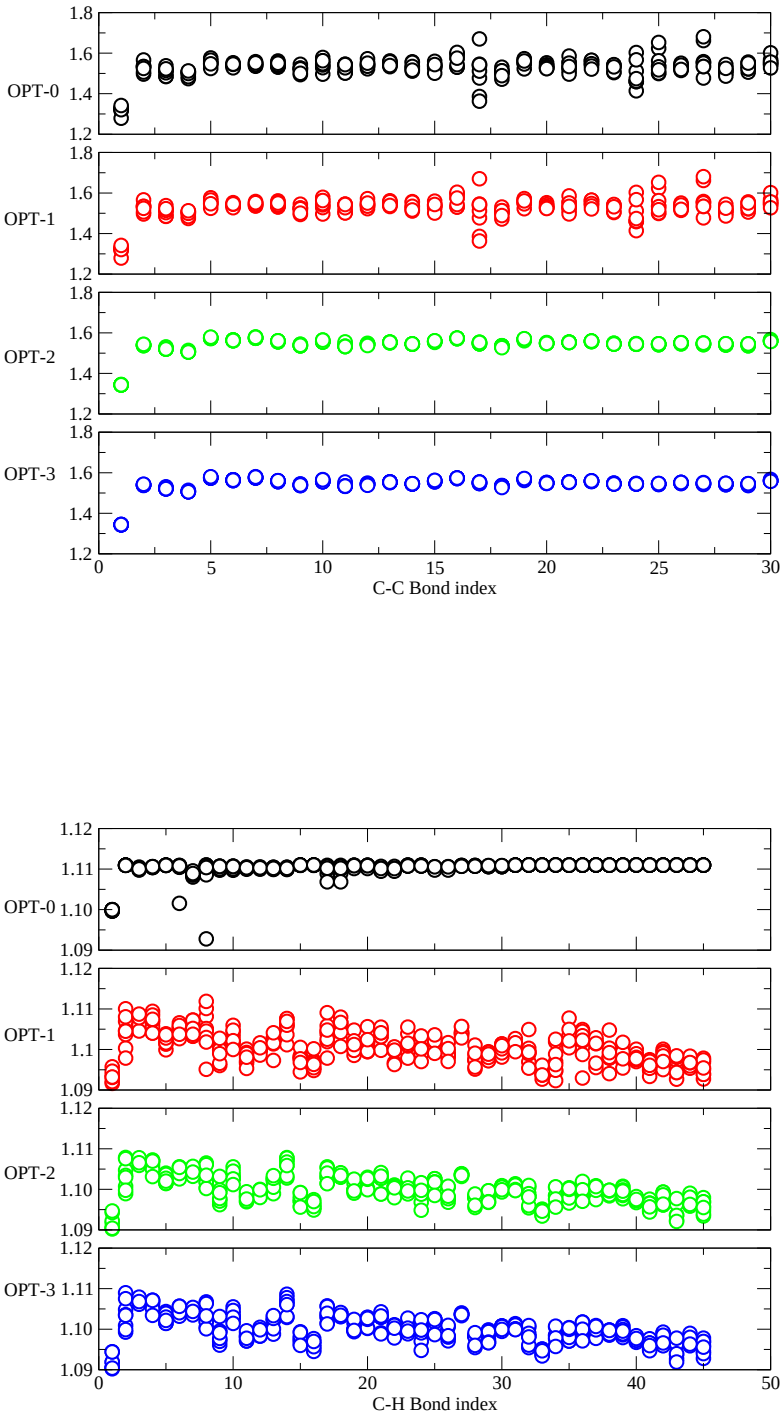


Figure 3.8: C-non H (top) and C-H (bottom) bond lengths in ChAl phase for all levels of optimization considered. Bond lengths belonging to all 8 molecules are shown for a given bond index. See Fig. 3.6 for labeling.

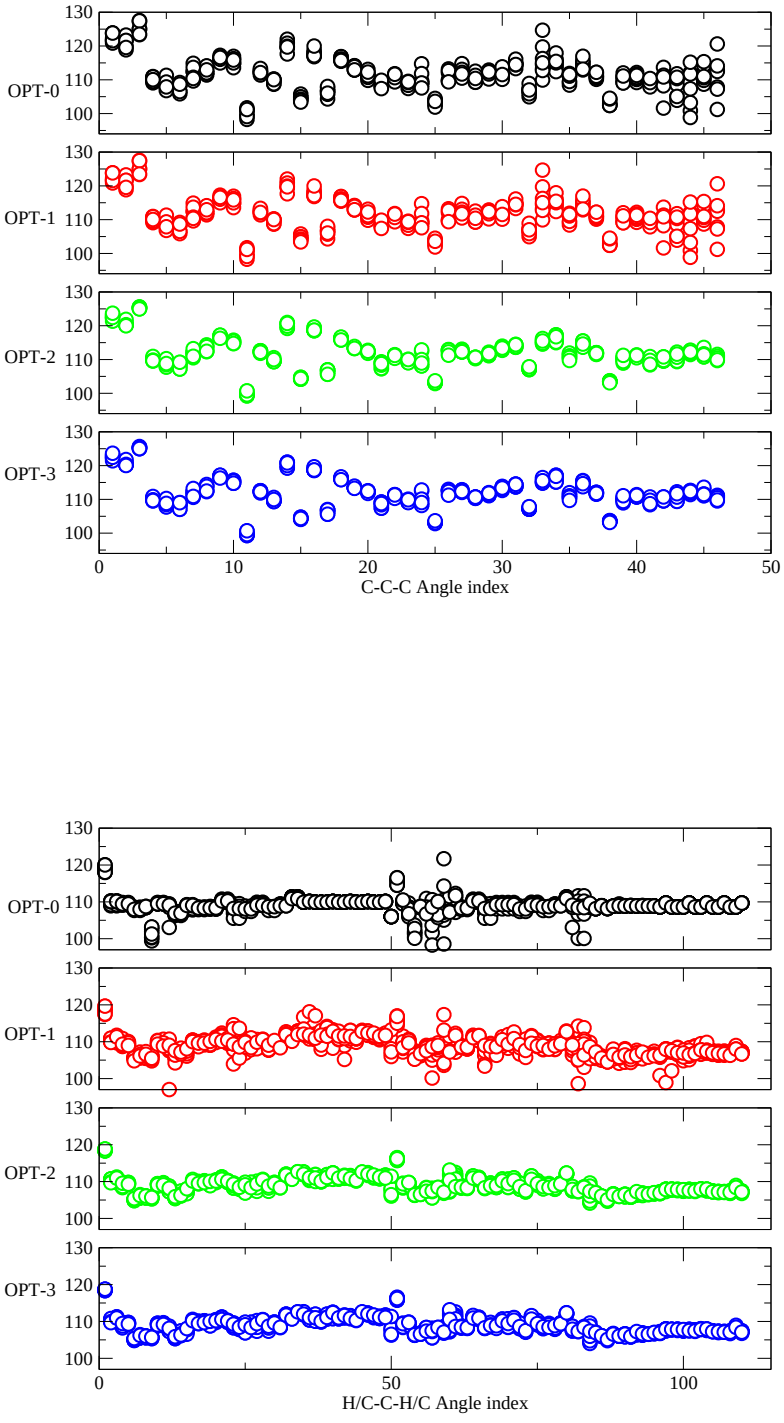


Figure 3.9: C-C-C(top) and C/H-C-C/H (bottom) angles (in degrees) for ChAl phase for all levels of optimization considered. Angles belonging to all 8 molecules are shown for a given angle index. Since these data are presented for qualitative comparison only, the long list of angle index is not given.

3.3.3 Anhydrous Cholesterol-High Temperature Phase

Thanks to the high-precision of a comparably recent X-ray study, both hydrogen and carbon positions are available for ChAh phase. Therefore we start our relaxation procedure treating all atoms on the same level, skipping the separate hydrogen position relaxation (OPT-1).

Our static calculation reveals that at experimental positions, the *ab initio* forces on the atoms reach as much as 0.3 Ry/au, as in the case of ChM and ChAl phases, therefore no improvement is observed between theoretical and experimental results, even with recent, technically superior, X-ray experiments.

To understand the source of this phenomenon, let us turn our attention to C-H bond lengths reported for ChAh phase (Fig. 3.11). We see that the C-H bonds are considerably longer in our optimized structure. We suggest that the main reason for this discrepancy is the extreme thermal motion of the molecules. In the harmonic motion of C-H bonds dominated by bond angle fluctuations, the average bond length would appear smaller than the actual bond length (see Fig. 3.10).

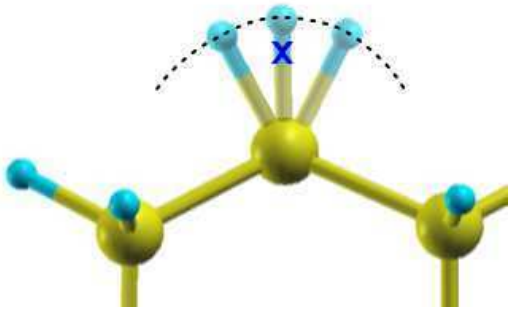


Figure 3.10: Due to the strong wagging motion of the atoms at finite temperature, average bond lengths measured based on average atomic positions may appear shorter. This is shown above for a C-H bond where the path of H atom during the wagging motion and the estimated average position of H atom are indicated.

With a similar argument, Ref. [9] mentions that C-C bond lengths are shorter in experimental results due to the wagging motion of the molecule. Since we do not consider thermal motion, *ab initio* determined bond lengths should be larger than the experimental ones. However a quick glance to Fig. 3.6, 3.8 and 3.11 might appear to suggest the opposite, theoretically obtained bond lengths are slightly shorter than the experimental ones. However a more careful look would reveal that the lack of the wagging motion

of the hydrocarbon tail in theoretical calculations can still be observed on selected bonds where it is expected to be more apparent: Let us examine the average C25-C26 bond length, terminal of hydrocarbon tail. Experimental results for ChM, ChAl, and ChAh phases are respectively 1.5143 , 1.5236 , 1.4838 Å. The same average from theory is (in the same order) 1.5431, 1.5445, 1.5452 Å. Notice that, as expected, experimental results underestimate the average bond length, furthermore the difference between theory and experiment follows the order of strength of the thermal motion in different phases (ChAh > ChM > ChAl). This analysis once more emphasizes the importance of performing *ab initio* relaxation and the handicaps of room temperature experiments for structures with extreme thermal motion.

Continuing with the cell relaxation, also in this case, we see that relaxing cell parameters does not cause any significant change in the details of the molecular structure, as in other phases. The cell parameters change insignificantly (a=27.37 Å, b=38.93 Å, c=10.74 Å, α=93.59°, β=90.80°, $\gamma = 117.73°$)and the optimized volume is overestimated by +0.34%.

3.3.4 Use of force fields in determining structural properties

We have shown the importance of *ab initio* relaxation to obtain a consistent set of atomic positions for cholesterol crystals. However, these structures are involved in complex biological systems such as mammal plasma membranes, or as precursors in enzymatic reactions [55]. For such calculations, *ab initio* methods are still limited, instead, force field methods are feasible. Here we test the accuracy of a force field that is proposed for cholesterol molecules [45], by comparing the optimized positions for these two methods.

In Fig. 3.13 and 3.14, the differences between *ab initio* and FF optimized bond lengths and angles are given for ChM and ChAl phases. Note that the comparison is only made between OPT-2 and cOPT-2 levels, as explained in 3.2.

Note that the majority of C-C distances are underestimated ($\approx 1\%$) by the FF optimization while C-H ones are overestimated ($\approx 3\%$) for both phases. A more significant picture appears, however, if we consider the differences in angles. As can be seen, with the exception of molecule G (in Fig. 3.2), the differences are in the same order for both phases. In the monohydrate ChM phase, considerable rotation of several bonds for molecule G can be seen both in C-C-C and in the corresponding C/H-C-C/H angles. ChM phase calculated by this FF, due to this rotation, no more displays the well known pseudosymmetry, that instead is recovered well by first principles relaxation.

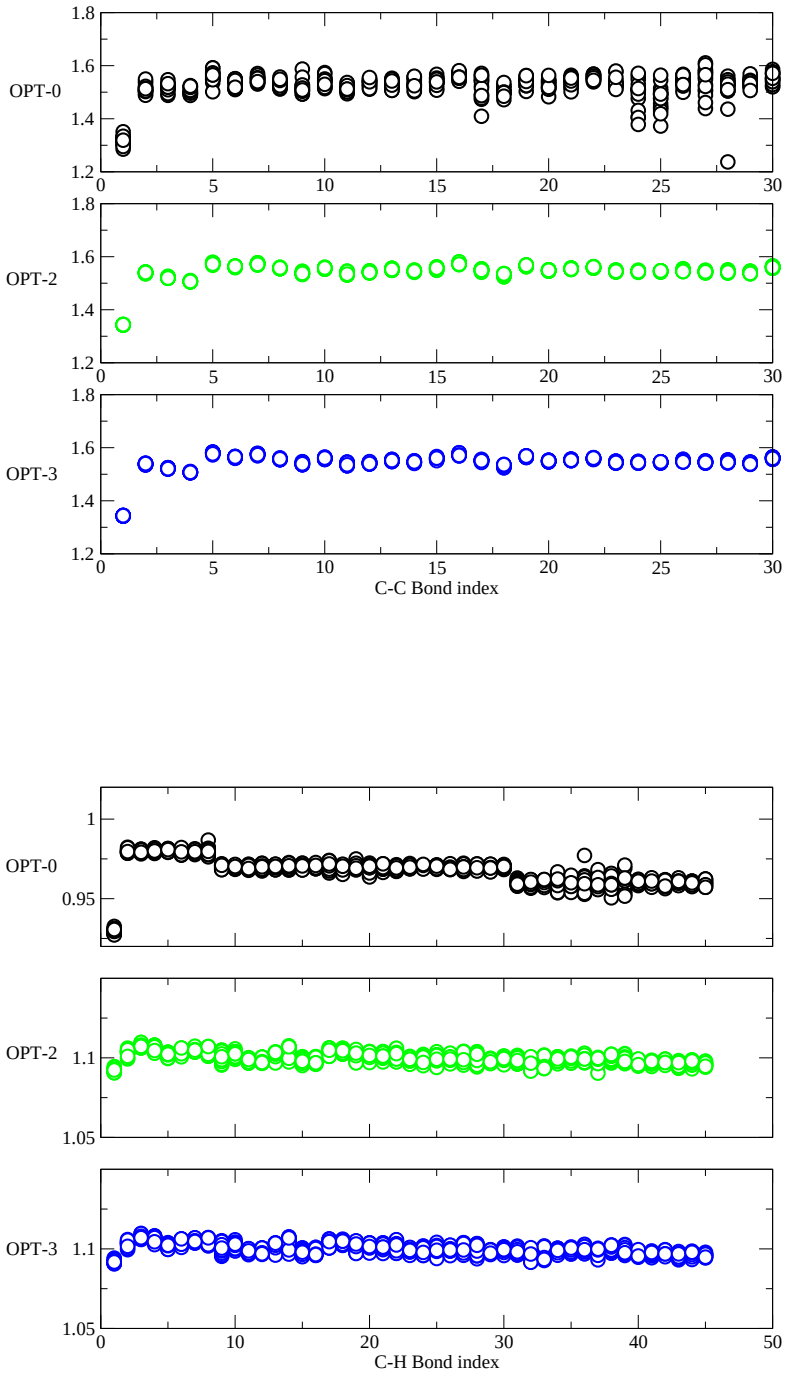


Figure 3.11: C-non H (top) and C-H (bottom) lengths in ChAh phase for all levels of optimization considered. Bond lengths belonging to all 16 molecules are shown for a given bond index. Bond index 1 corresponds to the double bond between C5-C6 atoms (top) and the bond between C6-H atoms (bottom). Note that theoretical and experimental C-H bond lengths fall in different ranges.

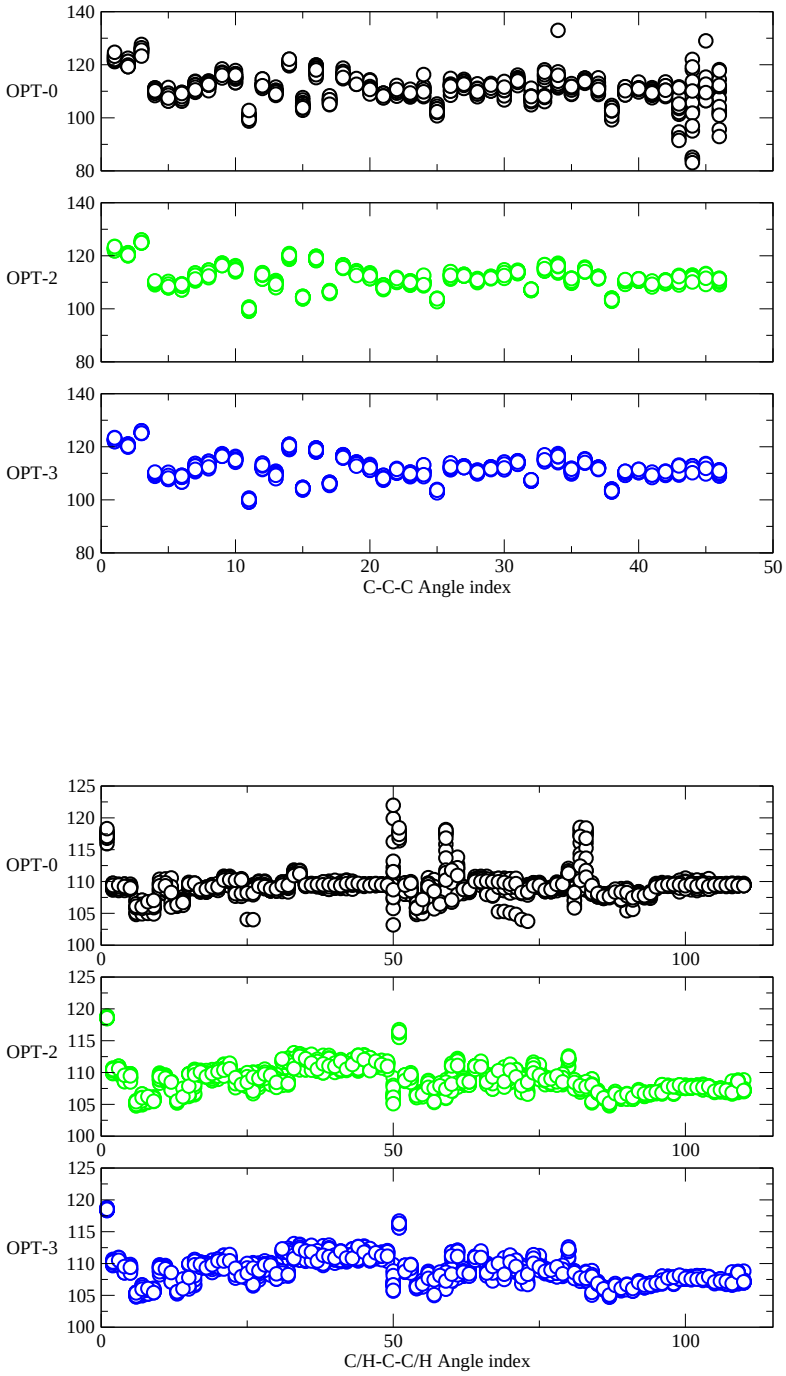


Figure 3.12: C-C-C(top) and C/H-C-C/H (bottom) angles (in degrees) for ChAh phase for all levels of optimization considered. Angles belonging to all 16 molecules are shown for a given angle index. Since these data are presented for qualitative comparison, the long list of angle index is not given.

We investigated the reasons for this phenomenon exploring the following possibilities: *i)* The relaxation procedure with force fields can give pathological results due to discontinuity in non-bonded potential. However several other trials with different minimization schemes yield the same rotation behavior, therefore we believe there should be another factor enhancing the pathology, *ii)* The total energy landscape of this structure holds several local minima with similar energies, one of them being the rotated molecule case. However starting from different initial positions did not cure the appearance of this feature. *iii)* As explained previously in section 3.2 we have employed the T3PIP potential for water molecules in the bilayer. This potential is constructed for water in liquid form however it has been used in intermolecular layers in a previous study [54]. The inadequacy of this potential for the intermolecular water layer can result in an incorrect description of energy ordering and ground state prediction, especially when several minima close in energy are present. Note also that no rotation is observed for the anhydrous phase. *iv)* It is also plausible that this force field, being generated for a single cholesterol molecule, might behave differently in the sparse ChM phase and in the dense ChAl phase.

Therefore we conclude that, a more suitable force field should be used for water molecules for a better description of hydrated cholesterol condensed phases, keeping in mind that the inadequacy of the description of water layer, although trapped between layers, can have effects extending throughout the crystal. Furthermore, a FF with good agreement at the molecular level does not necessarily represent the crystal properties, especially in complex crystals such as cholesterol where several phases exist. Our results also indicate that structure determination for such systems should be performed with caution, when *a priori* knowledge on the phase does not exist.

For ChAl phase, the agreement between ab initio and force field method does not display drastic differences. Further test of the quality of structure prediction with force fields will be made in sec. 3.3.4 by using a sensitive probe, NMR chemical shifts.

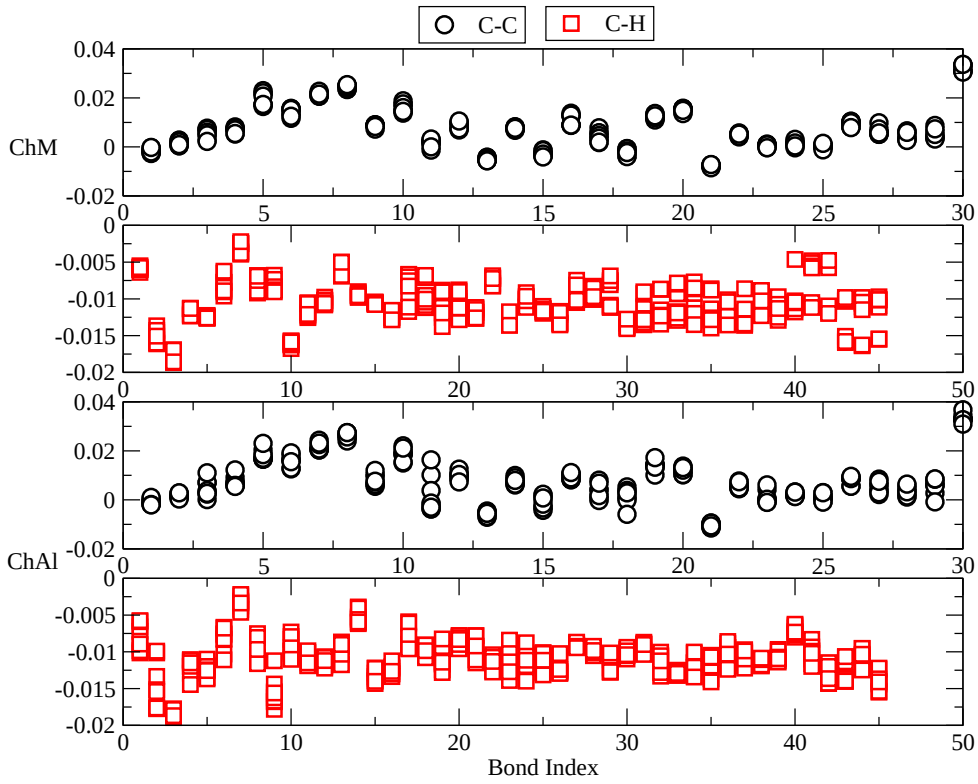


Figure 3.13: Differences (ab initio -FF) between ab initio relaxation at OPT2 level and FF calculations for C-C and C-H bond lengths in ChM and ChAl phases (in Å)

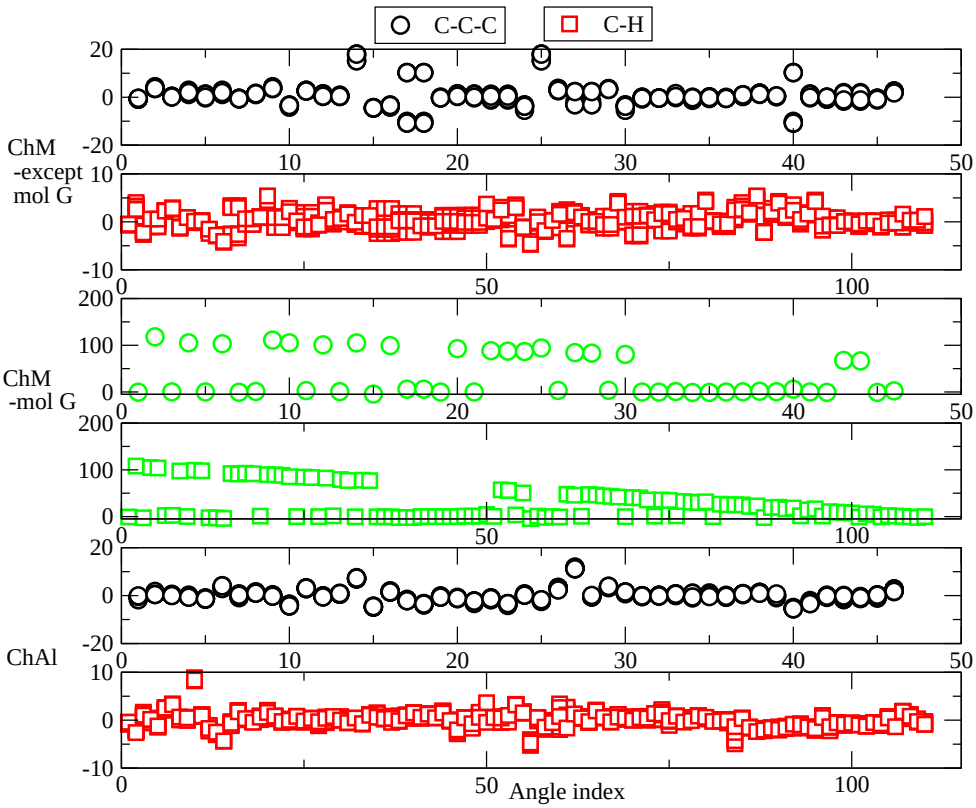


Figure 3.14: Differences (ab initio -FF) between ab initio relaxation at OPT2 level and FF calculations for C-C-C and only C and H including angles in ChM and ChAl phases in the units of degrees. Note that for clarity, data for molecule G in ChM phase is given separately (see text).

Chapter 4

NMR Study of Cholesterol Single Crystals

Nuclear Magnetic Resonance (NMR) is a powerful technique in structure determination and it is widely used in several fields ranging from solid state physics to molecular biology. Unfortunately there is no generally valid, simple "recipe" which uniquely assigns a structure to a given spectrum. For simple systems one often refers to empirical databases to interpret the measured spectra in terms of the underlying structure. However as the NMR methods advance, more complex systems can be studied and highly predictive tools are required for an unambiguous mapping of spectra to the correct structure.

In this direction, several methods have been developed to determine the NMR spectrum from first principles. Among them, quantum chemistry methods have been extensively used for the interpretation of solution-state NMR spectra, as they are limited to studies of molecules and finite systems. For periodic systems, a solid-state physics approach, Gauge Including Projector Augmented Waves, has been widely employed and has proved to successfully reproduce experimental Solid State NMR chemical shifts in a wide variety of systems.

In this chapter we use our new implementation of GIPAW with PAW pseudopotentials described in chapter 2 to obtain NMR chemical shifts for the three known phases of cholesterol crystals and compare them with the available experimental data. We demonstrate that, with our implementation, calculation of NMR parameters for such complex molecular crystals is now feasible with sufficient accuracy to distinguish the different crystalline polymorphs. Then we analyze the effect of certain theoretical factors on the obtained results such as structural relaxation and exchange-correlation functionals. Additionally we address the presence of some systematic error in GIPAW calculations and suggest a solution. After achieving increased

accuracy thanks to the correction procedure we introduce, we attempt to obtain an unambiguous peak assignment for these complex spectra, which has not been available experimentally. Furthermore we will also explore the intermolecular interactions in these crystals based on our results.

4.1 Computational Details

Thanks to the new full implementation of the PAW formalism within the GIPAW method, generation of new pseudopotentials for chemical shift calculations are no more necessary. The same pseudopotentials used for geometry optimization can directly be used for the determination of NMR parameters. First principles calculations yield the absolute chemical shielding tensor for each nucleus, $\vec{\sigma}(r)$. Isotropic chemical shielding, $\sigma_{iso} = Tr[\vec{\sigma}/3]$, is compared to the experimental isotropic chemical shifts by using the standard expression: $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$. In this work we take σ_{ref} as 167.5 ppm based on our study on cyclobutane (see section 2.5). The theoretical spectra are drawn using normalized Lorentzian distribution centered at the chemical shifts with an arbitrary broadening corresponding to a FWHM of 0.2 ppm to achieve better visual comparison with the experimental spectra. For the preceding self consistent field calculations the same setting of section 3.2 is used with a tighter convergence threshold on the density for an accurate representation of one electron wavefunctions.

4.2 *Ab initio* NMR Spectra of Cholesterol Crystals

Experimental ^{13}C NMR chemical shifts of cholesterol crystals have been previously reported by Guo and Hamilton [56] by cross-polarization and magic angle spinning (CPMAS) method with high power decoupling to prevent the dipolar interactions due to H atoms interfere with the ^{13}C spectra. The chemical shifts were referenced to the carbonyl carbon resonance of glycine (176.06 ppm from tetramethylsilane (TMS)). In these experiments, it is reported that peaks separated as little as 0.2 ppm are resolved. Spectra obtained from these experiments are extracted from Fig.1 of Ref. [56] manually, compromising accuracy. Thanks to the research group of Dr. Neeraj Sinha at Centre of Biomedical Magnetic Resonance, Lucknow, India, we also have access to a set of recent, unpublished experimental spectra, again with CP-MAS method and higher resolution[57]. For these experiments, digital data are available and for the sake of accuracy, we will be using these results for

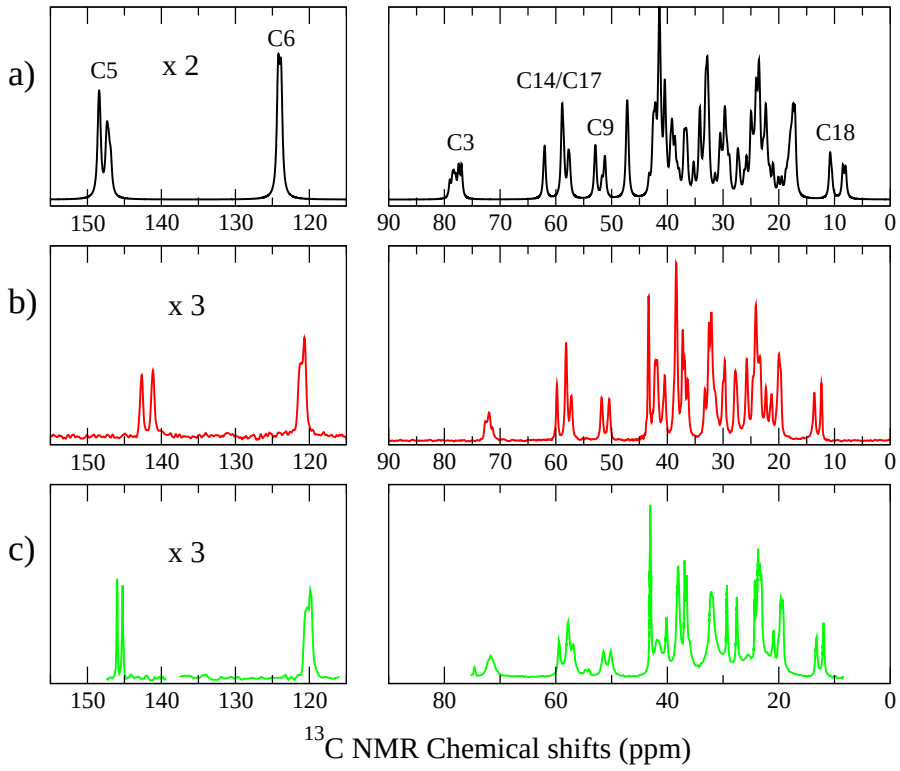


Figure 4.1: ^{13}C NMR chemical shifts spectrum for ChM phase a) this study b) Ref. [57], c) Ref. [56]. The discontinuity in this particular spectrum is due to manual extraction of the data. The high-shift end of spectra are rescaled for clarity.

the comparison between experiment and theory.

Let us familiarize ourselves with the chemical shift spectrum of a cholesterol crystal, for example ChM phase, based on the experimental data (see Fig. 4.1b). As described before in section 3.1, there are eight molecules in the unit cell of ChM. In the NMR spectrum, multiple signals of the same carbon corresponding to inequivalent molecules in the crystal unit cell can be resolved for the isolated peaks associated to C5, C6, C3, C14/17, C9 and C18 (see Fig. 3.1 for carbon atom labeling). The other carbons might also give multiple signals but their overlap prevents further resolution. These isolated peaks are also the only ones that have been assigned experimentally. Unfortunately, the majority of the resonances fall in the 10-50 ppm overlap range. Due to cross-polarization, intensities of peaks belonging to carbons with different hydrogen environment are not directly comparable, as cross polarization typically results in higher intensity for carbons with more hydrogen neighbors.

In Fig. 4.1, 4.2 and 4.3, we report our theoretical spectra together with the experimental ones for ChM, ChAl and ChAh phases respectively. Considering the complexity of these structures, the agreement between experiment and theory is remarkable. Note that the differences between phases involve shifts in peak positions as small as 2 ppm and these are correctly resolved by our calculations.

For ChM phase, pairing of molecules due to pseudosymmetry, therefore occurrence of double peaks, can be seen for C5, C9, C18 and, albeit slightly, for C6 atoms. The C3 peak is broad as in the experimental findings, suggesting that the broadening might originate from structural differences between molecules, in addition to the proposed vibrational motion [56]. The composite peak of C14/17 is also in very good agreement with the experimental observation. The isolated peak around 47.2 ppm in our calculations belongs to C13, which we propose to identify with the sharp feature at 43.4 ppm in experiments. This discrepancy can be understood considering that GIPAW is known to over-(under-) estimate the high-(low-) ppm shifts [58]. Also, due to the same reason, the overall shape of the 10-50 ppm region is expected to agree with experiment only qualitatively. We observe this behavior clearly for the lowest shielding peak, C5, which has the highest variation from experiment, with an overestimation of 6.0 ppm. This systematic error in the results obtained via GIPAW calculations will be addressed in subsection 4.4. Nevertheless, we see that the agreement of the overall NMR peak shapes is apparent and multiplicities are correctly represented, implying that both the optimized structure and the implementation of GIPAW method are satisfactorily accurate.

Examining the NMR theoretical spectrum for the less symmetric ChAl

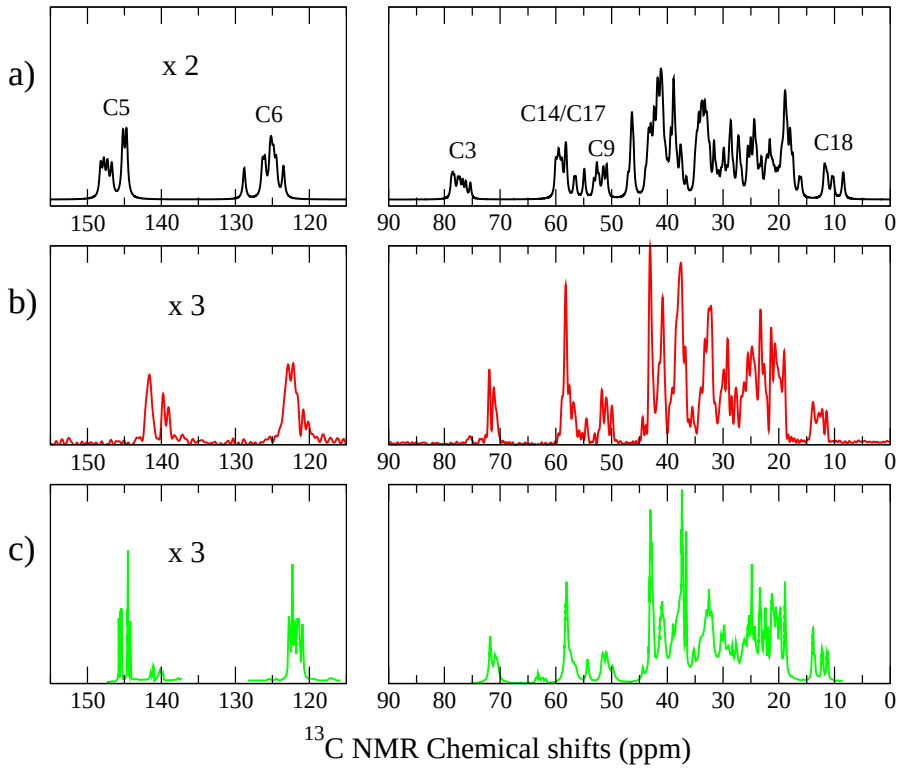


Figure 4.2: ^{13}C NMR chemical shifts spectrum for ChAl phase a) this study b) Ref. [57], c) Ref. [56]. The discontinuity in this particular spectrum is due to manual extraction of the data. The high-shift end of spectra are rescaled for clarity.

phase, we see that sharp double peak features, originating from pairing of molecules, are missing for most of the carbon atoms. This is perfectly in line with the results of our structural analysis (see subsection 3.3.2), and in very good agreement with the experimental findings. Note that, as predicted, rather weak pairings between molecules can be seen in C5, (displaying a line shape characterized by two groups) and C18 (three groups). For this phase, the differences between the two experiments should also be noted, such as the one for C6. Comparing our spectrum with experimental ones, we observe a satisfactory agreement, the main discrepancies occurring for C3 peak, where our results report higher multiplicity than in the experimental one. A possible explanation for this can be attributed to large thermal motion of the molecule head and possible vibrational averaging of the resonances originating from C3 in two distinct groups. For a better understanding of this issue, NMR calculations at room temperature would be necessary. Such calculations are not trivial as the thermal average spectrum should be obtained by the superposition of several representative spectra obtained from shielding tensors averaged over times of the order of Larmor period (nanoseconds). It should also be noted that the error bar of the calculations, 0.5 ppm, may fail to represent complex narrow line shapes such as the one of C3. Comparing our results for this phase with the one of ChM, a clear distinction is visible. Therefore, in spite of the mentioned discrepancies, our calculations can distinguish these two phases as much as the experiments can.

ChAh phase is the largest system where *ab initio* NMR calculations have been reported to date. The result of this both computationally and fundamentally challenging system is given in Fig. 4.3. Note that the increase in the number of molecules in the unit cell can be observed comparing the distinguishable peaks with respect to ChAl phase, both in the experiment and in our theoretical spectra. Again, except C3 peak, the overall NMR spectrum is in good agreement with the experiments, although small discrepancies are more apparent compared to other phases. There is a clear resemblance between ChAl and ChAh phases in terms of chemical shifts, in agreement with the experiments and the structural analysis.

We observe that the C18 peak shows the most characteristic variation between phases: double peak in ChM, triple in ChAl and composite in ChAh. It therefore acts as a clear fingerprint for distinguishing cholesterol crystal phases from one another. Our results for this peak and in general for the overall spectra show that *ab initio* calculations can be used to distinguish different phases. This observation is of uttermost importance for the emerging fields of NMR crystallography for organic systems and diagnostics based on NMR spectra as it gives confidence in the use of GIPAW calculations in aiding NMR crystallography and structure determination based on NMR

chemical shifts.

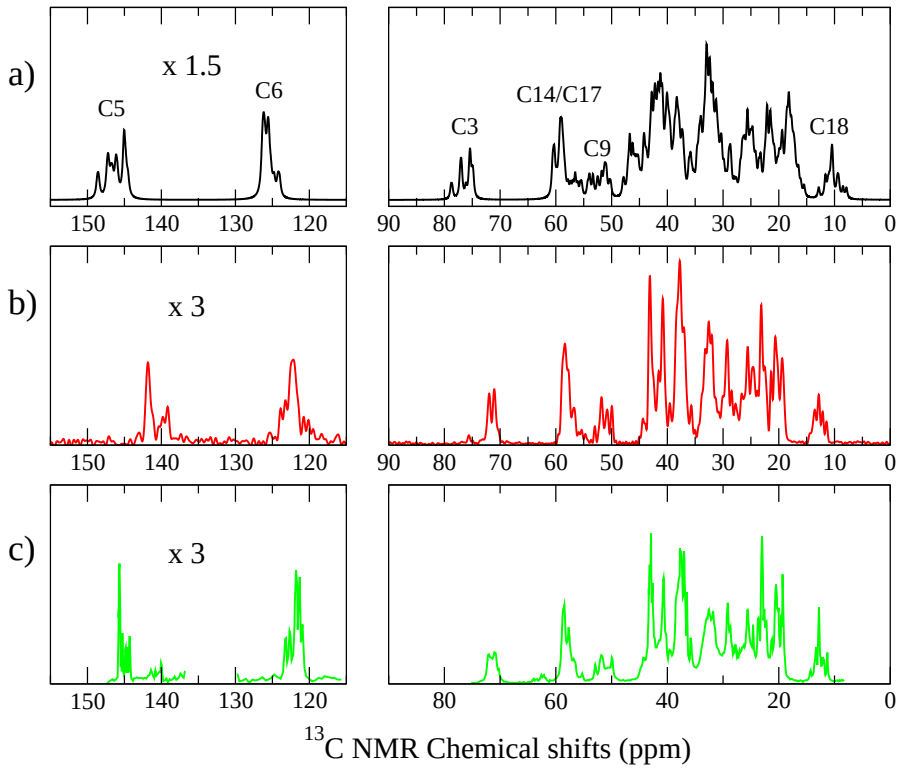


Figure 4.3: ^{13}C NMR chemical shifts spectrum for ChAh phase a) this study b) Ref. [57], c) Ref. [56]. The discontinuity in this particular spectrum is due to manual extraction of the data. The high-shift end of spectra are rescaled for clarity.

4.3 Factors Affecting Theoretical NMR Spectra

The satisfactory agreement between theoretical and experimental NMR spectra of CLR crystals reported in the previous section, critically depends on a number of factors that we briefly review here.

4.3.1 Effect of Structure Optimization on *ab initio* NMR Spectrum - ChM

NMR is a powerful probe of structure however, as emphasized before, the relationship between chemical shifts and the underlying structure is not simple. For *ab initio* calculations, this prevents us to estimate the error propagation in NMR calculations due to the accuracy and precision of the optimization schemes used, i.e. there is no simple relationship that estimates the error in a certain NMR chemical shift due to error in the bond lengths that involve the resonating nucleus etc. However we believe that it is important to have an estimate of the effect of the quality of the optimization scheme used, especially to appreciate the soundness of common practices such as directly using experimental positions as they are or relaxing only the hydrogen atoms when X-ray results are available. To demonstrate this effect, here we discuss in the case of ChM the NMR spectrum corresponding to each of the optimization steps introduced in section 3.2.3. As we have already seen in the previous section, our NMR calculations at the fully relaxed positions (OPT-3) agree well with the experiment. By comparing the NMR spectrum at each optimization level to OPT-3 results, we can comment on their adequacy.

OPT-0

We have seen in section 3.3 that at OPT-0 stage, where experimental positions were used for non-hydrogen positions and FF methods for hydrogen ones, our calculations result in large forces on atoms, suggesting that further relaxation is necessary. We test if the disagreement between experiment and theory is an artifact of the theoretical methods by calculating the NMR spectrum at OPT-0 positions (see top panel of Fig. 4.4). Comparing theoretically obtained spectra of this structure and experimental results, we see that within the sensitivity of NMR, experimental atomic positions do not provide sufficiently accurate structural information. Double-bonded carbons C5 and C6 are still distinguishable from the rest due to their large shifts; however the lack of symmetry that have been observed in section 3.3 results in separate peaks for all eight molecules. It is also the chemical shifts of these atoms that show the largest deviation from the experimental ones, followed by C3. The chemical shifts of these atoms vary over 30 ppm among molecules.

OPT-1

Further optimization of the hydrogen atom positions by *ab initio* methods results in changes up to 2% in C-H bond lengths and up to 4 ppm in chemical shifts. A general trend of increase in the shifts can be observed with carbons in hydrogen proximity. This can be correlated with a general trend of decrease in C-H bond lengths (see Fig. 3.6). However the overall improvement in the spectrum is insignificant, indicating once more the complexity

of the relationship between NMR spectra and the underlying structure.

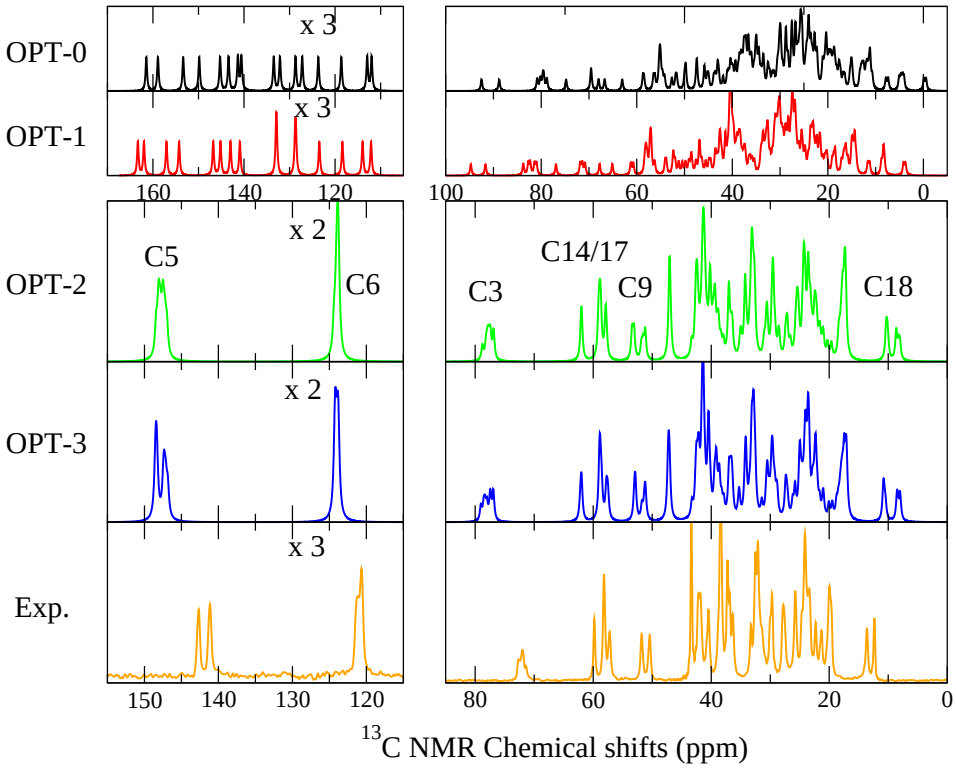


Figure 4.4: ^{13}C NMR chemical shifts spectrum for ChM phase at several optimization levels compared to experimental results of Ref. [57]. Note that in OPT-0 and OPT-1 levels, spectral peaks agree with the experimental ones poorly and show a wide spread, as a result, a different scaling was necessary for the presentation of these spectra.

OPT-2

We have seen that after this stage of relaxation, the bond lengths and angles change profoundly, i.e the differences between molecules decrease, individual C-C bond lengths varying less than 0.5% among the molecules. This homogenization profoundly improves the spectrum. Chemical shifts corresponding to C5, C6, C3, C14/17, C9 and C18 can be easily distinguished and can be seen to have the same features as the experimental results. The sharpness of the C6 peak however does not agree with the experiment.

OPT-3

Allowing the cell parameters to relax, we have seen that the resulting unit cell is in good agreement with the experiment, therefore the spectrum is not expected to change significantly. We observe that even a small change in cell parameters allows the water molecules and the tail region to rearrange such that bond lengths vary within 0.3% with respect to the OPT-2 values. This rearrangement further improves the spectrum, where C5 double peak is resolved clearly and instead of the sharpness of C6 peak that was observed in OPT-2 level, now the shoulder feature can be observed.

Effect of structure optimization on other phases

Fig. 4.5 and 4.6 show a similar analysis for the other two crystalline phases of Cholesterol. We see that the use of experimental positions without further optimization results in poor NMR spectra for these phases too. We can see that in the case of ChAh, even the most recent, high precision experimental results do not prove to be accurate enough when the sensitivity of NMR parameters is considered. Note also that, similar to ChM phase where from OPT-0 to OPT-1 relaxation the increase in C-H bond lengths was accompanied by a slight increase in the shifts, in ChAh phase too, experimentally underestimated C-H bond lengths in OPT-0 give rise to a large underestimation in shifts. This suggests that a considerable correlation between bond lengths and chemical shifts exists, however only observable when bond lengths change drastically and homogeneously throughout the structure. This correlation may be useful in detecting some systematic errors in calculations/experiments but is not enough for an estimate of chemical shifts based on bond lengths only.

In summary, the dependence of the NMR chemical shifts on underlying structure is very strong yet very complicated. Therefore calculations of chemical shifts require a rigorous structure optimization of all atoms and cell parameters. The precision obtained with *ab initio* calculations is found to be not only sufficient for an accurate description of NMR spectra, but also essential as NMR spectroscopy is sensitive to structural details to a higher

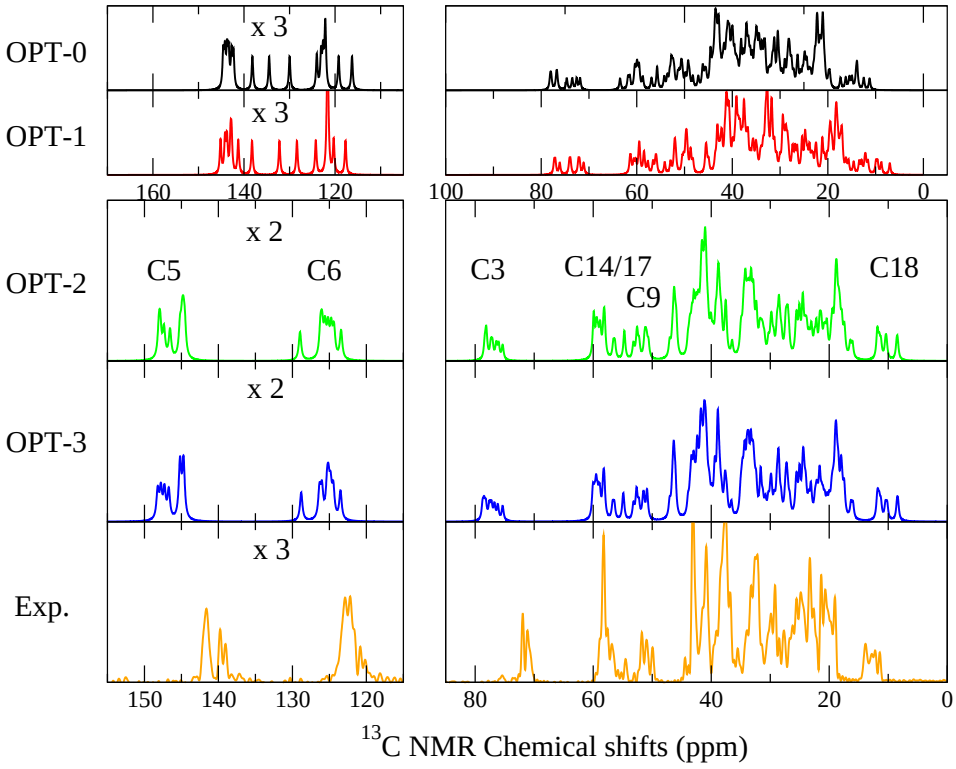


Figure 4.5: ^{13}C NMR chemical shifts spectrum for ChAl phase at several optimization levels compared to experimental results of Ref. [57]. Although not as severe as in the case of ChM, the agreement between OPT-0, OPT-2 spectra with the experiment is evident. The scaling for these spectra is kept same as the one of ChM, Fig.4.4, for a better comparison.

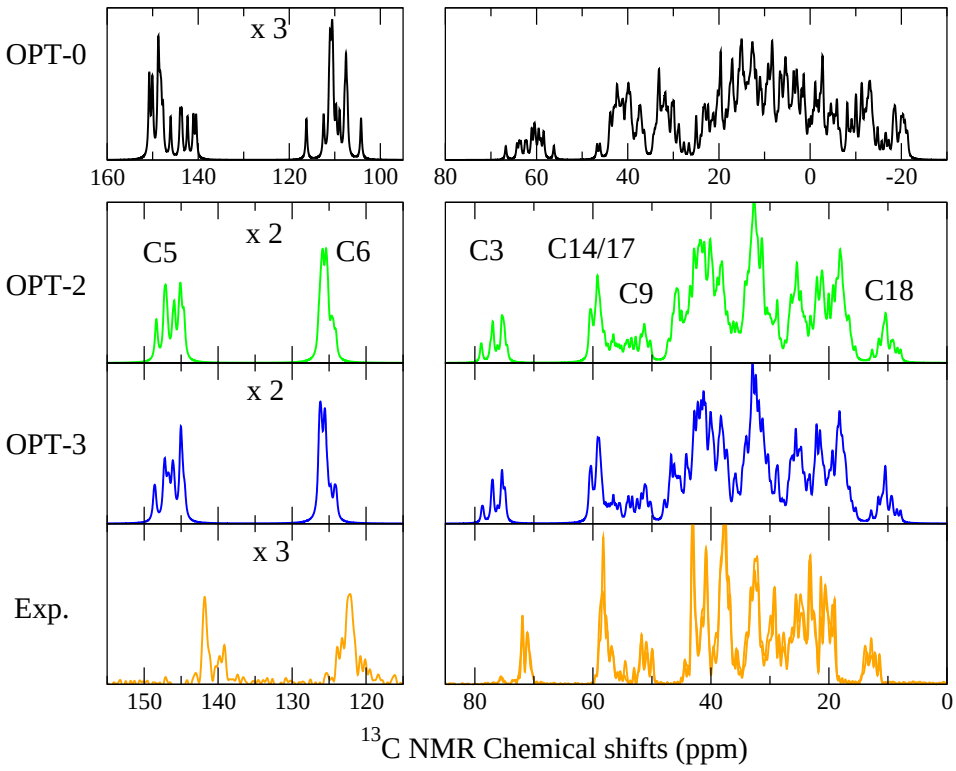


Figure 4.6: ^{13}C NMR chemical shifts spectrum for ChAh phase at several optimization levels compared to experimental results of Ref. [57]. Note that the OPT-0 spectrum shows poor agreement with the experimental results and is spread over a very broad range. As a result, spectrum corresponding to this optimization level is given with a different scaling.

degree than can be currently determined experimentally.

4.3.2 cOPT-2 or Can we obtain accurate enough optimized structures using FFs?

Classical all-atom force field calculations are commonly employed for systems with complex long range interactions and over thousand electrons. The force field studies of cholesterol in the literature [45, 59] reproduce the experimental volume, bond lengths and angles satisfactorily. Based on these reports, and in view of the importance of accurate structural optimization discussed previously, we performed cOPT2 level optimization using a classical force field as explained in section 3.2.2 and afterwards performed NMR calculations at the determined positions to test whether atomic positions obtained with force field methods are accurate enough to achieve a reliable modeling of NMR parameters.

The resulting spectrum is reported in Fig. 4.7. Note that the lineshape of C5, C3, C14/17 peaks are different from experimental and *ab initio* results. Although the differences in chemical shifts do not exceed the difference between *ab initio* and experimental results, an important characteristic of this phase, the presence of double peaks, is not observed for these carbons. One of the reasons for this discrepancy can be the choice of water potential. The T3PIP water interaction potential is derived for liquid water, which differs from the water in the middle of the bilayers. To eliminate this bias, we compare the spectra obtained by *ab initio* and force field methods for ChAl phase (lower panel of Fig. 4.7). Qualitatively, the agreement between the two is closer than it is in the case of ChM, revealing the importance of the water interaction potential. It can be concluded that pure cholesterol crystals can be simulated using this force field for a qualitative modeling of the NMR parameters.

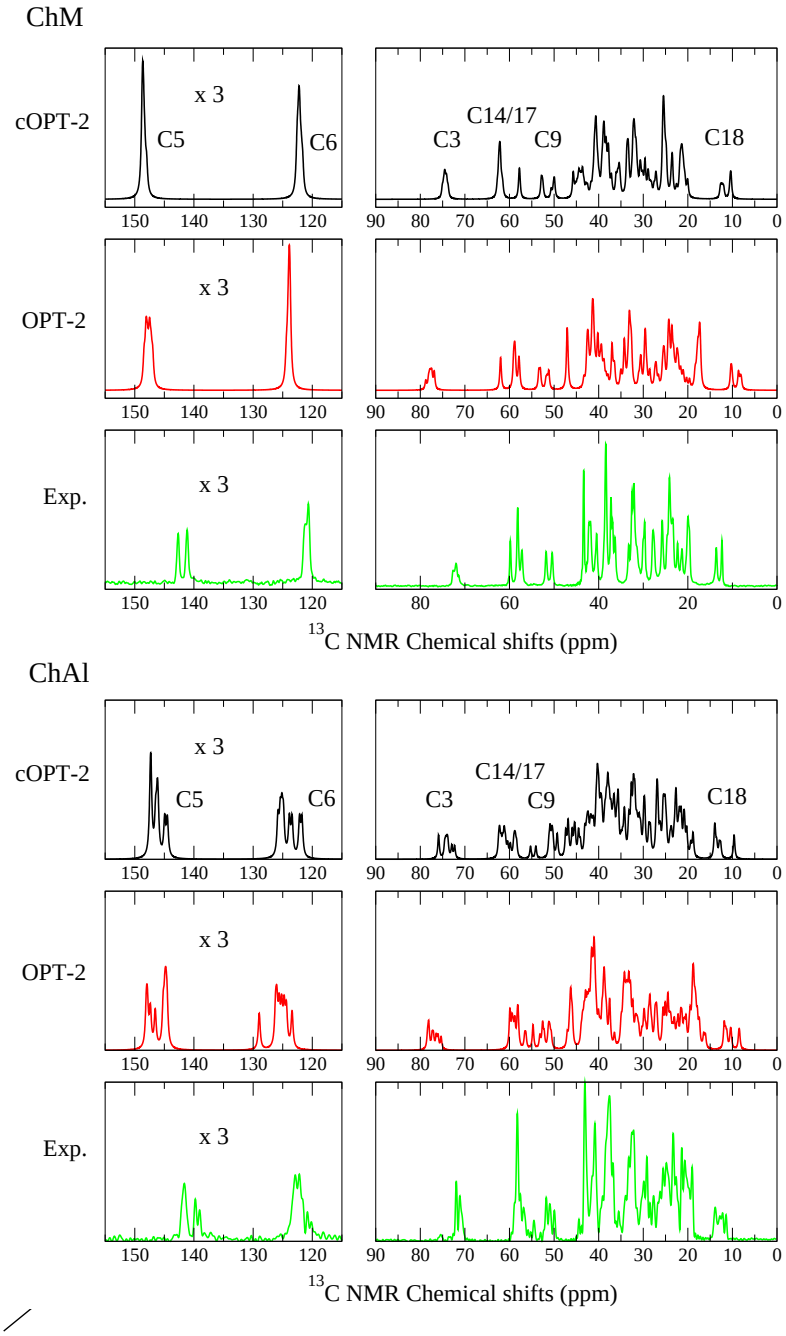


Figure 4.7: ^{13}C NMR chemical shifts spectrum for ChM (top) and ChAl (bottom) phases. The results of cOPT-2 level optimization, where atomic positions obtained via FF methods are used, are compared to the ones of OPT-2 and experiments from Ref. [57].

4.3.3 Effects of exchange-correlation functionals

Since the cholesterol crystals are held together mostly by van der Waals forces, unit cell volume and intermolecular distances are expected to depend greatly on inclusion of this interaction. In addition, based on our OPT3 results, we observe that the changes due to cell relaxation do not cause great variations in the calculated NMR spectrum. Considering these results, one may argue that van der Waals forces might be omitted as long as the experimental cell is employed in the calculations. In fact, this had been a common practice in the past to overcome the difficulties of employing van der Waals interactions in DFT. Here we repeat our OPT2 calculations, at experimentally determined cell parameters, with revPBE and PBE functionals to test the accuracy of this practice for NMR parameters. The resulting spectra (Fig. 4.8) are similar to the experimental one in terms of their main features. As can be seen easily, the multiplicity of the C5, C6, C18 peaks are represented well, indicating that the correct symmetry is found and chemical shifts are robust with respect to the exchange-correlation functional used. However the overall spectrum is shifted towards higher shielding and expands to a larger range than the experimental one. Examining the structural changes among different exchange correlation functionals, we observe that changes in bond lengths are homogeneous throughout the molecule, which explains the preserved symmetry and qualitatively adequate spectra that is obtained (Fig. 4.9). While moving from vdW-DF to revPBE and PBE, C-C bond lengths decrease and C-H bond lengths increase throughout the molecule and the chemical shifts decrease more or less uniformly. It should be noted that a similar observation was made in the previous section, 4.3.1, where a uniform increase of C-H and/or C-C bond lengths were accompanied by an increase in chemical shifts. However this time C-C bond lengths decrease while C-H bond lengths increase and this results in lower shifts, implying that there is no simple correlation between chemical shifts and bond lengths.

We further repeat our calculations at OPT-0 positions for ChM using PBE XC functional. Since in OPT-0 no relaxation is performed, the quality of the resulting spectra is poor if compared with the experiments, however by this procedure we can establish whether the dominant reason behind the discrepancies between NMR parameters obtained using different XC functionals rests in the difference in wavefunctions or in the optimized structure. Due to different XC functionals, the one-electron solutions are different for PBE and vdW-DF. As can be seen in Fig4.10, NMR parameters are sensitive to these differences on the order of several ppm in peak positions. Therefore we conclude that even in the cases where structural relaxation is not per-

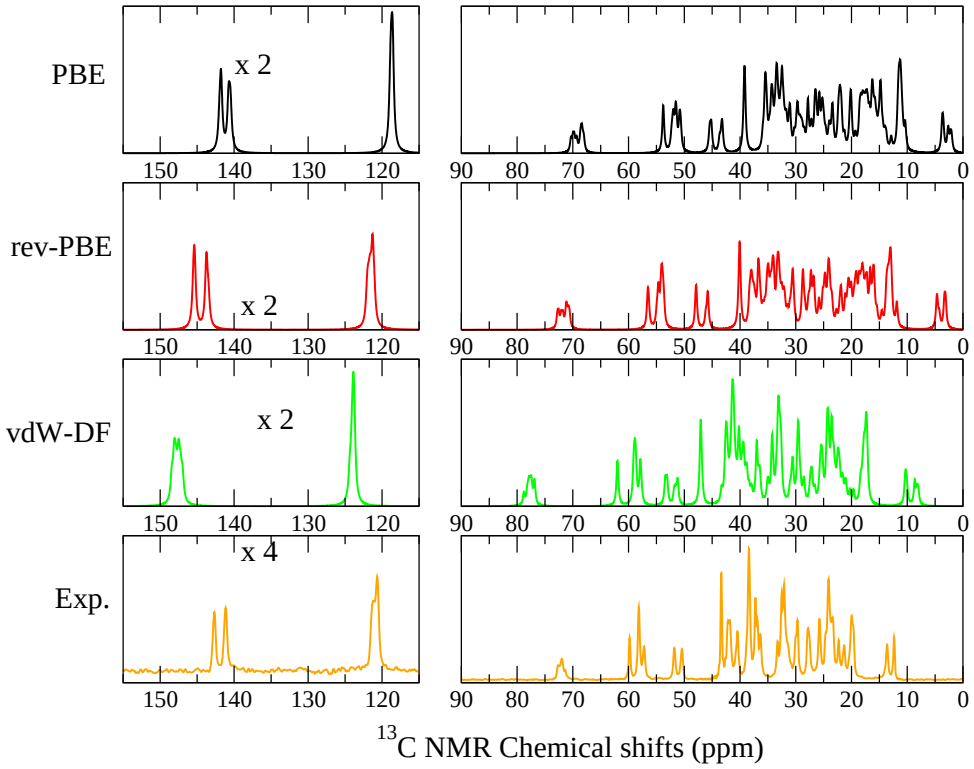


Figure 4.8: Comparison of ChM ^{13}C NMR chemical shift spectra using different XC functionals with OPT2 level geometry optimization. The experimental results are taken from Ref. [57]

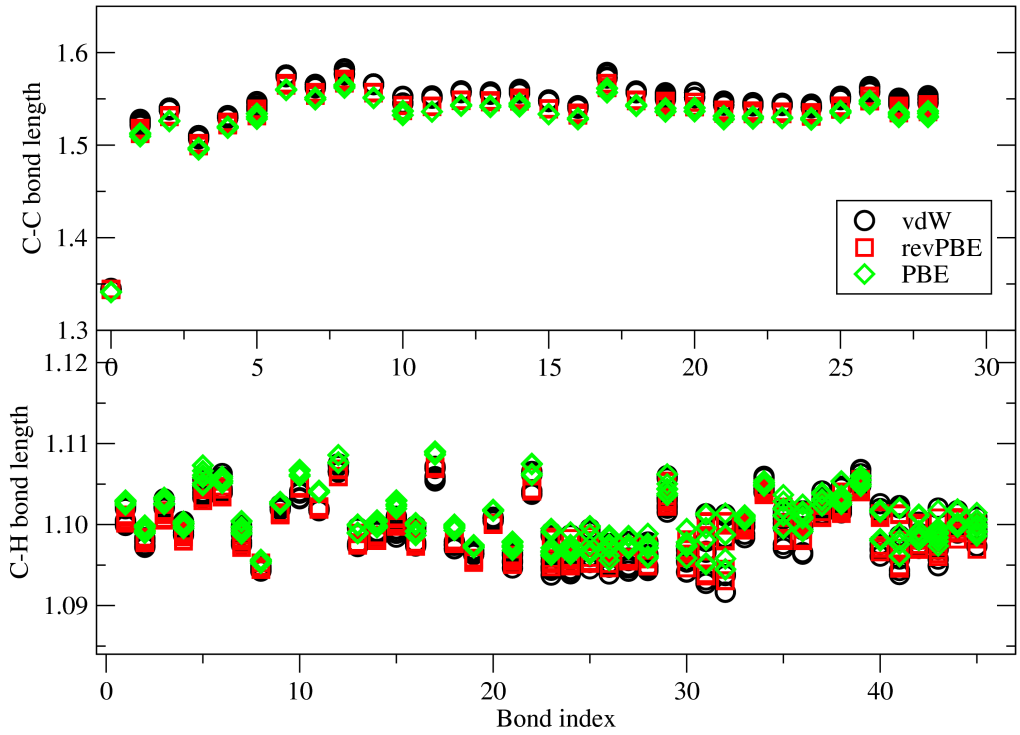


Figure 4.9: Bond lengths (in Å) with different XC at OPT-2 level geometry optimization for ChM phase. The bond labeling is given in Fig. 3.6.

formed, XC functionals taking into account the vdW interactions should be used for molecules strongly affected by these interactions. However we also note that the shape of resonance peaks are rather stable for both XC functionals, allowing one to use PBE for a qualitative insight into the spectral features.

4.4 A Systematic Error in GIPAW calculations and a Possible Solution

As it has been previously reported [58], GIPAW tends to over(under)-estimate the high(low)-ppm resonances with respect to experiment. In order to improve the comparison, a separately determined reference can be used for each carbon group associated to the isolated peaks in the NMR spectrum, following a procedure proposed in Ref. [58]. Remember that the original referencing

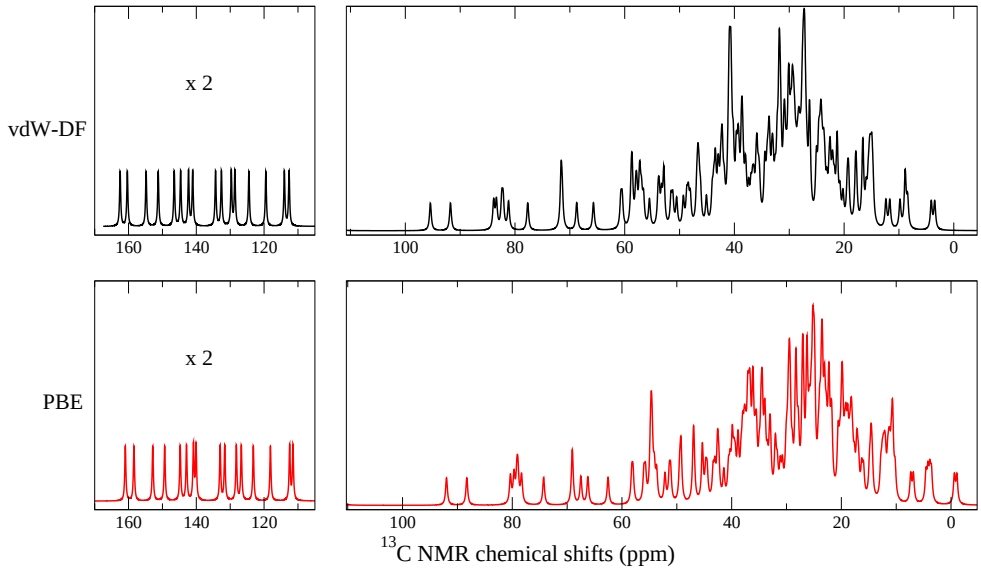


Figure 4.10: Comparison of ChM spectra using vdW-DF (top) and PBE (bottom) XC functionals at OPT-0 positions.

was done using the expression : $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$ where $\sigma_{ref} = 167.5$ ppm. For these isolated peaks, the reference is corrected with the difference between experiment and theory such that the mean of the two peak positions match: $\sigma_{ref} \rightarrow \sigma_{ref} - c$, c correction values used are -6.0, -3.1, -5.8, -1.0, -1.0, +2.5 ppm for the isolated peaks of C5, C6, C3, C14/17, C9 and C18 respectively. The result of this referencing correction for the isolated peaks of the three crystal polymorphs is given in Fig.4.11.

However the majority of carbon resonance peaks in cholesterol fall into the range of 10-50 ppm, where such practice is not possible. For these peaks we explore the physically more satisfactory possibility to define corrections to our reference for each carbon group based on its local chemical environment, defined by the number of direct bonds to hydrogen atoms, i.e. quaternary (CH0), single H (CH1), double H (CH2) and triple H (CH3) groups are shifted with different references. To obtain the correction for CH0 group, $c_{CH0} = -4$ ppm, we use the value that assigns the peak of C13 to the sharp feature at the lower edge of the overlap region (see the discussion in section 4.2). For CH1 group we have three carbon signals for C14/17 and C9 and a correction value $c_{CH1} = -1$ ppm gives good agreement with the experimental peak positions for all of them as previously discussed (see

Fig.4.11). We add no correction for CH2 group as our reference structure is cyclobutane, which is made up of CH2 type of carbons. For CH3 group we use the isolated C18 and the rather sharp C19 peaks (assigned to the sharp feature at the low-shift range of the overlap region) which yield $c_{CH3} = +2.5$ ppm (see again Fig.4.11). Note that we expect this correction to be phase independent, and use the same values for all structures considered. We show in Fig. 4.12 the comparison for the complete spectra between experiments and the theoretical results with the corrected referencing. An improvement of the agreement in the 10-50 ppm region can be seen with respect to the not-shifted spectrum presented in Fig. 4.2. Looking more closely at the spectra in Fig

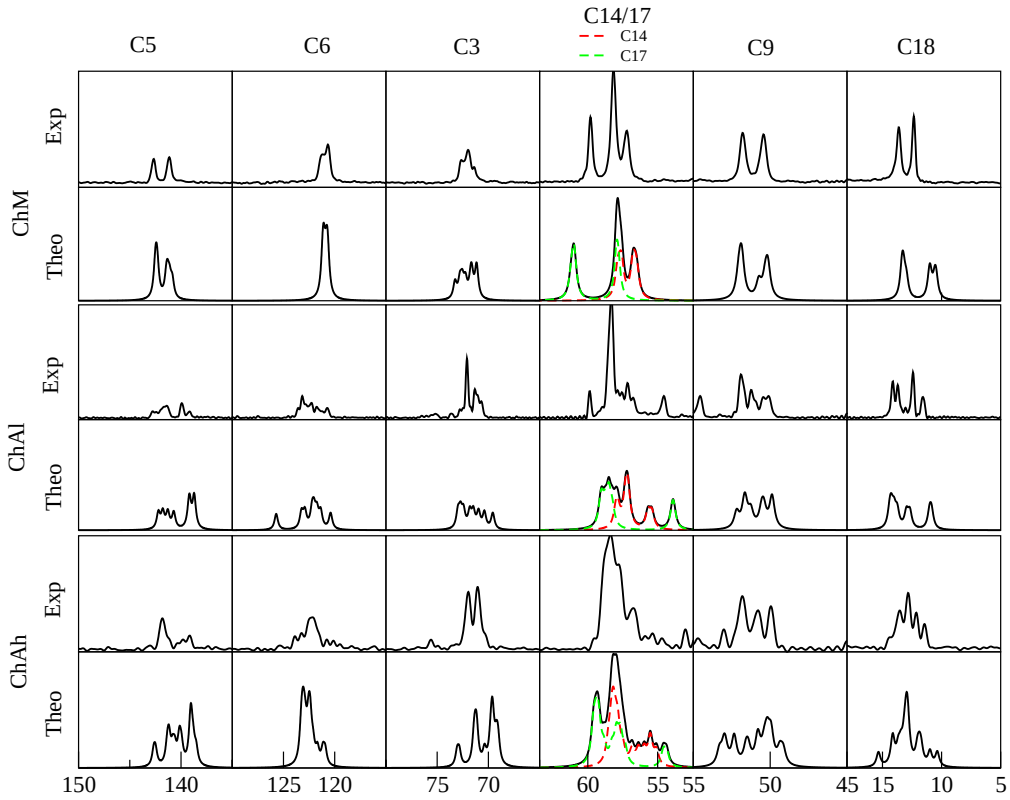


Figure 4.11: Isolated peaks for all structures after reference correction. Experimental data is taken from Ref. [57].

4.11, especially at the C14/17 and C18 peaks, we see that theoretical peaks are rather 'stretched' with respect to experiment, suggesting a correction of reference more complicated than a shift, perhaps a linear correction in the form of $\sigma_{ref} \rightarrow a \times \sigma_{ref} - c$ where both a and c are to be determined. However

with few available isolated peaks and without further theoretical insight, we follow the simpler procedure. Yet this correction scheme is still more fundamentally grounded than just shifting all the peaks one by one to match the experimental ones and has the potential to be more predictive. In order to

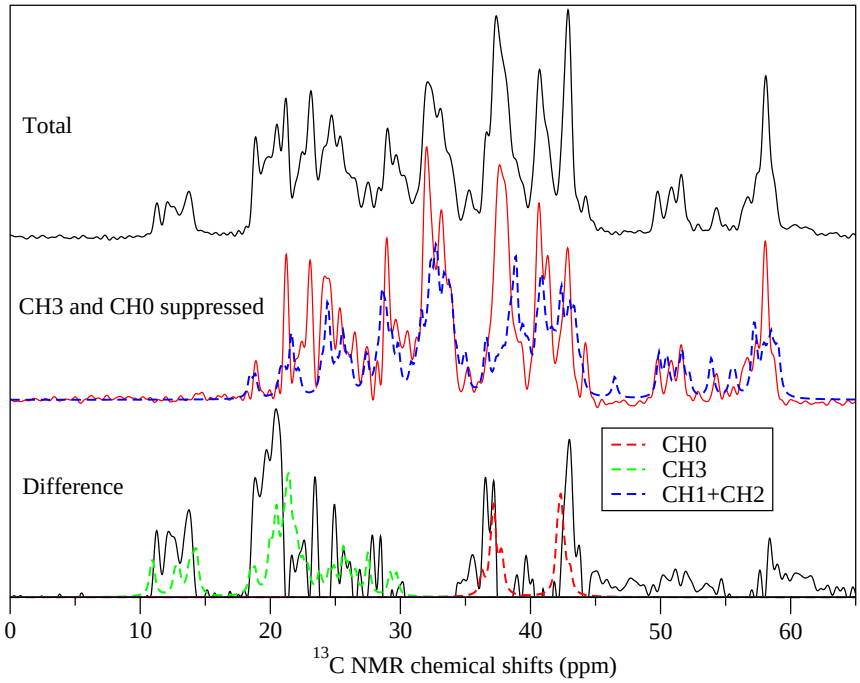


Figure 4.12: Total and edited experimental spectra for ChAl phase. The positive difference is due to signal from CH3 and CH0 groups. The theoretical data for these groups and the sum of CH1 and CH2 groups after reference correction are given for comparison.

confirm the validity of the proposed correction and strengthen our analysis we suggested our collaborators in the research group of Dr. Neeraj Sinha at Centre of Biomedical Magnetic Resonance, Lucknow, India, to perform challenging spectral editing experiments. In the spectral editing experiments that they are performing, the spectra of ChAl and ChAh phases will be resolved with respect to carbon groups based on the number of direct bonds to H atoms, i.e. CH0, CH1, CH2 and CH3 groups can be distinguished. Some comparison of these preliminary edited experimental spectra and theoretical ones is given in the following.

In Fig. 4.12, the total experimental spectrum for ChAl phase and the spectrum where CH0 and CH3 signals are mostly suppressed are given. We report, in the middle panel of Fig. 4.12, the total of theoretical CH1+CH2 spectrum. Using only the available spectral editing experimental results, it is not possible to distinguish CH1 and CH2 groups as they overlap. However an overall agreement, given the uncertainties in these preliminary experimental results, is observed for the sum of CH1 and CH2 signals. The positive difference between the total spectrum and the edited spectrum, $(\delta_{Total} - \delta_{edited})$, is mainly due to signals from CH0 and CH3 groups. There is also some negative difference (not shown) which is within the experimental error range, therefore discarded. As can be seen in the bottom panel, after the reference correction on these groups, the peak positions agree to a very high degree. The possible need for a higher order correction to prevent the 'stretching' of theoretical spectrum, as mentioned earlier, can also be observed. Still, we see that a simple shift based on H-content, and determined using only a few carbon signals (C18 and C19 for CH3 and C13 for CH0), results in very good agreement with the experiment, validating our approach.

With the confidence gained from the agreement between theory and experiment, both for the total spectrum and the edited spectra, we conclude that the reference correction scheme proposed in this study is a simple yet effective way to correct for the systematic error in GIPAW calculations. It should be emphasized that the grouping we have performed based on H-environment is a physically motivated and well-defined criterion. The objectivity of this scheme is also demonstrated: *i)* The correction values are the same for all phases considered. *ii)* Uncorrected CH2 shifts, originally referenced to another CH2-group molecule, cyclobutane, yield good agreement between theory and experiment, although cholesterol and cyclobutane are very different structures.

4.5 Peak assignment

An unambiguous peak assignment is of uttermost importance for the NMR experiments. It can not only help determine the molecules in the sample but also enables experimentalists to track certain interactions, reactions and conformational changes that the material of interest may undergo, for example, the binding site of a catalyst can be revealed by the changes in the NMR spectrum if all the sites are correctly assigned. For the future studies concerning cholesterol crystals, we expect that peak assignment can reveal important information on crystallization from bile samples in gallstones or on the structure determination of new stable phases and intermediates that

appear during crystallization. Before we start the peak assignment, let us

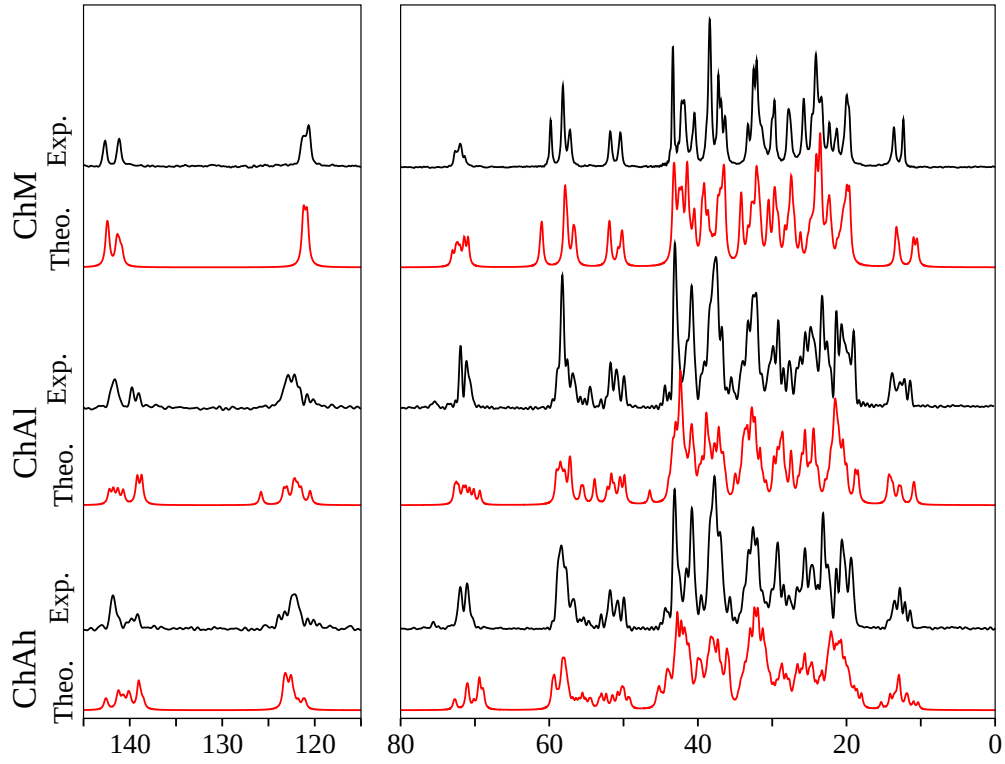


Figure 4.13: ^{13}C NMR chemical shifts for all phases after referencing correction compared to experimental results of Ref.[57]

revisit the total spectra for all the phases after the reference correction in Fig. 4.13. Comparing the theoretical spectra with the experimental one, we see a strong resemblance not only for isolated peaks but also in the overlap region 10-50 ppm. Now we attempt to use this resemblance to perform peak assignment for ChM phase. For the sake of simplicity, we examine the overlap region of ChM spectrum in three parts: a) 10-26 ppm, b) 26-36 ppm, c) 36-50 ppm regions. We use both the calculated and reference corrected peak positions and the lineshape to assign the experimental peaks to carbon atoms. All CH3 group signals, from atoms C18, C19, C21, C26, C27, fall in region a) (Fig. 4.14). Peaks of atoms closer to ring region show less dispersion, and can therefore be assigned with higher confidence than the ones at the tail region of the molecule. This is also true for the CH2 group signals in this region: C11, C15 and C23. For many atoms the double peak feature

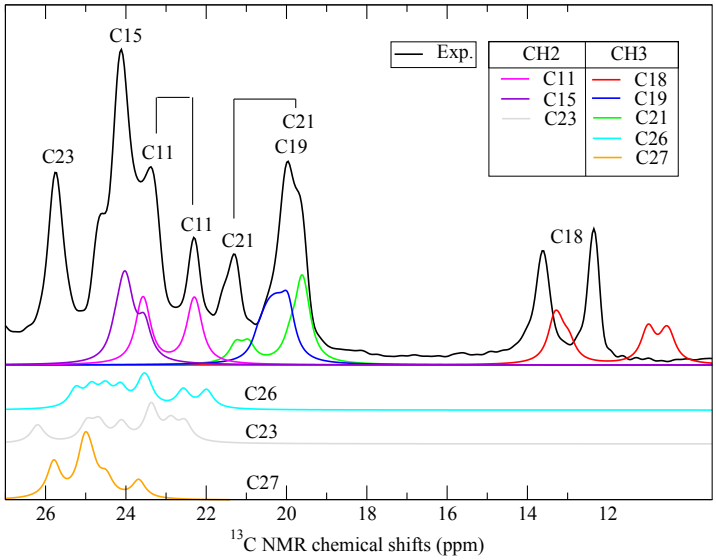


Figure 4.14: Experimental ^{13}C NMR chemical shifts in 10-26 ppm for ChM phase region are assigned to carbon atoms based on our theoretical results, which are also given in detail.

originating from pseudosymmetry can be observed. During assignment we keep in mind that a 'stretching' behavior is still observed in the theoretical results: the lowest ppm signal from group CH3, namely C18 signal, is underestimated. Therefore it is possible that the highest signal of this group, the one of C27, may be overestimated by our calculations. Additionally, as previously discussed temporal averaging of NMR spectra at finite temperature may have important effects on the resonance signal of this tail atom. For these reasons, we choose to assign the 26 ppm peak to C23 atom rather than C27.

In region b), the peaks are from both CH1 (C8,C20,C25) and CH2(C2,C7,C16) groups. For C20, only 4 out of 8 molecules fall in this region. The tail atom C25 shows a large dispersion in the 27-33 ppm range, so it can be suggested to contribute to any/all of the three peaks in this range. Neighbor C16 and C17 (as seen earlier in Fig. 4.11) atoms show double peaks demonstrating the two different conformations for the penta- ring of cholesterol molecule. It is worth to note that C2 atom, which is bonded to C3 atom on the molecule head where thermal motion is strong, displays much less scattered peak than C3 (see also Fig. 4.13). In region c), the scattered peaks of tail atoms C20, C23 and C22 result in loosely assigned peaks for these atoms, in the range

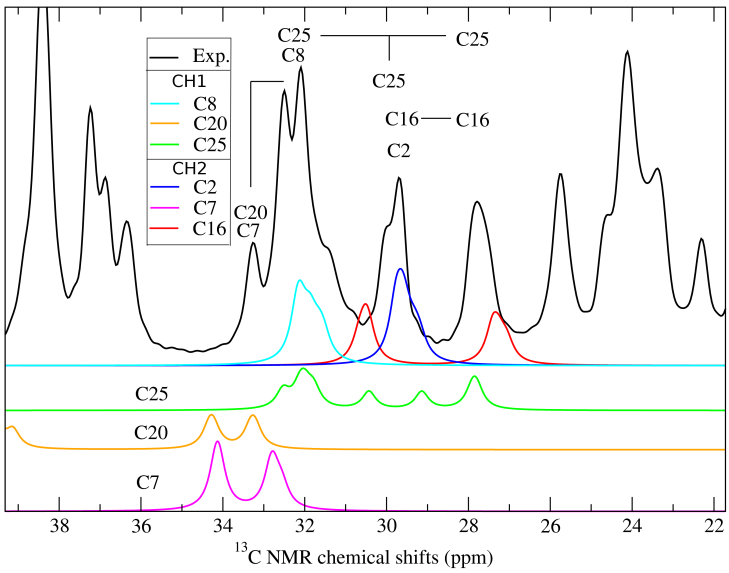


Figure 4.15: Experimental ^{13}C NMR chemical shifts in 26-36 ppm region for ChM phase are assigned to carbon atoms based on our theoretical results, which are also given in detail.

of 36-42 ppm. Considering our 0.5 ppm uncertainty in chemical shifts, we claim that the single yet broad peak at 40.5 may have contributions from C12 atom as well as C24 and C20. Single peak of C13 is assigned with high confidence, as it is the only atom with such sharp resonance signal in this region. In Fig.4.17, we report the whole spectrum with assigned peak positions for ChM phase.

For ChAl phase the peak assignment is more challenging due to the lower symmetry of this phase. Spectral editing experiments with high signal-to-noise ratio that we have requested for further clarity are still ongoing, however the assignment is already very promising. In the 10-25 ppm range (Fig. 4.18), at variance with ChM phase our calculations have estimated a low shift C23 signal at 18.5 ppm, which is validated by the spectral editing experiments in which CH3 signal is suppressed. These low shift signals originate from the molecules C and G in the unit cell (see Fig. 3.4) where molecule tails are sharply bent at C22-C23 bond. The fact that our calculations and experimental results agree on this peak implies that there is no conformational change in these molecules within the relaxation time of the nuclear magnetic moment. The preliminary results of the suggested spectral editing experiments also show that C26 and C27 peaks are scattered also under experimental conditions, in agreement with our 0K calculations. In the 25-35 ppm region

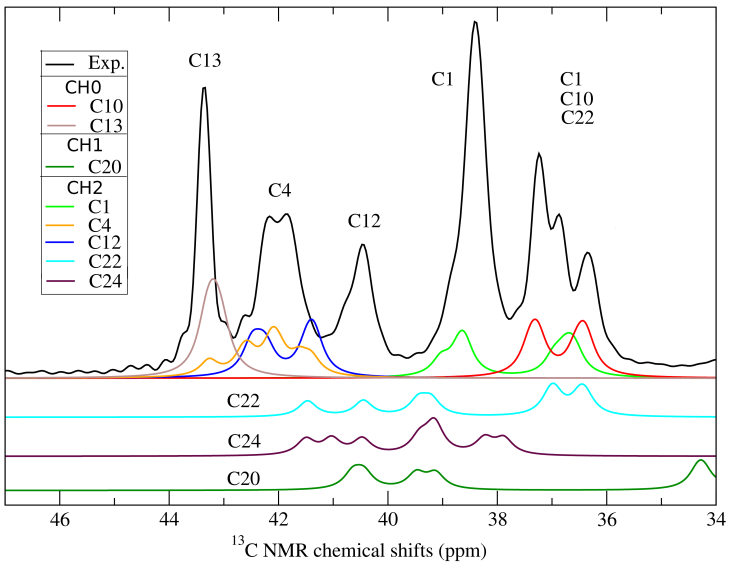


Figure 4.16: Experimental ^{13}C NMR chemical shifts in 36-46 ppm region for ChM phase are assigned to carbon atoms based on our theoretical results, which are also given in detail.

(Fig. 4.19) there are signals only from CH1 and CH2 groups, therefore spectral editing experiments that invert the sign of these groups (CH2-CH1), in addition to suppressed CH3 and CH0 signals, were requested. The difference between the sum and difference of these groups indicates that the disperse C25 signal is correctly predicted by our calculations.In the 35-45 ppm region (Fig. 4.20) we see that several peaks can be assigned with confidence, even details such as the fragment of C4 at the high-ppm end.

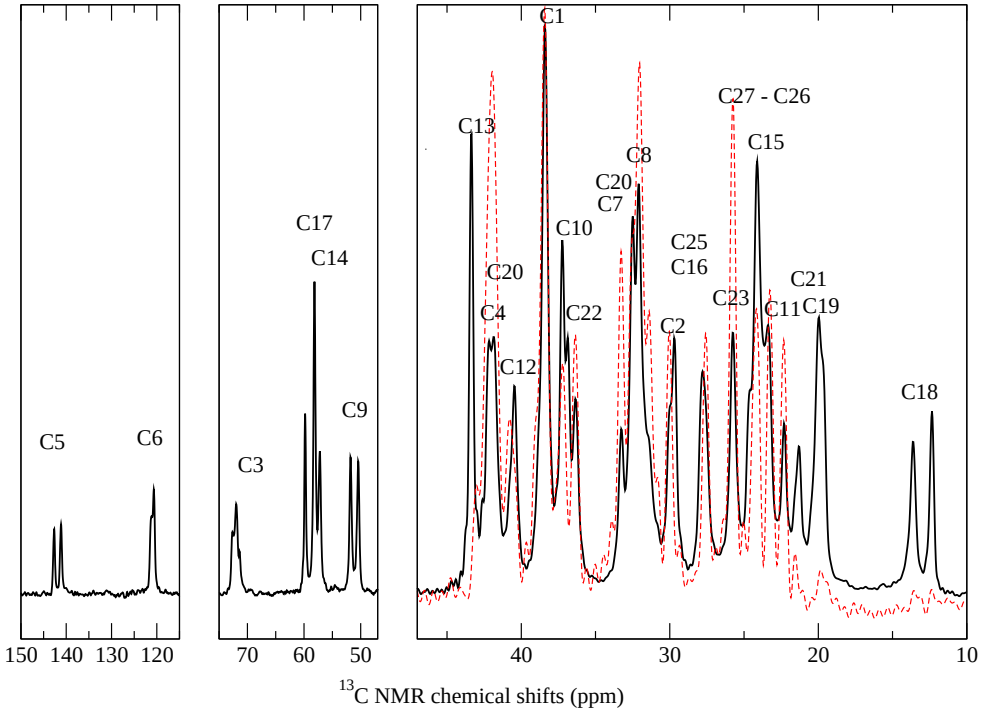


Figure 4.17: Experimental ^{13}C NMR chemical shifts for ChM phase are assigned to carbon atoms as shown. The CH3 and CH0 suppressed editing results are also given for comparison.

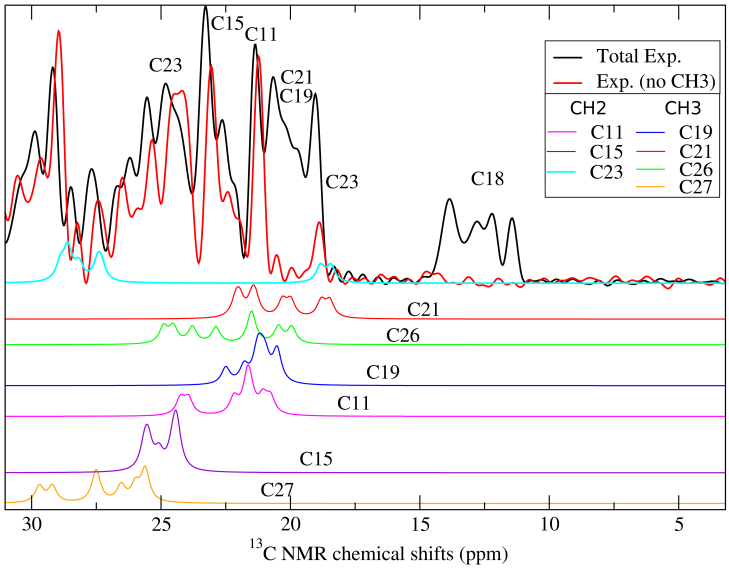


Figure 4.18: Experimental ^{13}C NMR chemical shifts in 10-25 ppm region for ChAl phase are assigned to carbon atoms based on our theoretical results, which are also given in detail.

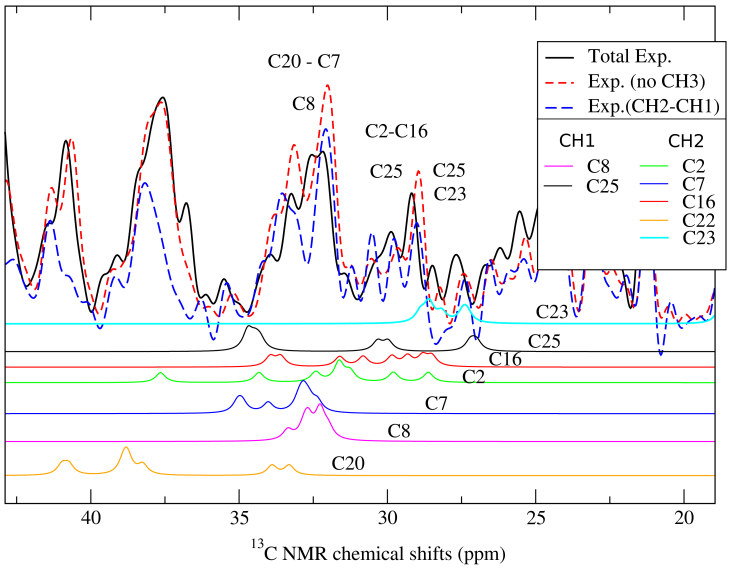


Figure 4.19: Experimental ^{13}C NMR chemical shifts in 25-35 ppm region for ChAl phase are assigned to carbon atoms based on our theoretical results, which are also given in detail.

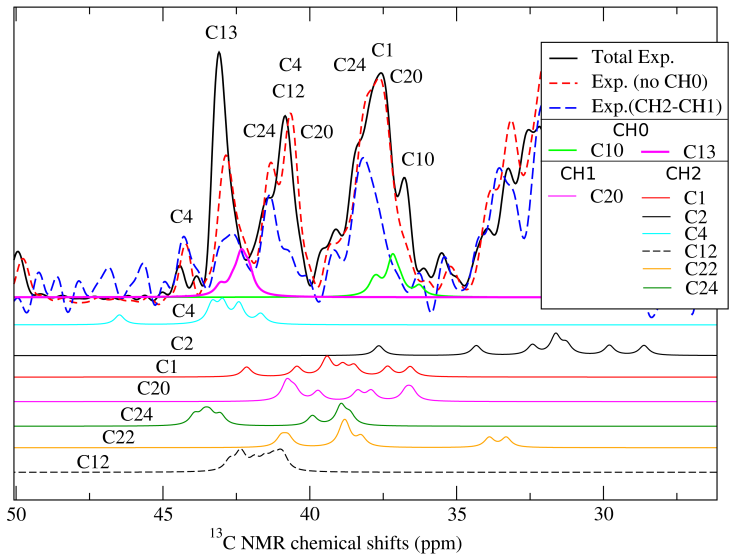


Figure 4.20: Experimental ^{13}C NMR chemical shifts in 35-45 ppm region for ChAl phase are assigned to carbon atoms based on our theoretical results, which are also given in detail.

4.6 Intermolecular interactions

With the help of chemical shift calculations, valuable insight on intermolecular interactions can be obtained. To demonstrate that, we performed chemical shift calculations separately on eight distinct molecules in the asymmetric unit cell of ChAl, at the positions extracted from crystal without further geometry optimization. The comparison between the sum of isolated molecule spectra and full crystal one is reported in Fig. 4.21. It is evident that the two spectra display significant differences: Full crystal carbon peaks are more spread, as in experimental findings, while the majority of peaks in the sum of isolated-molecule spectra has narrower lineshapes. Fig. 4.22 shows a more

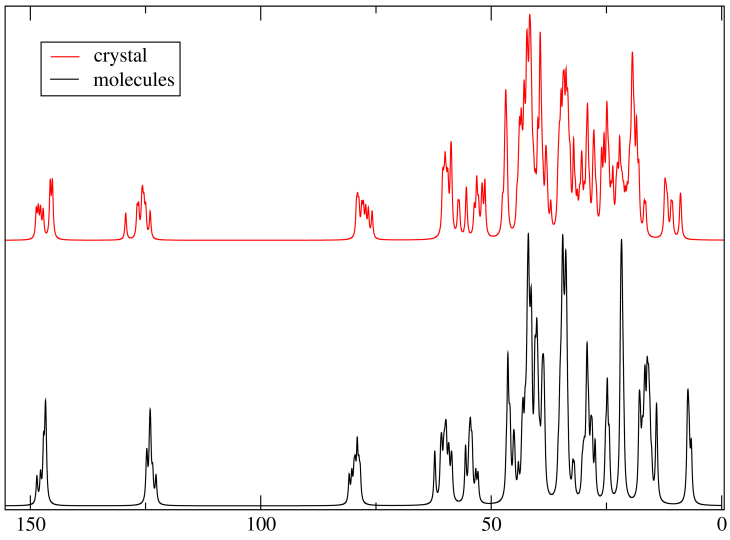


Figure 4.21: ^{13}C NMR chemical shifts for isolated molecules vs. crystal: ChAl spectrum obtained as the sum of eight isolated molecules at the atomic positions extracted from the crystal (black solid line) is compared to original full crystal spectrum (red solid line).

detailed comparison for the isolated peaks C5, C6, C3, C14/17, C9 and C18, and additionally C26 and C2 (see Fig. 3.1 for labeling of atoms). The spread due to crystal environment is especially pronounced in peaks belonging to C18 which points to the strong interlayer interaction between molecules on the same side of the bilayer. Previously we have seen that C26 peak shows a wide spread in the crystal spectrum (Fig. 4.18). It is known that the tail region of cholesterol molecules is flexible, giving rise to several conformations that are close in energy. From the spread peak feature in the crystal, one

could argue that our calculations might not have found the right geometry for C26 atoms, due to not resolving small energy differences precisely. However, in Fig. 4.22 we see that the sum of isolated molecules results in a narrower spread for this peak, suggesting that the spread is due to intermolecular interactions rather than different conformations of individual molecules.

The differences in peak shapes and number for C3 and C2 in the case of isolated molecules with respect to the full crystal stems from the loss of hydrogen bond between the molecules on opposite sides of the bilayer, which are arranged in head-to-head conformation, as well as the loss of aromatic $\pi - \pi$ interaction between the rings of the molecules on the same side of the bilayer, since the rings of different molecules are best aligned at the C3 end of the molecules.

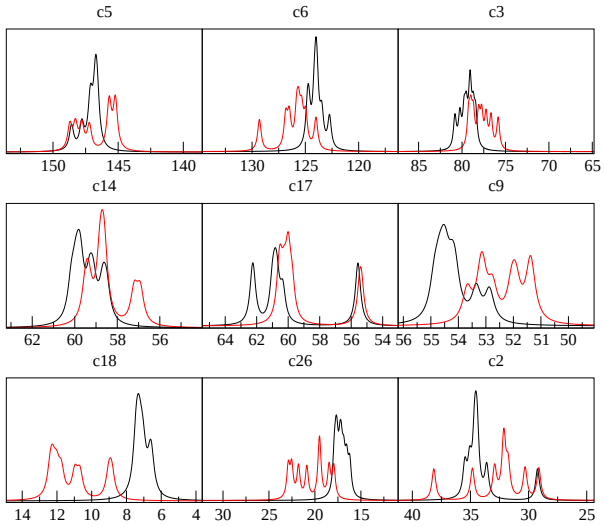


Figure 4.22: ^{13}C NMR chemical shifts for selected peaks. ChAl spectrum obtained as the sum of eight isolated molecules extracted from the crystal (black solid line) compared to original crystal spectrum (red solid line).

The analysis above exemplify the type of modifications that crystal packing and intermolecular interactions can induce on the molecular spectra. Combined with the peak assignment that we have performed previously, these observations highlight the importance of *ab initio* NMR calculations in aiding NMR experiments to obtain a better understanding of sample structure. In particular comparison with the isolated molecule calculations show the

the atomic sites where intermolecular interactions are present. In further studies of surfaces, absorbates, or chemical interactions between molecules, this information can prove useful in reliably determining binding sites and detecting conformational changes.

Chapter 5

Conclusion

In this work we have focused on molecular crystals of cholesterol, motivated by their medical significance: Cholesterol crystals are the main constituents in gallbladder stones and play an important role in the formation of artheromatous plaques. In the experimental NMR studies of these crystals, only a portion of the spectrum can be assigned to the underlying structure due to the complexity of the spectra. We aimed at improving this situation by *ab initio* NMR calculations.

In our studies we used the well established GIPAW method to calculate the NMR chemical shielding of cholesterol crystals. To be able to work on such big systems, we have implemented the use of PAW pseudopotentials in GIPAW modules of Quantum Espresso. By doing so, we were able to use PAW formalism consistently throughout our calculations and reduced considerably the computational cost of these expensive calculations. In our computational tests on small molecules and glycine crystals, we have verified the accuracy of our implementation. With the confidence gained from these smaller systems, we have investigated all three known phases of cholesterol crystals, namely cholesterol monohydrate and low- and high-temperature phases of anhydrous cholesterol.

First we have investigated the structural properties: The results of *ab initio* calculations agree with the experimental ones within the uncertainty of the experiments and correctly reproduce the observed structural features such as pseudosymmetry. Furthermore, we found that experimental positions with partial or no further optimization may not be accurate enough due to uncertainties associated with, for example, strong thermal motion. This feature was most obvious in the case of high temperature anhydrous phase. In fact we have come to the conclusion that a full first principles geometry optimization is necessary to accurately determine any observable that is structure dependent.

Since cholesterol crystals are parts of big organic systems such as membranes, which are still beyond the scope of DFT calculations, applications relying on force fields might be necessary in future studies. For this reason we have examined the structures determined with a force field. Comparisons showed that for the hydrated phase of cholesterol, ChM, the choice of force field for bilayer water molecules greatly affects the obtained structure and T3PIP, potential in conjunction with the cholesterol molecule force field we have used, resulted in a different conformation from experimental and *ab initio* ones, to a degree outside the experimental uncertainties. Repeating the calculations for the low-temperature anhydrous phase we see that agreement is improved, suggesting that force-field methods can be used to gain insight into the structure of pure cholesterol crystals.

Next, we turned our attention to the NMR spectra of cholesterol crystals. Our results showed that experimental NMR spectra can be reproduced accurately by first principles. We have seen that different phases can be distinguished through *ab initio* NMR spectra, even though differences in resonance signals between these phases are as small as 2 ppm. This indicates the high accuracy of our calculations and, in general, the potential of first principles calculations in aiding NMR spectroscopy experiments for structure determination.

We then examined some important factors affecting the theoretically obtained spectrum. The significant impact of an accurate, full *ab initio* geometry optimization was observed. In the case of force fields, we have seen, again, that attention to the water potential is necessary for the hydrated phase, while the anhydrous crystals can be treated by force fields to have qualitative information on the spectral features.

An important observation was made while investigating the effects of exchange-correlation functionals. Cholesterol crystals are held together mostly by van der Waals interactions, which we have taken into account by using the vdW-DF functional. We saw that spectral features are robust with respect to the exchange correlation functionals but an accurate NMR spectrum could only be obtained in the case of vdW-DF, and not for the other functionals considered, PBE and revPBE. Our calculations also showed that the difference between spectra obtained with different functionals originate not only from the variations of resulting structure through geometry optimization, but also from the differences in electronic structure, given fixed positions. Therefore in systems with strong vdW interactions, we suggest that not only a correct vdW-including structure should be used but also an appropriate exchange correlation functional should be employed in the GIPAW calculations.

Throughout our first principles NMR studies we have observed the com-

monly known systematic error of GIPAW calculations where high(low) shifts are over-(under-) estimated. We propose a way of correcting for this error based on correcting the shielding reference. The use of different shielding references for different peaks have already been proposed in the literature but no systematic criterion has been suggested. Previous method relied heavily on the experimental assignment of the peaks to source nuclei and was therefore hardly predictive. Our correction, by linking the reference shifting to the number of hydrogen atoms bonded to the carbon atom can instead be applied even to crowded spectral regions. By comparing to the results of spectral editing experiments, we showed that our correction is valid, and phase and molecule independent, thus demonstrating its predictive power.

With the help of this correction, we further improved the agreement between our results and the experiment, so that we could perform a full peak assignment for the NMR spectra of cholesterol crystals. Based on our results we were able to propose spectral editing experiments to confirm and further improve the confidence of these assignments. Preliminary experimental findings confirm our assignment and further experiments are on the way.

Finally we have taken advantage of the power of simulation, by comparing the NMR spectra of single molecules, at atomic positions extracted from crystals and transfered in vacuum without further relaxation, with the relaxed full crystal spectrum. This comparison reveals the dominant sites for inter-molecular interactions. We have seen that the carbons C18 and C19, poking out of the ring plane, are strongly affected by the presence of other molecules. This is a rather expected observation considering the stacking of molecules in the crystals. However we have also seen a more surprising result: the terminal end of the tail, C26 atom shows a disperse signal in crystal. This could easily be assumed to be caused by the flexible nature of the tail, and the presence of several close minima for the position of this atom, and one could argue that our calculations could just not resolve the differences between these minima. Instead we see that the dispersion is reduced once the crystal environment is eliminated, suggesting that intermolecular interactions are the main reason behind the observed dispersion.

Appendix A: vdW-DF and Glycine

Throughout this thesis we have employed the vdW-DF exchange correlation functional to describe the long range interactions in the molecular crystals of cholesterol. In this chapter, we perform tests on a simpler system, glycine crystals, to investigate the adequacy of this functional in describing the structure accurately. We explore the effect of high pressure on glycine crystals and compare structure and stability with the experimental findings.

Introduction

The effects of high pressure on organic molecular systems and proteins are of interest in various areas, from the industrially relevant issue of pressure treatment of food products to academically significant studies of extremophile bacteria that live in earth's crust or in the depths of the oceans. High-pressure studies of molecular systems are also important while exploring new phases of molecular crystals and their stabilities as in the case of crystal structure prediction.

The use of *ab initio* DFT methods in investigating the high pressure effects and stability is hampered due to two important reasons: *i)* In the case of molecular crystals, a large number of stable phases lies in a very small energy range, amongst several plausible structure candidates. *ii)* Due to the limitations of standard exchange-correlation functionals in modeling the van der Waals (vdW) interactions, weakly bound molecular crystals cannot be accurately described in standard DFT studies.

There have been recent developments, both theoretically [27, 29, 30] and algorithmically [28], that address the latter of these concerns and led to new vdW-aware exchange-correlation functionals [27, 29, 30]. In this work we investigate the performance of both standard (PBE and revPBE) and newly introduced XC functionals (vdW-DF and c09x) in predicting the correct stability of molecular crystal phases as a function of pressure, in the case of

the simplest aminoacid, glycine.

Glycine, $H_3N^+CH_2COO^-$ in its zwitterionic form (see section 2.5), has three polymorhps under ambient conditions: α, β and γ, the most stable one being the γ phase. The evaporation of aqueous solutions yields the α phase. Crystallization of glycine from water/ethanol solutions, instead, results in the β phase. At ambient conditions, the order of stability is $\gamma > \alpha > \beta$. The γ phase transforms to the α phase upon heating, at around 170°C, depending on the thermal history of the sample. High humidity drives a phase transition from α to γ phase. The β phase, metastable in dry air, rapidly transforms to α or γ phases in the presence of moisture at room temperature.

Under pressure, the α phase is found to be stable up to 23 GPa, the β phase is shown to undergo a phase transition to the δ phase at 0.76 GPa. Two independent experimental studies showed that the γ phase transforms to a new glycine phase called ϵ, starting from 1.9 or 2.74 GPa. The experimental studies by Boldyreva and co-workers [60] resulted in a proposed structure for the ϵ phase based on the data obtained at 7.85 GPa. An alternative model for this phase is given by the works of Dawson and co-workers [61] at 4.3 GPa. In our studies we model the ϵ phase based on the latter structure. All known phases of glycine crystals are monoclinic, with the exception of trigonal γ form.

Computational Details

Ab initio calculations are performed in the framework of Density Functional Theory as implemented in the Quantum Espresso distribution [18]. The exchange-correlation functionals used are the standard Perdew-Burke-Ernzerhof (PBE) [25], revised-PBE (revPBE)[44], the original vdW-DF formalism using revPBE exchange [27] and vdW-DF-c09x [30] that uses c09x functional introduced by Cooper to be used in conjunction with the nonlocal vdW-DF functional. Ultrasoft pseudopotentials from the PSLibrary project [39] are used without further modification. A kinetic energy cutoff of 80 Ry and charge density cutoff of 560 Ry are used to achieve pressure convergence within less than 1 kbar. Monkhorst-Pack k-point grids of $4 \times 2 \times 4$, $3 \times 3 \times 3$, $3 \times 3 \times 3$, $2 \times 3 \times 3$, $3 \times 3 \times 3$ are used for α, β, γ, δ and ϵ phases respectively. With this setting, a tight convergence of less than 0.1 mRy in total energy is achieved. The cell parameters and atomic positions are fully optimized.

XC	a (Å)	b (Å)	c (Å)	β (deg)	no. of molecules	volume/ molecule (Å^3)
α-glycine						
PBE	5.203	12.717	5.476	109.64	4	85.32
revPBE	5.359	14.000	5.525	109.38	4	97.76
vdW-DF	5.244	12.289	5.570	111.07	4	83.74
c09x	5.034	11.567	5.449	112.93	4	73.06
Exp.[a]	5.083	11.820	5.460	111.93	4	76.0
β-glycine						
PBE	5.167	6.768	5.411	111.97	2	87.75
revPBE	5.414	8.231	5.546	120.04	2	106.98
vdW-DF	5.216	6.410	5.503	112.45	2	85.03
c09x	5.017	5.990	5.399	113.85	2	74.20
Exp.[b]	5.094	6.286	5.383	113.21	2	79.2
γ-glycine						
PBE	7.252	7.251	5.520	90	3	83.81
revPBE	7.868	7.854	5.496	90*	3	97.94
vdW-DF	7.236	7.235	5.590	90	3	84.46
c09x	6.892	6.892	5.459	90	3	74.84
Exp.[c]	6.975	6.975	5.437	90	3	76.9

Table 1: Optimized cell parameters at zero pressure for glycine phases stable under ambient conditions.
[a]Ref.[62],[b]Ref.[63],[c]Ref.[64]
*Upon releasing the constraint on symmetry, the angle β becomes 89.93°

Results

Structure at Zero Pressure

Lattice parameters a, b, c and β calculated at ambient pressure for the three stable phases α, β and γ are given in Table 1. Also listed are experimental values determined for these structures.

Calculated volumes are found to be sensitive to the exchange-correlation functionals used. In average, PBE is found to overestimate the volumes by +10.7%, revPBE by +30.4%, vdW-DF by +9.1%, where c09x is found to underestimate the volume by -4.3%. The difference between the highly overestimated volume obtained with revPBE and improved volume estimate with vdw-DF reveals the effect of including the long range interactions. We see that the vdW-aware c09x functional consistently underestimates the experimental volume outside the typical GGA behavior and gives the closest results

bonds	PBE	revPBE	vdW-DF	c09x	Ref. [62]
C1-O1	1.26	1.27	1.27	1.27	1.25
C1-O2	1.27	1.27	1.27	1.27	1.25
C2-N	1.49	1.50	1.51	1.48	1.48
C1-C2	1.53	1.55	1.55	1.53	1.52
C2-H1	1.10	1.10	1.09	1.10	1.05
C2-H2	1.09	1.09	1.09	1.10	1.04
N-H3	1.05	1.04	1.04	1.06	1.03
N-H4	1.07	1.07	1.06	1.07	1.09
N-H5	1.04	1.04	1.03	1.05	1.09
torsion					
N-C2-C1-O1	30.23	37.61	21.51	20.05	18.6

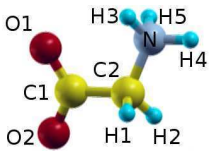


Table 2: Optimized bond lengths (Å) and torsion angle (degrees) for α glycine at zero pressure. The numbering of the atoms is given on the side.

to the experiment. From our results we conclude that crystal densities are strongly dependent on the choice of XC functional.

Next, we explore the performance of these functionals in reproducing the structure of the single molecules in the crystal. In Table 2, we compare the bond lengths and torsion angle for α glycine at ambient pressure.

We see that in terms of bond lengths, all functionals describe the molecule equally well, within the 2% range of experimental values for bonds not involving hydrogen atoms, and 6% for the bonds with hydrogen atoms. The effect of functional becomes more apparent in the torsion angle where only vdW-including functionals show a very good agreement with the experiment. We observe the same trend also for the other phases which are stable at ambient pressure.

Stability and High Pressure

Addressing the relative stability of glycine polymorphs is difficult due to the very small energy differences between the phases (less than 1 kcal/mol). Calculated differences are also sensitive to the exchange-correlation functionals used. In Fig. 1 we report the enthalpy as a function of pressure, referenced with respect to the α phase, for all exchange correlation functionals employed. At zero pressure, vdW-DF is the only functional that predicts the stability order $\gamma > \alpha > \beta$ correctly, while c09x also shows very similar enthalpy for γ and α phases. Since the energy differences are extremely small, we can regard these functionals as performing equally well at zero pressure.

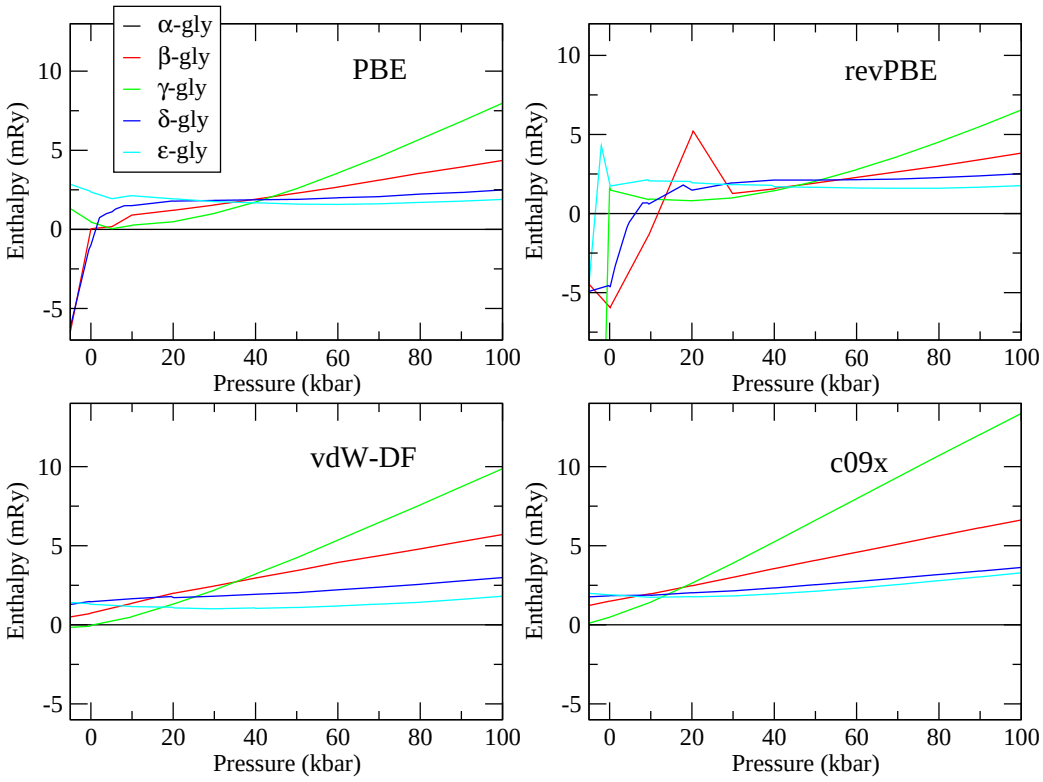


Figure 1: Enthalpy referenced with respect to the α phase as a function of pressure for all known phases of glycine up to 100 kbar, for all XC functionals considered. Calculations are performed with target pressure -5 kbar, 0 kbar and every 10 kbar up to 100 kbar, a total of 12 points. Additional calculations were performed for high-pressure phases at around ambient pressure for better convergence.

For both functionals, the β phase is well separated from the α and γ phases as expected for a metastable state, however this is not observed for PBE functional. We see that revPBE displays a highly erroneous stability order both at zero pressure and at higher pressures, therefore is neglected in the following discussion [1].

In agreement with the experimental observations, the α phase is stable up to 100 kbar for PBE, vdW-DF, c09x and after 11 kbar for revPBE. The phase transition from β to δ occurs at 7.6 kbar experimentally. The δ phase is calculated to be thermodynamically stable after around 40 kbar, 15 kbar and 7 kbar for PBE, vdW-DF and c09x respectively. Considering the precision of our calculations, the estimates from vdW-DF and c09x are equally satisfactory while PBE overestimates the transition point.

The phase transition of the γ form is more complicated, as single crystals undergo a phase transition to a polycrystalline phase already at 19 kbar but the identification of the ϵ phase could only be obtained at 43 kbar. Our calculations predict that the ϵ phase becomes thermodynamically more stable than the γ one at around 40 kbar, 18 kbar and 12 kbar for PBE, vdW-DF and c09x respectively. As in the case of $\beta \rightarrow \delta$ transition, we see that vdW-DF and c09x perform equally well in predicting the single crystal transition of γ form, while PBE overestimates the transition pressure. The stability order remains the same for all XC functionals at high pressures: $\alpha > \epsilon > \delta > \beta > \gamma$.

Conclusion

Our calculations show that energetics and structural properties are sensitive to the choice of exchange correlation functionals in the case of glycine crystals at ambient and higher pressures. Van der Waals interactions play an important role in intermolecular distances in glycine crystals. This is confirmed by the fact that XC functionals including these interactions result in better estimates for stability order and crystal density than the functionals that lack this interaction. As a trend, c09x is found to underestimate the volume

[1] The sharp increase in enthalpy for the β phase at 20 kbar with revPBE functional is due to the reorganization of the hydrogen bond network taking place from the 10 kbar configuration to the 30 kbar one. At 10 kbar, the two molecular layers in the unit cell are stacked on top of each other, so that linear hydrogen bonds between molecules of the two layers are formed, normal to the layer plane. At 30 kbar, the layer stacking changes such that a molecule of top layer sits in the void between two molecules of the lower layer, forming a trigonal hydrogen bond network. The 20 kbar configuration is the transition point between these two, resulting in higher enthalpy than both. The on-top stacking of 10 kbar is only observed with revPBE functional

while vdW-DF overestimates it. For the stability order at ambient pressure and transition pressures, both functionals perform equally well and better than PBE and revPBE, suggesting that using vdW-aware XC functionals can greatly improve the energetics and structural properties of molecular crystals obtained by *ab initio* methods.

Appendix B: Quadrupolar interactions with PAW-GIPAW method: Validation with Alumina phases.

In our studies of cholesterol crystals, we focused on NMR chemical shift spectrum of ^{13}C *nuclei which have spin* $I = 1/2$*. Instead, nuclei with spin* $I \geq 1$*, such as* ^{27}Al*, have an electric quadrupole moment so that their energy levels are split by the electric field gradient created by the local electronic environment. The effect of this quadrupolar interaction can be observed experimentally in NMR spectra and is usually parametrized in terms of quadrupolar coupling constant* C_Q *and asymmetry parameter* η*.*

In this chapter, we present our ab initio *NMR studies of alumina crystals, which validate our implementation of PAW pseudopotentials in GIPAW method also in the case of quadrupolar interaction parameters. The accuracy of this method is also exploited to investigate the adequacy of several structural candidates for* γ*-alumina, where experimentally obtained information is not conclusive. The preprint of the paper resulting from this study is given below [65].*

Ab initio ^{27}Al NMR chemical shifts and quadrupolar parameters for Al_2O_3 phases and their precursors

Ary R. Ferreira,[1] Emine Küçükbenli,[2,3] Alexandre A. Leitão,[1] and Stefano de Gironcoli[2,3]

[1] *Universidade Federal de Juiz de Fora (UFJF), Department of Chemistry, Juiz de Fora, MG, 36036-330, Brazil*
[2] *Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, I-34136 Trieste, Italy*
[3] *CNR-IOM DEMOCRITOS Simulation Center, Via Bonomea 265, I-34136 Trieste, Italy*

The Gauge-Including Projector Augmented Wave (GIPAW) method, within the Density Functional Theory (DFT) Generalized Gradient Approximation (GGA) framework, is applied to compute solid state NMR parameters for ^{27}Al in the α, θ, and κ aluminium oxide phases and their gibbsite and boehmite precursors. The results for well-established crystalline phases compare very well with available experimental data and provide confidence in the accuracy of the method. For γ-alumina, four structural models proposed in the literature are discussed in terms of their ability to reproduce the experimental spectra also reported in the literature. Among the considered models, the $Fd\overline{3}m$ structure proposed by Paglia *et al.* [Phys. Rev. B **71**, 224115 (2005)] shows the best agreement. We attempt to link the theoretical NMR parameters to the local geometry. Chemical shifts depend on coordination number but no further correlation is found with geometrical parameters. Instead our calculations reveal that, within a given coordination number, a linear correlation exists between chemical shifts and Born effective charges.

I. INTRODUCTION

Aluminium oxide (Al_2O_3), also known as alumina, is one of the most important oxides because of its many industrial applications[1]. Corundum (α-Al_2O_3), the most stable and common crystalline form of alumina, is the final product of the calcination of hydroxides or oxyhydroxides of aluminium at temperatures above 1273 K, which can vary according to the precursor. The transformation from aluminium hydrates to the final oxide is not direct and a variety of relatively stable intermediate phases can be detected for each combination of starting compound and thermal treatment.

The most commonly used precursors in the synthesis of the various alumina phases are gibbsite [γ-$Al(OH)_3$], bayerite [α-$Al(OH)_3$], and boehmite [γ-$AlO(OH)$]. The transformation of gibbsite to corundum can proceed through a sequence of hexagonal close packed aluminas (χ and κ) or an alternative sequence in which boehmite is initially formed and the α phase is achieved via the cubic spinel transition phases (γ, δ, and θ)[2]. Bayerite can also follow this same sequence through boehmite to form corundum[3] or an alternative path transforming to η and θ phases[4].

Among the transition aluminas, the γ phase is highly valued for industrial applications due to its textural properties (surface area, pore volume, pore size), which makes it an important material in many industrial processes acting as an adsorbent, a catalyst and/or catalyst support. In petroleum and petrochemical industries γ-alumina is used as catalyst support for transition-metal sulfides Co(Ni)MoS in hydrotreatment catalysts and metallic alloys in reforming catalysts[6,7]. Due to its low crystallinity and the consequent difficulty in characterization, the debate on the structure of γ-alumina remains open and a series of theoretical and experimental works concerning this subject have been published along the decades[8–17].

Since the transition between the distinct intermediate phases is a gradual process, the precise temperature at which each phase is obtained with a high degree of purity can not be determined by X-ray diffraction (XRD) experiments only. Solid-State Nuclear Magnetic Resonance (SS-NMR) is an important technique for material characterization. Long-range order is not a prerequisite to distinguish different phases and the knowledge of ^{27}Al NMR data can allow the detection of the onset of phase changes during alumina calcination[18], permitting a discussion of the transition mechanisms[19]. A detailed interpretation of the results remains, however, a challenge.

The advent of theoretical techniques such as the Gauge-Including Projector Augmented Wave (GIPAW) method[20,21] enables the *ab initio* calculation of isotropic chemical shielding, σ_{iso}, quadrupolar coupling constant, C_Q, and asymmetry parameter, η_Q, in solids. First principles simulations of NMR spectra of structural models for transition aluminas are now possible[22,23] and can be compared with available experimental data.

In this paper we apply the GIPAW method to compute solid state Magic Angle Spinning (MAS) NMR parameters for a number of well characterized aluminium oxide phases and for their boehmite and gibbsite precursors. The results are compared with available experimental data to validate the method. In order to contribute to a better characterization of γ-alumina, a number of structural models, proposed in the literature, are examined and their simulated spectra compared to experimental ones, thus revealing their adequacy.

We then examine possible correlations of the predicted chemical shifts with local atomic geometry or local electronic structure, described through Bader analysis[24] and Born dynamical effective charges[25], finding significant correlations.

The rest of the paper is organized as follows: in Sec. II we describe the theoretical methodology and the structural models used. In Sec. III we present our calculated

NMR results, compare them with experiments and discuss the resulting correlations. Sec. IV contains our conclusions.

II. DETAILS OF CALCULATIONS

A. Electronic structure

All *ab initio* calculations in this study were performed using the codes available within the Quantum ESPRESSO distribution[26], which implements the DFT[27] framework using a plane waves basis set to expand the one-electron wavefunctions of Kohn-Sham equations[28]. The effect of exchange-correlation (XC) potential was explored by comparing the results of different descriptions for this term: Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation[29] and its revision (revPBE) by Zhang and Yang[30]. Furthermore, we also considered a van der Waals-aware density functional (vdW-DF)[31,32] recently implemented in Quantum ESPRESSO. Interaction of valence electrons with nuclei and core electrons were treated by the Projector Augmented-Wave (PAW)[33] method.

The plane wave kinetic energy cut-off and k-points sampling were adjusted to yield less than 1 mRy/atom convergence in total energy for all models. A kinetic energy cut-off of 45 Ry and expansion of augmentation charges up to 220 Ry was sufficient to ensure this criterion. Integration in the Brillouin zone were determined by the Monkhorst-Pack[34] procedure[35]. Both atomic positions and cell vectors were fully optimized. NMR chemical shieldings were converged within less than 1 ppm.

The Born effective charge tensor $Z^*_{\kappa,\alpha\beta}$ is defined[25] by the macroscopic polarization induced, in direction β and under conditions of zero macroscopic electric field, by a zone center phonon displacing atomic sublattice κ in direction α. Effective charges were calculated within density functional perturbation theory (DFPT)[36,37] and the values presented in the following sections correspond to their isotropic component obtained by $Z^*_\kappa = Tr[Z^*_{\kappa,\alpha\beta}/3]$.

B. NMR chemical shifts and quadrupolar parameters

First principles GIPAW calculations[20,21] yield the absolute chemical shielding tensors for each nucleus, $\vec{\sigma}(r)$. Isotropic chemical shieldings, $\sigma_{iso} = Tr[\vec{\sigma}/3]$, are compared to the experimental isotropic chemical shifts by using the standard expression: $\delta_{iso} = \sigma_{ref} - \sigma_{iso}$. In this work we choose corundum as reference such that ^{27}Al shift of the α phase is aligned to the experimental one at 0 ppm.

The resulting *ab initio* NMR spectra were obtained by using the QuadFit program[38] with the theoretically calculated chemical shifts and quadrupolar interaction parameters, using the experimental magnetic field intensity, and normalizing each spectral component to reflect the relative number of aluminium types. A Lorentzian broadening was added to each spectral feature to obtain the best comparison with the experimental lineshape.

C. Structural models

Among the oxide phases with well characterized structures, we studied the final product α and the transitional aluminas θ and κ.

The structural model for α-alumina, used as a reference for all simulated NMR spectra in this work, was published by Ishizawa *et al.*[39] in an XRD study. The structure was reported as a corundum-type, with an hexagonal crystal system and $R\overline{3}c$ space group. The crystallographic cell contains six Al_2O_3 units in which all aluminium sites are coordinated by six oxygens.

The θ phase is present along the $\gamma \rightarrow \alpha$ transition in different dehydration paths and its structure has been characterized by Zhou and Snyder[9], with a monoclinic crystal system and $C2/m$ space group. In that work the structure was refined from Rietveld analysis resulting in a crystallographic cell with four Al_2O_3 units, in which half of the aluminium atoms are octahedrally (Al_{oct}) and half are tetraedrally (Al_{tet}) coordinated.

The κ phase is one of the intermediate products of the dehydration path from gibbsite to corundum. Ollivier *et al.*[40] describes the κ phase with an orthorhombic system and $Pna2_1$ space group. The cell contains six Al_2O_3 units and 25% of the Al^{3+} sites are tetrahedral, 50% octahedral and 25% in a very distorted octahedral.

We have also studied the NMR spectra of precursor phases gibbsite and boehmite. For gibbsite we started from the structure resolved by Saalfeld and Wedde in a single-crystal XRD study[41]. The structure is monoclinic with $P2_1/n$ space group where the Al^{3+} cations are octahedrally coordinated by 6 OH^- groups forming double layers and occupy two thirds of the octahedral holes in alternate layers[42]. The interlayer cohesion is granted by hydrogen bonds between these OH^- groups.

In the also layered boehmite structure, described by Christensen *et al.*[43], each Al^{3+} cation is octahedrally coordinated by 2 OH^- groups and 4 intralayer O^{2-} anions in an orthorhombic system with $Cmcm$ space group. As in the gibbsite structure, the boehmite double layers interact with each other via hydrogen bonds, which are exclusively interlayer. In this phase, hydrogen bonds are organized in chains along [001] direction and different bond networks are possible depending on the relative orientation of neighboring chains. In our calculations we considered all possible combinations compatible with a 2×2 supercell. The structure with the lowest energy was found to be the one where nearest neighbor chains are antiparallel. However energy differences among various combinations are found to be less than 1 mRy/cell which is consistent with the experimentally observed disorder

in this phase[44].

As one of the aims of this work is to contribute to the characterization of γ-alumina phase, four different γ-alumina structural models from published theoretical works have been studied.

The first γ-phase model, γ-$Al_2O_3(A)$, considered here was proposed by Gutierrez *et al.*[45,46] This model, also called *defect spinel* or *spinel-like* structure, consists of a cell with 8 Al_2O_3 units, in which 37.5% of the cations are Al_{tet} and 62.5% are Al_{oct}. Among the 24 O atoms, 12 are four-fold coordinated (O_{4-fold}) and 12 are three-fold coordinated (O_{3-fold}) to aluminium atoms[47]. It is important to note that in this model only spinel sites are occupied by the Al^{3+} cations.

The next γ-phase model, γ-$Al_2O_3(B)$, used in this work was published by Digne *et al.*[7] and proposed by Krokidis *et al.*[14]. The model has 8 Al_2O_3 units in the cell, 25% of all aluminium atoms are Al_{tet} sites, in a sublattice of O^{2-} anions. The crystal system is monoclinic, but very close to an orthorhombic one with the $P2_1/m$ space group. In this model cations occupy also non-spinel sites.

Two other γ-phase models were published by Paglia *et al.*[48,49]. The unit cells of these models contain a large number of atoms, 32 Al_2O_3 units, and were generated from an extensive search on all structural possibilities of the γ-Al_2O_3 structures using $Fd\overline{3}m$ and $I4_1/amd$ space groups. In the $Fd\overline{3}m$ model, γ-$Al_2O_3(C)$, among the 64 Al^{3+} sites 22 are Al_{tet}, 41 are Al_{oct} and 1 is Al_{pen} (five-coordinated), while in the $I4_1/amd$ model, γ-$Al_2O_3(D)$, there are 21 Al_{tet} and 43 Al_{oct} sites. In these two models, due to the breaking of the local symmetry by the variations in cation occupancies and related distortions in octahedral and tetrahedral sites, the symmetry is actually P_1[50].

III. RESULTS AND DISCUSSION

A. Structure optimization and approximations on exhange-correlation functional

Geometry optimization was performed for all phases, allowing both the atomic positions and cell vectors to relax keeping the group symmetry fixed. To investigate the effect of approximations on exhange-correlation functionals, each geometry optimization was repeated with PBE, rev-PBE, and the vdW-DF functionals as described in Sec. II A. All XC functionals considered overestimate the experimental volume, PBE by about 1.8% on average, revPBE by 4.2%. The use of van der Waals functional was found to have negligible effect, inducing a tiny further expansion, on all structures except for gibbsite (γ-$Al(OH)_3$) [see Table I] where a slight contraction was observed. The exception in the case of gibbsite can be understood considering its layered and open structure.

Since *ab initio* NMR calculations are very sensitive to structural details, we performed all NMR calculations at optimized positions for each XC functional. In spite of the discrepancy in calculated equilibrium volume, the average difference in σ_{iso}, $|C_Q|$ and η_Q between PBE and revPBE were calculated as 1.15 ppm, 0.14 MHz and 0.03, respectively. These values were found to have negligible effect on total spectra and, except for gibbsite, only the PBE spectra will be shown. The numerical values for the three XC functionals are reported in Table II.

TABLE I. Volumes of the experimental and optimized cells with PBE, rev-PBE, and vdW-DF functionals for the five phases with well characterized experimental structures considered in this work.

	Volume (Å^3)			
Structure	Experiment	PBE	rev-PBE	vdW-DF
α-Al_2O_3	254.25[a]	261.84	266.30	268.06
θ-Al_2O_3	187.92 [b]	192.53	195.54	196.34
κ-Al_2O_3	361.31[c]	369.19	375.21	377.75
γ-$AlO(OH)$	261.13[d]	262.61	270.24	271.65
γ-$Al(OH)_3$	427.98[e]	431.57	449.51	447.30

[a]Ref. [39] [b]Ref. [9]. [c]Ref. [40]. [d]Ref. [43]. [e]Ref. [41].

B. NMR parameters and spectra for well characterized structures

In this section, we present a detailed comparison of our *ab initio* calculated chemical shifts and quadrupolar interaction parameters for all the well characterized structural phases mentioned in Sec. II C with experimental data, as well as with very recent theoretical results from the literature. All relevant data are shown in Table II.

1. α and θ phases

Corundum, α-Al_2O_3, shows a single well defined NMR peak that we take as reference when comparing spectra for other structures.

For the θ phase, we compare our results with the ones reported in Ref. [19], where a mixture of α and θ phases were examined, and the NMR parameters (see Table II) for these two coexisting phases were extracted from the experimental spectrum by lineshape fitting.

The overall good agreement between experimental and theoretical spectra can be seen in Fig. 1, upper curves. The decomposition in individual contributions is also given. The main difference between experiment and theory is in the asymmetry parameter η_Q for the Al_{tet} site. For this peak, it should be noted that a recent theoretical work[23] has also reported a value similar to ours.

TABLE II. Comparison between chemical shifts and quadrupolar coupling parameters calculated in this work and experimental and theoretical data from the literature.

	δ_{iso} (ppm)			C_Q (MHz)			η_Q		
Structure	previous		this work	previous		this work	previous		this work
Al site	Exp.	Th.	PBE/rev-PBE/vdW-DF	Exp.	Th.	PBE/rev-PBE/vdW-DF	Exp.	Th.	PBE/rev-PBE/vdW-DF
α-Al_2O_3									
Al_{oct}	0.0 (1)[a]	0.0[b]	0.0	2.38[a]	2.33[b]	2.05/1.98/2.10	0.00[a]	0[b]	0.04/0.04/0.03
θ-Al_2O_3									
Al_{oct}	-3.0 (1.0)[a]	-5.9[b]	-4.9/-4.8/-5.0	3.50 (0.30)[a]	3.44[b]	2.91/2.81/2.71	0.00 (0.10)[a]	0.18[b]	0.26/0.36/0.29
Al_{tet}	66.5 (1.0)[a]	62.3[b]	62.6/60.2/61.1	6.40 (0.10)[a]	6.34[b]	5.63/5.53/5.37	0.65 (0.02)[a]	0.42[b]	0.43/0.37/0.42
κ-Al_2O_3									
$Al_{oct}(1)$[f]	0.0[c]	0.3[b]	2.0/2.3/2.0	0.50[c]	-9.98[b]	-9.53/-9.45/-9.41	–	0.33[b]	0.29/0.29/0.28
$Al_{V+I}(2)$[f]	–	1.4[b]	7.1/6.7/6.8	> 1.5[c]	5.20[b]	4.43/4.33/4.27	–	0.94[b]	0.76/0.77/0.68
$Al_{tet}(3)$[f]	68.5[c]	60.7[b]	62.6/60.2/61.2	0.76[c]	-5.53[b]	-4.83/-4.67/-4.49	0.3[c]	0.33[b]	0.33/0.31/0.32
$Al_{oct}(4)$[f]	5.0[c]	4.4[b]	3.8/3.6/3.5	0.85[c]	4.51[b]	4.99/-4.96/5.06	–	0.77[b]	0.99/1.00/1.00
γ-$Al(OH)_3$ Gibbsite									
$Al_{oct}(1)$	-5.6[d,g]	2.2[e]	2.2/2.3/2.1[h]	4.70±0.20[d]	5.10[e]	-5.30/-4.74/-4.43	1.00±0.05[d]	0.35[e]	0.32/0.35/0.35
$Al_{oct}(2)$	0.0[d,g]	0.0[e]	0.0[h]	2.20±0.20[d]	2.80[e]	2.33/2.06/1.81	0.75±0.05[d]	0.66[e]	0.81/0.71/0.66
γ-$AlO(OH)$ Boehmite									
Al_{oct}	-1.0[d,g]	–	-0.9/-0.5/-0.2[h]	1.8 - 2.8[d]	–	2.16/2.08/2.27	0.5 - 1.0[d]	–	0.45/0.46/0.67

[a]Ref. [19]; [b]Ref. [23]; [c]Ref. [40]; [d]Ref. [44]; [e]Ref. [51].

[f] Aluminium sites are labeled according to Ref. [40] and the sites symmetry, see Table III.

[g] The experimental $\delta_{iso}(Al_{oct}(2)) = 11.5\pm0.2$ ppm from gibbsite is taken as reference.

[h] The calculated $\sigma_{iso}(Al_{oct}(2)) = 562.72$ ppm from gibbsite is taken as reference.

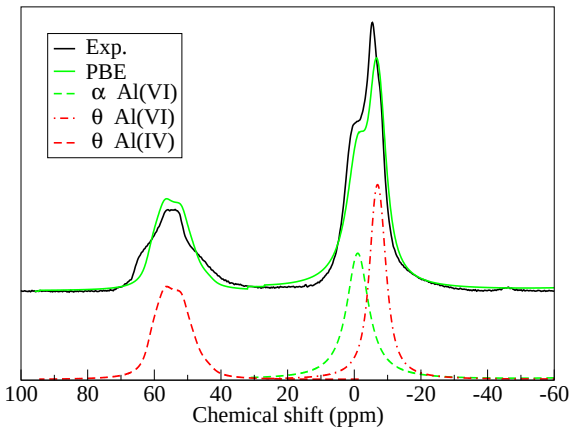


FIG. 1. (color on-line) Comparison between our theoretical results and the experimental ^{27}Al MAS NMR spectra from Ref. [19] for a mixture of the α and θ alumina phases (upper curves). Decomposition of the theoretical spectrum in its individual components (lower curves).

TABLE III. Optimized $Al-O$ distances for the κ-Al_2O_3 structure. Additionally, for the experimental[40] and the optimized structures, we present the average absolute deviation in the distances, $D_{dist.}$, and in the $O-Al-O$ angles, $D_{ang.}$, according to equations (1) and (2).

	$Al_{oct}(1)$	$Al_{V+I}(2)$	$Al_{tet}(3)$	$Al_{oct}(4)$
Distances (Å)	1.962	1.991	1.797	1.879
	1.961	2.269	1.772	1.959
	1.936	1.848	1.756	1.821
	1.820	1.829	1.784	2.026
	1.917	2.038	–	1.838
	1.968	1.844	–	2.215
$D_{dist.}$ (Å)				
Th.	0.039	0.130	0.013	0.111
Exp.	0.058	0.171	0.025	0.105
$D_{ang.}$ (°)				
Th.	5.028	8.392	3.572	7.305
Exp.	5.150	8.590	5.081	6.888

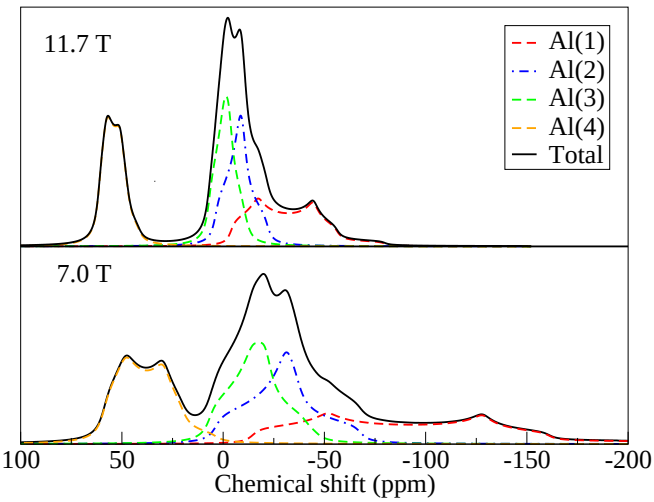


FIG. 2. (color on-line) Simulated ^{27}Al MAS NMR spectra for κ-alumina at 11.7 T (upper graph) and 7.0 T (lower graph). Decomposition of the spectra in individual components is also given.

2. κ phase

The crystal structure of κ-alumina has been reported in a work by Ollivier *et al.*[40], in which ^{27}Al MAS NMR and multiple quantum magic angle spinning (3Q MQ MAS) ^{27}Al NMR experiments were also performed and interpreted on the basis of the local structure. Three distinct resonance signals were indentified in the spectrum obtained with magnetic field intensity of 7.0 T, two of them from Al_{oct} sites. At higher magnetic field of 11.7 T, further resolution revealed an additional peak for this site at high shielding, resulting in the interpretation of the complex lineshape for Al_{oct} peak region as the overlap of three distinct peaks: two with high, one with low quadrupolar coupling. The high coupling component was assigned to the strongly distorted Al_{oct} site, $Al(2)$ in the published crystallographic description, and was labeled as Al_{V+I} to stress the presence of a strongly elongated $Al-O$ bond.

In order to discuss the local distortion of the aluminium sites we calculated the average absolute deviation in distances

$$D_{dist.} = \sum_{i=1}^{n} \frac{|d_i - \overline{d}|}{n} \qquad (1)$$

where n = 4 or 6 for Al_{tet} and Al_{oct} sites, respectively. The angular deviation of the $O-Al-O$ angles from their ideal values in octahedra and tetrahedra are calculated as

$$D_{ang.} = \sum_{i=1}^{k} \frac{|\beta_i - \beta_{ref}|}{k} \qquad (2)$$

where k = 6 and $\beta_{ref} = 109.47°$ for the Al_{tet} sites or k = 12 and $\beta_{ref} = 90°$ for the Al_{oct} sites.

As it can be seen in Table III, our calculations confirm the order of increasing distortion in octahedral sites suggested in the experimental analysis and in particular site $Al_{V+I}(2)$ is found to be the most distorted.

The simulated total spectra, reported in Fig. 2, obtained with our calculated δ_{iso}, C_Q and n_Q, are in very good qualitative agreement with the ^{27}Al MAS NMR experimental spectrum of Ref. [40] at both magnetic field intensities.

However, from the theoretical decomposition of the composite octahedral peak, it can be seen that the high shielding feature around -50 ppm, which was experimentally observed at high magnetic field intensity and was assigned to be Al(2) on the basis of the distortion data, is actually part of a bimodal peak with high quadrupolar coupling and low asymmetry parameter, which belongs to Al(1). Morever, the experimental detection of this particular feature only at high field intensity can also be understood considering the *ab initio* spectrum for low field intensity, reported in the lower part of Fig. 2. As can be seen, at low magnetic field intensity, Al(1) peak broadens considerably due to its high quadrupolar coupling constant, making it diffcicult to observe over the background in the experiment. This demonstrates that, even though experiments can provide accurate structural information, *ab initio* NMR calculations might be essential for an unambiguous peak assignment.

3. gibbsite and boehmite phases

The structures and NMR properties of gibbsite and boehmite are experimentally well characterized.[44,51–53].

As can be seen in Fig. 3, the gibbsite NMR spectrum simulated with the theoretically determined parameters agrees very satisfactorily with the experimentally obtained spectrum by Hill *et al.*[18]. This structure is the only one for which we observed a noticeable dependence of the theoretical spectra on the exchange correlation functional used, showing improved results when using vdW-DF[31] functional (see inset in Fig. 3). In Table II, we present a quantitative comparison between our calculated NMR parameters and the experimental results reported by Damodaran *et al.*[44]. Our results for C_Q and η_Q are in good agreement with both of these experiments and can accurately reproduce the asymmetry of the peak.

In agreement with the previous theoretical work by Vyalikh *et al.*[51], in Fig. 3 we show that this resonance profile can be decomposed as the superposition of two distinct peaks with different second-order quadrupolar structure, belonging to two distinct Al_{oct} sites in gibbsite.

As described in Ref. [51] the two sites differ in the OH groups surrounding them. This can be seen in Fig. 4 where a single $Al(OH)_3$ layer is drawn showing that among the six OH groups surrounding $Al_{oct}(1)$, two participate in interlayer hydrogen-bonds as donors while the remaining four are oriented in-plane and participate to interlayer hydrogen-bonds as acceptors. For $Al_{oct}(2)$ the

opposite occours.

We further characterize the two aluminium sites reporting, in Table IV, the ^{17}O chemical shifts, the Born effective charges and the hydrogen-bond connectivity of the six oxygen types (labeled in Fig. 4) surrounding them. This analysis reveals that the two aluminium sites are mostly surrounded by the same types of oxygens and the distinction is based on just two different oxygen environments (O_c and O_d), each one neighboring only one type of aluminium site.

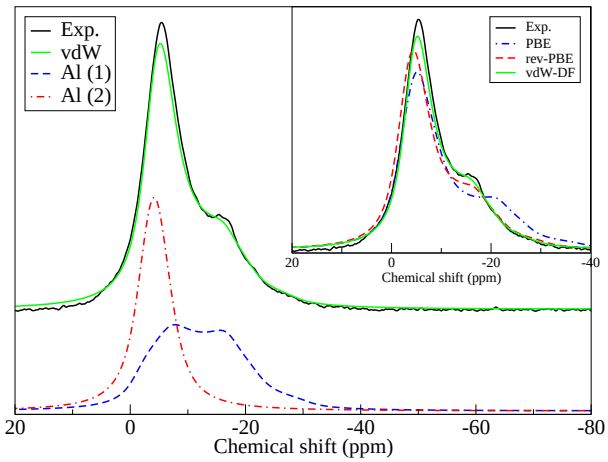


FIG. 3. (color on-line) Comparison between the experimental ^{27}Al MAS NMR spectra from Ref. [18] for gibbsite and the theoretical spectrum obtained with the vdW-DF functional. Decomposition of the spectrum in individual components is also given. The effect of different exchange and correlation functionals on the simulated spectrum is shown in the inset.

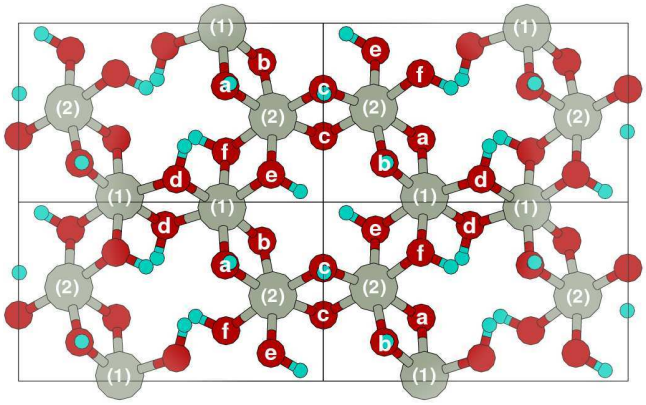
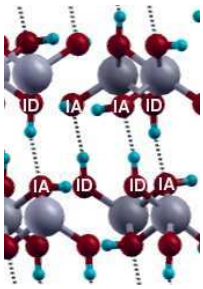


FIG. 4. (color on-line) View from direction (001) of a single $Al(OH)_3$ layer in gibbsite showing the arrangement of the different types of hydroxyl groups around the aluminium cations (large gray atoms). Oxygen (small red atoms) are labeled according to their different environments as referred in Table IV.

The oxyhydroxide polymorph boehmite, γ-$AlO(OH)$, was detected by XRD in the experiment by Hill *et al.*[18] in the temperature range from 200 to 400°C in a mixture with gibbsite or the transitional χ phase. A direct comparison between a simulated NMR spectrum for this phase with that experiment is therefore not possible. In the experiment by Damodaran *et al.*[44] the lack of high resolution in the boehmite NMR spectrum is suggested to be due to the disorder in Al positions in the sample used, leading to small variations in the isotropic chemical shift and quadrupolar couplings. Nevertheless, it was possible to extract experimental estimates for C_Q and η_Q, which are in good agreement with our calculated values, as presented in Table II.

TABLE IV. Effective charges, chemichal shifts and hydrogen-bonding properties (ID = interlayer donor; IA = interlayer acceptor. See side figure) for the different types of oxygen atoms in gibbsite. Labelling according to Fig. 4.

Site	Z^*	σ_{iso}	H-Bond
O_a	-1.46	238.2	ID
O_b	-1.45	242.4	ID
O_c	-1.43	246.2	ID
O_d	-1.41	229.4	IA
O_e	-1.38	238.4	IA
O_f	-1.36	228.0	IA



C. The γ-Al_2O_3 phase and its structural models

As previously mentioned, γ-alumina structure is not yet well characterized in the literature and a vast discussion about this subject persists. Experimental NMR spectra of γ-alumina have been reported in a number of studies. In the work by O'Dell *et al.*[19], the samples obtained with calcination temperatures from 600 °C to 1000 °C are attributed to mixtures of cubic spinel transitional phases, as indicated by the presence of a tetrahedral peak around 60 ppm (taking the α phase as a reference). The fivefold-coordinated peak that has been observed in this phase has been attributed to inherent disorder[19] or surface atom contributions[54]. The MAS NMR study of Pecharromán *et al.*[55] reported a small percentage of AlO_5 sites as well, along with an occupation of 76.3% and 21% for Al_{oct} and Al_{tet} sites, respectively. However, no evidence of a fivefold-coordinated Al peak is present in some other experiments, such as the one of Ref. [56], obtained from a hydrogenated boehmite precursor calcinated at 600 °C for several hours and the one of Ref. [18], in which the sample was obtained from the calcination of gibbsite at 700 °C through the formation of boehmite at 300 °C.

To gain further insight on the structure of this phase, we have calculated the *ab initio* NMR parameters and obtained the simulated spectrum for each one of the four γ-alumina structural models that have been mentioned in Section II C . From the comparison of the calculated spectra with the experimental ones obtained for calcina-

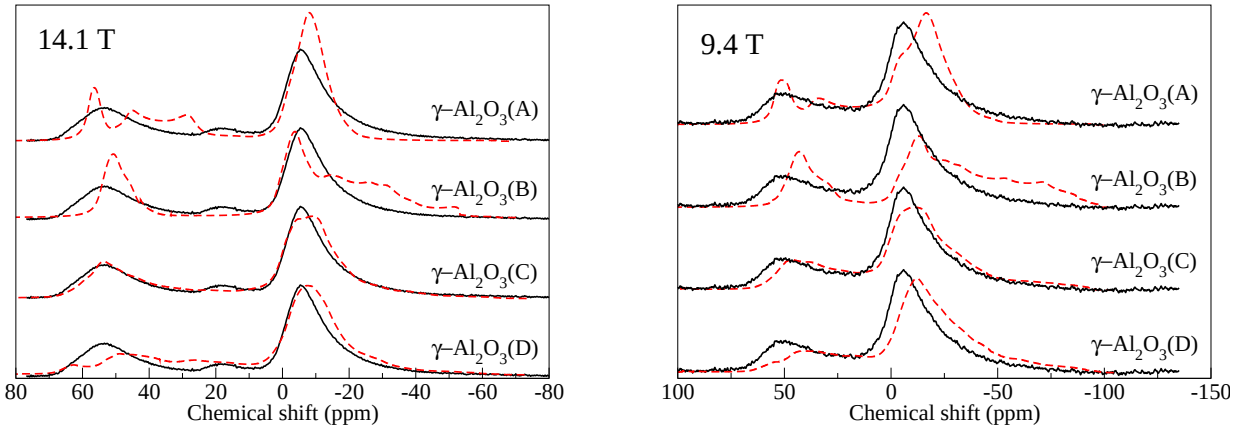


FIG. 5. (color on-line) Comparison of the theoretical spectra (dashed red lines) for the four γ-alumina models considered in this study and the experimental ^{27}Al MAS NMR spectra (black solid lines) from Ref. [19] (left panel) and Ref. [18] (right panel). A Lorentzian broadening of 0.5 KHz was added to each component which were properly normalized to reflect relative site abundance.

tion conditions associated to γ-alumina, an evaluation of the adequacy of these structural models will be possible. Since NMR is a sensitive probe of the local structural and chemical environment, comparisons based on NMR parameters will provide complementary constraints with respect to X-ray and neutron diffraction methods used, for instance, in the detailed structural search of Ref. [49]

In Fig. 5 we compare the theoretical spectra with the experimental data by O' Dell *et al.*[19] (left panel, 14.1 T), and by Hill *et al.*[18] (right panel, 9.4 T). We also compared with the 9.4 T spectrum by Pecharromán *et al.*[55] (not shown) which agrees very well with the one by Hill *et al.*.

From Fig. 5, it is clear that the simulated MAS NMR spectra for the γ-$Al_2O_3(C)$ structural model proposed by Paglia *et al.*[49] best reproduces all the experimental results considered. The agreement is very satisfactory and even better for the more precise spectrum at the higher magnetic field [left panel of Fig. 5].

It may appear natural that the two models that display lesser agreement with the experimental spectra are the two (models A and B) whose unit cells only contain 8 Al_2O_3 formula units and a reduced number of non-equivalent aluminium environments, while model C contains a large number of Al_2O_3 formula units allowing a distribution of NMR parameters, which is more suitable for the representation of the broad features observed experimentally. Notice however that having a large number of non-equivalent aluminium environments may be considered a necessary, but not a sufficient condition to properly reproduce the experimental NMR spectra. In fact both Paglia's structural models (C and D) contain 64 Al_2O_3 formula units and reproduce neutron diffraction data[49] equally well but only model C (the one associated to $Fd\overline{3}m$ symmetry) satisfactorily reproduces the spectral region of the tetrahedrally coordinated Al atoms

TABLE V. Average values and corresponding standard deviations for ^{27}Al chemical shifts, absolute quadrupolar coupling, asymmetry parameter, and site occupation, decomposed according to Al coordination number, for the four γ-alumina models considered in this work.

	δ_{iso} (ppm)	$\lvert C_Q\rvert$ (MHz)	η_Q	%
γ-$Al_2O_3(A)$				
AlO_6	-0.6±1.8	4.90±0.47	0.86 ± 0.12	62.5
AlO_4	52.1±3.0	6.36±2.65	0.14 ± 0.08	37.5
γ-$Al_2O_3(B)$				
AlO_6	-0.7±4.9	7.98±3.25	0.41±0.23	75
AlO_4	52.0±0.1	4.51±0.91	0.47±0.35	25
γ-$Al_2O_3(C)$				
AlO_6	-0.8±4.3	5.93±2.45	0.52±0.26	64.0
AlO_5	15.9	7.43	0.34	1.6
AlO_4	59.9±4.7	7.70±3.00	0.64±0.25	34.4
γ-$Al_2O_3(D)$				
AlO_6	-0.1±4.2	6.25±2.38	0.57±0.23	67.2
AlO_4	52.7±6.5	8.27±2.92	0.58±0.27	32.8

and even for this model the sharpness of the octahedral-Al peak is not completely satisfactory. Table V shows that, on average, model C gives a higher average value for δ_{iso} for the Al_{tet} peak while the other models display rather similar values that underestimate the experiment.

Furthermore, in the γ-Al_2O_3 structural model C, a truly AlO_5 site is evident, and its calculated δ_{iso} of 15.9 ppm is easily separated from the peaks of octahedrally (-9.4 to 9.4 ppm) and tetraedrally (42.6 to 68.0 ppm) coordinated sites. Although the small occupation of this site in model C cannot reproduce the AlO_5 peak observed by some experiments, the δ_{iso} is in the experimentally

measured range, implying that an increase in AlO_5 sites due to defects and/or surface effects would explain the observed feature.

It should be noted that in the study by Pecharromán *et al.*[55], the complex NMR spectrum in the γ-alumina region was analyzed as superposition of a small number of peaks for which the quadrupolar interaction parameters were estimated for tetrahedral ($C_Q = 4.7$ - 4.9 MHz) and octahedral ($C_Q = 3.6$ - 3.9 MHz) sites. These estimated values do not agree with the ones obtained for model C. Since model C shows very good agreement for the total spectrum, this implies that extracting NMR parameters from a complex spectrum, in absence of further experimental or theoretical charaterization, is likely an unreliable procedure as it can result in widely different distributions of NMR parameters for the same spectrum.

D. Correlation analysis

In the literature there have been several attempts to correlate features in NMR spectra with local structural details such as coordination numbers[22], local atomic bondlength and angular distortion[57], shortest bondlength[58], Mulliken charge population[22] etc.

Our calculations confirm the dependence of chemical shifts on aluminium coordination numbers, such that 4-, 5- and 6-coordinated aluminium sites show chemical shifts in well separated ranges: the lower the coordination number the larger is the chemical shift.

However, the internal distribution within a given coordination number does not correlate with local geometric descriptors such as local bondlength, angular distortions[57], or shortest bondlength[58]. To explain this we performed a simple test in the *non-spinel* model B for γ-alumina by displacing an oxygen atom while keeping the rest fixed and we observed that the NMR parameters were affected in a region extending up to the third coordination shell, thus demonstrating the sensitivity of NMR to non-local structural details.

The only properties that we found to display a significant correlation with the NMR shieldings were aluminium Bader charges and Born dynamical effective charges. For both charges the correlation is linear and structure independent as demostrated by the linear fits obeyed equally well by all phases, as shown in Fig. 6.

It should be mentioned that Bader charges and Born effective charges correlate very strongly with each other so that no additional information can be gained by considering the two charges together. This is because γ-alumina is a crystal with a high degree of ionicity, where these two quantities are strongly related.

Furthemore, we can understand the correlation of NMR shieldings with Born effective charges because they both originate from local electronic susceptibility, one being determined from the current induced by the magnetic field, the other measuring the charge flow associated to vibrational motion.

As an hystorical curiosity, we mention that proposing a correlation between NMR chemical shifts and effective charges is not a complete novelty since in a couple of papers[59,60] in the late seventies a correlation was reported to exist for binary semiconductors among Szigeti's effective charges and ^{27}Al and ^{31}P NMR chemical shifts.

IV. CONCLUSIONS

We investigated from first principles the ^{27}Al NMR properties of several well characterized crystalline phases of Al_2O_3 and of two of its calcination precursor phases, obtaining very good agreement with available experimental results. New insight for the peak assignments in the spectra were proposed for some structures.

This gave confidence in the theoretical approach and allowed us to address the open problem of the structural characterization of the technologically important γ-alumina phase by comparing the experimental spectra with the theoretical predictions calculated for four structural models recently appeared in the literature: our study supports the model structure with $Fd\overline{3}m$ symmetry proposed by Paglia *et al.* in Ref. [49] as the one that best reproduces the NMR experimental results in the bulk.

Calculations confirm that chemical shifts strongly depend on coordination number. Moreover, within a given coordination number, a linear correlation exists between chemical shifts and Born effective charges or Bader charges.

V. ACKNOWLEDGMENTS

We are grateful to Dr. L.A. O'Dell, Dr. M.E. Smith, and Dr. T.J. Bastow for providing their experimental NMR spectra and to Dr. Gonzalo Gutiérrez and Dr. Eduardo Menéndez-Proupin for providing their structural model. SdG and EK like to thank Davide Ceresoli for useful discussions and for suggesting the use of QuadFit code for the simulation of spectra including quadrupolar interaction. ARF wishes to thank SISSA for the support and facilities during the obtention of the results. Calculations have been performed on the Sp6-IBM machine at CINECA in Bologna and on the HPC cluster at SISSA. This work was also supported by Petrobras S.A. and brazilian agencies CAPES, FAPEMIG and CNPq.

[1] S. H. Cai, S. N. Rashkeev, S. T. Pantelides, and K. Sohlberg, Phys. Rev. B 67, 224104 (2003).

[2] K. J. D. MacKenzie, J. Temuujin, and K. Okada, Thermochimica Acta **327**, 103 (1999).

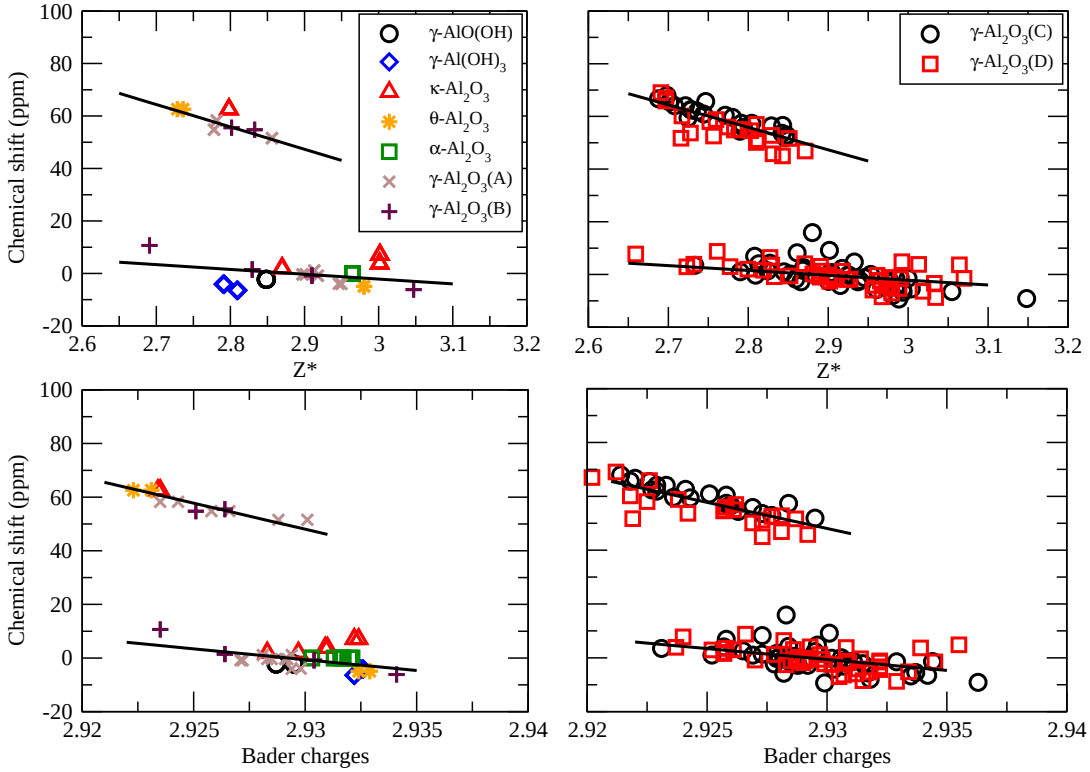


FIG. 6. (color on-line) Correlations between chemical shifts and effective charges, Z^*, (above) or Bader charges (below). Left panels collect the data for all the structures considered in this work except the γ-alumina models C and D, that contain many different Al sites. For these models the data are reported separately in the right panels. Lines correspond to linear fit to Al_{oct} and Al_{tet} data for all the structures.

[3] Y. Cesteros, P. Salagre, F. Medina, and J. E. Sueiras, Chem. Mater. **11**, 123 (1999).
[4] J.E. Brown, D. Clark, W.W. Elliott, J. Chem. Soc. **13**,84 (1953).
[5] D. B. Tilley and R. A. Eggleton, Clays and Clay Min. **44**, 658 (1996).
[6] G. Ertl, H. Knözinger, J. Weitkamp, The Handbook of Heterogeneous Catalysis, Wiley-VCH, Weinheim, 1997
[7] M. Digne, P. Sautet, P. Raybaud, P. Euzen, and H. Toulhoat, J. Catal. **226**, 54 (2004).
[8] H. C. Stumpf, A. S. Russell, J. W. Newsome, and C. M. Tucker, Ind. Eng. Chem. **42**, 1398 (1950).
[9] R. Zhou, and R. L. Snyder, Acta Cryst. **B47**, 617 (1991).
[10] M.-H. Lee, C.-F. Feng, V. Heine, and J. Klinowki, Chem. Phys. Lett. **265**, 673 (1997).
[11] J.A. Wang, X. Bokhimi, A. Morales, O. Novaro, T. López, R. Gómez, J. Phys. Chem. B **B103**, 299 (1999).
[12] C. Wolverton, K.C. Hass, Phys. Rev B **63**, 024102 (2000)..
[13] K. Sohlberg, S.J. Pennycook, S.T. Pantelides, Chem. Eng. Comm. **181**, 107 (2000).
[14] X. Krokidis, P. Raybaud, A. Gobichon, B. Rebours, P. Euzen, and H. Toulhoat, J. Phys. Chem. B **105**, 5121 (2001).
[15] P. Raybaud, M. Digne, R. Iftimie, W. Wellens, P. Euzen, and H. Toulhoat, J. Catal. **201**, 236 (2001).
[16] Ľ. Smrčok, V. Langer, J. Křesťan, Acta Cryst. C62, i83 (2006).
[17] A.R. Ferreira, M.J.F. Martins, E. Konstantinova, R.B. Capaz, W.F. Souza, S.X. Chiaro, and A.A. Leitão, J. Solid State Chem. **184**, 1105 (2011).
[18] M. R. Hill, T. J. Timothy, J. Bastow, S. Celotto, and A. J. Hill, Chem. Mater. **19**, 2877 (2007).
[19] L. A. O'Dell, S. L. P. Savin, A. V. Chadwick, and M. E. Smith, Solid State Nucl. Magn. Reson. **31**, 169 (2007).
[20] C. J. Pickard and F. Mauri, Phys. Rev. B **63**, 245101 (2001).
[21] J. R. Yates, C. J. Pickard, and F. Mauri, Phys. Rev. B **76**, 024401 (2007).
[22] M. Choi, K. Matsunaga, F. Oba, and I. Tanaka, J. Phys. Chem. C **113**, 3869 (2009).
[23] R. Lizárraga, E. Holmström, S. C. Parker, and C. Arrouvel, Phys. Rev. B **83**, 094201 (2011).
[24] R. Bader, Atoms in Molecules: A Quantum Theory; Oxford University Press; New York, 1990. W. Tang, E. Sanville, and G. Henkelman, J. Phys.: Condens. Matter **21**, 084204 (2009).
[25] M. Born, and K. Huang, Dynamical Theory of Crystal Lattices (Oxford University Press, Oxford), 1954; H. Böttger, Principles of the Theory of Lattice Dynamics (Physik-Verlag, Weinheim), 1983.
[26] P. Giannozzi, *et al.*, J. Phys. Cond. Matt. **21**, 395502 (2009). Web site: http://www.quantum-espresso.org.
[27] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[28] W. Kohn, and L.J. Sham, Phys. Rev. **140**, A1133 (1965).
[29] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[30] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).
[31] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).

[32] T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard, and D. C. Langreth, Phys. Rev. B **76**, 125112 (2007).

[33] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[34] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[35] The k-point meshes for each structure are as follows: Boehmtie 3x1x3; Gibbsite 2x2x2; α-Al_2O_3 2x2x1; κ-Al_2O_3 3x2x2; θ-Al_2O_3 1x4x2; γ-Al_2O_3(A) 2x2x1; γ-Al_2O_3(B) 2x2x1; γ-Al_2O_3(C) 2x2x1; γ-Al_2O_3(D) 1x2x1. All meshes were shifted by half of the mesh interval in each direction.

[36] P. Giannozzi, S. de Gironcoli, P. Pavone, and S. Baroni Phys. Rev. B **43**, 7231 (1991).

[37] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[38] T. F. Kempa and M. E. Smith, Solid State Nucl. Magn. Reson. **35**, 243 (2009).

[39] N. Ishizawa, T. Miyata, J. Minato, F. Marumo, and S. Iwai, Acta Cryst. B **36**, 228 (1980).

[40] B. Ollivier, R. Retoux, P. Lacorre, D. Massiot, and G. Férey, J. Mater. Chem. **7**, 1049 (1997).

[41] H. Saalfeld and M. Wedde, Z. Kristallogr. **139**, 129 (1974).

[42] A. Ramos-Gallardo and A. Vegas. Z. Kristallogr. **211**, 299 (1996).

[43] A. N. Christensen, M. S. Lehmann, and P. Convert, Alta Chem. Scand **A36**, 303 (1982).

[44] K. Damodaran, P. R. Rajamohanan, and D. Chakrabarty, J. Am. Chem. Soc. **124**, 3200 (2002).

[45] G. Gutiérrez, A. Taga, and B. Johansson, Phys. Rev. B **65**, 012101 (2001).

[46] E. Menéndez-Proupin and G. Gutiérrez, Phys. Rev. B **72**, 035116 (2005).

[47] W. Y. Ching, L. Ouyang, P. Rulis, and H. Yao, Phys. Rev. B **78**, 014106 (2008).

[48] G. Paglia (2004), Determination of the structure of c-alumina using empirical and first principles calculations combined with supporting experiments. Ph.D. thesis, Curtin University of Technology.

[49] G. Paglia, A. L. Rohl, C. E. Buckley, and J. D. Gale, Phys. Rev. B **71**, 224115 (2005).

[50] C. Loyola, E. Menéndez-Proupin, and G. Gutiérrez, J. Mater. Sci. **45**, 5094 (2009).

[51] A. Vyalikh, K. Zesewitz, and U. Scheler, Magn. Reson. Chem. **48**, 877 (2010).

[52] R. C. Slade, J. C. Southern, and I. M. Thompson, J. Mater. Chem. **1**, 563 (1991).

[53] S. E. Ashbrook, J. McManus, K. J. D. MacKenzie, and S. Wimperis, J. Phys. Chem. B **104**, 6408 (2000).

[54] J. H. Kwak, J. Z. Hu, D. H. Kim, J. Szanyi, and C. H. F. Peden, J. Catal. **251**, 189 (2007).

[55] C. Pecharromán, I. Sobrados, J. E. Iglesias, T. González-Carreño, and J. Sanz, J. Phys. Chem. B **103**, 6160 (1999).

[56] G. Paglia, C. E. Buckley, A. L. Rohl, B. A. Hunter, R. D. Hart, J. V. Hanna, and L. T. Byrne, Phys. Rev. B **68**, 144110 (2003).

[57] D. Mao, E. J. Walter, and H. Krakauer, Phys. Rev. B **76**, 014105 (2007).

[58] D. L. Pechkis, E. J. Walter, and H. Krakauer, http://arxiv.org/abs/1107.4611 (2007).

[59] R. E. J. Sears, Phys. Rev. B **18**, 3054 (1978).

[60] R. E. J. Sears, Phys. Rev. B **22**, 1135 (1980).

Bibliography

[1] SSNMR spectrometers with ultra-high magnetic field of 21.1 T are available.

[2] V. Ladizhansky, Solid State Nucl. Magn. Reson. **36**, 119 (2009).

[3] K. Jayalakshmi, K. Sonkar, A. Behari, V.K. Kapoor and N. Sinha, Solid State Nucl. Magn. Reson. **36**, 60 (2009).

[4] D. M. Small and G. G. Shipley, Science **185**, 222 (1974).

[5] D. Heuman, W. Wassef, Z. Vlahcevic, *Gallstones in Shiff's Diseases of the Liver* (eds. E. R. Shiff, M. F. Sorrell, W. C. Maddrey) 631-645 (Lippincott-Raven, Philadelphia, 1999).

[6] B. M. Craven, Nature **260**, 727 (1976).

[7] H. S. Shieh, L. G. Hoard and C. E. Nordman, Nature **267**, 287 (1977); H. S. Shieh, L. G. Hoard and C. E. Nordman, Acta Cryst. B **37** 1538 (1981).

[8] L. Y. Hsu and C. E. Nordman, Science **220**, 604 (1983).

[9] L. Y. Hsu, J. W. Kampf and C. E. Nordman, Acta Cryst. B **58**, 260 (2002).

[10] C. J. Pickard and F. Mauri, Phys. Rev. B **63**, 245101 (2001).

[11] M. dAvezac, N. Marzari and F. Mauri. Phys. Rev. B **76**, 165122 (2007).

[12] M. Benoit, M. Profeta, F. Mauri, C. J. Pickard and M. E. Tuckerman, J. Phys. Chem. B **109**, 6052 (2005).

[13] J. R. Yates, T. N. Pham, C. J. Pickard, F. Mauri, A. M. Amado, A. M. Gil and S. P. Brown, J. Am. Chem. Soc. **127**, 10216 (2005).

[14] S. Baroni, S. de Gironcoli, A. dal Corso and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).

[15] P. E. Blochl, Phys. Rev. B **50**, 17953 (1994).

[16] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[17] J. R. Yates, C. J. Pickard and F. Mauri, Phys. Rev. B **76**, 24401 (2007).

[18] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni et al., J. Phys.: Condens. Matter **21**, 395502 (2009), URL: http://www.quantum-espresso.org.

[19] P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864 (1964).

[20] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).

[21] I. Štich, R. Car, M. Parinello and S. Baroni, Phys. Rev. B **39**, 4997 (1989).

[22] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).

[23] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[24] J. P. Perdew and Y. Wang. Phys. Rev. B **45**, 13244 (1992).

[25] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[26] Y. Zhang and W. Yang Phys. Rev. Lett. **80**, 890 (1998).

[27] M. Dion, H. Rydberg, E. Schroder, D. C. Langreth, and B. I. Lundqvist. Phys. Rev. Lett. **92**, 246401 (2004).

[28] G. Román-Pérez and J. M. Soler, Phys. Rev. Lett. **103**, 096102 (2009).

[29] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist and D. C. Langreth1, Phys. Rev. B **82**, 081101 (2010).

[30] V. R. Cooper, Phys. Rev. B **81**, 161104(R) (2010).

[31] O. Krogh Andersen, Phys. Rev. B **12**, 3060 (1975).

[32] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[33] L. Paulatto, Ammonia Synthesis on Proton-enriched Palladium Substrate, SISSA (PhD. Thesis), Trieste, Italy, 2009.

[34] Encyclopedia of NMR, edited by D.M. Grant and R.K. Harris (Wiley, London, 1996).

[35] See e.g., J.A. Tossell, Phys. Chem. Miner. **27**, 70 (1999); Chem. Phys. Lett. **303**, 435 (1999); G. Valerio and A. Goursot, J. Phys. Chem. **103**, 51 (1999).

[36] See e.g., P. Florian, F. Fayon and D. Massiot, J. Phys. Chem. C **113**, 2562 (2009); S. Cadars, D.H. Brouwer and B.F. Chmelka, Phys. Chem. Chem. Phys. **11**, 1825 (2009); T.H. Sefzik, T.M. Clark and P.J. Grandinetti, Solid State Nucl. Magn. Reson. **32**, 16,(2007).

[37] F. Mauri, B. G. Pfrommer, and S. G. Louie, Phys. Rev. Lett. **77**, 5300 (1996); F. Mauri and S.G. Louie, Phys. Rev. Lett. **76**, 4246 (1996).

[38] T. Gregor, F. Mauri and R. Car, J. Chem. Phys. **111**, 1815 (1999).

[39] A. A. Adllan and A. Dal Corso, J. Phys.: Condens. Matter **23**, 425501 (2011). Publicly available via: http://qe-forge.org/projects/pslibrary/

[40] T. Gregor, F. Mauri, and R. Car, J. Chem. Phys. **111**, 1815 (1999).

[41] T. A. Keith and R. F. W. Bader, Chem. Phys. Lett. **194**, 1 (1992); T. A. Keith and R. F. W. Bader, Chem. Phys. Lett. **210**, 223 (1993).

[42] NMRShiftDB : Open source database:
http://nmrshiftdb.nmr.uni-koeln.de/portal and
http://www.chem.wisc.edu/areas/reich/Handouts/nmr-c13/cdata.htm.
Since cyclobutane is a commonly used molecule in NMR experiments. The variations between experiments can be attributed to experimental settings and solutions used.

[43] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[44] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).

[45] Z. Cournia, J. C. Smith and G. M. Ullmann, J. Comput. Chem. **26**, 1383 (2005).

[46] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[47] S. Plimpton, J. Comp. Phys. **117**, 1 (1995).
URL: http://lammps.sandia.gov.

[48] MacKerell, Bashford, Bellott, Dunbrack, Evanseck, Field, Fischer, Gao, Guo, Ha, et al, J Phys Chem, 102, 3586 (1998)

[49] Lennard-Jones and Coulombic interactions are computed with an additional switching function that ramps the energy and force smoothly to zero between an inner and outer cutoff, which are chosen as 10 Åand 12 Årespectively.

[50] P. J. in 't Veld, A.E. Ismail, and G. S. Grest, J. Chem. Phys. **127**, 144711 (2007).

[51] T. Schneider and E. Stoll, Phys Rev B **17**, 1302 (1978).

[52] Steepest descent, damped dynamics method (as described in D. Sheppard, R. Terrell and G. Henkelman, J. Chem. Phys. **128**, 134106 (2008)) and Hessian-free truncated Newton algorithm were used as implemented in LAMMPS code.

[53] A. C. Vaiana, A. Schulz, J. Worfrum, M. Sauer, J. C. Smith, J. Comput. Chem. **24**, 632 (2003).

[54] M. C. Frincu, S. D. Fleming, A. L. Rohl and J. A. Swift, J. Am. Chem. Soc. **126**, 7915 (2004).

[55] D. W. Russell, K. D. Setschell, Biochemistry **31**, 4737 (1992).

[56] W. Guo and J. A. Hamilton, Biophys. J. **71** 2857 (1996).

[57] EK, K. Sonkar, N. Sinha and S. de Gironcoli, in preparation.

[58] A. L. Webber, L. Emsley, R. M. Claramunt and S. P. Brown, J. Phys. Chem. A **114**, 10435 (2010).

[59] C. Liang, L. Yan, J. R. Hill, C. S. Ewig, T. R. Stouch and A. T. Hagler, J. Comput. Chem. **16**, 883 (1995).

[60] E. V. Boldyreva, S. N. Ivashevskaya, H. Sowa, H. Ahsbahs and H. P. Weber, Dokl. Phys. Chem. **396**, 111 (2004).

[61] A. Dawson, D. R. Allan, S. A. Belmonte, S. J. Clark, W. I. F. David, P. A. McGregor, S. Parsons, C. R. Pulham and L. Sawyer, Cryst. Growth Des. **5**, 1415 (2005).

[62] J. P. Legros, A. Kvick, Acta Crystallogr. B **36**, 3052 (1980).

[63] G. L. Perlovich, L. K. Hansen, Brauer-Brandl, A. J. Therm. Anal. Calorim. **66**, 699 (2001).

[64] A. Kvick, Acta Crystallogr. B **36**, 115 (1980).

[65] A. Ferrera, EK, A. Leitao and S. de Gironcoli, submitted to Phys. Rev. B.