

PREMIO TESI DI DOTTORATO

- 43 -

PREMIO TESI DI DOTTORATO
Commissione giudicatrice, anno 2013

Luigi Lotti, *presidente della Commissione*

Tito Arecchi, *Area Scientifica*

Franco Cambi, *Area Umanistica*

Paolo Felli, *Area Tecnologica*

Michele Arcangelo Feo, *Area Umanistica*

Roberto Genesisio, *Area Tecnologica*

Luigi Lotti, *Area Scienze Sociali*

Mario Pio Marzocchi, *Area Scientifica*

Adolfo Pazzagli, *Area Biomedica*

Mario Giuseppe Rossi, *Area Umanistica*

Salvatore Ruggieri, *Area Biomedica*

Saulo Sirigatti, *Area Biomedica*

Piero Tani, *Area Scienze Sociali*

Fiorenzo Cesare Ugolini, *Area Tecnologica*

Vincenzo Varano, *Area Scienze Sociali*

Graziella Vescovini, *Area Umanistica*

Francesco Muniz Miranda

Modelling of spectroscopic and structural properties using molecular dynamics

Firenze University Press
2014

Modelling of spectroscopic and structural properties using molecular dynamics / Francesco Muniz Miranda. – Firenze : Firenze University Press, 2014. (Premio Tesi di Dottorato; 43)

<http://digital.casalini.it/9788866556909>

ISBN 978-88-6655-690-9 (online)

Peer Review Process

All publications are submitted to an external refereeing process under the responsibility of the FUP Editorial Board and the Scientific Committees of the individual series. The works published in the FUP catalogue are evaluated and approved by the Editorial Board of the publishing house. For a more detailed description of the refereeing process we refer to the official documents published in the online catalogue of the FUP (www.fupress.com).

Firenze University Press Editorial Board

G. Nigro (Co-ordinator), M.T. Bartoli, M. Boddi, R. Casalbuoni, C. Ciappei, R. Del Punta, A. Dolfi, V. Fargion, S. Ferrone, M. Garzaniti, P. Guarnieri, A. Mariani, M. Marini, A. Novelli, M. Verga, A. Zorzi.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0: <http://creativecommons.org/licenses/by/4.0/>)

CC 2014 Firenze University Press
Università degli Studi di Firenze
Firenze University Press
Borgo Albizi, 28, 50122 Firenze, Italy
www.fupress.com

The training is nothing!
Will is Everything!

Ra's al Ghul', Batman Begins, 2005

*Actually, this is a spoiler. But after almost ten years from the premiere, I suppose I can be forgiven.

Table of contents

Abstract	11
Introduction	13
I Theory and Models	21
1 Computational “Tools” in a Nutshell	23
2 Computational Chemistry	27
2.1 <i>Ab Initio</i> Calculations	27
2.1.1 Molecular Quantum Mechanics	27
2.1.2 Electron Density	30
2.1.3 Hohenberg-Kohn Theorem	31
2.1.4 Kohn-Sham Equations	31
2.1.5 Basis Sets	34
2.1.6 Hartree-Fock VS Kohn-Sham	35
2.2 Classical Molecular Dynamics	37
2.2.1 Equations of Motion	38
2.2.2 Integration Scheme	40
2.2.3 Periodic Boundary Conditions	41
2.2.4 Ergodic Hypothesis	42
2.2.5 Force Fields	43
2.2.6 Long Range Interactions	45
2.2.7 Effectiveness and Shortcomings of the Method	45
2.3 <i>Ab Initio</i> Molecular Dynamics	47
2.3.1 Born-Oppenheimer Molecular Dynamics	47
2.3.2 Car-Parrinello Molecular Dynamics	48
2.3.3 Explanations	50
2.3.4 Implementation	53

2.3.5	Hydrogen vs Deuterium	55
3	Molecular Properties	57
3.1	Structure	57
3.1.1	Pair Radial Distribution Functions	57
3.1.2	What Information can be recovered from $g(r)$?	60
3.1.3	Angular Distribution and H-bond Functions	61
3.2	Electronic Structure	62
3.3	Vibrations: Some Rules of Thumb	64
3.4	Vibrations: Linear Response Theory	66
3.5	Vibrations: Spectra from MD Simulations	68
3.5.1	Vibrational Density of States	68
3.5.2	IR and Raman Spectra	70
3.5.3	Time-Sampling and Spectra: Part I	71
4	Fourier and Wavelet Analysis	73
4.1	Fourier Transform	73
4.1.1	Time-Sampling and Spectra: Part II	76
4.1.2	What the Fourier Transform Just Cannot Do	77
4.2	Wavelet Transform	79
4.2.1	Explanations	83
4.2.2	Some Examples	84
4.2.3	Implementations	88
II	Case Studies	93
5	Anchor Dipeptide into Phospholipid Bilayer	97
5.1	Computational Protocols	99
5.2	Brief Survey of Experimental Methods	104
5.3	Findings	104
5.3.1	Experimental	104
5.3.2	<i>Ab Initio</i> Calculations and Molecular Dynamics	106
6	Simple Diols in Water and Acetonitrile	123
6.1	Propanediol and Ethylene Glycol in Water	124
6.1.1	Computational Chemistry Details	124
6.1.2	Structural and electronic properties	126
6.1.3	Time-frequency analysis	131
6.2	Propanediol in Acetonitrile	139

6.2.1	Experimental Setup	139
6.2.2	Computational Chemistry Details	139
6.2.3	Intra-molecular H-Bond of Propanediol	141
6.2.4	... and What About the Nitrile Stretch Blueshift?	154
7	Methyl Acetate in Protic Solvents	163
7.1	Methyl Acetate in Methanol	164
7.1.1	Computational Chemistry Details	164
7.1.2	Structural Data and Spectrum Interpretation	164
7.2	Methyl Acetate in Water	173
7.2.1	Spectral Interpretation	173
7.3	Discussion	176
8	Water Solvation of Thiazole	177
8.1	<i>Ab Initio</i> Molecular Dynamics Details	177
8.2	Simulation Analysis	179
8.2.1	Discussion	181
9	Bifurcated Hydrogen-Bond of Confined Water	183
9.1	Computational Details	186
9.1.1	<i>Ab Initio</i> Molecular Dynamics Simulation	186
9.1.2	Dipole Moment Calculation	187
9.2	Analysis of the Trajectory	188
III	Concluding Remarks	197
	Bibliography	203
	Appendixes	215
	Appendixes	215
A	Some DFT Details	217
A.1	Proof of Hohenberg-Kohn Theorems	217
A.2	Discussion about Kohn-Sham Equations	218
A.3	Exchange-Correlation Functionals	219

B	Some MD Details	223
B.1	Liouville Operator and Integration Algorithms	223
B.2	Replica Exchange	225
C	Some <i>Ab Initio</i> MD Details	227
C.1	Hellman-Feynmann Theorem	227
C.2	Pulay Forces and BSSE	228
C.3	Pseudopotentials	228
C.4	Van der Waals Forces	229
	Acknowledgements	230

Abstract

The following dissertation presents the study that I performed at the Chemistry Department of the University of Florence and at the European Laboratory for Non-Linear Spectroscopy (LENS) to recover and elucidate structural, dynamical, and spectroscopic molecular features adopting computer simulations. In particular, here *ab initio* molecular dynamics simulations and time-frequency analysis are the most employed “tools”, in order to have a better understanding of the origins of vibrational features. Hydrogen-bonding is the main intermolecular interaction that can affect the vibrational spectra, and occurs in all the systems studied here. It is shown how the hydrogen-bonding usually induces a redshift on the vibrational frequencies of the groups engaged in it, and how this redshift nicely correlates with the strength of this type of bond, as obtained from (so-called) “first principles simulations”. While the redshift induced by the hydrogen-bond is the most common occurrence, I have also studied a system where hydrogen-bonding induces a blueshift. Recovering and explaining this somewhat unusual effect, well established experimentally, required particular computational efforts. Hydrogen-bonding can manifest itself also as a bifurcated interaction between one donor and two acceptor centres. This bifurcated configuration is usually seen just as a very brief intermediate step occurring in water during breaking and creation of “true” hydrogen-bonds, but, in confined water, it has longer lifetimes, thus allowing it to be studied by both spectroscopic and computational means. The computational protocols implemented and adopted in this dissertation allow a direct comparison between structural features and vibrational spectrum, highlighting how the formers influence the latter.

Introduction: Vibrations, H-Bonding, and Structure

Vibrations occurring in molecules are a powerful tool to investigate other relevant properties. The time scale of the molecular vibrations ranges from tens of femtoseconds to picoseconds, allowing to probe the dynamics and the structural features of both simple and complex systems. As a matter of fact, even small changes of relative conformations and environment interactions can have deep consequences in the vibrational spectra.

In particular, molecular vibrations are affected by the presence of the hydrogen-bond [Jef]. The hydrogen-bond is an inter-molecular interaction of type $A-H \cdots B$, where A and B contains atoms with high electronegativity,² such as nitrogen and oxygen. The hydrogen atom H is covalently bound to the donor atom of A, the other atom from B acts as the acceptor: these concepts are briefly illustrated in fig. 1.

The hydrogen-bond is partially electrostatic in origin. In fact, high partial charges often lead to large static dipole moments, and dipole-dipole interactions are one of the main contributions of this inter-molecular bond. However, the hydrogen-bond also exhibits a covalent character [ISP⁺99], and, due to this fact, it is highly directional and can modulate the acceptor and donor receptivity of forming more H-bonds, allowing the creation of H-bond networks.

Hydrogen-bonding is the “strongest” inter-molecular interaction due to its (partial) covalent character; typical binding energy of hydrogen and other inter-molecular bonds are reported in table 1. Due to its high binding energy, the hydrogen-bond is the paramount inter-molecular interaction of the systems where it is present, and has deep implications on the thermodynamic and dynamical properties. One of the most recurrent examples is liquid water, whose peculiar properties (which put liquid wa-

²The electronegative atom of A is covalently bonded to H, whereas the electronegative atom of B is a first neighbour of H. The usual distances of an hydrogen-bond are in the range 1.7 – 2.5 Å.

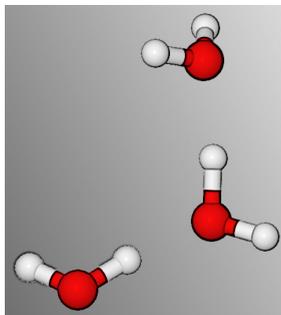


Figure 1: Three water molecules engaged in hydrogen-bonds. The lowest molecule acts as hydrogen-bond donor to that placed in the middle, which therefore is an acceptor. Moreover, the middle molecule also acts as donor towards the highest one.

	covalent	hydrogen	dipole-dipole	van der Waals
Binding energy (kcal) :	$\simeq 400$	$\simeq 15$	$\simeq 1$	< 1

Table 1: Binding energy of selected molecular and inter-molecular interactions. Values are merely indicative of the relative order of magnitude, and are taken from Ref. [Ege].

ter into a completely different league compared to other liquids) are due to its network of hydrogen-bonds. Hydrogen-bonding is also essential to elucidate the behaviour of large bio-molecules, e.g., proteins and ribonucleic acids. The mechanism beneath protein folding and unfolding is driven by breaking and making of hydrogen-bonds, and base pairing and base recognition mechanism in DNA is also due to hydrogen-bonding creation. In fact, it is exactly the intermediate “character” of this type of bond (strong compared to other inter-molecular interactions, weak compared to intra-molecular bonds), that makes it so versatile: it is strong enough to stabilise double helix structures that store the genetic code in DNA, but at the same time it is also weak enough to allow opening of this double helix during cell division.

Thus, the hydrogen-bond effects comprise wide and different fields ranging from biology to matter physics, and a deep understanding of these effects is required to elucidate complex processes at the molecular level.

The relative strength, lifetime and stability of the hydrogen-bond modulate the molecular vibrations of both acceptor and donor groups, and are the main causes of frequency shift and broadening of the vibrational bands.

Due to this relationship, the effect of the hydrogen-bond upon the vibrational spectrum can also be exploited to study the behaviour of the molecular system. Lifetimes of the hydrogen-bond provide information about the molecular dynamics, and structural changes can be inferred from the shifts and the broadening of the peaks.

From an experimental point of view, infrared and Raman spectroscopy are the preferred techniques to approach molecular vibrations. These spectroscopies have benefited in the last decades of great progresses with the development of multi-dimensional techniques that allow more precise and comprehensive data, e.g. the possibility to distinguish homogeneous and inhomogeneous broadening.

Nevertheless, the experimental techniques are not the only possible approaches to this topic. Quantum chemistry (“*ab initio*”) calculations nowadays are able to optimise the geometrical structure of molecules and clusters, and to calculate reliable vibrational density of states (along with infrared and Raman spectra): thus, straightforward correlations between structural and spectroscopic features can be analysed.

Classical molecular dynamics allows obtaining trajectories whose sta-

tistical relevance can elucidate the stability of different conformations. Furthermore, structural changes can be monitored to predict or understand the working of active sites and other inter-molecular phenomena.

Ab initio molecular dynamics, on the other hand, takes the best of both worlds: it allows obtaining both structural and dynamic properties, and also spectroscopic features, via the linear response theory.³ Within this computational framework, the electronic (time independent) problem is solved and quantum forces are used to propagate the dynamics of classical nuclei.

Unfortunately, this scheme also requires huge resources, as both computational power and memory space, so practical calculations are limited in the size of the molecular systems. If classical molecular dynamics simulations of tens of thousands atoms for nanoseconds are nowadays routine, *ab initio* molecular dynamics simulations are actually limited to hundreds or (at most!) a thousands atoms for (at most!) hundreds picoseconds. These limitations basically rule out this approach as a mean to study the slow dynamics of large bio-molecular systems. Still, the possibility to study simultaneously the dynamics of the systems, the structural changes, and the vibrational spectra, allows accurate and reliable observations, even if just on medium sized systems for tens picoseconds. Moreover, due to the continuous increase in both computing power and memory space, these limitations are doomed to be ever less relevant in the future. So, the *ab initio* molecular dynamics framework can be viewed as “proving grounds” for development of new approaches that (hopefully) could be soon adopted to perform reliable “*in silico* experiments” even on large systems. Besides, it also allows to find and elucidate correlations and subtle effects that are important on their own, even in not-so-large molecular systems.

My doctoral dissertation is devoted to study both vibrational spectra and structural properties adopting molecular dynamics simulations. The main link between these two aspects of the molecular behaviour, at

³Linear response theory could also be employed with classical molecular dynamics trajectories, but it would give very little meaningful information due to the use of classical force fields. Polarisable force fields can be developed in order to obtain accurate results, but they cannot be transferred from one molecular system to another: basically, a polarisable force field should be developed for every molecular system under study.

least for the systems simulated for the present dissertation, is hydrogen-bond (H-bond). The rationale behind the choice of simulating H-bonded systems, is that H-bonding is well reproduced by present *ab initio* approaches, and it accounts for most of the induced frequency shifts onto localised vibrational modes. This is well exemplified by fig. 2, taken from Ref. [CLS04], which shows how a vibrational frequency changes due to the variation of a structural parameter (the H-bond distance). This two-

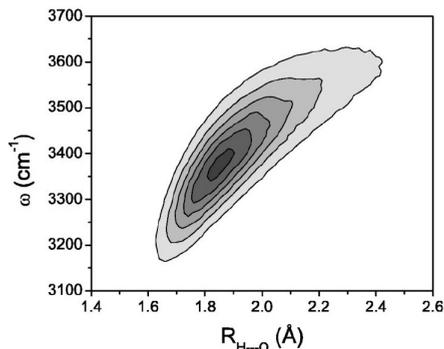


Figure 2: Changing of the OH stretching frequency of HOD molecules, as a function of the inter-molecular H \cdots O distance. Distances are taken from a (classical) molecular dynamics simulation, while vibrational frequencies are subsequently obtained with a separate *ab initio* calculation. Figure is taken from Ref. [CLS04].

dimensional contour plot has been extracted from a classical molecular dynamics trajectory employing a heuristic model; in this dissertation analogous contour plots shall be presented obtained at higher levels of theory with much more refined structure-frequency correlation tools.

The relevance of this work, focused on H-bonding, can be better understood remembering the words of Prof. Mark Ratner in a 2004 seminar [Bec14]:

chemistry of the XX century was about intra-molecular interactions; chemistry of the XXI century will be about inter-molecular interactions.

Overview of the Dissertation

Chapter 1 aims to give a quick and comprehensive view of the adopted computational techniques. The price to pay for said simplicity is the complete absence of details, references, and all those things that hinder a “laid

back” reading of the text. Details about these subjects are given in the appendixes and in chapters 2, 3, and 4.

A brief description of the classical and quantum frameworks employed in this dissertation is given in Chapter 2. In Chapter 3 the computational procedures to calculate structural and time-dependent properties are illustrated, making use of the tools introduced in the previous Chapter. This dissertation is particularly focused onto the correlation between structural and vibrational features: so, to find and elucidate new correlations, a computational approach based on wavelet transform has been developed, allowing accurate time-frequency and structure-frequency analysis, as shown in Chapter 4.

Part II of this dissertation is devoted to illustrate the various case studies where these computational means have been used to extract new information. All the systems studied share the H-bonding feature. Links between theory, computation and experimental results are given in Part II specifically for each case study.

In Chapter 5, a classical molecular dynamics trajectory of a substituted dipeptide in a phospholipide bilayer has been analysed, and the H-bond dynamics has been directly correlated with the measurements of two dimensional IR spectroscopy. Relevant information about dipeptide penetration into the bilayer have been extracted from the simulated dataset.

In Chapter 6, I present the results obtained from many *ab initio* molecular dynamics simulations, all them regarding a glycol solute in a polar solvent (water or acetonitrile). Different spectroscopic responses arise comparing simulations of different solutes, whereas in acetonitrile an unusual structural conformation occurs. Moreover, I have carried out additional *ab initio* (static) calculations to correctly reproduce an unusual frequency shift in the vibrational spectrum of acetonitrile.

In Chapter 7, I report the results of the wavelet analysis of two *ab initio* molecular dynamics trajectories. This approach proved useful into retrieving a peculiar splitting of a localised mode of methyl acetate (*i.e.* C=O stretching) due to H-bond; moreover, using wavelets I have been able to correctly assign the two components of the doublet to their own corresponding structural situation.

Chapter 8 reports the results of the application of wavelet analysis to probe the response of the water environment to a polar heterocyclic solute as emerging from an *ab initio* molecular dynamics simulation. The water response to H-bond acceptor has been investigated also in the crystal of

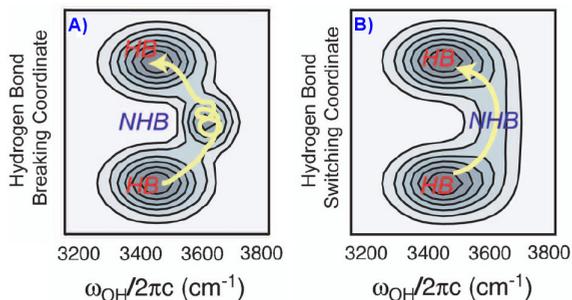


Figure 3: OH stretching frequency distribution spread along the fictitious “H-bond switching/breaking coordinate”. Figure is taken from Ref. [LRT06b].

lithium nitrate trihydrate, as reported in Chapter 9. The simulated crystal is characterised by the presence of an unusual “bifurcated” H-bond, which is still subject of much speculation in the literature. In fact, Loparo *et al.* proposed two competitive models to describe the mechanism of H-bond breaking and formation [LRT06b]: in a first model (fig.3, panel A), this rearrangement of H-bonds goes through a reaction intermediate, whereas in the second one the reaction proceeds through a transition state (fig.3, panel B). Loparo *et al.* in figure 3 schematised these two possible mechanisms spreading the infrared activity along a fictitious “H-bond switching/breaking coordinate”. In Chapter 9 I presents results that rule out one of these proposed models and, in turn, corroborate the other.

Finally, a comprehensive general discussion about the findings of this dissertation is given in Part III, along with some concluding remarks.

List of Some Recurring Abbreviations (sorted alphabetically)

ACN	Acetonitrile
BSSE	Basis Set Superposition Error
CPMD	Car-Parrinello Molecular Dynamics
DFT	Density Functional Theory
EG	Ethane-1,2-diol (a.k.a. ethylene glycol)
FT / FFT	Fourier Transform / Fast Fourier Transform
H-bond	Hydrogen-bond
HF	Hartree-Fock
MA	Methyl acetate
MD	(generic) Molecular Dynamics
MG	Morlet-Gabor (wavelet)
MLWF	Maximally Localised Wannier Function (centre)
MrG	N-myristoylated-methyl-glycine
PBC	Periodic Boundary Conditions
PDO	Propane-1,3-diol
PLPC	1-palmitoyl-2-linoleyl phosphatidylcholine
REM	Replica Exchange Methods
SCF	Self Consistent Fields
SDF	Spatial distribution function
VDoS	Vibrational Density of States
WFT	Wave-function Theory
WT	Wavelet Transform

I
Theory and Models

1. Computational “Tools” in a Nutshell

This very short Chapter is a quick presentation of the computational methods employed in this dissertation. References, and a much deeper and verbose discussion, can be found in the following chapters of this 1st Part of the dissertation, as well as in the appendixes. As I said, this Chapter just aims to give an idea of what these computational techniques do and what could be obtained by them. Here I am going to tell just the absolute minimum of information required to approach the research case studies presented in the 2nd Part.

Ab Initio (Static) Calculations

Ab initio (static) calculation basic goal is to solve the (time-independent) Schrödinger equation (eqn. 2.2) for molecules and, from the obtained wavefunctions, calculate some molecular properties. The procedure is iterative and, depending on the specific level of theory adopted and the size of the investigated system, calculations can be more or less accurate and took minutes or weeks of computer time.

(Classical) Molecular Dynamics

Molecular Dynamics can move molecules according to (more or less) realistic forces. Basically, the computer simulates the molecular behaviour: translations, rotations, and vibrations are the main motions performed by a molecule.

In classical molecular dynamics, forces are calculated by adopting a semi-empirical force field (see Section 2.2.5). This allows simulating a huge number of atoms for long timescales; anyway, reliability of the force field is not something to be assumed and should not be taken for granted.

Moreover, being based on crude harmonic approximations, it gives little information about molecular vibrations.

Since the frontier research in classical molecular dynamics is simulating even larger systems, a number of techniques have been developed to speed up the conformational space sampling, like for example the “replica exchange” method (see Appendix B.2). These techniques, however, have some side-effects that hinder some time-dependent analysis of the simulated trajectories, including analysis used to study molecular vibrations and spectra.

***Ab Initio* Molecular Dynamics**

Ab initio molecular dynamics moves the molecules like classical molecular dynamics, but it does not employ a force-field. So, what moves molecules? In *ab initio* molecular dynamics the forces are obtained solving “on the fly” the molecular Schrödinger equation for electrons and applying them to the nuclei (treated as classical particles). To achieve this goal, some mathematical and computational tricks have to be adopted (*i.e.* Car-Parrinello method, see Section 2.3.2), which have the side-effect to render impractical the study of systems containing more than few hundreds of atoms (up to almost a thousand on very powerful computer workstations).

Nevertheless, freeing the simulation from the force field crude approximations, a lot of molecular features can be studied, including vibrations.

Structural and Electronic Properties

Typical calculations involving (static) *ab initio* calculations are the search for the equilibrium geometry of molecules, the electronic structure (*i.e.* the charge distribution). These features can be obtained also by *ab initio* molecular dynamics, but at an higher computational cost.

With (static) *ab initio* calculations vibrational spectra are usually calculated at the harmonic approximation level, and often neglecting the effects due to the environment. Since breaking and formation of H-bonds (when present) modulate the vibrational frequencies, this is a problematic approximation.

From a molecular dynamics trajectory it is possible to extract information about the structure of a much larger sample of molecules. The pair radial distribution function (see Section 3.1.1) gives the so-called structure of a liquid and can be used in combination with angular distribution functions to obtain a picture of the relative arrangement of molecules.

Time-Frequency Analysis

While vibrational spectra can be obtained with Fourier analysis of autocorrelation functions, wavelet transforms (see Section 4.2) allow obtaining time-dependent spectroscopic information. These can be, in turn, correlated with time-dependent structural information to have a comprehensive picture of how the environment modulates the vibrational frequencies. In particular, this type of analysis has been adopted for the strongest of the (weak) inter-molecular forces, the H-bonding.

Some Hints for the Interpretation

A brief discussion about ideas and concepts used to elucidate vibrational spectra with Fourier and wavelet transforms of molecular dynamics trajectories can be found in Section 3.3. Basically, there it is explained how to recover spectra from simulated trajectories, and also some considerations about the predictive power of the harmonic oscillator approximation are reported. In fact, harmonic reasoning still proves useful also to have an immediate interpretation of anharmonic effects, since these latter can often be *understood* as due to a *change* in the value of the harmonic force *constant*.

2. Computational Chemistry

In this Chapter the main principles of density functional theory and molecular dynamics are presented.

The former is a quantum chemistry theoretical framework (discussed in Section 2.1) that allows reliable electronic structure calculations and many other molecular properties, with a much lower computational effort than other *ab initio* methods of similar accuracy.

The latter is a computational approach (presented in Section 2.2) that aims to reproduce the time-dependent behaviour of molecules, which is pivotal in elucidating their vibrational properties.

Ab initio molecular dynamics (introduced in Section 2.3) combines both the two previous approaches: it adopts density functional theory to solve the electronic problem and to calculate quantum forces that, in turn, are used to “move” the nuclei (treated as classical particles).

2.1 *Ab Initio* Calculations

2.1.1 Molecular Quantum Mechanics

The main aim of quantum chemistry is to solve the quantum equations of motion for a many-electrons system, to describe accurately the molecular properties. For a non-relativistic system, the equation to solve is the Schrödinger equation for the molecular state vectors (“kets”) $|\Psi^i\rangle$:

$$i\hbar \frac{\partial}{\partial t} |\Psi^i\rangle = \hat{H} |\Psi^i\rangle \quad ; \quad (2.1)$$

here \hat{H} is the many-body molecular Hamiltonian operator. The formal solution for this equation is $|\Psi^i(t)\rangle = e^{-i\hbar\hat{H}t} |\Psi^i(0)\rangle$. If the Hamiltonian operator is time-independent (as usually assumed in quantum chemistry), then $|\Psi^i\rangle = e^{-i\hbar E_i t} |\psi^i\rangle$ and the time-independent ket $|\psi^i\rangle$ is an eigenvector of the molecular Hamiltonian. This leads to the following time-

independent Schrödinger equation for the time-independent ket $|\psi^i\rangle$:

$$\hat{H}|\psi^i\rangle = E_i|\psi^i\rangle \quad , \quad (2.2)$$

where E_i is the eigenvalue of the Hamiltonian and represents the energy of many-body quantum state $|\psi^i\rangle$. Assuming that no energy degeneration is present, there is a one-to-one correspondence between energy levels and ket vectors. Molecular ket vectors can be viewed as tensor products of a nuclear ($|\chi\rangle$) and an electronic ket ($|\phi\rangle$):

$$|\psi\rangle = |\chi\rangle \otimes |\phi\rangle = |\chi\rangle|\phi\rangle \quad .$$

Since nuclear masses are much greater than electron mass (*i.e.* electrons kinetic energy \gg nuclear kinetic energy), the Born-Oppenheimer approximation is usually adopted, which leads to a partitioning of the molecular Hamiltonian as a sum of two terms:

$$\hat{H}^{\text{total}} = \hat{H}^{\text{nuclear}} + \hat{H}^{\text{electronic}} \quad ,$$

and

$$\hat{H}^{\text{nuclear}}|\chi^r\rangle = E_r^{\text{nuclear}}|\chi^r\rangle \quad (2.3)$$

$$\hat{H}^{\text{electronic}}|\phi^k\rangle = E_k^{\text{electronic}}|\phi^k\rangle \quad , \quad (2.4)$$

with $E^{\text{total}} = E^{\text{nuclear}} + E^{\text{electronic}}$. Within this approximation, the molecular state space can be viewed as a direct sum of nuclei-only and electrons-only vector spaces

$$\mathcal{V}_{\text{total}} = \mathcal{V}_{\text{nuclei}} \oplus \mathcal{V}_{\text{electrons}} \quad .$$

The rationale behind this approximation is that electrons and nuclei, being so much different in mass, move at very different time-scales and so their equations of motion can be solved separately, as if electrons react immediately to changes of the relative position of slow moving nuclei. This could also be explained saying that, within Born-Oppenheimer approximation, electronic states depend only parametrically from the geometry of nuclei. So, nuclear coordinates $\{\mathbf{R}_N\}$ are merely considered parameters when the electronic problem is solved. This means that, when the nuclear geometry changes, then the analytical form of the electronic wave-functions $\phi^i(\mathbf{r}_j, s) = \langle \mathbf{r}_j, s_z | \phi^i \rangle$ changes too. Within Born-Oppenheimer approximation, the electronic Hamiltonian is

$$\hat{H}^e = \hat{T}^e(\{\mathbf{r}_i\}) + \hat{V}^{Ne}(\{\mathbf{R}_{iN}\}) + \hat{V}^{ee}(\{\mathbf{r}_{ij}\}) \quad , \quad (2.5)$$

where $\{\mathbf{R}_{iN} = \mathbf{r}_i - \mathbf{R}_N\}$ is the difference between nuclear and electronic coordinates, $\{\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j\}$ is the difference between pair of electronic coordinates, N and e are indexes to address nuclear or electronic variables, respectively, and \hat{T} and \hat{V} denote kinetic and potential energies, respectively. Adopting atomic units, Schrödinger representation and projecting operators and kets onto the coordinate basis (*i.e.* adopting wavefunction representation), eqn. 2.5 becomes

$$\hat{H}^e = -\frac{1}{2} \sum_i^{\text{electrons}} \nabla_i^2 - \sum_i^{\text{electrons}} \sum_N \frac{Z_N}{\mathbf{R}_{iN}} + \sum_i^{\text{electrons}} \sum_{j<i} \frac{1}{\mathbf{r}_{ij}} . \quad (2.6)$$

The presence of electron-electron interaction ($\hat{V}^{ee}(\mathbf{r}_{ij})$) in eqn. 2.6 prevents finding an analytical solution, so an approximated approach has to be used instead. Approximation methods for eqn. 2.6 can be traced back to two main approaches:

1. perturbative methods, where the electron-electron interaction is considered as a perturbation on a single-electron Hamiltonian ($\hat{H}^e - \hat{V}^{ee}(\mathbf{r}_{ij})$); unfortunately, repulsion energy $\hat{V}^{ee}(\mathbf{r}_{ij})$ does not respect this assumption;
2. variational methods, as application of the variational calculus onto some functional of the energy. The variational theorem in quantum mechanics provides the consequence that, if

$$\hat{H}|\phi_0\rangle = E_0|\phi_0\rangle ,$$

then using a “test-state” $|\tilde{\phi}\rangle$ that is not (at least, not necessarily) an eigenstate of \hat{H} , then

$$\frac{\langle \tilde{\phi} | \hat{H} | \tilde{\phi} \rangle}{\langle \tilde{\phi} | \tilde{\phi} \rangle} \geq E_0 , \quad (2.7)$$

with the equality standing true only if $|\tilde{\phi}\rangle$ happens to be a true eigenstate of \hat{H} . Hartree-Fock method (HF) and density functional theory (DFT) represent the most used variational methods in quantum chemistry.

Post Hartree-Fock methods are at an higher “level of theory” compared to DFT methods, but they require greater computational resources, so they are not implemented in the *ab initio* molecular dynamics packages adopted in this dissertation (*i.e.* CPMD [CPM]).

2.1.2 Electron Density

Due to the fact that electrons have an half-integer spin angular momentum (*i.e.* they are fermions), a couple of them cannot occupy the same eigenstate of both energy and component of spin (say \hat{S}_z). Thus, the electrons in a molecule are described by a set of kets $|k\rangle$ (k here is just a generic index comprehensive of both energy and spin state), each of them having occupation probability p_k . It is useful to use a “density operator” $\hat{\rho}$, defined as an external product of a ket and bra vectors, as:

$$\hat{\rho} = \sum_k p_k |\phi_k\rangle \langle \phi^k| \quad . \quad (2.8)$$

Since wavefunctions and their complex conjugates can be defined in terms of bras and kets as

$$\phi(\mathbf{r}) = \langle \mathbf{r} | \phi \rangle \quad (2.9)$$

and

$$\phi(\mathbf{r})^* = \langle \phi | \mathbf{r} \rangle \quad , \quad (2.10)$$

then the density operator can be rewritten as

$$\langle \mathbf{r}'' | \hat{\rho} | \mathbf{r}' \rangle = \langle \mathbf{r}'' | \left(\sum_k p_k |\phi_k\rangle \langle \phi^k| \right) | \mathbf{r}' \rangle = \sum_k p_k \phi^k(\mathbf{r}'') \phi_k^*(\mathbf{r}') = \rho(\mathbf{r}'', \mathbf{r}') \quad . \quad (2.11)$$

Defining with $\langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n | \varphi \rangle$ the many-electron state of a time-independent system with n electrons, with \mathbf{x}_k the whole coordinates of the k^{th} electron (both spatial and spin coordinates), then following eqn. 2.11 the density $\rho(\mathbf{r})$ for the electron #1 can be written as

$$\rho(\mathbf{r}_1) = n \int \dots \int ds_1 dx_2 \dots dx_N |\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)|^2 \quad . \quad (2.12)$$

Still, because all electrons are equal, then it has to stand true also that

$$\int d\mathbf{r} \rho(\mathbf{r}) = n \quad ; \quad (2.13)$$

thus, the electron density of 2.13 is a non-negative function of just the three spatial variables ($\mathbf{r} = (x, y, z)$).

2.1.3 Hohenberg-Kohn Theorem

A molecular system of n electrons whose Hamiltonian is defined by eqn. 2.5, is completely determined (at least from a mathematical perspective) by the potential V^{Ne} (also referred to as “external potential” in [PY89]). We can say that the pair n and V^{Ne} determine all the properties of the eigenfunctions and all the features of the ground state of the system. In place of n and V^{Ne} , the first Hohenberg-Kohn theorem considers the electronic density $\rho(\mathbf{r})$ as the paramount quantity, as the following states:

V^{Ne} is determined, albeit a additive constant such as the inter-nuclear potential V^{NN} (it is considered a parameter in this case), by $\rho(\mathbf{r})$. Since $\rho(\mathbf{r})$ determines n (see eqn. 2.13), then $\rho(\mathbf{r})$ completely determines the electronic ground state and all its properties.

The “*ab absurdo*” demonstration of this relevant statement is reported in Appendix A.1 along with observations on the second Hohenberg-Kohn theorem.

2.1.4 Kohn-Sham Equations

Equation A.6 (see Appendix A.1) implies the existence of a variational principle of the form

$$\delta E[\rho] = 0 \quad , \quad (2.14)$$

which, in turn, has as a consequence eqn. 2.7. Since the density ρ is constrained by the condition $\int \rho(\mathbf{r})d\mathbf{r} - n = 0$ (eqn. 2.13), the variational principle takes the form

$$\delta \left\{ E[\rho] - \mu \left[\int \rho(\mathbf{r})d\mathbf{r} - n \right] \right\} = 0 \quad . \quad (2.15)$$

The μ coefficient can easily be obtained noting that

$$\delta \left\{ E[\rho] - \mu \left[\int \rho(\mathbf{r})d\mathbf{r} - n \right] \right\} = \delta E[\rho] - \mu \delta \left[\int \rho(\mathbf{r})d\mathbf{r} - n \right] = 0 \quad .$$

For many-variables functions, the differential can be written as

$$df(q^1, q^2, \dots, q^i, \dots) = \sum_{i=1}^n \frac{\partial f}{\partial q^i} dq^i \quad .$$

Exploiting the analogy between functionals like $F[f(q^1, q^2, \dots, q^i, \dots)]$ and functions such as $f(q^1, q^2, \dots, q^i, \dots, q^n)$, it is possible to write the following equality

$$\delta F = \int \dots \int \frac{\delta F[f]}{\delta f} dq^1 dq^2 \dots dq^i \dots dq^n \quad ;$$

then, inserting $\delta E[\rho]$ into the previous equation we obtain

$$\int \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} \delta \rho(\mathbf{r}) d\mathbf{r} - \mu \int \delta \rho(\mathbf{r}) d\mathbf{r} = \int \left\{ \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} - \mu \right\} \delta \rho(\mathbf{r}) d\mathbf{r} = 0 \quad .$$

Moreover, we also have

$$\frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} - \mu = 0 \implies \mu = \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} \quad ; \quad (2.16)$$

here the definition of the coefficient μ matches that of chemical potential, and explicit calculations leads to

$$\mu = \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} = V^{Ne}(\mathbf{r}) + \frac{\delta F^{HK}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} = V^{Ne}(\mathbf{r}) + \frac{\delta \{T^e[\rho(\mathbf{r})] + V^{ee}[\rho(\mathbf{r})]\}}{\delta \rho(\mathbf{r})} \quad . \quad (2.17)$$

Since in Schrödinger representation the \hat{V}^{Ne} operator is just a power of the electronic coordinates and does not contains mixed terms like $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, we have

$$\langle \varphi | V^{\hat{N}e} | \varphi \rangle = V^{Ne}[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) V^{Ne}(\{\mathbf{R}_N\} - \mathbf{r}) \quad .$$

Anyway, it must be pointed out that the explicit exact form of functionals $T^e[\rho]$ and $V^{ee}[\rho]$ is unknown and cannot be deduced analytically. In fact, $T^e[\rho]$ and $V^{ee}[\rho]$ functionals contain operators that act on the wavefunctions changing these latter up to the point that an analytical expression for the density $\rho(\mathbf{r})$ cannot be obtained. Thus, Kohn and Sham did several assumptions to make the equations more manageable.

1. they introduced in eqn. 2.17 a kinetic energy functional that is given by the average kinetic energy of non-interacting electrons $\bar{T}[\rho]$;
2. they adopted a purely Coulombian functionals in place of $V^{ee}[\rho]$ obtained by classical electrodynamics (the so called "Hartree poten-

tial")¹ that is basically the bi-electronic Coulomb integral of the HF method:

$$J_{ij} = \iint \phi^i(\mathbf{x}_1) \phi_i^*(\mathbf{x}_1) \frac{1}{r_{12}} \phi_j^*(\mathbf{x}_2) \phi^j(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 = \\ = V^{Hartree} = \iint \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 \quad ;$$

3. they put all corrections to the previous simplifying assumptions into a new functional $E_{xc}[\rho]$, called "Exchange-Correlation Energy".

With this approach, it is correct to write

$$E[\rho] = \bar{T}[\rho] + \int d\mathbf{r} \rho(\mathbf{r}) V^{Ne} + \underbrace{\frac{1}{2} \iint \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'}_{\text{Hartree potential}} + E_{xc}[\rho] \quad . \quad (2.18)$$

Putting this expression for $E[\rho]$ into eqn. 2.17 we have

$$\mu = \frac{\delta E[\rho]}{\delta \rho(\mathbf{r})} = \frac{\delta \bar{T}[\rho]}{\delta \rho(\mathbf{r})} + V^{Ne}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}[\rho]}{\delta \rho(\mathbf{r})} \quad .$$

It is useful to define two new potentials, V_{xc} and V^{eff} , usually called "Exchange-Correlation Potential" and "Effective Potential" respectively, which depend from \mathbf{r} and $\rho(\mathbf{r})$ and are defined as:

$$V_{xc} = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \quad (2.19)$$

and

$$V^{\text{eff}} = V^{Ne}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{xc} \quad . \quad (2.20)$$

¹Adopting a plane wave basis, it is better using the Laplace equation

$$\nabla^2 V^{Hartree} = -\frac{1}{\epsilon} \rho(\mathbf{r}) \quad ,$$

and then projecting it into the Fourier space

$$V^{Hartree}(\mathbf{k}) = \frac{4\pi}{\mathbf{k}^2} \mathcal{F}[\rho] \quad ;$$

here with $\mathcal{F}[\rho]$ we refer to the Fourier transform of $\hat{\rho}(\mathbf{k}) \propto \int d\mathbf{r} \rho(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}}$ of the density $\rho(\mathbf{r})$. Calculation of this Fourier transform is quite easy in plane wave basis, as adopted in the CPMD [CPM] program.

Still, a “recipe” to calculate the electronic density still has to be presented. To overcome this issue, Kohn and Sham assumed [KS65] that the density ρ could be calculated from single-electron wavefunctions ϕ^i of the following eigenvalue equation (which is obtained in Appendix A.2)

$$\hat{h}_{(i)}^{\text{eff}}\phi^i = \hat{h}_{(i)}^{\text{KS}}\phi^i = \left[-\frac{1}{2}\nabla_{(i)}^2 + V^{\text{eff}} \right] \phi^i = \epsilon_i\phi^i \quad . \quad (2.21)$$

Here the total density is calculated from the many-electrons wavefunctions φ , which, in turn, are constructed as a Slater determinant of the single-electron wavefunctions:

$$\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{1}{\sqrt{n!}} \det \left[\phi^1\phi^2 \dots \phi^n \right] \quad (2.22)$$

$$\rho(\mathbf{r}) = |\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)|^2 = \sum_j |\phi^j(\mathbf{x})|^2 \quad . \quad (2.23)$$

The last equality is due to the fact that the square modulus of the Slater determinant of eqn. 2.22 $\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ is a sum of single-electron orbitals. Thus it is obtained

$$\sum_i \epsilon_i = \sum_i \langle \phi_i | \left[-\frac{1}{2}\nabla_{(i)}^2 + V^{\text{eff}} \right] | \phi_i \rangle = \bar{T}[\rho] + \int V^{\text{eff}}(\mathbf{r})\rho(\mathbf{r})\mathbf{d}\mathbf{r} \quad . \quad (2.24)$$

Remembering equations 2.20 and 2.18, the total electronic energy can finally be expressed as

$$E = E[\rho] = \sum_i \epsilon_i - \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathbf{d}\mathbf{r}\mathbf{d}\mathbf{r}' - \int V_{xc}(\mathbf{r})\rho(\mathbf{r})\mathbf{d}\mathbf{r} + E_{xc} \quad . \quad (2.25)$$

2.1.5 Basis Sets

Talking about basis sets in quantum chemistry eventually leads to talking about Gaussians. Why Gaussian functions are so widely employed in electronic structure calculations? Basically for three very important reasons:

1. the overlap of two Gaussians is a Gaussian;
2. the Fourier transform of a Gaussian is a Gaussian (this property shall be invoked again in section 4.2.1);

3. with the “fast Fourier transform” algorithm, performing Fourier transforms consumes very little computational resources.

Before computer-era, most quantum chemistry calculations were done employing Slater-type orbitals (STOs), which are basically hydrogen-like eigenfunctions with an effective nuclear charge.

For the three reasons presented before, today it is normal practice to simulate a Slater-type orbital with a linear combination of Gaussians. For example, with so called “STO- n G basis”, the n value represents the number of Gaussian primitive functions comprising a single Slater-type orbital.

In this dissertation, the most widely employed basis for static *ab initio* calculation is the 6-311++G(d,p) basis set (whereas, as discussed in Section 2.3.4, the native basis set for CPMD are the plane waves). This confusing abbreviation means that 6 Gaussians are used for every core electron Slater-type orbital, whereas the three digit number “311” means that the valence electrons are described by a combination of three orbitals, the first simulated by 3 Gaussians and the other two by a single Gaussian each. Moreover, polarisation orbitals are added as virtual $|d\rangle$ and $|p\rangle$ orbitals, as well as diffuse functions (denoted by the “++” symbol). The 6-311++G(d,p) is a rather accurate basis set currently employed in quantum chemistry calculations.

2.1.6 Hartree-Fock VS Kohn-Sham

The Kohn-Sham approach can be schematised as follows:

1. single-electron orbitals ϕ^i are guessed;
2. eqn. 2.21 is tentatively solved;
3. eigenvalues ϵ_i and density $\rho = \sum_i |\phi^i|^2$ are calculated;
4. with ϵ_i and $\rho = \sum_i |\phi^i|^2$, the functional E is calculated adopting eqn. 2.25;
5. E is minimised changing some parameters of the original orbitals ϕ^i ;
6. return to step #2 and continue until the ϵ_i eigenvalues of tentatives C and $C + 1$ are as close as desired.

This is basically the standard “recipe” of any self consistent field (SCF) approach. DFT has both pros and cons compared to the HF procedure.

- In HF theory, a single Slater determinant wavefunction is build. Then, at the post Hartree-Fock level ², correction to it (to recover the so called “correlation energy”) are introduced considering excited states and, overall, introducing additional Slater determinants to the many-electrons wavefunction. Hartree-Fock as well as post Hartree-Fock methods are nowadays collectively called “wavefunction method/theory” (WFT) [Bec14].

The rationale behind the Kohn-Sham approach is different: calculations are performed onto reference system of electrons that interact with each other only through the V_{xc} term in eqn. 2.20 (it is often said that they are “non-interacting electrons”, meaning that no term in the Hamiltonian of eqn. 2.21 couples their coordinates). This reference system is represented only by a single Slater determinant.

Hence, it can be proved (see ref. [Bec14] for more references) that if V_{xc} is exact, then the density of the reference systems is equal to the true many-electrons density. This is the main connection between WFT and DFT: the theories have different hamiltonians, but should yield the same densities.

Basically, if V_{xc} is exact DFT yields true ground state properties without expanding the many-electrons wavefunction into a space composed by many Slater determinants: just one is enough.

- From a theoretical point of view, the DFT *actual implementation* seldom are truly “*ab initio*”. Even if Kohn-Sham equations (eqn. 2.21) are exact, the adopted expressions for V_{xc} in chemical relevant calculations usually contain terms and parameters that are empirically derived and not theoretically justified. On the contrary, within HF theory, only mathematical approximations are allowed;
- DFT equations are much more easily solved than post Hartree-Fock ones: if n is the number of electrons in the system, then DFT calculations can be proportional up to n^3 , whereas a corresponding post-HF calculation can be proportional up to n^6 . So, doubling the number

²Post Hartree-Fock calculations basically start with a simple HF calculations and then introduce corrections with some perturbative method (e.g. “Many Body Møller-Plesset Perturbation Theory”) or with Configuration Interaction (e.g. CI or Coupled Cluster methods.).

of electrons a DFT calculation becomes 8 times slower, but a post-HF one becomes 64 times slower!

- DFT does not suffer from lack of correlation energy, which, conversely, is absent into HF method. To recover correlation energy within a HF-based scheme one has to introduce further corrections (“post-Hartree-Fock”) that make the calculations much slower;
- a final comment: to derive the HF equations, one starts from an exact many-electrons Born-Oppenheimer Hamiltonian and arrives to an approximate single-electron Fock Hamiltonian. Then one can correct this result with some post-Hartree-Fock approach to recover electron correlation (assuming one has the computational power to do so) and come close at pleasure to the desired result.

The DFT approach is quite the opposite! Kohn-Sham equations are exact, but the exchange-correlation functionals adopted in practice are not.

In HF-based methods one knows (at last, theoretically) how to improve results. To do the same within DFT, one should develop a new (and better working) exchange-correlation functional. It does not surprise that the development of new functionals is still an active research field in quantum chemistry.

Indeed, a wise choice of

$$V_{xc} = \frac{\delta E_{xc}}{\delta \rho(\mathbf{r})}$$

determines the accuracy of the DFT calculation. For a system of free and (almost) non-interacting electrons, V_{xc} depends only on the electron density ρ . Anyway, for real electrons, V_{xc} depends not only on ρ but also on further derivatives of it, so that

$$V_{xc} = V_{xc}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \nabla(\nabla\rho(\mathbf{r})), \dots] \quad .$$

A discussion on functional forms and properties is reported in Appendix A.3.

2.2 Classical Molecular Dynamics

Molecular Dynamics (MD) calculations allow simulating the behaviour of molecules at atomic scale, solving numerically the proper equations of

motion. When these equations are Newton's, the method is called classical Molecular Dynamics. In this Section the principles of classical and general MD are presented.

Since both classical and *ab initio* MD share a lot of features, many aspects of the former still apply to the latter and will not be discussed again in Section 2.3.

2.2.1 Equations of Motion

Classical MD has the aim to solve the Newton equation

$$\mathbf{F}_k = m_k \mathbf{a}_k \quad (2.26)$$

for all N particles of the system (which are represented by the k index). In a classical MD simulation, the atoms (nuclei and electrons) are the particles that move and on which the calculated forces are applied. The forces derive from the so called "force fields" (see Section 2.2.5), which take into accounts both electrostatic and dispersive contributions. The electrostatic forces are due to the fact that, even in (neutral) molecules, local charge can exceed zero and so partial charges are attributed to the atoms (or pseudo-atoms³). Dispersive forces are due to higher multipole interactions and weak polarisation effects. H-bond is not accurately described with a standard classical force field, due to the inability of the latter to correctly reproduce the covalent character of this type of bond. Still, the bulk of the electrostatic H-bonding interaction is retained, enough to roughly recover at least the structure of H-bonded systems. Further details on force fields are given in Section 2.2.5. Calculating energies, forces and displacements represent the main computational burden of a MD simulation. Anyway, before discussing this issue, it is useful to put the Newton equation in a more convenient form. In fact, eqn. 2.26 could be rewritten in Cartesian coordinates as:

$$F_i = m \ddot{x}_i = \frac{dp_i}{dt} \quad , \text{ with } i = 1, \dots, 3N \quad . \quad (2.27)$$

³Pseudo-atoms are basically excesses of charge permanently put on positions around the molecules where there are not true atoms. For example, in the so called TIP4 force field for liquid water [JCM⁺83], an additional negative charged dummy atom is placed near the centre of mass of the molecule (close to the oxygen atom). This procedure is used to parametrise complex electrostatic responses when simply putting charges onto real atoms cannot reproduce known effects.

This form is not invariant with respect to a change of coordinates, which will be relevant in presenting the Car-Parrinello method, so it is better to adopt Lagrangian and Hamiltonian formulations in order to express the equations of motion in generalised coordinates. The Lagrangian \mathcal{L} is a function that depends only on the generalised coordinates $\{q^i\}$ and on the generalised velocities $\{\dot{q}^i\}$ of the system (with $i = 1, \dots, 3N$). The Lagrangian satisfies the following equations

$$\frac{\partial \mathcal{L}}{\partial q^i} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}^i} = 0 \quad , \quad (2.28)$$

which are obviously called ‘‘Lagrange equations’’. They represent a system of $3N$ partial differential equations, one equation for every i coordinate. When only conservative forces are present (*i.e.* systems where $\mathbf{F} = -\nabla V$, with V potential energy of the system depending only on the coordinates), the classical Lagrangian takes the form

$$\mathcal{L} = T(\dot{\mathbf{q}}) - V(\mathbf{q}) \quad ,$$

with T kinetic energy of the system. The momentum is defined as

$$p_i = \frac{\partial \mathcal{L}}{\partial \dot{q}^i} \quad , \quad (2.29)$$

and, inserting eqn. 2.29 into eqn. 2.28, its time-derivative assumes the form

$$\dot{p}_i = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}^i} = \frac{\partial \mathcal{L}}{\partial q^i} \quad . \quad (2.30)$$

It is particularly helpful to express the equations of motion as dependent on a function that coincides with the total energy of the system. Moreover, equations 2.28 constitutes a system of second-order partial differential equations, whereas in many fields it would be simpler to approach first-order partial differential equations. The required function for these equations is the Hamiltonian \mathcal{H} , which is defined as the Legendre-transform of the Lagrangian⁴ :

$$\mathcal{H} = \sum_i^{3N} \frac{\partial \mathcal{L}}{\partial \dot{q}^i} \dot{q}^i - \mathcal{L}(q^i, \dot{q}^i) \quad . \quad (2.31)$$

⁴Formally speaking, the \mathcal{H} function of eqn. 2.31, which is a function of coordinates and velocities, is not the same function of equations 2.32 and 2.33, which depends on coordinates and momenta. The two functions are numerically equal (at least for conservative forces), so it is not really misleading to adopt the same symbol for both of them.

With some manipulation, and exploiting equations 2.29 and 2.30, it is possible to demonstrate the following Hamilton equations

$$\frac{\partial \mathcal{H}}{\partial q^i} = -\dot{p}_i \quad (2.32)$$

and

$$\frac{\partial \mathcal{H}}{\partial p_i} = +\dot{q}^i \quad , \quad (2.33)$$

with

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) = T(\mathbf{p}) + V(\mathbf{q}) = E^{\text{total}} \quad .$$

Equations 2.32 and 2.33 form a system of $6N$ first-order partial differential equations, as required, and the Hamiltonian \mathcal{H} function is numerically equal to the total energy.

2.2.2 Integration Scheme

Hamilton equations 2.32 and 2.33 shall be helpful in the Appendix B.1, where a general approach to derive numerical integrators is presented. In this Section, I present a simplified approach (for example, see [Tuc10a]) to obtain just one class of numerical integrators, the so called “velocity verlet”.

The position \mathbf{q} of a particle at time $t + \Delta t$ can be expressed as a Taylor series of the position at time t . Dropping all terms of the Taylor expansion higher than second order in Δt , we obtain

$$\mathbf{q}(t + \Delta t) \approx \mathbf{q}(t) + \dot{\mathbf{q}}(t) \cdot \Delta t + \frac{1}{2} \ddot{\mathbf{q}}(t) \cdot (\Delta t)^2 \quad .$$

Knowing that $\dot{\mathbf{q}}(t) = \mathbf{v}(t)$ and $\ddot{\mathbf{q}}(t) = \mathbf{F}(t)/m$, the previous equation can be rewritten as

$$\mathbf{q}(t + \Delta t) \approx \mathbf{q}(t) + \mathbf{v}(t) \cdot \Delta t + \frac{1}{2m} \mathbf{F}(t) \cdot (\Delta t)^2 \quad . \quad (2.34)$$

Nothing prevents us from writing an analogous equation to recover $\mathbf{q}(t)$ from data at time $t + \Delta t$, with a time-leap equal to $-\Delta t$, obtaining

$$\mathbf{q}(t) \approx \mathbf{q}(t + \Delta t) + \mathbf{v}(t + \Delta t) \cdot (-\Delta t) + \frac{1}{2m} \mathbf{F}(t + \Delta t) \cdot (\Delta t)^2 \quad . \quad (2.35)$$

Now we can substitute eqn. 2.34 into eqn. 2.35, obtaining

$$0 = -\mathbf{v}(t + \Delta t) \cdot \Delta t + \mathbf{F}(t + \Delta t) \cdot (\Delta t)^2 + \mathbf{v}(t) \cdot \Delta t + \frac{1}{2m} \mathbf{F}(t) \cdot (\Delta t)^2$$

that can be rearranged as

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{1}{2m} [\mathbf{F}(t) + \mathbf{F}(t + \Delta t)] \cdot \Delta t \quad . \quad (2.36)$$

Since the state of a classical system is a point in the phase space, then the pair of eqn. 2.34 and 2.36 can propagate in time the state of the system to produce a dynamics. Here, eqn. 2.34 and 2.36 are expressed in terms of coordinates and velocities, but they could be reformulated in terms of coordinates and momenta to match eqn. B.8 as obtained in Appendix B.1, where a deeper discussion about the advantages of this particular integration approach can be found.

2.2.3 Periodic Boundary Conditions

In a MD simulation, atoms, molecules and ions are put into a box⁵ and their equations of motion are solved as explained in the previous Sections. Since in a simulation the number of particles N cannot be infinite, in order to reproduce the properties of a bulk and extensive environment, periodic boundary conditions (PBC) have to be employed; otherwise, only simulations in vacuum could be performed. With PBC, other cells are constructed around the simulation box, equal to the latter, and equations of motion are solved only for the “true” simulation box, whereas particles in the other cells are just images of the N particles inside the original box. For a cubic cell, this means that when a molecule exits from the true

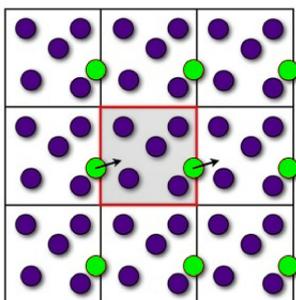


Figure 2.1: The green particle exits through the right side of the red cell, but its “image” enters from the left cell. This is repeated over all cells.

⁵The box does not necessarily has to be cubic.

box through the right side, then its image enters again into the same box through the left side, as showed in fig. 2.1. Adopting PBC has, however, some minor and yet undesired consequences:

- angular momentum is no more conserved. In fact, since the angular momentum \mathbf{L} is the vector product

$$\mathbf{L} = \sum_i^N \mathbf{r}_{(i)} \wedge \mathbf{p}_{(i)} \quad , \quad (2.37)$$

with i being particle-index, it is self evident that when a particle exits and its image re-enters the simulation box, its velocity has changed direction and this, in turn, changes the sign of the vector product;

- spurious correlation between particles and their images are introduced into the simulation. Nevertheless, if the simulation cell is large enough hopefully this effect would have little significance.

2.2.4 Ergodic Hypothesis

Since (\mathbf{q}, \mathbf{p}) is the classical $6N$ component vector of the N particle state, then the mean value of a quantity $\vartheta(\mathbf{q}, \mathbf{p})$ is

$$\langle \vartheta \rangle = \frac{\iint \vartheta(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \, d\mathbf{q} d\mathbf{p}}{\iint f(\mathbf{q}, \mathbf{p}) \, d\mathbf{q} d\mathbf{p}} \quad , \quad (2.38)$$

where $f(\mathbf{q}, \mathbf{p})$ is a function representing the density of states in the phase space. $f(\mathbf{q}, \mathbf{p})$ assumes different forms depending on the specific statistical ensemble. ⁶ Integrating on \mathbf{q} and \mathbf{p} in classical mechanics means taking the average on the states of the systems. Conversely, in quantum mechanics the classical integral has to be replaced by a sum of quantum states.

The so-called “ergodic hypothesis” allows to correlate simulation results with measurable thermodynamic properties (i.e. time independent

⁶For example, in the micro-canonical ensemble NVE , where the degree of freedoms (N), the volume (V) and the total energy of the system (E) are all constants, the ρ function is a Dirac’s delta $\delta(E_{tot} - \mathcal{H}(q^i, p_i))$, which means that states with energy different from the constant value E do not contribute.

ones). Returning to the general quantity ϑ , the ergodic hypothesis let us write its mean value as

$$\langle \vartheta \rangle = \lim_{t_{\max} \rightarrow +\infty} \frac{1}{t_{\max}} \int_0^{t_{\max}} \vartheta(t) dt \quad (2.39)$$

without performing the difficult task of integrating on classical phase-space states, as eqn. 2.38 would prescribe. Since in (equilibrium) MD one usually assumes that the ergodic hypothesis stands true, is helpful to pinpoint some relevant peculiarities of this hypothesis:

- the hypothesis imply that in MD the mean on the states is numerically equal to that on the time-steps, if the simulation is long enough;
- for long enough simulations the system completely explores the available phase-space;⁷
- there is no guarantee that a simulation run is long enough to have reached the ergodic limit.

Anyway, previous experiences and simulations broadly confirm that for fluid systems up to a thousand degrees of freedom (the typical value for an *ab initio* MD simulation) simulation run of tens of picoseconds are usually enough, and for systems up to a 100×10^3 degrees of freedom (indicative value for classical MD simulations) usually runs of tens or hundreds of nanoseconds are required. To speed up the MD achievement of the ergodic limit, many computational strategies were developed to improve the phase space sampling, like the replica exchange method (REM) discussed in Appendix B.2.

2.2.5 Force Fields

The reliability of a classical MD simulation depends crucially on the choice of the force field. Basically, the force field contains all the potentials of the many intra- and inter-molecular interactions, which are usually called “bonded” and “non-bonded” potentials.

Bonded potentials include stretching and bending interactions, both modelled with a quadratic potential of the appropriate coordinate. The

⁷For a *NVE* simulation, this imply that the available phase-space is such that the potential energy $V(\mathbf{q})$ is less (or equal) than the total energy $E(\mathbf{q}, \mathbf{p})$.

stretching interaction changes as $\sim k_s r_{ij}^2$, with r being the distance between two covalent bonded atoms i and j . The bending interaction changes as $\sim k_b \alpha_{ikj}^2$, with α being the angle between two atoms (i and j), both covalently bound to a third (k). Bonded potentials also includes torsional effects that are often modelled as $V_{torsion} \sim K_\phi [1 + \cos n\phi_{iklj}]$, where ϕ_{iklj} represents the torsional angle between the i and j atoms and n is an integer that usually assumes the value 3. The constants k_s , k_b , and K_ϕ have force-field specific values and are usually deduced empirically from spectroscopic and thermodynamic measurements. The total bonded potential is computed as a sum of two-atoms terms.

Non-bonded potentials determine the inter-molecular interactions and usually include electrostatic and Lennard-Jones terms:

$$V_{non-bonded} = \frac{1}{2} \sum_i \sum_{j \neq i} \left[\underbrace{k \frac{q_i q_j}{r_{ij}}}_{\text{electrostatic}} + \overbrace{4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right)}^{\text{Lennard-Jones}} \right]. \quad (2.40)$$

The electrostatic interaction shall be discussed in more detail in the following Section 2.2.5. The Lennard-Jones potential is adopted to model all inter-molecular interactions weaker than the electrostatic one, *i.e.* mainly van der Waals interactions. The σ_i parameter is approximately the van der Waals radius of the atom i , and usually σ_{ij} is calculated as the arithmetic mean

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2},$$

whereas ϵ_{ij} is often computed as a geometric mean

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j},$$

where $-\epsilon_i$ is the depth of the energy well at the minimum of the Lennard-Jones potential (which, in turn, is located at $r_{min} = 2^{1/6}\sigma_i$) for the atom i . The two previous equations are often referred to as "Lorentz-Berthelot mixing rules". Double sum $\frac{1}{2} \sum_i \sum_{j \neq i}$ is adopted to avoid 1-2 and 1-3 pairs into equation 2.40⁸ (*i.e.* to exclude stretching and bending related pairs

⁸The double sum $\frac{1}{2} \sum_i \sum_{j \neq i}$ is equivalent to $\sum_i \sum_{j > i}$.

from the non-bonded potentials); anyway, 1-4 pairs (related by torsional potentials) are somewhat considered at the border-line between bonded and non-bonded pairs. Some force fields, such as AMBER [HAO⁺06], take into account 1-4 pairs into eqn. 2.40, albeit with a dumping factor that reduces their relevance, whereas other force fields like CHARMM [FJ00] exclude these pairs in the sum of eqn. 2.40.

2.2.6 Long Range Interactions

As shown in Section 2.2.3, PBC allow performing bulk simulation avoiding border effects, but on the other hand the use of PBC implies a cut-off of all interactions. This is not a relevant problem for interactions modelled with a Lennard-Jones potential, since they decay quickly with the inter-atomic distance r_{ij} . This is equivalent to say that the series of Lennard-Jones potentials converges within the cutoff radius, but the same is not true for electrostatic interactions, so a cutoff of these could generate energy-conservation issues as well as severely unrealistic forces. To overcome this setback, different approaches were developed, including the Ewald Summation method. In the Ewald scheme, the electrostatic series is partitioned into two new different series (on more than one simulation cell) that are absolutely convergent (see for example Ref. [FS02], Chapter 12). Details of this approach are really tedious and can be found in the Appendix of Ref. [Tuc10a], pp.652-662. However, it should be pointed out that Ewald-based approaches introduce spurious correlations between particles, since the sum is extended to particles of different cells.

2.2.7 Effectiveness and Shortcomings of the Method

Classical MD is a powerful tool to reproduce the behaviour of many molecular systems. In fact, by means of MD calculations it is not only possible to recover the same structural data obtained with other simulation techniques (i.e. Monte-Carlo), but also to obtain time-dependent properties, as will be discussed in Chapter 3. This dissertation is mainly devoted to calculate vibrational properties, so the Monte-Carlo approach could not be employed.

It has to be pinpointed that the trajectories generated by classical MD are not necessarily close to the “real” molecular trajectories: force fields contains semi-empirical parameters and are mostly based on very crude approximations (*e.g.* harmonic approximation for bending and stretch-

ing potentials) that produce significant divergences. So, how it comes that classical MD is widely employed? Basically, what is most relevant in molecular simulations is to acquire not the “real” behaviour of molecules, but a “realistic” one that is sufficiently reliable to elucidate empirical datasets.

Serious shortcomings of classical MD simulations (particularly in relation to the purposes of the present dissertation) are listed in the followings.

- They cannot reproduce chemical reactions, which, on the contrary, can be obtained in *ab initio* MD trajectories; this is due to the fact that the force field adopted does not allow the alteration of molecular connectivity.
- Moreover, since the force fields determines the interactions once and for all, polarisation effects are largely lost, so anharmonic effects in vibrational spectra cannot be reproduced.
- As a consequence of the previous observations, even the effects of H-bonding (which retain a covalent character) cannot be correctly described.

It is true that with classical MD many observations regarding the statistical behaviour of the H-bond can still be made, but the effects of this bonding onto the vibrational spectra are mostly lost. It is possible to develop “polarisable force fields”, where the parameters are allowed to change according to the dynamic behaviour of molecules, but these force fields are optimised for the specific system to simulate and cannot be easily transferred from a system to another; besides, they are usually hard and time-consuming to be developed.

For these reasons, this dissertation mainly employs *ab initio* MD, which (in theory) provides the best of both static *ab initio* calculations and MD simulations. In fact, *ab initio* MD takes into accounts realistic electronic structure and is therefore able to retain polarisation effects, overcoming the three issues just made.

The price to pay to use *ab initio* MD is its larger computational cost. This is mainly due to the increase in the degrees of freedom of the system (as shall be shown in Section 2.3), which forces to simulate only a limited number of atoms and for relatively short time (up to a thousand of atoms and few hundreds of picoseconds at best, respectively).

So, *ab initio* MD is at a higher “level of theory” than classical MD, but the complexity of the systems that can be realistically simulated decreases. Indeed, classical MD is still the suitable computational tool to investigate large molecular aggregates, such as proteins and other large biomolecules, whereas *ab initio* MD is mainly employed to study in detail systems made up of hundreds of atoms.

2.3 *Ab Initio* Molecular Dynamics

This Section wants to show how DFT and classical MD combine to create a dynamics of nuclei that move according to realistic quantum forces.⁹ There are various *ab initio* MD approaches: in the following, I shall briefly present the Born-Oppenheimer MD, which provides much of the theoretical background of the whole subject, and then the Car-Parrinello MD, which is the method more extensively employed in this dissertation.

2.3.1 Born-Oppenheimer Molecular Dynamics

The core concept of *ab initio* MD is to treat nuclei like classical particles (just like in classical MD) but subject to forces obtained by *ab initio* calculations on electrons. Born-Oppenheimer MD is a conceptually simple approach to *ab initio* MD: at every time-step the *ab initio* equations for the electrons are solved (*e.g.* the Kohn-Sham equations of DFT) obtaining the eigenfunctions φ_i that minimise the electronic energy and calculating the Slater determinant φ ; then, semi-classical forces are obtained via the Hellman-Feynmann theorem adopting equation C.2,¹⁰ here reported for completeness,

$$\mathbf{F}_k = \dot{\mathbf{p}}_k = -\nabla_{\mathbf{R}_k} \langle \varphi | \hat{H}^e | \varphi \rangle \quad (2.41)$$

where $\nabla_{\mathbf{R}_N}$ represents the gradient with respect to the coordinates of nucleus k . Minimising the total electronic energy (at every MD time-step) means that the electronic system has to be onto the Born-Oppenheimer energy surface, which, in turn, depends parametrically onto the nuclear coordinates $\{\mathbf{R}_N\}$ and where the electronic ground states for each specific

⁹Obviously, with the term “quantum forces” I refer to semi-classical forces, analogue to those obtained via Hellman-Feynmann theorem. For an accurate discussion of this theorem, and its relation with *ab initio* MD, see Ref. [MH09], pp. 52-56.

¹⁰A short presentation of this theorem is reported in Appendix C.1.

nuclear configuration lie. An approach similar to this is adopted, for example, in the CP2K [VKM⁺05,CP2] *ab initio* MD code, which adopts a mix of plane waves and Gaussian functions to expand the electronic wavefunctions. Programs like CP2K have not been employed in this dissertation to perform *ab initio* MD, because they often show a sub-optimal energy conservation, with Gaussians associated to some issues (e.g. BSSE problems), as discussed in Appendix C.2.

In the eighties, M. Parrinello and R. Car developed an alternative approach that is computationally more efficient since, as explained soon, it requires just one wavefunction optimisation.

2.3.2 Car-Parrinello Molecular Dynamics

Parrinello and Car, in order to avoid to repeat the wavefunctions optimisation at every MD time-step, postulated in [CP85] the following Lagrangian \mathcal{L}^{CP} for a system of quantum electrons and classical nuclei:

$$\mathcal{L}^{CP} = \underbrace{\frac{1}{2} \sum_N M_N (\dot{R}_N)^2 + \frac{1}{2} \sum_i^{occ} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle}_{\text{kinetic energy}} - \overbrace{E[\rho; R_N]}^{\text{potential energy}} + \text{constraints} \quad . \quad (2.42)$$

In equation 2.42:

R_N are the nuclear coordinates;

M_N are the nuclear masses;

ϕ^i is the Kohn-Sham wavefunction for the i^{th} electron;

ρ is the total electronic density built from single-electron wavefunctions;

μ is a parameter that is called “electron fictitious mass”.

The “constraints” in eqn. 2.42 mean to assure normalisation and orthogonality of the wavefunctions and take the form

$$\sum_{ij} [\lambda_i^j \langle \phi_i | \phi^j \rangle - \delta_{ij}] \quad ,$$

where

λ_i^j is a Lagrange multiplier

$\langle \phi_i | \phi^j \rangle - \delta_{ij}$ is the ortho-normality constraint for single electron wavefunctions;

E is the total electronic energy $E[\rho; R_N]$ obtained through eqn. 2.25 (which depends parametrically on nuclear coordinates) and the total nuclear repulsion energy.

In eqn.2.42 the electronic wavefunctions take the role of system coordinates since their first derivatives $\dot{\phi}^i$ are part of the kinetic energy term of the Lagrangian.¹¹ Indeed, the term

$$\frac{1}{2} \sum_i^{occ} \mu_i \langle \dot{\phi}_i | \dot{\phi}^i \rangle = \frac{1}{2} \sum_i^{occ} \mu_i \int d\mathbf{r} \frac{d\phi_i^*}{dt} \frac{d\phi^i}{dt} = \frac{1}{2} \sum_i^{occ} \mu_i \int d\mathbf{r} |\dot{\phi}^i|^2$$

formally resembles the classic kinetic energy $T = \frac{1}{2} \sum_i m_i \dot{x}_i^2$ and hence is called “fictitious kinetic energy” of electrons. Adopting the Car-Parrinello Lagrangian, the corresponding equation of motion (eqn. 2.28) for the electronic wavefunctions becomes

$$\frac{d}{dt} \frac{\partial}{\partial \dot{\phi}_i^*} \mathcal{L}^{CP} = \frac{\partial}{\partial \phi_i^*} \mathcal{L}^{CP} \quad , \quad (2.43)$$

whereas for the nuclei

$$\frac{d}{dt} \frac{\partial}{\partial \dot{R}_N} \mathcal{L}^{CP} = \frac{\partial}{\partial R_N} \mathcal{L}^{CP} \quad . \quad (2.44)$$

Performing the calculations explicitly for the wavefunctions equation, we obtain

$$\frac{d}{dt} \mu_i \dot{\phi}^i(\mathbf{r}, t) = \mu_i \ddot{\phi}^i(\mathbf{r}, t) = - \frac{\partial E[\rho; R_N]}{\partial \phi_i^*} + \sum_j \lambda_j^i \phi^j(\mathbf{r}, t) \quad , \quad (2.45)$$

and, employing eqn. 2.25 and 2.24,

$$\frac{\partial E[\rho; R^N]}{\partial \phi_k^*} = \hat{H}^{KS} \phi^k(\mathbf{r}) \quad ,$$

thus, eqn. 2.45 reduces to the simple form

$$\mu_i \ddot{\phi}^i(\mathbf{r}, t) = - \hat{H}^{KS} \phi^i(\mathbf{r}, t) + \sum_j \lambda_j^i \phi^j(\mathbf{r}, t) \quad . \quad (2.46)$$

¹¹Sum \sum_i^{occ} is limited to occupied Kohn-Sham orbitals only.

For the nuclear coordinates we obtain

$$M_N \ddot{R}_N = - \frac{\partial E[\rho; R_N]}{\partial R_N} . \quad (2.47)$$

Equations 2.46 and 2.47 describe the time evolution of nuclear positions and electronic wavefunctions in Car-Parrinello *ab initio* molecular dynamics (CPMD) [CPM,CP85].

2.3.3 Explanations

As can be seen, CPMD approach is very different from Born-Oppenheimer MD and it is not immediately apparent why it should produce comparable results. This Section provides explanations of the method and shows some basic features of the CPMD procedure.

In Born-Oppenheimer MD the electronic quantum problem is solved whenever the nuclei are moved (*i.e.* at every MD time-step): the wavefunctions ϕ^i evolve in time only thanks to the SCF calculations that quench them again onto the Born-Oppenheimer energy surface. On the contrary, within the CPMD approach the electronic wavefunctions are time-dependent variables ¹²: they are calculated (and quenched onto the Born-Oppenheimer surface, *i.e.* optimised) just once before the dynamics starts, and then they evolve in time like classical coordinates according to eqn.

¹²It is necessary to specify that in the CPMD scheme the time-independent Schrödinger equation 2.2 is not solved at every time-step, nor the time dependent Schrödinger equation 2.1, but rather the algorithm simply propagates the wavefunction in time according to equation 2.46. To really solve eqn. 2.1, an approach like “time dependent density functional theory” (TD-DFT) should be employed, which is based on the following equation

$$\frac{\delta}{\delta \rho(\mathbf{r}, t)} \int_{t_1}^{t_2} \langle \Psi(t) | \left[i\hbar \frac{d}{dt} - \hat{H}^{\text{eff}} \right] | \Psi(t) \rangle dt = 0$$

so that

$$i\hbar \frac{d}{dt} \psi^i(t) = \hat{H}^{\text{eff}} \psi^i(t) = \left[-\frac{1}{2} \nabla^2 + V^{\text{ext}}[\rho(\mathbf{r}, t)] + \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\delta E_{xc}(\mathbf{r}, t)}{\delta \rho(\mathbf{r}, t)} \right] \psi^i(t)$$

where V^{ext} is a new time-dependent functional. TD-DFT still poses some theoretical (*e.g.* time evolution depends on both the density current $\mathbf{j} = -i\hbar[\Psi^*(t)\nabla\Psi(t) - (\nabla\Psi^*(t)\Psi(t)]/2m$ and the scalar density ρ) and practical challenges (can be very time-consuming, and the analytical expression of reliable time-dependent functionals is difficult to obtain). A brief discussion of this approach is reported in Ref. [Mar04], pp. 147-149.

2.46. In fact, eqn. 2.46 is integrated with a velocity verlet algorithm¹³ (see equations 2.36 and 2.34). Thus, the price to avoid optimisation at each time-step is to increase the number of degrees of freedom, as briefly anticipated in Section 2.2.7.

In order to assess the reliability of this approach, CPMD generated trajectories and other *ab initio* MD trajectories should be close to each other: in other words, the single electron wavefunctions ϕ^i have to evolve in time, remaining near the Born-Oppenheimer energy surface. Within the CPMD approach, this latter requirement is not always assured, but under some suitable conditions this requirement can be made true in practice.

The Lagrangian reported in eqn. 2.42 shows an “unusual” kinetic energy term,

$$\frac{1}{2} \sum_i^{occ} \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle ,$$

containing the fictitious mass of the electrons (μ parameter) and the “velocities” $\dot{\phi}^i$. The μ parameter has physical dimension of mass \times length and its actual value is (in theory) arbitrary, but it has profound effects on the dynamics. With a “large” value for μ , the dynamics of the wavefunctions will be particularly slow, which means that the wavefunctions will adapt slowly to the changes occurring in the nuclear configurations; conversely, if μ has a “small” value, wavefunctions will respond quickly to nuclear conformational changes, allowing the electronic system to remain near the Born-Oppenheimer surface.¹⁴ In other words, a large μ leads to a stiff wavefunction dynamics, whereas a small μ leads to a more flexible wavefunction dynamics that will adapt to the nuclear dynamics without depart too much from the electronic ground state.

If $\mu \approx M_N$, then the wavefunctions dynamics will be very dependent on the nuclear dynamics (*i.e.* electrons wavefunctions and nuclear positions are strongly coupled) and a flow of energy between nuclear and electronic degrees of freedom occurs. This has to be avoided since it undermines the dynamics of molecules. Instead, if $\mu_i \ll M_N$, then the dynamics for ϕ^i and R_N will be almost completely decoupled, which is a good thing since it allows to remain close to the electronic ground state.

¹³In the CPMD routines the velocity-verlet is coupled to an iterative scheme called “SHAKE” [RCB77] that ensures that wavefunctions remain \perp to each other.

¹⁴“Large” and “small” have to be understood in relation to the mass of the nuclei $\{M_N\}$.

However, there is a limit to how small μ can be chosen: if the dynamic of the wavefunction becomes too fast, the time-step to integrate the equations of motion should be made small in turn, which is a disadvantage since (as will be discussed in Section 2.3.4) the time-step in CPMD is already very small. Thus, the μ parameter assumes the role of an “adiabaticity parameter” (is referred so in Ref. [MH09], p. 37).

A computational prove of the adiabaticity of the CPMD trajectories will now be presented, and can be found with greater details in References [PSB91,BS98,MH09]. The electronic dynamics generated by eqn. 2.46 can be analysed in terms of correlations functions; in particular, for small (but finite) deviations from the electronic ground state (*i.e.* well made but real simulations), it has been proved that the Kohn-Sham orbital dynamics is given by a superposition of oscillators of frequency

$$\omega_{ij} = \sqrt{\frac{2(\epsilon_j - \epsilon_i)}{\mu}} \quad , \quad (2.48)$$

where ϵ_j and ϵ_i are eigenvalues of unoccupied (virtual) and occupied Kohn-Sham orbitals, respectively. This means that the energy of the wavefunctions oscillates about the expected ground state value, without departing significantly from it. Moreover, if with

$$C_{\dot{\phi}\dot{\phi}}(t) = \sum_i^{occ} \langle \dot{\phi}^i(0); \dot{\phi}^i(t) \rangle \quad (2.49)$$

we refer to the autocorrelation function of the occupied Kohn-Sham orbitals, then it is observed that the power spectrum obtained using eqn. 2.48 matches the Fourier transform of eqn. 2.49:

$$\begin{aligned} \Gamma(\omega) &= \int_{-\infty}^{\infty} e^{-i\omega t} C_{\dot{\phi}\dot{\phi}}(t) dt = \int_{-\infty}^{\infty} e^{-i\omega t} \sum_i^{occ} \langle \dot{\phi}^i(0); \dot{\phi}^i(t) \rangle dt = \\ &= \int_0^{\infty} \cos(\omega t) \sum_i^{occ} \langle \dot{\phi}^i(0); \dot{\phi}^i(t) \rangle dt \quad . \end{aligned} \quad (2.50)$$

Thus, the smallest frequency in the electronic system dynamics is

$$\omega_{\min} \propto \sqrt{\frac{E_{gap}}{\mu}} \quad ,$$

with E_{gap} being the difference between the lowest unoccupied (“LUMO”) and the highest occupied (“HOMO”) orbital. For a reference system with

$E_{gap} = 2\text{eV}$ ($\approx 7.35 \cdot 10^{-2}$ a.u.) and with $\mu = 300$ a.u. (a typical value for a CPMD simulation), we have that $\omega_{\min} = 30600 \text{ cm}^{-1}$ whereas the higher vibrational frequency of the nuclear motion will reasonably be $< 4000 \text{ cm}^{-1}$. This great difference between the characteristic frequency of (electronic) wavefunctions and nuclear motion ensures that, with a suitable choice of the parameter μ , the electronic dynamics shall follow adiabatically the nuclear motion, with just a negligible energy transfer between slow (nuclear) and fast (electronic) degrees of freedom.

2.3.4 Implementation

Within the CPMD program¹⁵ and in most DFT calculations on condensed phases, it is common practice to model core electrons with pseudopotentials (which are discussed in Appendix C.3) and to expand in plane waves the wavefunctions of valence electrons only.

The use of plane waves is particularly helpful and it is the natural choice of basis set to study periodic systems, but is also employed in CPMD calculations of non-periodic systems because the finite dimensions of the simulation cell as well as PBC by definition introduce a periodicity, albeit a spurious one.

In a system with periodicity \vec{x} , e.g. a molecular crystal, the electronic density is also periodic and we have

$$\rho(\mathbf{r}) = \rho(\mathbf{r} + \vec{x}) \quad . \quad (2.51)$$

In a non-periodic system simulated in a cubic cell adopting PBC, \vec{x} would correspond to the cell length. Such density can be decomposed into its Fourier components $\tilde{\rho}(\mathbf{k})$ in the reciprocal \mathbf{k} space, as

$$\rho(\mathbf{r}) = \int d\mathbf{k} \tilde{\rho}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} \quad .$$

The latter integral can be discretised and reduced to a simple sum by adopting the approximated expression

$$\rho(\mathbf{r}) \simeq \sum_{\mathbf{k}} \tilde{\rho}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} \quad .$$

¹⁵“CPMD” is the acronym of both the computational package (ref. [CPM]) and the abstract method (ref. [CP85]), but we risk no dangerous misunderstanding since it is always evident from the context what is what.

The Bloch theorem (see Ref. [JM11], pag.13, eqn. (1.4.9)) tell us that eigenfunctions have the same periodicity (\vec{x}) of the lattice:

$$\phi_{\mathbf{k}}(\mathbf{r} + \vec{x}) = e^{i\mathbf{k}\cdot\vec{x}} \phi_{\mathbf{k}}(\mathbf{r}) \quad . \quad (2.52)$$

Thus, we obtain

$$\phi^i(\mathbf{r}) \equiv \phi_{\mathbf{k}}^i(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_i(\mathbf{k}, \mathbf{r}) \quad (2.53)$$

$$u_i(\mathbf{k}, \mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}}^i(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \quad , \quad (2.54)$$

which combined lead to

$$\phi^i(\mathbf{r}) \equiv \phi_{\mathbf{k}}^i(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{G}} c_{\mathbf{k}}^i(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}} \quad (2.55)$$

To calculate the electron density and the total energy, a sum on all \mathbf{k} -points should be carried out:

$$\rho(\mathbf{r}) = \sum_{\mathbf{k}} w^{\mathbf{k}} \sum_i \left| \phi_{\mathbf{k}}^i(\mathbf{r}) \right|^2 \quad (2.56)$$

where $w^{\mathbf{k}}$ is the weight of the \mathbf{k} -point. Usually, the sum in eqn. 2.56 is limited to a small number of \mathbf{k} -points, chosen to reflect the symmetry properties of ρ . In crystals with significant lattice distortion (leading to a symmetry breaking) and non-periodic systems like gases and liquids, the sum can be limited to just the point $\Gamma = (0, 0, 0)$ without loss of generality, so that eqn. 2.55 reduces to

$$\phi^i(\mathbf{r}) = \sum_{\mathbf{k}} c_i e^{i\mathbf{k}\cdot\mathbf{r}} \quad . \quad (2.57)$$

A cut-off E_{cut} for the \mathbf{k} -sum is usually adopted, excluding from eqn. 2.57 plane waves with kinetic energy greater than E_{cut} :¹⁶

$$\frac{1}{2} |\mathbf{k} + \mathbf{G}|^2 \leq E_{\text{cut}}$$

The optimal value for E_{cut} is affected by various factors, first of all the type of chemical species to simulate. Finally, it should be noted that the use of plane wave in place of more widely adopted Gaussians (see Appendix 2.1.5) has the further advantage of avoiding undesired effects such as “basis set superposition errors” (BSSE) and Pulay forces, as discussed in Appendix C.2.

¹⁶Since plane waves are simultaneous eigenfunctions of both the momentum and the free-particle Hamiltonian, calculating the kinetic energy is particularly simple, $E_k = \frac{\mathbf{p}^2}{2m_e} = \frac{1}{2} \mathbf{k}^2$.

2.3.5 Hydrogen vs Deuterium

In this dissertation, as well as in literature (e.g. [MSC08a, MSC08b, KMM⁺04]), many simulations are performed on deuterated or partially deuterated samples. Why this unusual choice, considering that deuterium (D, or ²H) is much less common in nature than ¹H? There are at least three good reasons behind it.

First, ¹H is the lightest element relevant in chemistry, and deuterium mass is about two times greater. The fact that the mass of ¹H is so small represents a problem for Born-Oppenheimer MD, since quantum effects of the nuclear motion becomes really important in this case. Because in Born-Oppenheimer MD quantum effects of the nuclear motion are completely neglected, it is common practice to adopt an heavier isotope to mitigate this issue.

Furthermore, swapping ¹H with ²H strongly affects the reduced mass of the system (μ), thus affecting the vibrational frequency ν , which, within harmonic approximation, is expressed as

$$\nu \propto \frac{1}{\sqrt{\mu}} \quad ,$$

(with μ being the reduced mass of the vibrating group). This is useful to “change” the vibrational frequencies of overlapping bands, in order to sort out every specific contribution to the spectrum independently. For example, the characteristic vibrational frequency of the O–H stretching is found at about $\sim 3500 \text{ cm}^{-1}$, while that of the O–D stretching is much lower at $\sim 2500 \text{ cm}^{-1}$.

Finally, the dynamics of ¹H is much faster than that of ²H. This fact forces to perform MD simulations adopting a smaller time-step to correctly describe and integrate the equations of motion, decreasing the size of the sampled phase space. In fact, when ¹H is used, for the *ab initio* MD simulations performed in this dissertations we chose a time step of 4 a.u. (~ 0.096 femtoseconds), while a longer time-step of 5 a.u. (~ 0.12 femtoseconds) is adopted for fully deuterated simulations.

3. Molecular Properties

Simulating the molecular behaviour would mean nothing if, from this, one cannot extract some useful information. This latter can be correlated with experimental measurements and this allows interpreting or predicting the observed features. The information that can be extracted from *ab initio* MD can be roughly summarised in three categories, as following.

Structural data, such as pair distribution functions, give insights about the average behaviour of the molecular systems, and are also useful to check the reliability of the simulation.

The electronic structure calculations can elucidate empirical data, because they are useful to create suitable models of interpretation.

Time-dependent and spectroscopic data often combine the peculiarities of the previous two subjects, providing a direct comparison with experimental spectra.

In this Chapter a brief presentation of a selected number of properties shall be given, to elucidate the theoretical background for the research work presented in Part II.

3.1 Structure

3.1.1 Pair Radial Distribution Functions

Pair radial distribution function $g(r)$ is an important thermodynamic quantity that describes how density of matter changes around a specific origin. The obtained picture is an averaged one, but it is useful to understand the microscopic structure of a condensed phase system. Moreover, other macroscopic properties are accessible from $g(r)$. For example, the static structure factor $S(\mathbf{k})$ could be proved to be (see for example Ref. [Cha87],

p. 209) proportional to the Fourier transform of $g(r)$:

$$S(\mathbf{k}) = 1 - \rho \int_{-\infty}^{+\infty} \mathbf{d}\mathbf{r} g(r) e^{-i\mathbf{k}\cdot\mathbf{r}} ,$$

where $\rho = N/V$ is the density; or the energy $\langle E \rangle$ of a system of real interacting particles

$$\langle E \rangle = \frac{3}{2} N k_B T + \frac{1}{2} N \rho \int_{-\infty}^{+\infty} \mathbf{d}\mathbf{r} g(r) u(r) ,$$

where $u(r)$ is the pair interaction potential. To define $g(r)$, it is useful to start from the statistical distribution $F(\mathbf{q}, \mathbf{p})$ that gives the probability to find the system in a point of the $6N$ dimensional phase space (\mathbf{q}, \mathbf{p}) , and whose analytic expression in the canonical NVT ensemble is

$$F(\mathbf{q}, \mathbf{p}) = e^{-\beta\mathcal{H}} / \iint e^{-\beta\mathcal{H}} \mathbf{d}\mathbf{q} \mathbf{d}\mathbf{p} , \quad (3.1)$$

where \mathcal{H} is the system Hamiltonian and $\beta = 1/k_B T$.

Here I remind that with symbols \mathbf{q} and \mathbf{p} I refer to all coordinate and momenta components of all N particles and with $\int \mathbf{d}\mathbf{q} \mathbf{d}\mathbf{p}$ an integration in the $6N$ phase-space is implied; to refer to the three specific coordinates of the j^{th} particle, the symbol \mathbf{r}_j shall be adopted.

Since coordinates and momenta are decoupled (*i.e.* linearly independent), then $F(\mathbf{q}, \mathbf{p})$ can be splitted as a product of two terms, one purely configurational ($P(\mathbf{q})$) and one purely momenta-dependent ($\Phi(\mathbf{p})$):

$$F(\mathbf{q}, \mathbf{p}) = P(\mathbf{q}) \cdot \Phi(\mathbf{p}) ,$$

with

$$P(\mathbf{q}) = \frac{e^{-\beta V(\mathbf{q})}}{\int \mathbf{d}\mathbf{q} e^{-\beta V(\mathbf{q})}} ,$$

where $V(\mathbf{q})$ is the potential energy. $P(\mathbf{q})$ is the probability of finding the system in a specific point \mathbf{q} of the $3N$ dimensional coordinate space (which, in turn, is a subspace of the complete $6N$ dimensional phase-space). The probability distribution $P(\mathbf{q})$ cannot be partitioned in single particle terms, due to the potential $V(\mathbf{q})$, but it is possible to define a pair distribution function as

$$P^{(2/N)}(\mathbf{r}_1, \mathbf{r}_2) = \int_{-\infty}^{+\infty} \mathbf{d}\mathbf{r}_3 \int_{-\infty}^{+\infty} \mathbf{d}\mathbf{r}_4 \cdots \int_{-\infty}^{+\infty} \mathbf{d}\mathbf{r}_N P(\mathbf{q}) ,$$

which gives the probability of finding particle 1 in \mathbf{r}_1 and particle 2 in \mathbf{r}_2 simultaneously. The contribution of the other $N - 2$ particles (from 3 to N) has been averaged through integration. Obviously, particle #1 and 2 have no particular significance since all particles are indistinguishable, so we need another distribution function $\varrho^{(2/N)}(\mathbf{r}_1, \mathbf{r}_2)$ that would give the probability of finding whatever particle in \mathbf{r}_1 and whatever other particle in \mathbf{r}_2 . There are N possible choices for the first particle and $N - 1$ for the second, so

$$\varrho^{(2/N)}(\mathbf{r}_1, \mathbf{r}_2) = N(N - 1)P^{(2/N)}(\mathbf{r}_1, \mathbf{r}_2) \quad .$$

In the same way, it is possible to obtain two single-particle functions

$$P^{(1/N)}(\mathbf{r}_1) = \int_{-\infty}^{+\infty} d\mathbf{r}_2 \int_{-\infty}^{+\infty} d\mathbf{r}_3 \cdots \int_{-\infty}^{+\infty} d\mathbf{r}_N P(\mathbf{q}) \quad ,$$

$$\varrho^{(1/N)} = N \cdot P^{(1/N)}(\mathbf{r}_1) \quad .$$

Due to the fact that $\varrho^{(1/N)} = N/V$, this quantity is numerically equal to the standard definition of particle density ρ .

A pair radial distribution function can be defined as

$$g(r) = g(\mathbf{r}_1, \mathbf{r}_2) = \frac{\varrho^{(2/N)}(\mathbf{r}_1, \mathbf{r}_2)}{\rho} \quad ; \quad (3.2)$$

this quantity is essential to describe the so-called “structure” of non-solid systems, and allows obtaining quantitative information about the solvation sphere around a specific site. ¹ In fact, if we choose a specific site to be the origin of r coordinate, we have

$$n(r') = 4\pi\rho \int_0^{r'} r^2 g(r) dr \quad , \quad (3.3)$$

which is the number of particles that lie within a cut-off distance r' from the site. When $r' \mapsto +\infty$ we have

$$4\pi\rho \int_0^{+\infty} r^2 g(r) dr = N \quad , \quad (3.4)$$

with N being the total number of particles.

¹I have dropped any vector glyph for the quantity r since, in this case, it is just the scalar distance $|\mathbf{r}_1 - \mathbf{r}_2|$.

Equation 3.2 is the statistical mechanical expression for $g(r)$. It appears difficult to be computed ($\rho^{(2/N)}(\mathbf{r}_1, \mathbf{r}_2)$ requires multi-dimensional integral calculation), but it is very easy to obtain it in MD. It is sufficient to calculate at every time-step the distances between the site and the particles and create an histogram of these contacts. From this sort of histograms we obtain the probability of finding a particle at a specific distance from the site of interest. In fact, for a non-continuum system eqn. 3.4 becomes

$$4\pi r^2 g(r) \Delta r = \frac{\Delta n(r)}{\rho} ,$$

which, since $\rho = N/V$, can be rewritten as

$$g(r) = \frac{\Delta n(r)}{\rho 4\pi r^2 \Delta r} = \frac{\Delta n(r) V}{N 4\pi r^2 \Delta r} ,$$

a much more computationally feasible definition than eqn. 3.2.

3.1.2 What Information can be recovered from $g(r)$?

The $g(r)$ function gives averaged properties, yet they are relevant to understand the behaviour of molecules.

The maxima of the $g(r)$ function show the position of the solvation shells, whereas the minima show the boundaries of those shells. The integral

$$\int_0^{r_{min}^1} g(r) dr \tag{3.5}$$

is the number of particle of the first solvation shell if r_{min}^1 is the first minimum of $g(r)$. Usually, the molecules of the first solvation shell are those that most interact with the site at the origin of the r coordinate, and, if H-bonding is possible, the integral of eqn. 3.5 coincides with the number of molecules H-bonded to the site.

The value of the $g(r)$ minima is also very important: the more it is small, the less mobility there is in the system. A value of zero for the first minimum means that the molecules in the first and second solvation shell are actually separated by an impenetrable “wall”, with no interchange.² What type of wall am I talking about? A potential energy barrier that allows or impairs the molecular mobility between solvation shells.

²This usually means that the simulation is quite far from reaching ergodic limit.

This view of the potential energy barrier is supported by the so-called “reversible work theorem” (see [Cha87], p. 202), which states that for a fluid ³

$$-\left\langle \frac{\partial}{\partial r} V(\vec{r}_N) \right\rangle = k_B T \frac{\partial}{\partial r} \ln g(r) \quad ; \quad (3.6)$$

the quantity $-\left\langle \frac{\partial}{\partial r} V(\vec{r}_N) \right\rangle$ is the mean force acting on particles expressed as a function of the distance r from the origin. The integral of a mean force is a reversible work, so the “mean force potential”

$$w(r) = k_B T \ln g(r) \quad (3.7)$$

also expresses the reversible work to be spent in order to move a particle along the r coordinate. Due to the fact that the $\ln(x)$ function increases monotonically, high values of the mean force potential $w(r)$ correspond to minima in the $g(r)$ function. Obviously, all these considerations stand only near the ergodic limit of simulation long enough to sample the available phase space.

3.1.3 Angular Distribution and H-bond Functions

The angular distribution function $g(\vartheta)$ is computed analogously to the $g(r)$ function; however, the contacts are reported not as a function of some interatomic distance but as a function of a specific angle ϑ (defined by three atoms). Moreover, the increase of ϑ is associated to the increase of solid angle values, so with $g(\vartheta)$ the sampling bin is not constant. Thus, a specific normalisation function has to be employed and $g(\vartheta)$ behaves like

$$g(\vartheta) \sim \frac{\text{contacts}(\vartheta)}{\sin(\vartheta)} \quad .$$

The dependence on both the distance r and the angle ϑ can be combined into a single function $g(r, \vartheta)$. One of the possible choices for such function is the so-called “H-bond function” $F^{\text{HB}}(r, \vartheta)$, presented in Ref. [PCRS03] by Pagliai *et al.*, and here reported for sake of completeness:

$$F^{\text{HB}}(r, \vartheta) = A(r(t)) \cdot B(\vartheta(t)) \quad , \quad (3.8)$$

³In a crystal, the $g(r)$ function is exactly 0 for most values of the r coordinate even at thermodynamic equilibrium, thus giving divergences if eqn. 3.6 is employed.

with

$$A(r(t)) = \begin{cases} \exp(-(r_e - r(t))^2/2\sigma_r^2) & \text{if } (r_e - r(t)) < 0 \\ 1 & \text{if } (r_e - r(t)) \geq 0 \end{cases} \quad (3.9)$$

and

$$B(\vartheta(t)) = \begin{cases} \exp(-(\vartheta_e - \vartheta(t))^2/2\sigma_\vartheta^2) & \text{if } (\vartheta_e - \vartheta(t)) < 0 \\ 1 & \text{if } (\vartheta_e - \vartheta(t)) \geq 0 \end{cases} . \quad (3.10)$$

Defining the H-bond geometry schematically as A–H··B, as reported in fig 3.1, this function can “weight” the H-bond strength when the coordi-

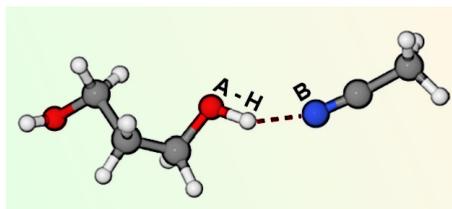


Figure 3.1: An example of H-bond between a donor diol (on the left) and an acceptor acetonitrile (right). The H-bond distance is represented with a dashed line.

nate r corresponds to the H··B distance and the angle ϑ to the A–H··B angle. Parameters r_e and ϑ_e are the equilibrium values for r and ϑ , respectively, and are extrapolated from the unnormalised pair distribution functions. Conversely, parameters σ_r and σ_ϑ are the half width at half height of the first peak of those distributions.

The F^{HB} function has already been employed, as well as in this dissertation, to study the H-bond network of methanol [PCRS03], of halides in methanol [PCS05,FPCS06], and of pyridine in water [PBMM⁺]. In particular, this function has been developed to correctly describe even the lifetime of the H-bonding in methanol [PCRS03], therefore it is a suitable choice to understand and probe the dynamic behaviour of protic species. The only drawback is that it cannot be computed “on the fly”, because it needs data from both the $g(r)$ and $g(\theta)$ functions to be computed.

3.2 Electronic Structure

Ab initio calculation allows obtaining information about the electronic structure, because this latter (in the form of electron density or molecular orbital approach) is actually optimised.

To quantitatively describe the electronic structure, suitable parameters have to be used: the simplest ones are the partial charges associated to atoms in molecules. The story of partitioning the total molecular charge in contributions due to its component atoms is quite long, and still far from being at the end (see for example [Bad91]). Anyway, even the simple Mulliken partitioning [Mul55] of charge (“Mulliken population analysis”) can provide useful insights on what happens to electrons in molecule, where they gather and from where they withdraw.

When orbitals are constructed as combinations of plane waves, specific methods to localise the electronic charge have to be adopted. Two methods to achieve this goal are the Electron Localisation Functions and the Wannier centres analysis, which have been both employed in this dissertation and are briefly presented in the following.

Electron Localisation Functions

Electron Localisation Functions (ELFs) have been developed by Edgecombe and Becke [BE90], who proposed to consider the following function:

$$\eta(\mathbf{r}) = \frac{1}{1 + \frac{D_\sigma}{D_\sigma^0}} \quad . \quad (3.11)$$

The function $\eta(\mathbf{r})$ has value close to 1 where Pauli repulsion is weaker than in a uniform electron gas of the same density, thus giving an high localisation; on the contrary, where spin coupling is stronger $\eta(\mathbf{r})$ shall tend to the value of 0.5, which corresponds to that of an homogeneous electron gas (which, by definition, is completely non-localised). In fact, the functions D_σ and D_σ^0 are the so-called “Fermi-curvatures” of a real system and of a uniform electron gas, respectively; Fermi-curvature is a gradient function of the spin density (ρ_σ):

$$D_\sigma \sim -\frac{(\nabla\rho_\sigma)^2}{\rho_\sigma} \quad ;$$

the more D_σ is small, the more the electron is localised. The maxima of the $\eta(\mathbf{r})$ function can be mapped onto the real space of the molecule to give a topological view of electronic rearrangement. Another way to look at the ELFs was proposed by Savin and Silvi [SS94]: the ELFs represent the bosonic character of electrons, with pairs of antiparallel electrons showing some bosonic behaviour and thus having high values of $\eta(\mathbf{r})$.

Wannier Centres

Maximally localised Wannier function (MLWF) centres represent another method to localise and visualise electrons in molecules. They are constructed by employing unitary translation operators. Translations in quantum mechanics are represented by unitary operators that commute with the Hamiltonian, when the translation \mathbf{R} is compatible with the system symmetry.

Wannier functions [MV97, SMVP98] can be proved to coincide with the Fourier transform of the Bloch orbitals $\phi_{\mathbf{k}}(\mathbf{r})$ (see eqn. 2.52):

$$\omega_{\mathbf{R}}(\mathbf{r}) \sim \int d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{R}} \phi_{\mathbf{k}}(\mathbf{r}) \quad .$$

These functions are still not localised. To reduce the space-spread of the Wannier functions, it is required to minimise the mean square displacements of the expectation values of the position operators; this minimisation is permitted because it is arbitrary how to set the phase of the wavefunction.

Once obtained these new Wannier functions $\omega(\mathbf{r})$, their centres are determined as [MV97]:

$$\mathbf{r}_c \sim -\Im\{\ln\langle\omega|e^{i\mathbf{k}\mathbf{r}}|\omega\rangle\} \quad .$$

The position of the centres gives a straightforward picture of the electron pairs distribution in molecules, and it happens to closely resemble the qualitatively diagrams often drawn “by hand” employing simple VSEPR concepts.

3.3 Vibrations: Some Rules of Thumb

Molecular vibrations happen in a large time and energy scale. For the aim of this dissertation, the vibrations are those occurring in the 1500–4000 cm^{-1} wavenumbers region, corresponding to energies of about 0.2–0.5 eV. In this frequency region, vibrations are often associated to fast and very localised molecular motions, whereas under the 1000 cm^{-1} threshold there are many slow and delocalised modes, such as the collective lattice modes in a crystal solid.

This spectral range can be probed experimentally with a number of techniques, in particular IR and Raman spectroscopies. These two types

of spectroscopies are somewhat complementary both regarding selection rules and the choice of solvent (for spectroscopy in liquid media). As a rule of thumb, we could say that the IR activity is associated to a change of the molecular dipole moment caused by an appropriate vibration, whereas Raman activity is due to a change in the polarisability tensor. In fact, the interaction of a system of slow-moving charges and an electromagnetic radiation can be reduced, for classical fields, to the Hamiltonian of some electric dipoles with the electric field $\mathbf{E}(t)$:

$$\mathcal{H}(t) = \vec{\mu}(t) \cdot \mathbf{E}(t) + \frac{1}{2} \overleftrightarrow{Q}(t) \cdot \nabla \mathbf{E}(t) + \dots, \quad (3.12)$$

where $\vec{\mu}(t)$ represent the dipole moment vector and $\overleftrightarrow{Q}(t)$ the quadrupole moment 2^{nd} rank tensor. Limiting ourselves to the first term, thus disregarding multipoles higher than the dipole, we can see that the latter can be expressed as made up of a permanent ($\vec{\mu}_0$) and induced terms:

$$\vec{\mu}(t) = \vec{\mu}_0 + \underbrace{\overleftrightarrow{\alpha} \cdot \mathbf{E}(t)}_{1^{st} \text{ induced dipole}} + \overleftrightarrow{\beta} \cdot \mathbf{E}(t) \mathbf{E}(t) + \dots \quad (3.13)$$

The first addend determines the IR activity, while the second one is related to the Raman activity. The electric field can polarise the charge distribution creating an induced dipole even in molecules without a permanent dipole, and this new dipole is proportional to the polarisability tensor $\overleftrightarrow{\alpha}$.

Moreover, it is useful to rationalise molecular vibrations on the basis of the harmonic approximation. This approximation is a rather crude one, but it is simple and powerful enough to clarify a lot of features with very little and straightforward reasoning.

In particular, the harmonic frequency ν for a classical oscillator is expressed as

$$\nu \propto \sqrt{\frac{k}{\mu}}, \quad (3.14)$$

where k is the so-called force constant and μ is the reduced mass of the system.⁴ The force constant is the Laplacian of the potential energy of

⁴The reduced mass for a two mass system m_1 and m_2 is

$$\mu = \frac{m_1 \cdot m_2}{m_1 + m_2} .$$

the harmonic oscillator U_{harm} :

$$k = \nabla^2 U_{\text{harm}} \quad (3.15)$$

The H-bond induces, as shown in the following chapters, anharmonic effects, *i.e.* effects that cannot be completely understood with a simple harmonic oscillator model and which should require a more extended potential Taylor expansion around the equilibrium position:

$$U(x) = U_0 + \underbrace{\frac{1}{2}kx^2}_{\text{harmonic term}} + \frac{1}{3!}\nabla^3 Ux^3 + \dots$$

Nevertheless, the harmonic oscillator model still proves useful to have an immediate understanding of anharmonic effects, since these latter can often be *interpreted* as due to a *change* in the value of the force *constant* k . This way of rationalising H-bond effects onto vibrational spectra are recurrent in this dissertation.

3.4 Vibrations: Linear Response Theory

Linear response theory is the conceptual framework to describe systems slightly displaced from an equilibrium state, close enough to it to allow a linear approximation to work. Linear response theory is closely related to the so called “fluctuation-dissipation theorem”, which says that the fluctuations of a system at equilibrium are related to the response of a system slightly perturbed out of equilibrium. Linear response theory is thus relevant to obtain information about vibrational spectra: to obtain experimental spectra, the system is perturbed out of equilibrium by radiation absorption, and to recover this effect we have to study the fluctuations of our simulated molecular system (which is, hopefully, close to equilibrium).

It has been shown in Section 2.2.4 that the equilibrium average of a quantity $\vartheta(\mathbf{q}, \mathbf{p})$ can be expressed as

$$\langle \vartheta \rangle = \frac{\iint \vartheta(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \, d\mathbf{q}d\mathbf{p}}{\iint f(\mathbf{q}, \mathbf{p}) \, d\mathbf{q}d\mathbf{p}} ; \quad (3.16)$$

in a canonical ensemble NVT , to be consistent with eqn. 3.1, the density of states $f(\mathbf{q}, \mathbf{p})$ is

$$f(\mathbf{q}, \mathbf{p}) = e^{-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})} \quad ,$$

with $\mathcal{H}(\mathbf{q}, \mathbf{p})$ representing the equilibrium Hamiltonian.

Now let us suppose that at time $t < 0$ the system was not at equilibrium, due to a perturbation of the type

$$\Delta = -f \vartheta \quad ,$$

with f a coupling term with property $\vartheta(\mathbf{q}, \mathbf{p})$, and that the perturbation is switched off at $t = 0$. Thus, for $t > 0$ the system should relax towards equilibrium, but at $t = 0$ the density $f(\mathbf{q}, \mathbf{p})$ is still

$$f = e^{-\beta(\mathcal{H} + \Delta)} \quad . \quad (3.17)$$

Equation 3.16 can be used to calculate an instantaneous average of ϑ even for a non-equilibrium state, which shall be indicated as $\bar{\vartheta}$ to distinguish it from the equilibrium average $\langle \vartheta \rangle$. At $t = 0$ the average $\bar{\vartheta}$ is

$$\bar{\vartheta}(t = 0) = \frac{\iint \vartheta e^{-\beta(\mathcal{H} + \Delta)} \, d\mathbf{q}d\mathbf{p}}{\iint e^{-\beta(\mathcal{H} + \Delta)} \, d\mathbf{q}d\mathbf{p}} \quad , \quad (3.18)$$

and for $t \rightarrow +\infty$ we have $\bar{\vartheta} = \langle \vartheta \rangle$. At $t > 0$ the expression for $\bar{\vartheta}$ is equal to the previous equation 3.18; however, it must be pointed out that in this case the dynamics of the ϑ function is governed by the Hamiltonian \mathcal{H} , and not by $\mathcal{H} + \Delta$, because the perturbation has been shut off at $t = 0$. The difference between the two averages $\bar{\vartheta}$ and $\langle \vartheta \rangle$ is due to the perturbation and, if this perturbation is small enough, this allows expanding the non-equilibrium density f of eqn. 3.17 as

$$f = e^{-\beta(\mathcal{H} + \Delta)} = e^{-\beta \mathcal{H}} (1 - \beta \Delta + \dots)$$

Substituting this result in eqn. 3.18 and remembering that $\langle \vartheta \rangle = \langle \vartheta(t) \rangle$, it can be proved that

$$\bar{\vartheta} = \langle \vartheta \rangle - \beta [\langle \Delta \vartheta(t) \rangle - \langle \vartheta \rangle \langle \Delta \rangle] + \dots$$

The difference between the averages $\bar{\vartheta}$ and $\langle \vartheta \rangle$ tell us how the system returns to equilibrium after the perturbation is switched off, and can be expressed as

$$\bar{\vartheta}(t) - \langle \vartheta \rangle = \beta f \langle [\vartheta(0) - \langle \vartheta \rangle] \cdot [\vartheta(t) - \langle \vartheta \rangle] \rangle \quad ,$$

where the averaged quantity, the product of differences

$$C_{\vartheta,\vartheta}(t) = \langle [\vartheta(0) - \langle \vartheta \rangle] \cdot [\vartheta(t) - \langle \vartheta \rangle] \rangle \quad (3.19)$$

is called “autocorrelation function” and plays a pivotal role in obtaining time-dependent information from MD simulations.

The autocorrelation function presented in eqn. 3.19 can be rewritten as

$$C(t)_{\vartheta,\vartheta} = \langle \vartheta(0) \vartheta(t) \rangle - \langle \vartheta \rangle^2$$

and has many interesting properties; for example, near time $t = 0$ we have that

$$C(\sim 0)_{\vartheta,\vartheta} = \langle [\vartheta(0) - \langle \vartheta \rangle]^2 \rangle \quad ,$$

whereas at large times the fluctuations become uncorrelated with the original value $\vartheta(0)$ and, as $t \rightarrow +\infty$ we eventually have that $C(t)_{\vartheta,\vartheta} = 0$.

The autocorrelation function is easily accessible by MD simulations, since it can be calculated by substituting the ensemble with the time average as in eqn. 2.39:

$$C_{\vartheta,\vartheta}(t) = \frac{1}{t_{max}} \int_0^{t_{max}} d\tau \vartheta(\tau) \cdot \vartheta(\tau + t) \quad .$$

3.5 Vibrations: Spectra from MD Simulations

3.5.1 Vibrational Density of States

The vibrational density of states (VDoS) can be seen as representing the collection of all possible vibrations in a molecular system. It is a quantity accessible to inelastic neutron scattering and, in principle, contains all information also present in IR and Raman vibrational spectra. It is particularly simple to be calculated from MD simulations, exploiting the autocorrelation function formalism.

Let us assume that the molecular system is made up of a collection of (classical) harmonic oscillators; thus, the coordinates of the i^{th} atom can be written as

$$x_i(t) = A_i \cos(\omega_i t + \phi_i) \quad , \quad (3.20)$$

which can be differentiated with respect to time to obtain

$$v_i(t) = \frac{d}{dt} x_i(t) = -A_i \omega_i \sin(\omega_i t + \phi_i) \quad ;$$

here the constants are not very important, so we can limit to

$$v_i(t) \sim \sin(\omega_i t + \phi_i) \quad , \quad (3.21)$$

and the velocity autocorrelation function takes the form

$$C_{vv}(t) = \sum_i \langle v_i(0)v_i(t) \rangle \sim \sum_i \langle \sin(\phi_i) \sin(\omega_i t + \phi_i) \rangle \quad .$$

Exploiting the equality

$$\sin a \cdot \sin b = \frac{1}{2} [\cos(a - b) - \cos(a + b)]$$

we obtain

$$C_{vv}(t) \sim \sum_i \langle \cos(\omega_i t) - \cos(\omega_i t + 2\phi_i) \rangle \quad .$$

The second term can be proved to be null, provided that phases ϕ_i are random distributed (which is true at thermodynamic equilibrium), due to the following equality

$$\cos(a + b) = \cos a \cos b - \sin a \sin b \quad ,$$

so that

$$\int_0^{2\pi} d\phi \cos(\omega t + 2\phi) = \int_0^{2\pi} d\phi [\cos(\omega t) \cos(2\phi) - \sin(\omega t) \sin(2\phi)] = 0 \quad .$$

Considering the only surviving term, we eventually arrive to

$$C_{vv}(t) \sim \sum_i \langle \cos(\omega_i t) \rangle \quad . \quad (3.22)$$

If we define $\Gamma(\omega)$ as the Fourier transform of $C_{vv}(t)$

$$\Gamma(\omega) \sim \int_{-\infty}^{+\infty} C_{vv}(t) e^{-i\omega t} dt \quad (3.23)$$

and consider the time average by following the ergodic principle, like in eqn. 2.39, then eqn. 3.22 can be rewritten as an integral in frequency space

$$C_{vv}(t) \sim \sum_i \int_0^{+\infty} \Gamma(\omega) \cos(\omega_i t) d\omega_i \quad ,$$

where the integral corresponds to the usual definition of real space Fourier transform. In this way, it becomes evident that $\Gamma(\omega)$ represents the number of oscillators in the system of frequency ω , *i.e.* the quantity $\Gamma(\omega)$ is the density of vibrational states.

The VDoS is a collective property of the system, and can be calculated as the Fourier transform of the velocity autocorrelation function as in eqn. 3.23, because single-particle displacements can be viewed as superimposition of all normal modes of the system. It has to be pointed out that the complete VDoS calculated via eqn. 3.23 should match the sum of all normal modes of the system, but VDoS of specific pairs of atoms (*i.e.* $\Gamma(\omega_i)$) do not necessarily match any given normal mode.

In this dissertation, the VDoS of the various cases of study has been calculated with a more direct formula than eqn. 3.23, based on direct Fourier transform of the coordinate displacements. This should not be surprising, since the analytic expressions of the displacements $x(t)$ in eqn. 3.20 and of velocities $v(t)$ in eqn. 3.21 are exactly the same, but for a phase factor to convert *sin* into *cos*. Moreover, in Ref. [Tuc10b], p. 546, it is shown that quantum and classical autocorrelation functions for a system of harmonic oscillators are directly proportional:

$$C(t)_{\text{quantum}} = \beta\hbar\omega \tanh\left(\frac{\beta\hbar\omega}{2}\right) C(t)_{\text{classical}} \quad .$$

3.5.2 IR and Raman Spectra

As shown in eqn. 3.12 and 3.13, the first term determining the interaction between matter and light is

$$\mathcal{H}(t) \simeq \vec{\mu}_0(t) \cdot \mathbf{E}(t)$$

and it determines the IR spectrum. In Ref. [Tuc10b], at page 547, it is shown that, just like the VDoS can be calculated as Fourier transform of the autocorrelation function of the displacements, the IR activity can be calculated as the Fourier transform of the autocorrelation function of the molecular dipole moment ($C_{\mu\mu}(t)$):

$$\text{IR}(\omega) = \frac{\omega}{3\hbar} \tanh\left(\frac{\beta\hbar\omega}{2}\right) \int_{-\infty}^{+\infty} dt e^{-i\omega t} C_{\mu\mu}(t) \quad . \quad (3.24)$$

As in the case of VDoS calculation, this results is usually exploited by adopting the *classical* dipole moment autocorrelation function.

In the present dissertation, I have observed that for high frequency vibrations (O–D and O–H stretchings) there is very little difference between VDoS and IR spectra. Calculating IR spectra from *ab initio* MD trajectories is significantly more computationally expensive than calculating

VDoS, since methods based on MLWF centres have to be employed to localise electrons, in order to calculate reliable dipole moments. In practice, after having collected a trajectory, a new calculation has to be performed to optimise the wavefunctions and employ MLWF centres analysis.

Calculating Raman spectra requires even more computation, due to the fact that it is necessary to calculate not simply the molecular dipoles but also the six independent components of the polarisability tensor. I have not performed such calculations for this dissertation, but details about the implementation of Raman calculation in CPMD [CPM] can be found in Ref. [PCC⁺08].

3.5.3 Time-Sampling and Spectra: Part I

I find appropriate to conclude this methodological chapter with some brief remarks about time-sampling and how it affects the computed spectra.

In fact, the more the time-spacing is small (*i.e.* the more the time-sampling is dense), the higher is the maximum accessible frequency obtainable by means of eqn. 3.23. Moreover, the longer is the MD simulation, the higher is the resolution of the computed spectrum. This is obvious: to sample high frequencies, a short time-interval sampling is needed. And, to improve resolution, repeated sampling has to be carried out.

A more quantitative discussion about these issues shall be presented in Section 4.1.1. To quantify these statements here and now, I pinpoint that even a huge time-step of ~ 3 femtoseconds provides a maximum accessible frequency of $\sim 5500 \text{ cm}^{-1}$, and even a not-so-long trajectory of ~ 17.5 picoseconds gives a frequency resolution of less than 2 cm^{-1} (see Ref. [MMPCR12a]).

4. Fourier and Wavelet Analysis

In this Chapter a quite deep discussion about Fourier and wavelet transforms shall be presented. Fourier transforms have already been exploited in eqn.2.50 as well as in sections 3.5.1 and 3.5.2, but here they will be discussed from a more general point of view in order to introduce time-frequency analysis.

What is time-frequency analysis? We could say that music is conceptually closely related to it: sheet music, something that every musician is used to read, associates a specific acoustic frequency (note) to the specific instant when it should be played. In time-frequency analysis correlation



Figure 4.1: Sheet music. The first 9 measures of the *Canon in D* of baroque composer Johann Pachelbel.

plots similar to sheet music are obtained.

There are various mathematical “tools” to perform time-frequency analysis, and the discussion shall be devoted mainly to short-time Fourier transform and wavelet analysis. Actually, the former is didactically helpful, the latter is much more accurate and useful.

4.1 Fourier Transform

The most used “tool” to obtain frequency information from a time-resolved signal is probably the Fourier transform (FT). There are many ways to in-

roduce this type of transform; in my humble opinion, adopting a quantum-mechanical formalism is the closer to a physical-chemist point-of-view.

Position eigenstates $|x\rangle$ can be used as a base for the state ket $|\phi\rangle$ producing the usual definition of wavefunction (already seen in eqn.2.9)

$$\phi(x) = \langle x|\phi\rangle \quad , \quad (4.1)$$

and of its complex conjugate (already seen in eqn.2.10)

$$\phi(x)^* = \langle \phi|x\rangle \quad .$$

In order to have a momentum-space wavefunction, we have to change the basis set employing the $\mathcal{L}^2(\mathbb{R})$ Hilbert space closure relation

$$\mathbb{I} = \int_{-\infty}^{+\infty} dp |p\rangle\langle p| \quad , \quad (4.2)$$

where the symbol \mathbb{I} is used to represent the identity operator that has the property

$$\langle \phi_1|\mathbb{I}|\phi_2\rangle = \langle \phi_1|\phi_2\rangle \quad .$$

This remarkable property can be exploited in eqn.4.1 to obtain:

$$\phi(x) = \langle x|\phi\rangle = \langle x|\mathbb{I}|\phi\rangle = \langle x|\left[\int_{-\infty}^{+\infty} dp |p\rangle\langle p|\right]|\phi\rangle = \int_{-\infty}^{+\infty} dp \langle x|p\rangle\langle p|\phi\rangle \quad . \quad (4.3)$$

The latter equation can be put into the following form

$$\phi(x) = \int_{-\infty}^{+\infty} dp c(x)\phi(p) \quad , \quad (4.4)$$

where the coefficient of the expansion $c(x)$ is

$$c(x) = \langle x|p\rangle \quad .$$

The $c(x)$ coefficient is the projection of a momentum eigenket into coordinate space: in fact, the momentum eigenket satisfies the following eigenvalue equation

$$\hat{P}|p\rangle = p|p\rangle \quad ,$$

which can itself be projected into momentum space with the usual operator substitution

$$\hat{P} \rightarrow -i\hbar \frac{\partial}{\partial x}$$

obtaining

$$-i\hbar \frac{\partial}{\partial x} \langle x|p\rangle = p \langle x|p\rangle \quad . \quad (4.5)$$

Equation 4.5 has as solution ¹

$$\langle x|p\rangle = e^{+i\hbar px} \quad ,$$

which enables us to rewrite eqn. 4.4 as

$$\phi(x) = \int_{-\infty}^{+\infty} dp e^{+i\hbar px} \phi(p) \quad . \quad (4.6)$$

By remembering that, if $\langle x|p\rangle = e^{+i\hbar px}$, then $\langle p|x\rangle = \langle x|p\rangle^* = e^{-i\hbar px}$, the $\phi(p) = \langle p|\phi\rangle$ function can be expanded as

$$\phi(p) = \int_{-\infty}^{+\infty} dx e^{-i\hbar px} \phi(x) \quad . \quad (4.7)$$

Defining $p = \hbar k$ and $k = p/\hbar$, equations 4.6 and 4.7 become

$$\begin{aligned} \phi(k) &= \int_{-\infty}^{+\infty} dx e^{-ikx} \phi(x) = \mathcal{F}[\phi(x)] \quad , \\ \phi(x) &= \int_{-\infty}^{+\infty} dk e^{+ikx} \phi(k) = \mathcal{F}^{-1}[\phi(k)] \quad . \end{aligned} \quad (4.8)$$

The pair of equations 4.8 defines the FT and its inverse. Moreover, because in the $\mathcal{L}^2(\mathbb{R})$ Hilbert space we have (in analogy with eqn. 4.1)

$$\phi(k) = \langle k|\phi\rangle \quad ,$$

then the \mathcal{F} -transform can be defined also as inner product:

$$\mathcal{F}[\phi(x)] = \langle k|\phi\rangle \quad . \quad (4.9)$$

Thus, the FT can be reduced to a “tool” that changes the vector basis and shifts information from coordinate to reciprocal space. Furthermore, the actual discussion has been made possible by the fact that coordinates

¹The complex conjugate $(\langle x|p\rangle)^* = \langle p|x\rangle = e^{-i\hbar px}$ is too a solution of eqn. 4.5, and in literature there is ambiguity regarding the sign of the exponent of the FT. In this dissertation the minus sign for the FT and the plus sign for the inverse have been consistently adopted .

and momenta are “conjugate variables” (*i.e.* $[\hat{x}, \hat{p}] \neq 0$) and their uncertainties are related by the Heisenberg principle:

$$\Delta x \Delta p \geq \frac{\hbar}{2} ;$$

this is a very relevant assumption because it enables to extend this way of reasoning to any other pairs of conjugated variables related by a similar uncertainty relationship. In fact, we have similar relationship for the energy-time pair: ²

$$\Delta t \Delta E \geq \frac{\hbar}{2} .$$

Remembering that $\Delta E = \hbar \Delta \omega$, the latter becomes

$$\Delta t \Delta \omega \geq \frac{1}{2} . \tag{4.10}$$

Thus, one can “jump” from time to frequency space adopting Fourier analysis.

4.1.1 Time-Sampling and Spectra: Part II

Inequality 4.10 is very important because it allows a deep discussion about time and frequency relationship. I have already anticipated that the time-sampling affects the power spectrum (*i.e.* the FT): this happens because of inequality 4.10. Here with Δx I refer to uncertainty in the x property, whereas with N I refer to the number of time-step Δt , *i.e.* the length of the signal expressed in units Δt .

Obviously, we have that

$$t_{max} = N \cdot \Delta t , \tag{4.11}$$

and, moreover,

$$\omega_{max} = N \cdot \Delta \omega . \tag{4.12}$$

With Δt fixed, increasing t_{max} has the effect of increasing N (eqn. 4.11). When N increases, since $\Delta \omega \propto 1/N$ (eqn. 4.12), the frequency uncertainty $\Delta \omega$ shrinks, thus the actual spectral resolution increases. So, from a

²However, I have to point out that time is not an observable in non-relativistic quantum mechanics, but a parameter.

longer MD simulation, we can obtain a vibrational spectrum with a higher resolution. In fact, it can be proved that

$$\Delta\omega(\text{cm}^{-1}) = \frac{1}{t_{max}(\text{sec}) \cdot c(\text{cm}/\text{sec})} ,$$

with c being the speed of light in vacuum. By using this equation one can obtain the numbers anticipated in Section 3.5.3.

Moreover, since $N = t_{max}/\Delta t$ (eqn. 4.11), inserting it in eqn. 4.12 we have

$$\omega_{max} \propto \frac{1}{\Delta t} ,$$

proving that smaller time steps allow reaching higher frequencies This explains why adopting smaller time-steps allow obtaining a larger spectral window.

4.1.2 What the Fourier Transform Just Cannot Do

Fourier transform can extract the power spectrum of a time-resolved signal, but it has some limitations.

First of all, very similar signals can have a very similar Fourier spectrum ³, as pictured in Ref. [MKO93] and also shown in fig.4.2. As can be seen, the $f(t)$ abruptly changes frequency of oscillation in the middle of the time axis; on the contrary, the $g(t)$ signal always shows two frequencies simultaneously. They are two very different signals, yet their Fourier transforms are almost indistinguishables because they are characterised by the same pair of frequencies.

Moreover, Fourier transform cannot be used to analyse signals “on the fly”, *i.e.* those not completely decayed to zero. This is due to the fact that the Fourier time-window spans from $-\infty$ to $+\infty$, and a signal has to decay to zero by a finite time for a reliable Fourier transformation. The time-window of the FT is so large because the function that gives the size

³Here I have to specify that, unless explicitly stated otherwise, every plotted FT of some function $f(t)$, actually represents the square modulus of the said FT, *i.e.*

$$\mathcal{F}[f(t)] \cdot \mathcal{F}^*[f(t)] = \int_{-\infty}^{+\infty} dt f(t)e^{-i\omega t} \cdot \int_{-\infty}^{+\infty} dt f^*(t)e^{+i\omega t} .$$

Moreover, every plotted FT reported here is obtained by the so called “Fast Fourier Transform” algorithm (FFT) adopting the free `fftw` package [FJ97,FJ98,JF08] or the `ESSL` package developed by `IBM`.

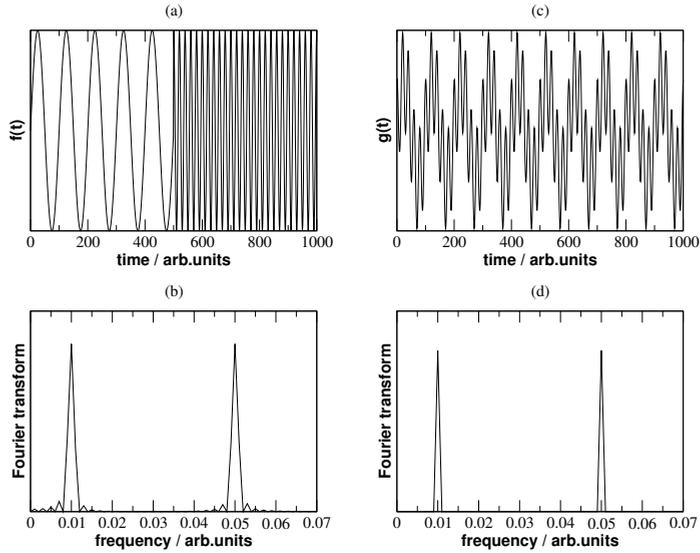


Figure 4.2: Two time-dependent signals $f(t)$ (panel a) and $g(t)$ (panel c) and their respective Fourier transforms (panels b and d).

of this window is a plane wave, which, as already discussed, has its square modulus equal to 1 over all its domain (which happens to be all \mathbb{R}). To overcome this issue, the so-called short-time Fourier transform has been introduced, defined by

$$\int_{t-\tau}^{t+\tau} dt f(t') e^{-i\omega t'} \quad ,$$

which limits the integration just to a finite time-window. The basic idea is that the computation is repeated in various time-windows of length 2τ in order to sample the entire signal. Notwithstanding this, a similar approach is affected by a very serious issue, called “aliasing”: in fact truncating the time series at the edge of the finite time-interval introduces into the power spectrum spurious high frequency signals. These signals alter the overall appearance of the spectrum and are seriously misleading.

Furthermore, there is the problem that a very compact and well resolved time-signal has its power spectrum spread over a very large interval of frequencies. This can be appreciated in fig. 4.3, where the FT of a Dirac’s delta function (panel a) is a completely delocalised function (panel b), whereas the transform of a less time-localised function (panel c)

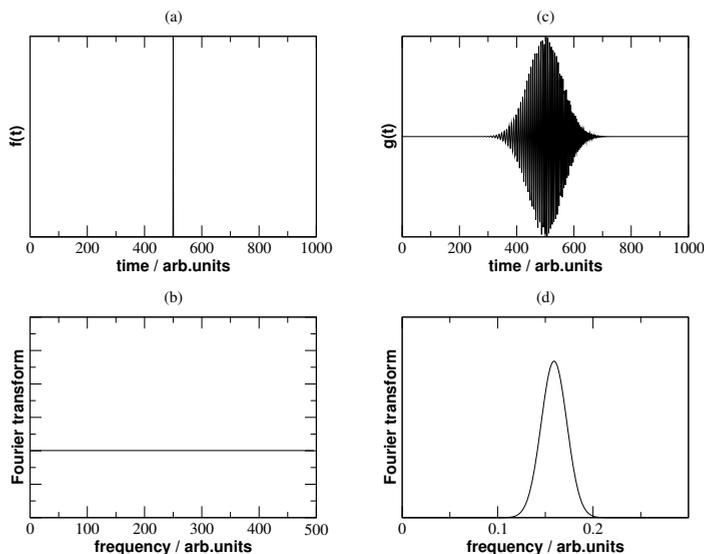


Figure 4.3: Two time-dependent signals $f(t)$ (panel a) and $g(t)$ (panel b) and their respective Fourier transforms (panels b and d).

is better localised in the frequency space (panel c). This can be considered a consequence of the time-frequency uncertainty relation (eqn. 4.10): a signal well localised in time (the Dirac's delta is the best example of this type of function) contributes to all possible frequencies of the spectrum.

To overcome these limitations and obtain good time-frequency localisations, many approaches have been developed, mainly based on the short-time Fourier transforms.

4.2 Wavelet Transform

FT is usually employed to extract the frequency content from a time-dependent signal, without a simultaneous localisation in both frequency and time domain. In order to obtain vibrational properties from MD trajectories, Fourier transforms are performed, [Tuc10b,BP76,D. 00] but it has been recently shown [MSC08a,MSC08b,RW05] that similar results can be obtained with a wavelet analysis [C. 92] approach.

We have seen that FT can be interpreted as the inner product (eqn. 4.9)

of the state vector onto a specific basis of plane waves $\langle \omega |$:

$$\mathcal{F}[\phi(t)] = \langle \omega | \phi \rangle = \int_{-\infty}^{+\infty} dt e^{-i\omega t} \phi(t) \quad .$$

The wavelet transform (WT) can be defined in an analogous way as the projection of the state vector $|\phi\rangle$ (which belongs to $\mathcal{L}^2(\mathbb{R})$) onto a basis of $\langle \psi_{s,\tau} |$ functions:

$$\mathcal{W}[\phi(t)] = \frac{1}{\sqrt{s}} \langle \psi_{s,\tau} | \phi \rangle = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} dt \psi_{s,\tau}^*(t) \phi(t) \quad ; \quad (4.13)$$

the $\psi_{s,\tau}(t)$ function takes the name of “wavelet” and the range of t values where $\psi_{s,\tau}(t) \neq 0$ represents the sampling window of the transform. The wavelet depends on two real parameters s and τ whose meaning shall soon be explained, and these tune the span of the wavelet.

The wavelet $\psi_{s,\tau}(t)$ can be obtained in the following way

$$\psi_{s,\tau}(t) = \psi \left(\frac{t - \tau}{s} \right) \quad (4.14)$$

from a so-called “mother wavelet” $\psi(t)$, which also has to satisfy the two following relationships

$$\int_{-\infty}^{+\infty} dt \psi(t) = 0 \quad ,$$

$$\int_{-\infty}^{+\infty} d\omega \frac{|\mathcal{F}[\psi(t)]|^2}{\omega} < \infty \quad .$$

The former relation implies that $\psi(t)$ should really be a wave, with an oscillatory behaviour, whereas the latter implies that $\psi(t)$ has no power spectrum around $\omega = 0$.

It can be proved (see Ref. [C. 92], pag 61) that the time-interval sampled by the WT is

$$[\tau + st_c - s\Delta_t, \tau + st_c + s\Delta_t] \quad ,$$

where t_c is the time corresponding to the centre of the wavelet and Δ_t is its radius. In the same way can be proved that the corresponding frequency interval is

$$\left[\frac{\omega_c}{s} - \frac{1}{s}\Delta_s, \frac{\omega_c}{s} + \frac{1}{s}\Delta_s \right] \quad ,$$

where ω_c is the centre of $\mathcal{F}[\psi(t)]$ and Δ_ω its radius. It can be seen that

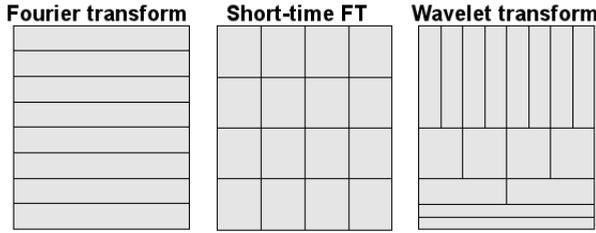


Figure 4.4: Schematic representation of the windows for the time sampling of Fourier, short time Fourier and wavelet transforms. On the x -axis is reported the time, on the y -axis the frequencies.

the time interval expands itself when s increases, whereas the contrary could be said for the frequency interval. Moreover, the τ parameter has the effect of translating the time window. Thus, while for the FT the time-sampling window spans all the time axis, for the WT this window has a rectangular shape in the time-frequency space and is tuned by the pair of s and τ parameters, as pictorially illustrated by fig. 4.4.

The sampling of the FT prevents a time localisation of the frequencies, whereas the sampling of the WT has an optimal time-frequency localisation capacity that, moreover, can be tuned changing some parameters to obtain the desired time-frequency accuracy. The time-frequency accuracy is, in fact, the result of a trade-off between time and frequency accuracy, respectively, due to a Heisenberg-like uncertainty theorem that stands true for WT too [C. 92]. Please notice that to extract high frequencies one uses a small time-window, which is intuitively reasonable since to sample high frequency oscillation just a small time-interval is needed; on the contrary, to extract low frequencies it is necessary to sample a larger time-interval to catch all the oscillations there. This a sort of “smart” sampling that short time Fourier transform just cannot do, since in that case the time-frequency sampling boxes are fixed once and for all (see fig 4.4, central panel).

Obviously, the results obtained with WT are very dependent on the choice of the functional form of the mother wavelet. One of the most employed mother wavelets is the so-called Morlet-Gabor (MG) function [CHT98], which is basically the product of a plane wave by a Gaussian

$$\psi(t) = \frac{1}{\pi^4} e^{+i\omega_0 t} e^{-t^2/2\sigma^2} \quad , \quad (4.15)$$

and whose plot is reported in fig. 4.5 along with those of its real and

imaginary parts.

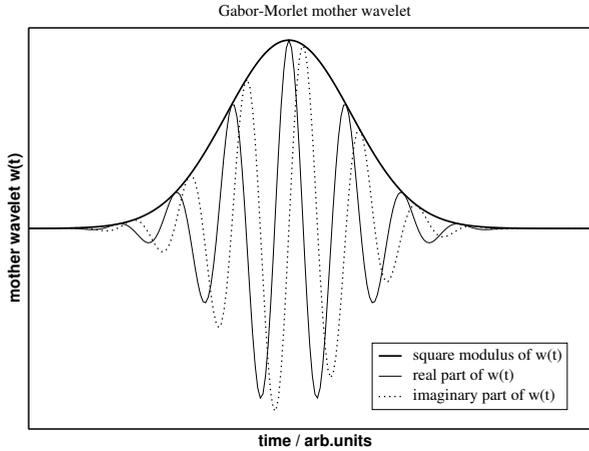


Figure 4.5: Morlet-Gabor mother wavelet.

In the literature other types of mother wavelet are also used: two of them are the Paul function of order m , defined as

$$\psi(t) = \frac{2^m i^m m!}{\sqrt{\pi(2m)!} (1 - it)^{-m+1}} \quad ,$$

and the DOG (“derivative of a Gaussian”) functions of order k

$$\psi(t) = \frac{(-1)^{m+1}}{\sqrt{\Gamma(m + 1/2)}} \frac{d^m}{dt^m} (e^{-t^2/2}) \quad .$$

Unfortunately, spectrograms obtained with those latter wavelets do not have a time-frequency resolution suitable to MD trajectory analysis [MMH04]: the Paul wavelet gives a spectrogram with poor accuracy in both time and frequency, the DOG wavelet gives discontinuous spectrograms that, moreover, are difficult to understand due to the presence of too many Fourier components. In fact, in Ref. [Kir05] it is shown that the MG wavelet best reproduces the Fourier spectrum. In this dissertation, as well as in References [RW05, MSC08a, MSC08b], only the MG function is adopted.

4.2.1 Explanations

The question is: how does the WT works? It is not so intuitive, in particular the meaning of all those parameters (s , τ , σ) appearing in the WT and into the wavelet function (ω_0). This Section shall try to answer this question.

In my opinion, the best way to understand the meaning of the s parameter is due to Meyers *et al.* [MKO93] that proposed to calculate the WT of a function $f(t)$ whose Fourier power spectrum was known, and searched for the value of s that maximises the wavelet power spectrum. Adopting the MG function as mother wavelet

$$\psi(t) \sim e^{i\omega_0 t} e^{-t^2/2\sigma^2} \quad ,$$

it is possible to prove that the s parameter and the λ Fourier wavelength are related by

$$s = \left[\frac{\omega_0 + \sqrt{(2 + \omega_0^2)}}{4\pi} \right] \lambda \quad ,$$

which can be expressed in term of frequency $\omega = \lambda^{-1}$ as

$$\omega = \frac{\omega_0 + \sqrt{(2 + \omega_0^2)}}{s4\pi} \quad . \quad (4.16)$$

This latter equation induces also the choice for the ω_0 real parameter: in fact, if we put $\omega_0 = 2\pi$ (as done in all wavelet computations of this dissertation), eqn. 4.16 reduces to the simple equation $\omega \simeq 1.01/s$. Therefore, it is a good approximation to think of $1/s$ as the actual frequency of the signal.

All this mathematical formulation can also be understood qualitatively. The s parameter stretches and shrink the mother wavelet $\psi(t)$ to generate the wavelet basis $\psi((t - \tau)/s)$. But, as already discussed about fig. 4.4, the value of s affects also the time-frequency interval used to sample the time-dependent signal, in such a way that lower frequencies are obtained when s increases, thus displaying the inverse relationship between these two quantities, as explained quantitatively by eqn. 4.16.

The meaning of the parameter ω_0 is also obvious due to eqn. 4.15: ω_0 is the main sampling frequency of the mother wavelet that acts as a window function. This means that WT only samples the ω_0 frequency? No, because the function is stretched both in time and frequency space

by s , as previously discussed, and the real sampling frequency of the “daughter” wavelets $\psi((t - \tau)/s)$ is given by eqn. 4.16.

Now also the meaning of the τ parameter should be clear: whereas s stretches and shrinks the time-frequency interval, τ translates in time the wavelets without affecting their shape. It is τ that allows obtaining a time-resolved spectrum; τ , in fact, has the same dimension of the time variable t .

The value of σ in the exponent of the MG wavelet (see eqn. 4.15) has a slightly more complex effect onto the spectrum. Basically, σ tunes the width of the Gaussian window of the MG function, and, thus, fixes the time-interval used to sample the signal. As has been already said, the s parameter shall stretch and shrink this interval, but starting from the default interval set by σ . Thus, a large value of sigma produces time-resolved spectra that have a very low time-resolution, albeit with optimal frequency accuracy. This is easily understandable, since the larger is the time-sampling interval, the more the spectrum obtained with WT becomes similar to that obtained with FT and the time-resolving capacity of the former is compromised. On the contrary, with smaller value for σ the time localisation capacity of the WT increases, but decreasing the frequency accuracy. Another way to picture this behaviour is the following: the MG wavelets obtained from eqn. 4.15 represent the sampling windows in the time-space; anyway, the sampling windows in frequency space are their Fourier transforms ⁴

$$\mathcal{F}[\psi(t)] = \hat{\psi}(\omega) \sim e^{-2\sigma^2(s\omega - \omega_0)^2} ; \quad (4.17)$$

thus, it is evident that the frequency interval spanned by $\hat{\psi}(\omega)$ decreases with the increase of σ , allowing a better frequency-sampling.

4.2.2 Some Examples

The WT can be used both to reproduce smoothed power spectra (this is one of the most employed applications of the WT in chemistry, see for example Refs. [KJ09, AWW⁺97, JDMN00, UP06, Ehr02]) and to obtain time-frequency spectrograms. Here I present a couple of examples and graphs to visualise both these possibilities provided by WT.

⁴Remember from Section 2.1.5: the FT of a Gaussian function is still a Gaussian function.

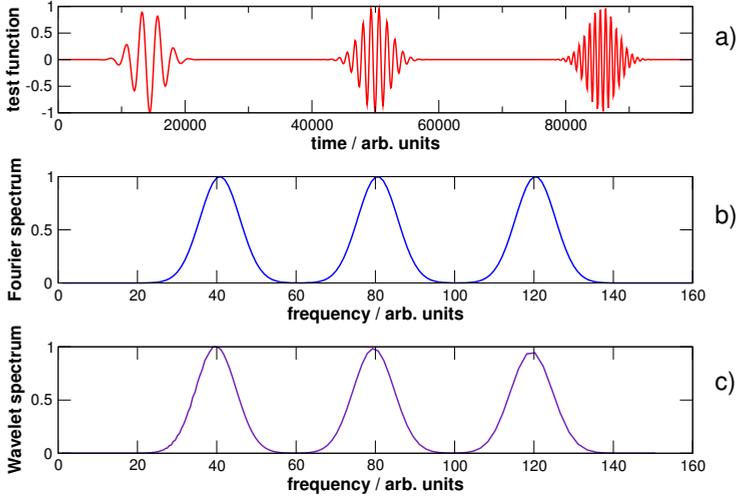


Figure 4.6: a) the test function defined in Eq. 4.18; b) its Fourier power spectrum; c) its wavelet spectrogram projected on the frequency axis (wavelet power spectrum). A value of 16 for the σ parameter has been adopted for a better frequency localisation.

The test function is a cosine oscillating at three frequencies, modulated by Gaussian windows:

$$\begin{aligned}
 f(t) = & \cos(3\omega t)e^{-(t-t_3)^2/\tau} + \\
 & \cos(2\omega t)e^{-(t-t_2)^2/\tau} + \\
 & \cos(\omega t)e^{-(t-t_1)^2/\tau} ,
 \end{aligned} \tag{4.18}$$

with $\tau = 10^4$, $\omega = 0.0025$ and $t_3 > t_2 > t_1$. In fig. 4.6 a comparison between Fourier and wavelet transforms of a test signal (a panel) is reported. The Fourier power spectrum and the wavelet spectrogram projected on the frequency axis (wavelet power spectrum) of $f(t)$ are reported in panel b and c of fig. 4.6, respectively. The wavelet transforms not only are able to reproduce with good accuracy the Fourier power spectrum, but they also give spectrograms like the one in fig. 4.7, localising the test signal in both time and frequency. This correlation plots is conceptually similar to sheet music (see fig. 4.1).

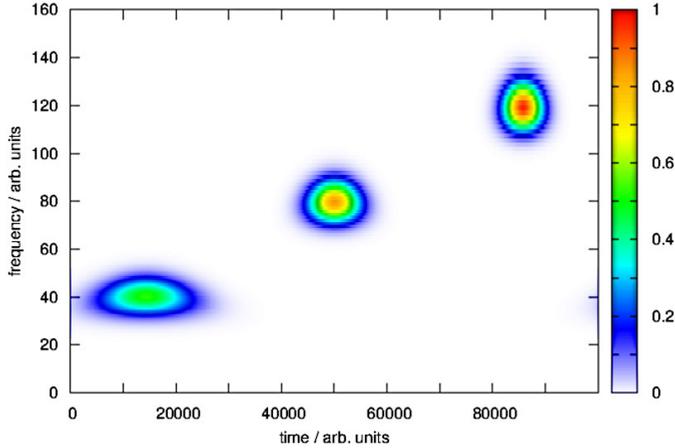


Figure 4.7: Wavelet spectrogram of the test function defined in Eq. 4.18. A value of 1 for the σ parameter has been adopted for a better time localisation.

In figures 4.6 and 4.7 I have used two different values for σ : a relatively huge value for the former, since in that case I wanted to reproduce the FT power spectrum, and a very small value for the latter, because I wanted to show the accuracy in time localisation of the three-frequency signal. To discuss the effects of σ , in fig. 4.8 the FFT of a signal coming from *ab initio* MD simulation (see Ref. [MMPCS11]) is compared with the WT at various values of the σ parameter, showing details of the band structure at increasing resolution. The WT appears to correctly reconstruct the Fourier spectrum with high values of parameter σ , showing that the truncation in the GM wavelet does not significantly affects the calculations. Moreover, it is also apparent why WT is frequently adopted for smoothing spectra [SLC03, Ehr02, SSZ⁺06]: for low values of σ , WT acts as a low-pass filter that removes high-frequency noise.⁵

For a test function like that of eqn. 4.18, the trend of the width at half-height (Γ) of its 1st frequency peak can be plotted with changing of σ . A roughly hyperbolic dependence of Γ on σ is obtained, as shown in fig. 4.9, meaning that the spectrum becomes more and more blurred with the decrease of σ . On the contrary, for a signal of this type (I remember that

⁵It has to be pointed out, however, that when WT is used only to smooth spectra, function such as the Haar wavelet [Haa10] and not the MG wavelet (whose primary use is time-frequency localisation) are usually employed as mother wavelets.

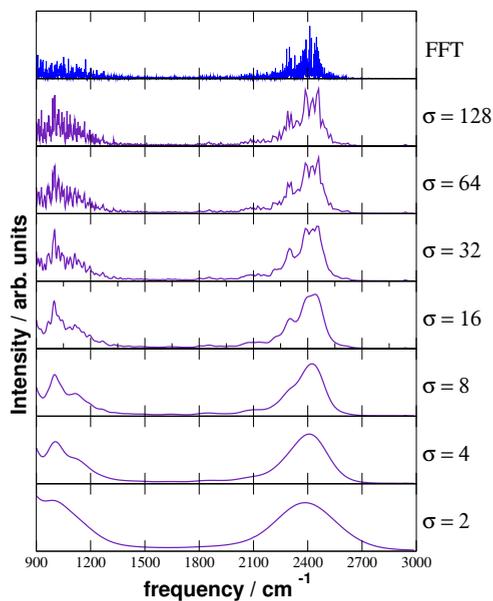


Figure 4.8: Fourier (blue) and wavelet power spectra (violet) of the displacement of inter-molecular $\Delta r_{O...D}(t)$ function of site #2 of ethylene glycol [MMPCS11] with changing of the value of σ .

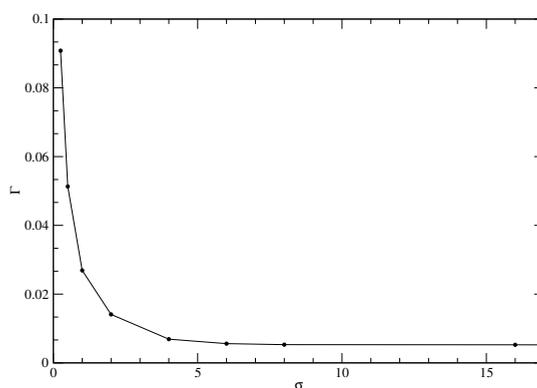


Figure 4.9: Spread of the wavelet spectrum as a function of the σ parameter.

this is an “artificial” signal, meaning that the analytic form for the input function is known and is reported in eqn. 4.18) the frequency-accuracy does not improve significantly for values of σ higher than ~ 10 .

4.2.3 Implementations

WT can be implemented in practice in at least two ways. There is a direct and straightforward approach, which basically consists in applying the following equation

$$\mathcal{W}[\phi(t)] = \frac{1}{\sqrt{s}} \langle \psi_{s,\tau} | \phi \rangle = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} dt \psi^* \left(\frac{t-\tau}{s} \right) \phi(t) \quad (4.19)$$

and calculate many time this integral for a lot of values of s and τ in order to map all the time and frequency space spanned by the input function $\phi(t)$ and its Fourier transform $\hat{\phi}(\omega)$, respectively. Otherwise, it is possible to calculate the bra(c)ket

$$\langle \psi_{s,\tau} | \phi \rangle$$

into the Fourier space. Since I have employed both methods in my research activity, I am going to briefly discuss them.

Direct Space Implementation

To uniform the notation to what can be found in literature, I adopt now the formalism of Torrence and Compo [TC98]. With this formalism, equations 4.13 and 4.19 can be rewritten as

$$\mathcal{W}(n,s) = s^{-1/2} \int_{-\infty}^{+\infty} dt' f(t') \psi_{n,s}^*(t') \quad , \quad (4.20)$$

where $\psi_{n,s}$ is the wavelet and n plays the role of the time-shifting parameter a in the previous equations. The discretised expression of eqn. 4.20 is (Ref. [TC98]):

$$\mathcal{W}(n,s) = s^{-1/2} \sum_{n'=0}^{N-1} f(n' \cdot \delta t) \psi^* \left[\frac{(n' - n) \cdot \delta t}{s} \right] \quad . \quad (4.21)$$

In eqn. 4.21, the product $n' \cdot \delta t$ represents the total time at the time-step # n' of the MD and localises the signal in time. Thus, the wavelet transform $\mathcal{W}(n,s)$ gives the frequency content of the signal $f(t)$ over a Gaussian time-window centred at $n \cdot \delta t$. Obviously, here n, n', N (the total number

of time-steps) are all natural numbers (hence the symbol change) and δt , being the MD time-step, has a finite non-zero value. The algorithm calculates the previous sum and looks for the value that maximises the modulus $|\mathcal{W}(n, s)|^2$ of the WT at a given time-step n' . The corresponding value of $1/s$ is taken as the “instantaneous frequency”. A cutoff of $\pm 3\sigma$ from the maximum of the GM wavelet function has been used to allow retaining more than 99.7% of the power spectrum energy and save computation time. The cutoff does not significantly affect the spectra [MMPCS11].

This “brute force approach” (*i.e.* just calculate the sum of eqn. 4.21) has the advantage of being relatively easy to understand and implement, but the resulting algorithm is very slow and memory consuming. Even adopting the $\pm 3\sigma$ cutoff for the time-sampling of the signal, it requires a lot of calculation. The main issue is given by the fact that variables (indexes in a Fortran90 program) n , n' and s have to be changed independently and the results stored into memory. Thus, the first version of this algorithm that I wrote was not able to run on a normal PC, if not for input function up to $N = 5 \cdot 10^4$ points (time-step). Since *ab initio* MD produces trajectories with hundreds of thousands time-snapshots, the code had to be parallelised ⁶ by adopting the MPI [GLS97] paradigm (and library of the same name) and run on a IBM power6 workstation. Typical calculations on this machine could last many days. This arrangement is less than satisfactory, since the computing time of the power6 workstation available at the Chemistry Dept. of the University of Florence is used for many other calculations.

Fourier Space Implementation

Having assessed that wavelets could be used to study *ab initio* MD trajectories [MMPCS11], I have developed ⁷ another code that computes WT in Fourier space and (partially) overcomes some of the shortcomings of the “brute force approach” described before.

⁶I acknowledge the invaluable help of Prof. Gianni Cardini (Chemistry Dept., University of Florence) for this task.

⁷I acknowledge the help of Dr. M. Pagliai (Chemistry Dept., Univ.of Florence) in developing my first version of the Fourier-space WT code. I also acknowledge Torrence and Compo [TC98] for providing an old Fortran77 code that I have used as a reference to develop my own. The code of Torrence and Compo is available free of charge via the internet at <http://paos.colorado.edu/research/wavelets/software.html>.

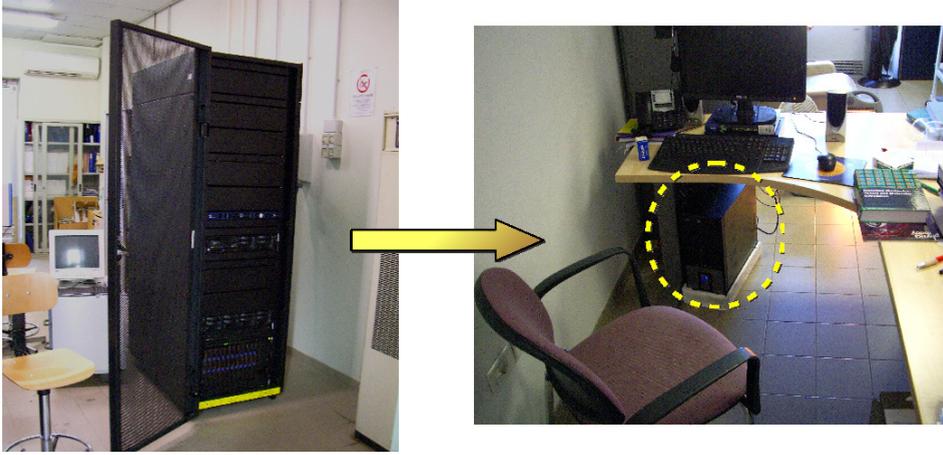


Figure 4.10: On the left, the IBM power6 workstation used to run the direct-space WT code is displayed: it has 12 so-called “blades” (*i.e.* computing node), each of them equipped with 4 cores and 8 GB of RAM. On the right there is a photo of the Athlon II-X4 PC (equipped with 4 GB of RAM) that I currently use to run the Fourier-space implementation of the code.

The algorithm adopted in subsequent research activity computes the WT of eqn.4.21 in frequency space as

$$\mathcal{W}_n(s) = s^{1/2} \sum_{k=0}^{N-1} \hat{f}_k \cdot \hat{\psi}^*(s\omega_k) e^{i\omega_k n \delta t} \quad , \quad (4.22)$$

where k is the frequency index, ω_k is the angular frequency and \hat{f}_k and $\hat{\psi}$ are the Fourier transforms of the time series f_n and of the adopted mother wavelet ψ , respectively.

Eqn. 4.22 can be better understood if put in it continuous form

$$\mathcal{W}_n(s) = s^{1/2} \int_{-\infty}^{+\infty} d\omega e^{+i\omega t} \hat{f}(\omega) \hat{\psi}_{n,s}^*(\omega) \quad ,$$

which can be put in a more succinct way as

$$\mathcal{W}_n(s) = s^{1/2} \mathcal{F}^{-1} \left[\hat{f}(\omega) \cdot \hat{\psi}_{n,s}^*(\omega) \right] \quad ; \quad (4.23)$$

this latter equation means that the WT of the signal $f(t)$ can be obtained computing the inverse FT of the product of the FT of the complex conjugate of wavelet $\psi_{n,s}^*(t)$ and of the signal $f(t)$ itself. In fact, defining the

convolution product of two function $a(t)$ and $b(t)$ as

$$a(t) * b(t) = \int_{-\infty}^{+\infty} dt' a(\eta) b(t - \eta) \quad , \quad (4.24)$$

and remembering the convolution-Parseval theorem of the Fourier analysis

$$\mathcal{F}[a * b] = \mathcal{F}[a] \times \mathcal{F}[b] \quad ,$$

we immediately obtain that eqn. 4.24 can be explicitly expressed using Fourier transforms and their inverse:

$$a(t) * b(t) = \mathcal{F}^{-1} [\mathcal{F}[f] \times \mathcal{F}[g]] \quad . \quad (4.25)$$

Following definition 4.24, it is evident that WT (eqn. 4.20) can be written as a convolution product

$$\mathcal{W}_n(s) = s^{-1/2} f(t) * \psi^*(t/s)$$

if we put $\eta = n/s$. Deriving eqn. 4.23 from the property of convolution product expressed by eqn. 4.25 is therefore straightforward.

The algorithm based upon eqn. 4.22 is a little more difficult to be understood, but it is much more faster and computationally feasible than the “brute force approach”. This is mainly due to two reasons. The first one is that in the Fortran90 code eqn. 4.22 can be implemented using just two variables (indexes) n and s , since the ω integration is demanded to the FFT routines. This saves both memory and computational time. The second one is that FFT routines, as already hinted, are nowadays truly fast (as their name implies) and computationally optimised.

II

Case Studies

In the following Part, details of simulation and analysis of hydrogen-bonded systems are reported. These are research case studies, and I have contributed to them at various levels. Here I present a summary of my contributions, explicitly stating what my contribution actually was and what was done by others.

Regarding the study of the dipeptide dynamics in a phospholipid membrane (Chapter 5), I mainly devoted myself to analyse the classical molecular dynamics simulation to extract structural properties and correlation function to probe the H-bonding behaviour. I did not perform the molecular dynamics simulation nor the experiments. I have performed the analysis of the simulated data, which has also been used to check the reliability of the simulation protocol, and this induced in turn to perform new simulations to rule out some possible sources of errors. I wrote part of the supplementary information of Ref. [VCMMR11] (mainly regarding the hydrogen-bonding dynamics), as well as creating most of the figures for the main paper.

I performed the calculation and analysis of data regarding propane-1,2-diol and ethylene glycol in water (Chapter 6), and I developed the original direct space wavelet code to study their dynamics. Subsequently, I decided to study the propane-1,2-diol dynamics in acetonitrile for a more straightforward comparison with experimental data. The research regarding propane-1,2-diol in acetonitrile was designed and performed by me, with the exception of the experimental work. I also wrote the draft of the articles [MMPCS11, MMPCR12b].

In the study about methyl acetate in water and methanol (Chapter 7), I performed the wavelet analysis of the CPMD trajectory. I did not perform the main CPMD simulation. I developed the computational scheme adopted to correlate frequency with structural data. I also wrote some parts of the papers (Ref.s [PMMC⁺10, PMMC⁺11]), in particular those

regarding the adopted wavelet protocols. This research was partially designed by me.

Regarding the behaviour of thiazole in water (Chapter 8) I performed wavelet analysis on the simulated data. I did not perform the main CPMD simulation. I performed *ab initio* static calculations to obtain information about charge transfer occurring during the simulated run. I wrote part of the paper (Ref. [MMPMMS11]) and large parts of the supplementary information, as well as creating most of the figures regarding the analysis of data.

Finally, I decided to investigate the highly unusual behaviour of one of the few example of a stable bifurcated H-bond in lithium nitrate trihydrate (Chapter 9). I performed the calculations and the analysis of data regarding this system. The research was designed by me and I also wrote the draft of the corresponding article (Ref. [MMPCR12a]).

Part of this thesis include research work already published in the previously mentioned papers [VCMR11,MMPCS11,MMPCR12b,PMMC⁺10,PMMC⁺11,MMPMMS11,MMPCR12a]. In particular, some paragraphs of the following Part of this dissertation, as well as most of the subsequent figures, are taken from the drafts of the aforementioned papers of which I am either first author or co-author.

5. Anchor Dipeptide into Phospholipid Bilayer

Linear and non-linear spectroscopic methods with classical MD simulations and (static) quantum mechanical calculations have been combined to probe the structural features of a simple dipeptide, namely N-myristoylated-methyl-glycine

(MrG), interacting with a phospholipid membrane fragments (namely 1-palmitoyl-2-linoleyl phosphatidylcholine, a.k.a. PLPC); both of them are schematically represented in fig. 5.1. This artificial dipeptide has been

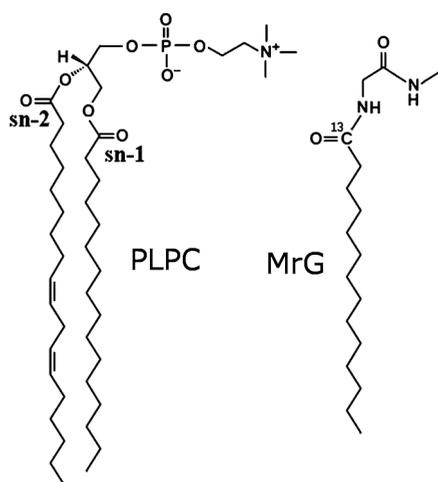


Figure 5.1: Schematic representation of PLPC phospholipid and MrG dipeptide molecules. MrG has been marked using ^{13}C in the shown position. Two C=O groups of PLPC are labelled as sn-1 and sn-2.

designed to mimic the natural anchor that can be found in a variety of membrane associated polypeptides [JO90].

With its myristoyl tail, MrG represents a typical example of proteins that associate by means of hydrophobic tails (the anchors) attached to certain conservative structural segments of the polypeptide [JO90]. In fact, anchoring often provides an opportunity for a stable, specific and (rela-

tively) controllable form of interaction of the polypeptide with the membrane surface (of certain composition and fluidity) on the time scale of cellular metabolism. Thus, a deeper understanding of the details of this peptide-membrane interaction at the molecular level, beside the general relevance of learning about the effective anchoring of proteins to phospholipid bilayers, recently became an important step in the realisation of novel anticancer remedies [dip06]. As a matter of fact, in the continuous search for new clinical treatments for cancer, small polypeptides can be employed as subjects in drug design studies [ACAR96]. Even nowadays, the selection process of substances for medical applications continues to be based essentially on clinical evaluations and tests that are mostly phenomenological in character. Moreover, the knowledge of their structural properties is often very limited, even for certified drugs of this type admitted to clinical practice, because, when injected into blood plasma, the molecule experiences a variety of different environments and can even be subject to partitioning into the phospholipid membrane.

Optical spectroscopy allows identifying the structural properties of molecules in these latter environments. Other experimental tools, such as calorimetric and electrochemical techniques, cannot to help in the identification of the relative arrangement of molecular moieties on a local scale, and X-ray and electron scattering experiments are not suitable to probe disordered system like the membrane samples in physiologic conditions. In fact, in membrane samples even NMR spectra are heavily broadened as a result of the slow molecular motions on the time scale of the sampling processes.

Conversely, methods based on nonlinear optical effects make it possible to probe both the intra- and inter-molecular structural correlations between chromophores, including their time-evolution on a subpicosecond time scale. As a matter of fact, third order nonlinear experiments in the IR spectral region [HLH98] allow probing certain components of the molecular susceptibility and, hence, the related structural properties on the time scale of the atomic motions. 2D-IR spectroscopy has been recently applied to characterise phospholipid membranes [VCZ⁺07] and to perform structural and dynamical analysis of several polypeptides (including those of potential anticancer activity) associated with phospholipid bilayers [VR09]. Nevertheless, the resolving power of 2D-IR spectroscopies can be limited, in some cases, by persisting ambiguities possibly due to the coexistence of several molecular conformations [WPH⁺02].

Thus, MD simulations can be proficiently used to explain and elucidate the fine features of the 2D-IR spectra [WH00, WPH⁺02].

In a previous paper [VR09], Volkov and Righini used two-colour 2D-IR spectroscopy to determine the depth of penetration of the MrG backbone into the polar fraction of a phospholipid bilayer made of 1-palmitoyl-2-linoleyl phosphatidylcholine (PLPC) molecules. The experiment consisted in exciting the water stretching modes and detecting the induced perturbations in the spectral region of the C=O vibrations.

In Ref. [VCMMR11] we specifically tried to obtain information about the MrG backbone conformation and its localisation in the PLPC bilayer with the essential support of (classical) MD simulations and (static) quantum mechanical calculations. With these combined techniques we had a sufficiently exhaustive picture of the structure of the guest molecule (MrG) and of its interaction with the neighbouring moieties and of the local restructuring of the host bilayer (PLPC).

5.1 Computational Protocols

Molecular Dynamics Simulation

In order to simulate a system composed of ~ 20000 atoms, classical MD has to be employed. The simulations have been carried out using the ORAC code [MSC⁺10], because of the implemented replica exchange method (REM) that speed up phase space sampling (see Appendix B.2). In order to simulate the behaviour of peptides and PLPC membranes near realistic biological conditions, MD simulations are performed in the isothermal-isobaric NPT ensemble using an orthorhombic simulation box with PBC. Temperature was set at 298 K with a Nosé-Hoover thermostat [Hoo85], and pressure at 0.1 MPa, adopting the Parrinello-Rahman Lagrangian [PR80] as implemented in ORAC [MP98] with fixed cell angles. Lorentz-Berthelot mixing rules (see Section 2.2.5) have been used to model Lennard-Jones interactions between different atom types. The AMBER99 [HAO⁺06] force field has been adopted. The phospholipids, dipeptides, and water molecules are completely flexible except for the covalent bonds involving hydrogen atoms, which are kept rigid to save computations. The equations of motion have been integrated using the multiple time-step technique [TBM92] and the Ewald method with the smooth particle mesh algorithm [EPB⁺95] is applied to compute the electrostatic interactions. The simulated sample consists of 2877 water molecules and of two layers,

each made of 63 PLPC molecules and one MrG molecule.

The sample has been prepared in two steps.

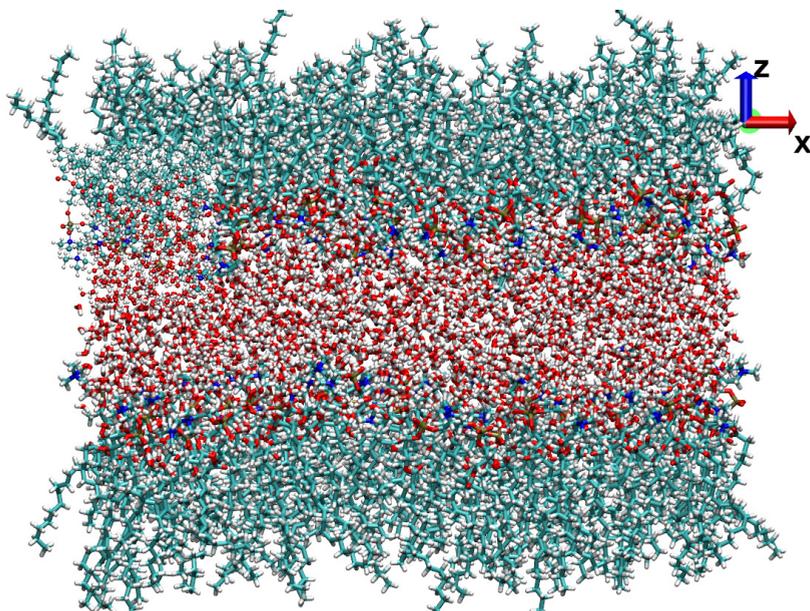


Figure 5.2: MD Simulation box. Cartesian axes are displayed on the upper-right corner.

In the first one an equilibrated sample of a pure PLPC bilayer has been produced, starting from a previously equilibrated configuration¹ of a hydrated PLPC bilayer made of 128 phospholipid molecules and 2460 water molecules. The box size of this configuration is a parallelepiped of volume $79.666 \times 54.478 \times 55.065 \text{ \AA}^3$.

To avoid possible inconsistencies due to unexpected peptide arrangement arising from periodic boundary conditions, the box dimension along the bilayer normal (hereafter, the *z*-axis) is increased by 3.5 \AA by adding 417 water molecules (to get a total of 2877 water molecules). The resulting hydrated PLPC bilayer is then equilibrated once again through a standard 5 nanoseconds MD simulation. Equilibration is verified by monitoring the potential energy, the simulation-box volume, and the stress tensor, which

¹The “equilibrated configuration” actually is the final configuration of a MD simulation performed by Herrenbauer *et al.* [HTP95], available at <http://moose.bio.ucalgary.ca>.

is found to be almost isotropic.

Then, Dr. R. Chelli (Chemistry Dept., Univ of Florence) built the initial configuration of the MrG-PLPC sample by replacing two PLPC molecules, randomly chosen on the two leaflets of the pure PLPC sample, with MrG molecules. The replacement of PLPC molecules with MrG molecules does not introduce relevant frustration in the system, because of the similarity between these two types of molecules. System equilibration is then achieved with a 2 nanosecond NPT MD simulation.

With this procedure the PLPC molecules can be assumed equilibrated, while the dipeptide molecules may still preserve memory of their initial configuration, due to their rugged energy landscape possessing several local minima where the peptide can be trapped during the simulation.

Therefore, in order to accelerate the conformational sampling of the dipeptide molecules, also considering the small number of MrG molecules, we employed the parallel tempering technique, whose MD version is known as “replica exchange method” (REM) [SO99]. The specific version implemented in ORAC and adopted for these simulations is the hamiltonian-REM, discussed in Appendix B.2. In these simulations, the hamiltonian-REM scaling is applied only to the intra-molecular potential energy of both MrG molecules, apart from the harmonic stretching and bending potential terms. 16 replicas as been adopted, swapping among the following sequence of states: $c = 1.000, 0.898, 0.807, 0.725, 0.651, 0.585, 0.525, 0.472, 0.424, 0.381, 0.342, 0.307, 0.276, 0.248, 0.223, \text{ and } 0.200$. Replica exchanges are attempted every 8 femtoseconds. The replica configurations characterised by $c = 1$ have unscaled potential energy and therefore correspond to configurations sampled at 298 K, *i.e.* room temperature. Instead, the replica configurations characterised by, *e.g.*, $c = 0.2$, correspond to states where the intra-molecular potential energy of the MrG molecules (excluding harmonic stretching and bending terms) have a “virtual” temperature of $T/c = 298/0.2 = 1490$ K. Since each replica can move through the c ensemble, the MrG molecules can easily overcome the internal energy barriers with significant improvement of the conformational sampling. Before performing the production REM run, an equilibration of 1.68 nanosecond is carried out. Atomic configurations are recorded every 0.4 picosecond during a production run of 5.4 nanosecond. The statistical analysis is done on the $c = 1$ system configurations. The Ramachandran plots I have calculated for $c = 0.2$ configurations (*i.e.*, at high virtual temperature) show that the sampling involves a much broader conformational

space for both MrG molecules, as shown in fig. 5.3.

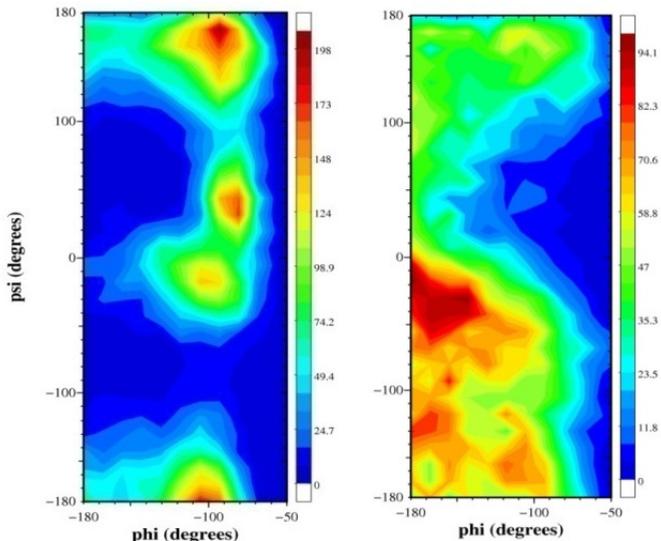


Figure 5.3: Left Panel: Ramachandran plot calculated using MrG configurations taken from replicas corresponding to $c = 1$ (the virtual temperature of the torsional and inter-molecular potential energies of MrG is $298/c = 298\text{K}$). Right panel: Ramachandran plot calculated using MrG configurations taken from replicas corresponding to $c = 0.2$ (the virtual temperature of the torsional and inter-molecular potential energies of MrG is $298/c = 1490\text{K}$). In this case the sampled conformational space is much wider.

I have used the distribution of Ramachandran Φ and Ψ angles² to check qualitatively the effectiveness of the MD parameters and of the equilibration protocol. Unfortunately, an entire hamiltonian-REM MD trajectory had to be discarded due to the evident non-equilibrium behaviour of the internal dynamics of the two dipeptide molecules (see Section 5.3.2 for more details). In fact, the dynamics is very dependent on the choice of the partial charges for the electrostatic potential, and the original protocol adopted to obtain these latter was unsuitable. The dipeptide was optimised with the Gaussian09 [FTS⁺b] program at the B3LYP/6-31++G(d,p) level of theory to obtain partial ESP charges, but the possibility of electronic rearrangement due to internal torsion was not considered by this

²The Φ angle represents torsions around the N(H)–C(O) bond, while Ψ represents torsions around the C(O)–C $_{\alpha}$ bond.

preliminary and inaccurate approach, and only the “folded” conformation reported in fig. 5.4 was found in the trajectory, thus invalidating the first hamiltonian-REM MD simulation. At a second attempt, partial charges

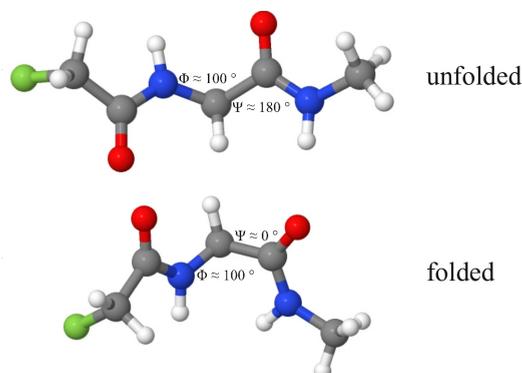


Figure 5.4: Two possible conformations of the dipeptide with Ramachandran torsional angles reported. Green atoms are used to represent the aliphatic chain.

obtained optimising the two constituent amino acids separately have been employed. Therefore, a second run of MD simulations has been carried out to achieve a correct equilibration and behaviour in the Ramachandran space.

Ab Initio Calculations

The quantum mechanical calculations are based on the density functional theory, as implemented in the Gaussian09 [FTS⁺b] program using the B3LYP functional and the 6-31++G(d,p) basis set. These calculations are due to Dr. V Volkov (then at LENS, now in Mainz, Max Planck Institut für Polymerforschung) To make these computations feasible, the myristoyl tail of the dipeptide was replaced with a methyl. The molecular conformation of the peptide backbone has been represented in terms of the Ramachandran Φ and Ψ angles ranging from $(-180^\circ, -180^\circ)$ to $(-100^\circ, 180^\circ)$, with steps of 20° for both angles. The structures of these configurations are optimised by keeping the Φ and Ψ angles fixed. For all optimised configurations, the coupling constants, the angle between the transition dipole moments, and the ratio between the IR intensities of the two amide I modes have been computed. These quantities are then given as functions of the Φ and Ψ angles within the Ramachandran space. To be specific, the

values of the coupling constants between interacting C=O oscillators are obtained from *ab initio* calculations by reconstructing the Hessian matrix for the normal modes in the subspace of the amide I modes [TT98]. The method described in Ref. [CHC03], based on the Cartesian displacements of the normal modes, has been adopted to retrieve the coupling.

5.2 Brief Survey of Experimental Methods

Fig. 5.5 represents the linear IR and nonlinear 2D-IR spectra of the sample in the spectral range of the C=O stretching modes, recorded by Dr. Volkov.

The 2D-IR spectra, in parallel and perpendicular polarisation, are recorded at 500 femtoseconds pump-probe delay time. The band centred at 1734 cm^{-1} is due to the vibration of the C=O groups of PLPC. The resonances of the ^{13}C -labelled and of the native C=O of MrG are clearly separated in frequency (1587 cm^{-1} and at 1632 cm^{-1} , respectively). The latter two bands are much narrower than the PLPC resonance, whose full width at half-maximum (FWHM) is about 40 cm^{-1} . For a similar phospholipid bilayer (dimyristoylphosphatidylcholine) [VCZ⁺07], the C=O stretching bandwidth was reduced by approximately 25% upon removal of the excitonic coupling by isotopic substitution. The 2D-IR spectrum shows a well-defined cross-peak between the two amide I bands of the dipeptide (pump centred at 1598 cm^{-1} , probe at 1632 cm^{-1}). In Ref. [VR09], Volkov and Righini showed that the MrG molecule is embedded in the polar layer of the membrane. The resulting inter-molecular host-guest coupling is confirmed by the appearance of cross-peaks between the two amide I transitions and the PLPC C=O stretching band (pump at 1598 and 1632 cm^{-1} , probe at 1740 cm^{-1}). The presence of rather intense intra- and inter-molecular cross-peaks allows exploring structural properties of MrG anchored to the membrane.

5.3 Findings

5.3.1 Experimental

The (almost) Lorentzian lineshapes of the two amide I bands in the FTIR spectrum correspond to the essentially vertical alignment of the same two resonances in the 2D-IR spectrum of fig. 5.5. No detectable inhomogeneous character shows up in the linear and nonlinear spectra. Actually,

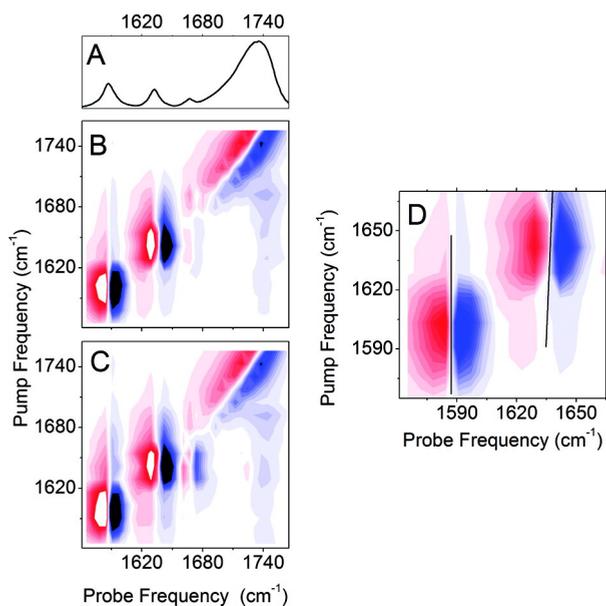


Figure 5.5: (A) IR spectrum of MrG in PLPC membrane. (B,C) Nonlinear 2D-IR spectra of MrG in PLPC membrane measured in parallel and perpendicular pump-probe polarisations, respectively. (D) Zoomed view of the amide I resonances of panel B.

in the analysis of the 2D-IR lineshape, one has to take into account the limitations due to the pump bandwidth (17 cm^{-1}) and to the actual spectral resolution of the detection system (5 cm^{-1}). On this basis, it cannot be excluded that some small inhomogeneous contribution, on the order of very few wavenumbers, is present, resulting in a hardly detectable tilt toward the diagonal of the nodal line separating the negative and positive lobes of the 2D-IR resonances. An additional element to be considered is the relatively long pump pulse used in the experiment. The 17 cm^{-1} bandwidth corresponds to a pulse duration of about 800 femtoseconds. With this low time resolution, a very fast (typically, < 300 femtoseconds) loss of memory of the vibrational excitation would lead to smearing out any trace of inhomogeneity, even at very short pump-probe delay times. In fig. 5.5, panel D, the lines represent the orientation of the nodal lines for the two amide I bands. Within the experimental sensitivity, the lines are essentially vertical. Under this respect, it is striking the difference with the spectral properties of, for instance, trialanine in water [WPH⁺02]. In fact, in that case, a partial inhomogeneous character is observed, attributed to the co-presence of different intra-molecular conformational states of trialanine. The essentially homogeneous character of the amide I resonances of MrG suggests that this dipeptide is present in the phospholipid environment with one specific, largely prevalent conformation. The different behaviours of MrG in the membrane and of trialanine in water is at first sight contra-intuitive. In fact, the smaller size of the two glycine groups of MrG should in principle give rise to a shallower backbone torsional potential and, consequently, to an easier access to different conformations. The very role of the anisotropic environment provided by phospholipid matrix and, possibly, the constraint due to the anchor tail, then come into play in “freezing” the conformational mobility of MrG. In the following, MD simulations and quantum mechanical calculations provide a substantial support to this interpretation.

5.3.2 *Ab Initio* Calculations and Molecular Dynamics

The coupling constant between the amide I modes of the MrG, obtained from the relative intensities of the diagonal resonances and the cross-peaks, is $5 \pm 1\text{ cm}^{-1}$ [HLDH99]. From the cross-peak anisotropy, following the method proposed by Hochstrasser *et al.* [HLDH99], a value of 43 (or $180-43$) $\pm 8^\circ$ for the angle between the transition dipole moments of ^{13}C -labelled and native amide I modes can be obtained, as reported in fig. 5.6.

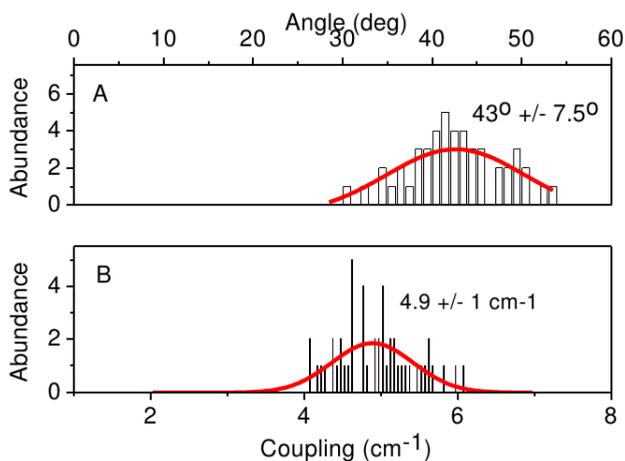


Figure 5.6: Panel A: distribution of the angle between the amide I transition dipole moments of MrG obtained from a set of 48 non-linear spectra obtained with different pump pulse frequencies and for 4 different samples. Panel B: distribution of the coupling constant between the amide I modes of MrG estimated from the same experiments.

The confidence limits for these quantities were derived from a statistical analysis of a set of 48 nonlinear spectra obtained with different pump pulse frequencies and for four different samples. Then, by comparing the experimentally derived data with the calculated distributions in the Ramachandran Φ and Ψ space, the compatible angular regions have been extracted.

The relative IR intensity of the two amide I bands is the additional property that has been used to restrict the range of possible backbone conformations. According to our *ab initio* calculated IR spectra, the intensity ratio of the two transitions changes with the Φ and Ψ angles. In the experimental IR spectrum, the intensity ratio of the ^{13}C and ^{12}C amide I bands is $I_{13}/I_{12} = 1.78$.

Note, however, that the calculations are done for a molecule in vacuum, since for this research *ab initio* MD trajectories have not been collected due to the complexity of the system. Considering that the effect of the membrane environment on the IR intensities may be not negligible, we accept all the conformations that satisfy the condition $I_{13}/I_{12} > 1$. Fig. 5.7 gives the calculated Φ/Ψ maps (panels A, B, and C) and the regions corresponding to compatible values of the coupling constant, of the an-

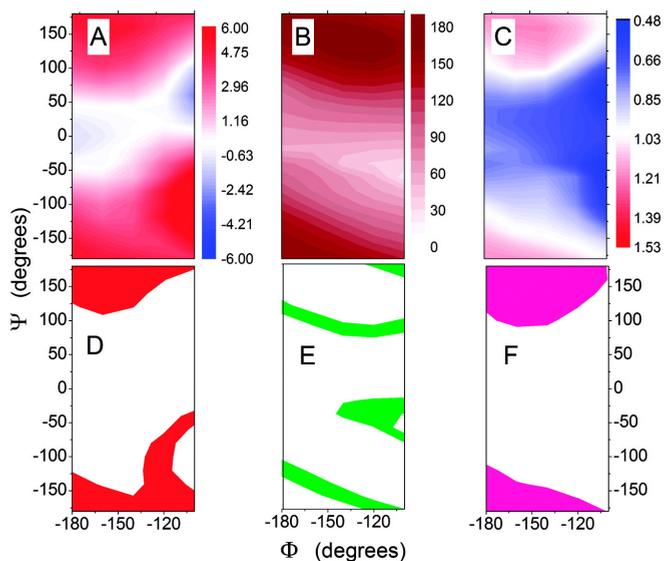


Figure 5.7: (A-C) Coupling constant, angle between the transition dipole moments, and IR intensity ratio between ^{13}C and ^{12}C amide I modes of MrG calculated *ab initio* as a function of the Φ and Ψ angles. (D-F) The coloured areas are the Φ/Ψ regions compatible with the experimental values of the coupling constant, the angle between the transition dipole moments, and the IR intensity ratio, respectively.

gle between transition dipole moments and of the intensity ratio (panels D, E, and F). Fig. 5.8, panel A, represents the result of the combination of the three constraints on the structural maps. The acceptable region is very limited and corresponds to a distribution of structures of essentially unfolded character, with Φ and Ψ angles varying in the degree range $[-100, -180]$ and $[-190, -120]$, respectively.

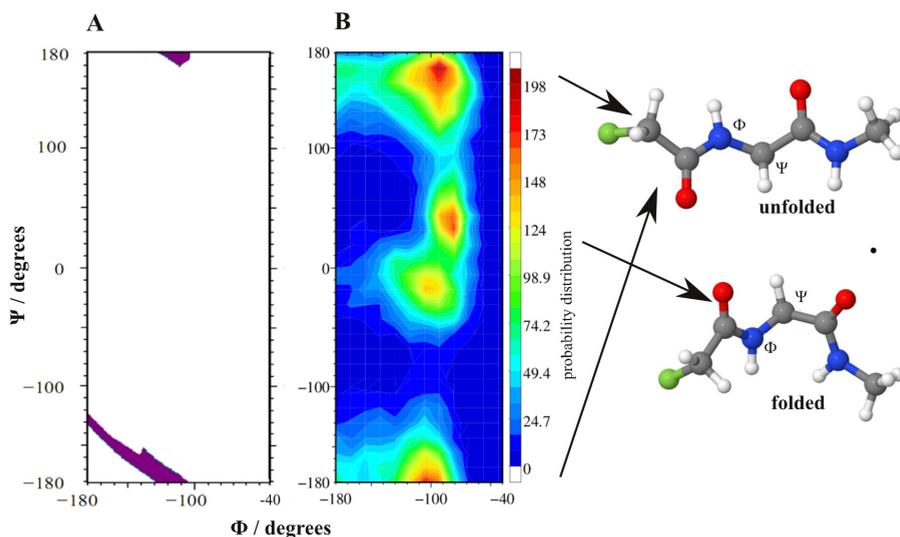


Figure 5.8: (A) Region of the Φ/Ψ space obtained from the intersection of the D, E, and F plots of fig. 5.7. (B) Probability distribution of the Φ and Ψ angles obtained from the REM-MD simulation. The two molecular structures on the right are representations of the two main backbone conformations (the green sphere represents the alkyl chain).

Fig. 5.8, panel B, shows the conformational distribution of MrG in terms of the Φ and Ψ angles according to the REM-MD simulation.

The Φ/Ψ distribution predicted by the simulation is broader. We have singled out two major structural configurations. The first one is essentially unfolded, with Φ and Ψ angles around $-90 \pm 20^\circ$ and $-180 \pm 20^\circ$, respectively. It is compatible with the values of the dihedral angles obtained from the experiments (fig. 5.8, panel A). The second structural motif is close to a $310/\alpha$ folded structure, with Φ and Ψ angles around $-95 \pm 20^\circ$ and $+10 \pm 50^\circ$, respectively. The unfolded structure has the higher probability (64%), compared to the folded one (36%). Representative structures of the two main conformers are given in fig. 5.8.

Dr. Volkov and Prof. Righini already noticed that both linear and 2D-

IR spectra in the amide I region are characterised by very low, practically undetectable, inhomogeneity.

In order to verify whether the conformational distribution predicted by the REMD simulation is compatible with the experimental results, we calculated *ab initio* the IR spectra of various folded and unfolded conformations of the dipeptide (using the Gaussian 09 code [FTS⁺b]). More precisely, we calculated the IR frequencies and intensities for a grid of points in the Φ/Ψ space (separated by 20°) according to the distribution of figure 5.8. Thus, each point in the Φ/Ψ grid contributes to the overall band profile with a Lorentzian line centred at the calculated frequency, with an intensity proportional to the product between the *ab initio* calculated intensity and the probability of the corresponding Φ/Ψ conformational state estimated in the REMD simulation. The spectral profiles of the folded and unfolded structures are then obtained as superposition of the relevant contributions. The calculated frequencies are scaled by 0.93 in order to fit the experimental spectrum.

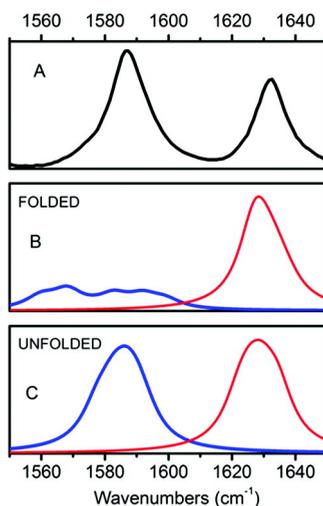


Figure 5.9: (A) IR spectrum of MrG in PLPC membrane (zoomed view of fig 5.5, panel A). (B) IR spectrum of the folded MrG conformations estimated with *ab initio* calculations. (C) IR spectrum of the unfolded MrG conformations estimated with *ab initio* calculations.

In fig. 5.9, panels B and C, the reconstructed spectra for the folded and unfolded conformations of MrG, respectively, are reported. The comparison of the two calculated spectral profiles shows that, in the case of the $310/\alpha$ folded structure, the two components have different intensities and,

most importantly, the low frequency band is substantially broader. This latter feature is a consequence of the much higher value of the vibrational coupling between the two C=O in the folded geometry. This calculated spectrum is inconsistent with the experimental profile.

On the contrary, the spectrum calculated for the unfolded structures (fig. 5.9, panel C) agrees well with the experiment. Thus, the REM-MD simulation agrees acceptably with the experiments in predicting the unfolded backbone conformation as the most probable one. An incorrect choice of the electrostatic potential charges forced us to reject entirely the results of a previous simulation because it just predicted the presence of the folded conformation.

The presence of about 36% peptide molecules with folded backbone conformation appears instead to be overestimated by the simulation. The profiles of the linear IR and 2D-IR spectra indicate that the abundance of this conformation should be much less. The weak shoulders that appear at the bases of the two main bands are compatible with an abundance of other conformers on the order of 10-15%. In evaluating this discrepancy, one has to consider that the actual conformational distribution is the result of the subtle competition of the intra-molecular (basically torsional) potential and of the inter-molecular peptide-phospholipid interactions. In our simulation we have chosen to adopt the AMBER-99 potential [HAO⁺06], developed and widely employed for peptides. In principle, it would be possible to remodel the intra-molecular potential of MrG to account for the conformational distribution estimated from the experiments. Anyway, the required computational effort would be well beyond the aim of this research, as already hinted in Section 2.2.7 about polarisable force fields.

The calculated spectrum of fig. 5.9, panel C, also provides additional arguments for interpreting the unusually narrow line-shapes of the two amide I transitions reported in fig. 5.5 (also reported and magnified in fig. 5.9, panel A). As already noticed, the positive and negative lobes of the 2D-IR bands are essentially vertical also at very early times, although the presence of a very small tilt toward the diagonal direction cannot be completely excluded, at least for the high frequency band. The bandwidth is about 14 cm^{-1} , in agreement with the value measured in the linear spectrum. On the other hand, fig. 5.9, panel C, shows that the calculated IR band-shapes, resulting from the frequency distribution of the unfolded conformation, are in qualitative agreement with the experiment, suggesting that the leading term determining the overall band shape is the (rather

narrow) distribution of the backbone dihedral angles of the unfolded conformers, with only a minor contribution from the environment.

The bandwidths of the calculated spectrum of fig. 5.9, panel C, are, however, broader (about 25 cm^{-1}) than the experimental ones. This discrepancy suggests that some motional narrowing effect has to be considered. With the time resolution of the experiment, very fast (say 300 femtoseconds, or less) spectral diffusion would make a limited inhomogeneous contribution (about 10 cm^{-1} , in this case) practically undetectable. In the assumption that only the unfolded conformation of MrG is appreciably populated, two main processes can provide the required fast frequency modulation: the breaking and making of H-bonds between the C=O groups and the water molecules present at the membrane interface, and the rapid small fluctuations of the dipeptide backbone around its equilibrium unfolded structure.

To check these possibilities, we have relied upon MD data analysis.

To this aim we have calculated several structural properties, especially to get information about the mutual arrangement of MrG and PLPC molecules in the membrane scaffold. In figure 5.10, panel A, I report the normalised distribution functions of the carbonyl-oxygen atoms and of the chain end-carbon atoms of MrG and PLPC along the z-axis. Overall, the observable features are consistent with the formation of a bilayer membrane where the hydrocarbon tails of MrG and PLPC are globally localised between the hydrophilic moieties represented by the carbonyl oxygen atoms. Moreover it is worthwhile to observe that, in spite of the significant structural differences of MrG and PLPC molecules, their distribution functions are similar in both, position of the maximum and broadening.

Additional insights are obtained from the distributions of the angle formed by the polar-head backbones and by the hydrocarbon tails with the z- axis (normal to the bilayer). We arbitrarily, but reasonably, defined the direction of the polar-head backbone of MrG as the vector linking the carbonyl- carbon atom next to the myristoyl tail to the other carbonyl-carbon atom. By analogy we defined the direction of the PLPC polar head as the vector linking the sn2 carbonyl-carbon atom to the sn1 carbonyl-carbon atom. The direction of the hydrocarbon tails is instead defined by the vector linking the end-carbon atom to the carbon atom bonded to the carbonyl group next to the chain. The normalised angle distribution functions are shown in figure 5.10, panel B. The distributions related to

the hydrocarbon tails are very similar, although the distribution related to the sn2-PLPC tail is slightly broader.

Two effects probably enhance the propensity of the sn2-PLPC tail to form folded structures: its length (17 carbon atoms against 15 and 13 carbon atoms of the sn1-PLPC and MrG tails, respectively) and the presence of non-conjugated sp² carbon atoms which may favor the conformational flexibility of the chain. In general we noted that all chains are preferentially distributed along the z-axis with average angles ranging from 32° (sn1-PLPC and MrG tails) to 37° (sn2-PLPC tail). The angle distributions of MrG and PLPC polar heads are instead very different, the former being globally shifted towards smaller angles. This is somehow expected by considering the strong difference in molecular structure. The relevant broadening of the MrG angle distribution can be ascribed to the smaller size of MrG polar head with respect to PLPC, which may reduce hindering effects and hence allow the access to larger configurational space. In fig. 5.10, panel C, I reported the distribution functions of the distances between the carbon atoms defining the hydrocarbon tails (see above for definition). The picture coming out from this quantity is consistent with a membrane made of globally unfolded molecules. The shift of the distribution functions is clearly due to the difference in chain length. The distance distribution related to the sn2-PLPC tail appears the broadest and thus showing (as previously remarked) a greater inclination to folding.

Short MD Simulations and Correlation Times

In order to characterise the H-bonding dynamics of the C=O groups, 40 MD simulations of 100 picoseconds each have been performed starting from samples extracted from the REMD simulation at 298K, chosen so that both dipeptides are in an unfolded conformation. Apart from the REM procedure, all other parameters are the same as those employed in the REM-MD production run previously described. Atomic coordinates have been recorded every 32 femtoseconds.

The H-bond dynamics between MrG/PLPC C=O groups and water has been probed through a simplified “Hydrogen-Bonding Function” f_{HB} which holds 1 for $0 \leq t \leq \tau$ and 0 otherwise, where τ is the lifetime of the H-bond established between a C=O group and a given water molecule. A H-bond is considered broken if it is not reformed after a time of 256 femtoseconds. This allows considering two H-bonding events separated by a short time gap with no interaction (a time comparable to inter-molecular

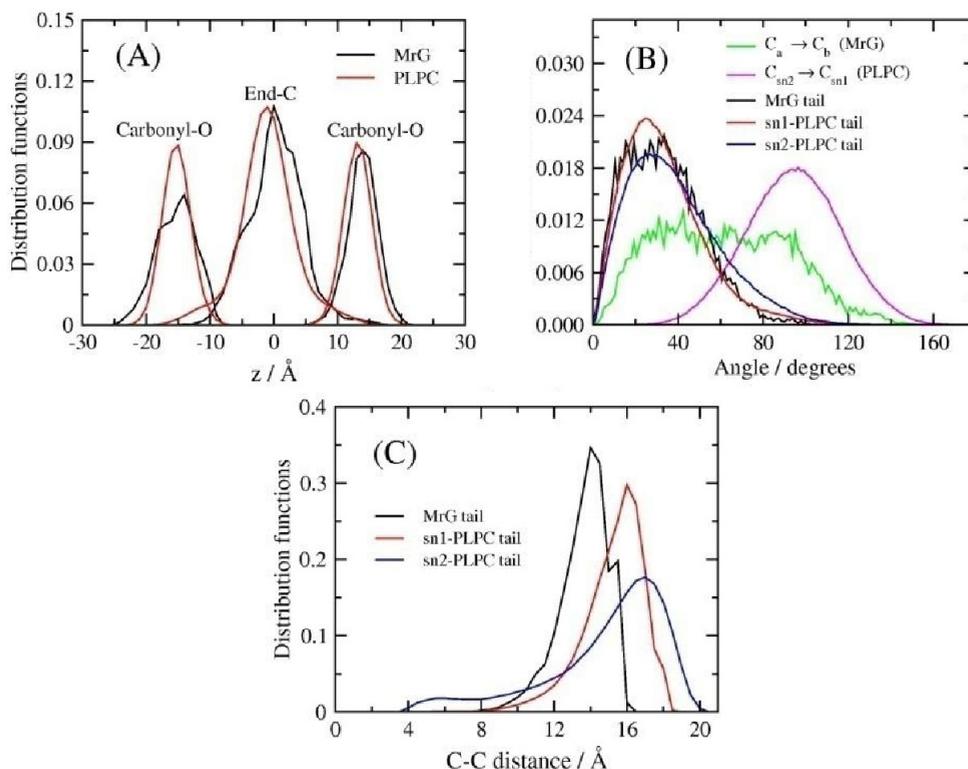


Figure 5.10: Panel A: normalised distribution functions of the carbonyl-oxygen atoms (carbonyl-O) and of the chain end- carbon atoms (End-C) of MrG and PLPC along the z-axis. Panel B: normalized distribution functions of the angle formed by the polar-head backbones and by the hydrocarbon tails of MrG and PLPC with the z-axis. The direction of the polar-head backbone of MrG corresponds to the vector linking the carbonyl-carbon atom next to the myristoyl tail (C_a) to the other carbonyl-carbon atom (C_b). The direction of the PLPC polar head corresponds to the vector linking the sn2 carbonyl-carbon atom (C_{sn2}) to the sn1 carbonyl-carbon atom (C_{sn1}). The direction of the hydrocarbon tails corresponds to the vector linking the end-carbon atom to the carbon atom bonded to the carbonyl group next to the hydrocarbon chain. Panel C: normalised distribution functions of the distance between the carbon atoms which define the hydrocarbon tails (see description of panel B).

vibrational motions) as a unique interaction event. The occurrence of a H-bond is established on the basis of a geometrical criterion, namely when the distance between the oxygen atom of the MrG/PLPC C=O group is smaller than 2.4 Å (this value corresponds approximately to the first minimum of the O...H pair distribution function). The function f_{HB} has been calculated for all C=O-water couples of all 40 MD simulations to achieve a good statistical significance. In order to get the H-bond lifetime, the normalised autocorrelation functions

$$\frac{\langle f_{\text{HB}}(t)f_{\text{HB}}(0) \rangle}{\langle f_{\text{HB}}^2 \rangle}$$

of all H-bonds have been averaged, obtaining the curves reported in fig. 5.11. The integral correlation times corresponding to the 3 types of H-

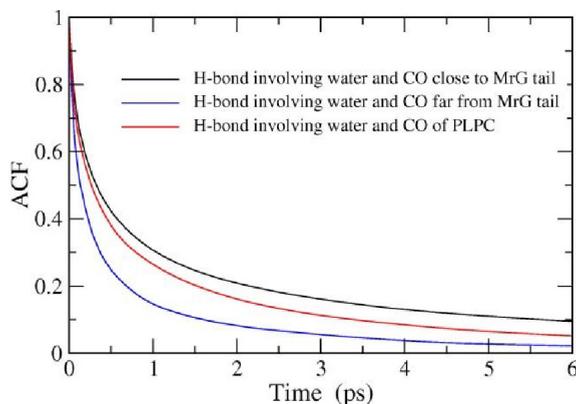


Figure 5.11: H-bond correlation functions for 3 types of H-bond involving the CO groups (of MrG or PLPC) and water (see legend).

bonds are 1.3 picoseconds (H-bond involving water and C=O of PLPC), 0.7 picoseconds (H-bond involving water and C=O far from the MrG tail) and 2.3 picoseconds (H-bond involving water and C=O close to the MrG tail).

Therefore, it seems unlikely that the hydrogen bond dynamics can be responsible for the rapid frequency fluctuations suggested by the experimental results. It seems more reasonable to attribute such fast modulation to small amplitude fluctuations of the MrG backbone. This process was found to contribute to the narrowing of the IR lineshape of trialanine amide I transition in water [WPH⁺02].

Dipeptide Penetration into the Membrane

The partially excitonic character of the delocalised C=O vibrations at the interface of the phospholipid membrane is a well known feature [VCZ⁺07]. In particular, as the dipeptide is present as an impurity in the membrane, and since the amide I modes are spectrally separated from the frequencies of the PLPC carbonyl modes, it can be securely stated that the spectral features of the cross-peaks between the amide I transitions and the PLPC C=O band can be interpreted in terms of inter-molecular pairwise interactions. In other words, the cross-peaks measured at the PLPC carbonyl stretching frequency upon excitation of the amide I modes provide information on the average angle between the transition dipole moments of MrG and of the neighboring PLPC molecules. From the anisotropy measurement that Dr. Volkov performed for the current research, the values of $21 \pm 14^\circ$ and $33 \pm 6^\circ$ for the ^{13}C and ^{12}C moieties have been obtained, respectively (or equivalently 159° and 147°). To rationalise these values, I calculated the weighted radial-angular distribution function related to MrG and PLPC carbonyl pairs, which are shown in fig. 5.12. For both

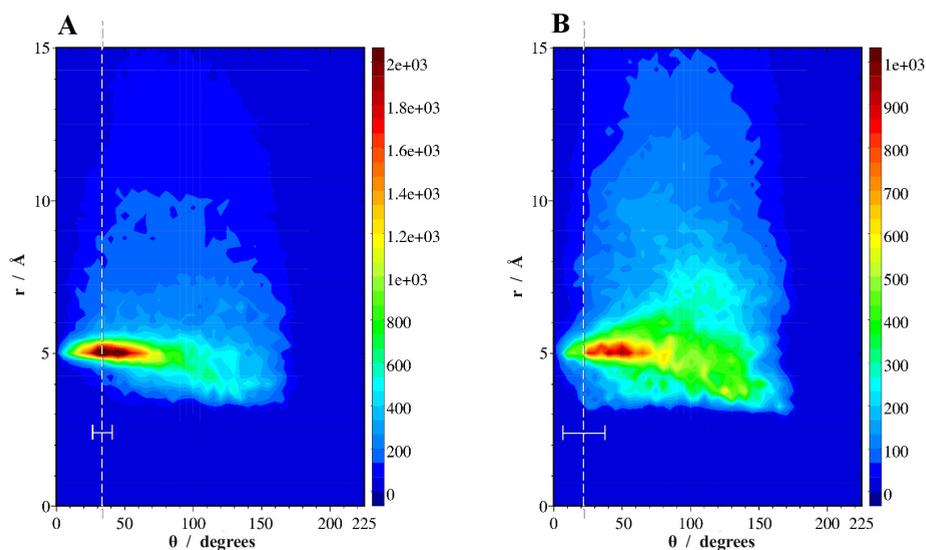


Figure 5.12: Weighted pair radial-angular distribution functions of the PLPC and MrG carbonyl groups. (A) Function related to PLPC C=O and MrG ^{13}C =O pairs. (B) Function related to PLPC C=O and MrG ^{12}C =O pairs. The vertical dashed lines represent the values of the angles derived from the 2D-IR experiment; the corresponding errors are indicated by the horizontal bars.

MrG amide groups, the function is sharply peaked at the distance of 5 Å. In the case of the carbonyl next to the anchor tail (fig. 5.12, panel A) also the angular distribution is quite narrow (between 20° and 50°). For the distal C=O instead, the angular distribution is definitely broader. The vertical dashed lines in the same figure correspond to the angles derived from the experiment. The calculated values are in general compatible with those obtained from the 2D-IR spectra, although the angular distribution calculated for the distal carbonyl is centred at a larger angle. In any case, it should be noted that the experimental error affecting the angular value of the distal carbonyl is much larger than that for the vicinal one (see error bars in fig. 5.12).

The depth of penetration of MrG into the PLPC membrane was the subject of a previous experimental work of Volkov and Righini [VR09]. That experiment consisted in exciting the water stretching modes and measuring the perturbation induced in the amide I transitions of MrG. The meter stick used for determining the vertical localisation of MrG is provided by the density profile of the water molecules across the membrane, which is known from neutron diffraction measurements [JW89]. In fact, the appearance of cross-peaks in the two-color 2D-IR spectra is a consequence of the hydrogen bonding between the C=O groups and the surrounding water molecules.

The intensity ratio between the cross-peaks involving the amide I resonances of MrG and that corresponding to the PLPC C=O (whose localisation in the bilayer thickness is known from neutron scattering data [JW89]) is taken as a measure of the relative degree of hydration of the MrG carbonyls with respect to those of PLPC. From the known vertical profile of the water density, one can then derive the localisation of the C=O groups of MrG. The results obtained following this procedure are summarised in fig. 5.13. The two C=O groups of the dipeptide are found to lay, in average, above those of PLPC molecules: more precisely, the vertical distances from the average position of PLPC carbonyls was 2 Å for the C=O next to the MrG tail, and 4.4 Å for the distal C=O.

Clearly, only the hydrated C=O can be detected by this method. The analysis of the REMD simulation described in the present paper provides a reliable test of this conclusion. In fig. 5.13, panel B, the blue and red lines are obtained by multiplying the calculated distribution of the oxygen atoms of MrG by the density of water oxygen atoms as obtained from the REMD simulation. The black curve is the distribution of all C=O

oxygens of the PLPC molecules. In agreement with the results of our two-color pump-probe experiment, the REMD simulation predicts for the two amide groups an average localisation above that of the PLPC carbonyls. The distances are slightly smaller (1.9 and 3.1 Å) than those deduced from the experiment. The method adopted in Ref. [VR09] and used here to calculate the curves in fig. 5.13, panel B, is based on the assumption that the water density depends only on the value of the z -coordinate across the membrane thickness. Still, from the simulation data we can exactly calculate the degree of hydration of the oxygen atoms of MrG and of the PLPC carbonyls in the sample.

Fig. 5.13, panel C, reports the distribution of the MrG oxygens hydrogen bonded to water molecules. Their average position is still above that of the PLPC carbonyls, but their distance from that reference altitude is even smaller: 0.8 and 2.3 Å. The obvious conclusion is that the assumption of the water density depending only on the z coordinate is not completely justified.

In order to verify this point, I have calculated the ratio between the water densities around MrG and PLPC molecules. In particular, I have considered the water density in xy planes parallel to the membrane surface, as a function of the altitude in the perpendicular (z) direction. The origin of the z axis is placed at the average altitude of the carbonyl oxygens; the abscissa represents the distance in the xy plane from the carbonyl oxygens of MrG/PLPC. The ratio of the two densities is given in fig. 5.14, panel A. The figure shows that there is definitely more water near the MrG oxygens than near the PLPC carbonyl oxygens.

The reason is made clearer from fig. 5.14, panel B: this plot is calculated with the same procedure of fig. 5.14, panel A, considering all atoms but excluding those belonging to water molecules. It is apparent that the two plots of fig. 5.14 are somehow complementary. The increased density of water molecules corresponds to areas of the membrane (those above the guest MrG molecules) where the density of the polar groups of PLPC (namely phosphate and choline) is lower. The MrG molecule entering the membrane polar surface creates the pathway for a deeper penetration of water inside the membrane. This results in a different hydration of the C=O groups, the amide groups of MrG being more hydrated than the C=O of PLPC located at the same altitude. This is confirmed by the mean hydrogen bond number calculated for the MrG and PLPC carbonyls: the values are 0.87 for the former and 0.54 for the latter. If this picture corre-

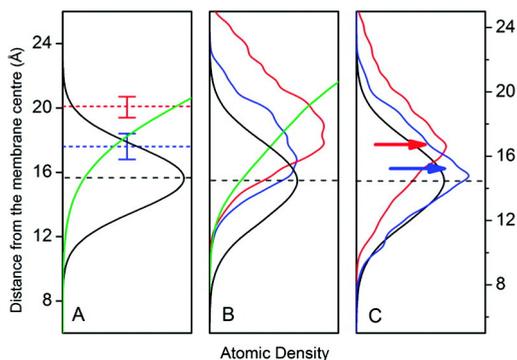


Figure 5.13: (A) Atomic densities of water (green line) and PLPC carbonyl (black line) from neutron scattering data. The blue and red dotted lines are the altitudes of the ^{13}C and ^{12}C carbonyls of MrG, as deduced from the experiment of Ref [VR09]. (B) Distribution of all PLPC carbonyl oxygen atoms obtained from the REMD simulation (black line); distribution of the water oxygen atoms from the REMD simulation (green line); blue line shows the function calculated from REMD simulation by multiplying the distribution of the oxygen atoms of the ^{13}C labeled carbonyl of MrG with the distribution of the water oxygen atoms; red line shows the same for the oxygen MrG native carbonyls. (C) Distribution of all PLPC carbonyl oxygen atoms obtained from the REMD simulation (black line); the blue and red curves represent the calculated distributions of the oxygens of the MRG carbonyls H-bonded to water molecules. The arrows indicate the average altitudes of the two distributions. For graphical reasons, in panels B and C the vertical axes are up-shifted to make the maxima of the experimental and calculated densities of the PLPC carbonyls coincident.

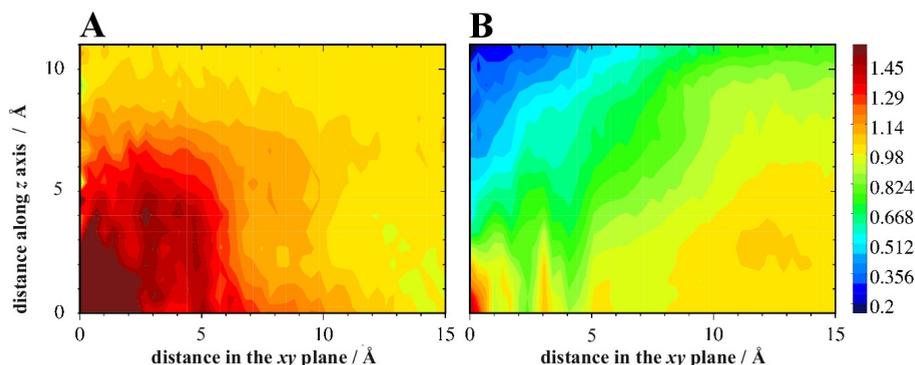


Figure 5.14: (A) Ratio between the densities of water oxygen atoms around MrG and PLPC carbonyl oxygens, respectively, as a function of the depth into the membrane of the MrG/PLPC carbonyl oxygen (distance along z axis) and of the distance in the membrane plane between water oxygen and MrG/PLPC carbonyl oxygen (distance in the xy plane). (B) Same ratio as in panel A, but considering all nonwater atoms.

sponds to the structure of the real membrane, one should expect that the method adopted in Ref. [VR09] overestimates the average distance of the dipeptide from the centre of the membrane.

The structure and conformation of the anchor tail of MrG, in relation to that of PLPC are not directly observable in Dr. Volkov's experiments. It is, however, of some interest to analyse, under this respect, the sample obtained from the REM-MD simulation, which basically tell us what the experiments keep for themselves. To this purpose, I consider several structural properties, concerning both the anchor chain conformation and the mutual arrangement of the aliphatic tails of MrG and PLPC molecules.

Fig. 5.15, panel A, shows the distribution functions of the carbonyl-oxygen atoms and of the chain end-carbon atoms of MrG and PLPC along the z -axis of the membrane during the simulation. It is noteworthy that the terminal carbon atom of the MrG anchor is centred near $z = 0$ and is very similar to that of the corresponding carbon atoms of the PLPC hydrocarbon tails. This implies that, on average, the anchor chain of MrG is parallel to the direction of the PLPC tails. This is confirmed by the plot of fig. 5.15, panel B, where the distribution functions of the distances between the first (next to the carbonyl group) and the last carbon atoms of the hydrocarbon tails of MrG and PLPC molecules (end-to-end distance) is shown. This is a simple criterion for quantifying the degree of folding

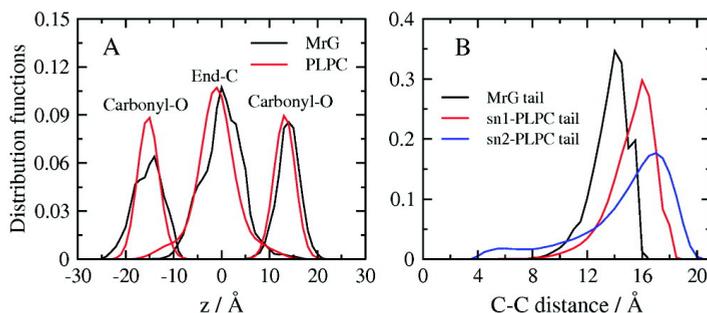


Figure 5.15: (A) Normalised distribution functions of the carbonyl-oxygen atoms (carbonyl-O) and of the chain end-carbon atoms (End-C) of MrG and PLPC along the z -axis of the membrane. (B) Normalised distribution functions of the distance between the first (next to the carbonyl group) and the last carbon atoms of the MrG, sn1-PLPC and sn2-PLPC hydrocarbon tails (see fig. 5.1 for *sn* labelling notation).

of the long aliphatic chains. The figure shows that only the linoleic (sn2) chain of PLPC, whose distribution function shows a non-negligible wing reaching values lower than 4 Å, is present in folded conformations. In fact, it is known that unsaturated hydrocarbon chains of lipidic bilayers can be found in folded geometries. On the contrary, the myristoyl chain of MrG and the palmitic (sn1) tail of PLPC have almost identical length distributions (apart from the obvious difference due to the different number of constituent carbon atoms), corresponding to prevalent unfolded structures.

Thus, as it is apparent from this research, understanding the structural properties of the association between a small peptide molecule and a phospholipid membrane is a complex task, which requires the coordinated use of a number of experimental and theoretical tools.

The results of the classical MD simulations are in acceptable agreement with the structural properties deduced from the spectroscopic investigation, thus substantiating and reinforcing the interpretation of the experimental data. Anyway, a lot of care should be given to correctly assemble the force field, since even minor inadvertence results in completely wrong trajectories, with no real physical meaning. The behaviour of the dipeptide, its depth of penetration into the membrane, its affinity for water and PLPC, all this is driven by H-bond breaking and formation. In fact, the autocorrelation function of a rather crude "H-bond function" allows

us to weight the relative stability of the different types of inter-molecular interactions.

The rather successful application of the diversified and coordinated investigation methods adopted here, including nonlinear IR spectroscopic techniques and molecular modelling, represents an indication that similar approaches can be used to tackle structural problems of higher complexity, such as those involving polypeptides of medical and pharmacological interest.

However, to obtain more detailed insights about the behaviour of molecules, *ab initio* MD proves a much more suitable probe, albeit at the price of a much more reduced molecular complexity. In particular, it is necessary to study molecular vibrations. The REM approach adopted for the dipeptide, for example, does not even generate a truly time-dependent trajectory (see Appendix B.2) to obtain a vibrational spectrum via Fourier or wavelet transforms.

6. Simple Diols in Water and Acetonitrile

Two simple diols in aqueous and organic solvent have been simulated at *ab initio* MD level of theory to be exploited as test case for the direct-space WT code (see Section 4.2.3), which I have developed to perform time-frequency analysis.

In fact, WT use in theoretical and computational chemistry is gradually increasing. In particular, it has been applied to electronic structure calculations [GI99,NG06,DAE03,CAJL93] and to MD [MSC08a,MSC08b,ACR96] also to obtain time-dependent vibrational frequencies [RW05].

In this research, WT has been mainly used to correlate vibrational frequencies of two molecular systems with other time-dependent structural properties. WT has been employed to improve the analysis of H-bond dynamics and its relation with spectroscopic properties. Since in MD simulations it is always possible to associate to each simulation step a structural quantity, distance-frequency correlation showing how the VDoS changes with a particular H-bond length has been obtained.

The two glycols simulated in aqueous environment are propane-1,3-diol (propanediol, PDO) and ethane-1,2-diol (ethylene glycol, EG), two homologous compounds of industrial interest that interact with the aqueous solvent mainly by inter-molecular H-bonds. PDO is a transparent, non-toxic liquid glycol that can be obtained by fermentation of sugars and can replace other glycols in formulations where petroleum-free [OHS08] ingredients are needed, which is relevant if the current hysteria about oil consumption, CO₂ emissions and so-called “global warming” is going to last. EG is the lower homologue of PDO and is mainly used as an antifreeze and a medium for convective heat transfer due to its low freezing point.

For both PDO and EG, a deep knowledge of the H-bond interactions and of the solvation dynamics is needed to improve their use in industrial applications. These glycols have two interaction sites with the water sol-

vent and the analysis of the H-bond features are challenging due to the overlap of the solvent-solvent and solute-solvent contributions.

For this reason, PDO has also been simulated in acetonitrile (ACN) solvent. In fact, a non-protic solvent like ACN allows measuring the OH stretching bands of PDO, thus providing an experimental reference. Moreover, the behaviour of PDO in ACN is significantly different from that in aqueous solvent and requires a more complex approach to be properly elucidated, due to the occurrence of intra-molecular H-bonds even between its two hydroxyl groups and not only with the solvent.

These two hydroxyl sites are often referred separately in the following, typically as “site # 1” and “site # 2”. While this choice has no great physical meaning for the aqueous simulations, it is very important for the simulation in ACN due to the different non-ergodic behaviour of the two sites in this latter case.

6.1 Propanediol and Ethylene Glycol in Water

6.1.1 Computational Chemistry Details

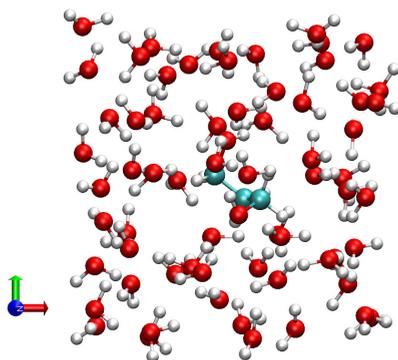


Figure 6.1: Simulation box of PDO in water.

Ab initio MD simulations have been performed for PDO and EG in water. These simulations have been carried out with the CPMD package [CPM] in the microcanonical ensemble (NVE) in conjunction with the BLYP [Bec88, LYP88] exchange and correlation functional, in cubic boxes of edge 12.7005 Å (for PDO) and 12.6819 Å (for EG), with periodic boundary conditions. The samples in the two simulation boxes are made of 1 solute molecule and 64 D₂O molecules (at the experimental density of deuterated water, $\sim 1.106 \text{ g}\cdot\text{cm}^{-3}$). To give a quick idea of the size of this

system, also in comparison to that of the dipeptides in PLPC (see fig. 5.2), I report fig. 6.1 that displays the simulation box with PDO and solvent.

Norm-conserving Martins-Troullier [TM91] pseudopotentials have been used along with Kleinman-Bylander [KB82] decomposition, and the plane waves expansions have been truncated at 85 Ry (this choice has been shown to be particularly effective in CPMD simulations [KMM⁺04]).

ethane-1,2-diol ^a	Ref. [HJK05]	BLYP ^b	B3LYP ^b	MP2 ^b	CPMD ^c
r (C–C)	1.514	1.527	1.517	1.513	1.523
r (C–O ₁)	1.433	1.454	1.434	1.431	1.457
r (C–O ₂)	1.421	1.440	1.421	1.418	1.444
r (O ₁ –H)	0.961	0.971	0.961	0.961	0.974
r (O ₂ –H)	0.964	0.975	0.965	0.964	0.977
r (O ₁ ···H)	2.331	2.434	2.399	2.323	2.418
θ (O ₁ –C–C)	106.1	106.7	106.8	106.2	106.8
θ (O ₂ –C–C)	111.2	112.2	112.0	111.1	112.2
ϕ (H–O ₁ –C–C)	-166.0	-164.8	-166.7	-164.0	-166.6
ϕ (H–O ₂ –C–C)	-51.5	-54.9	-53.6	-51.9	-52.1
propane-1,3-diol ^a	Ref. [KMMR82]	BLYP ^b	B3LYP ^b	MP2 ^b	CPMD ^c
r (C ₃ –C ₄)	1.514	1.533	1.523	1.518	1.528
r (C ₄ –C ₅)	1.514	1.543	1.532	1.527	1.539
r (C ₃ –O ₁)	1.410	1.461	1.439	1.435	1.465
r (C ₅ –O ₂)	1.410	1.441	1.422	1.420	1.445
r (O ₁ –H)	1.040	0.972	0.962	0.962	0.975
r (O ₂ –H)	0.980	0.977	0.967	0.965	0.979
r (O ₁ ···H)	1.753	2.051	2.034	2.000	2.052
θ (O ₁ –C ₃ –C ₄)	112.0	108.4	108.6	108.0	108.9
θ (O ₂ –C ₅ –C ₄)	108.0	113.4	113.4	113.0	114.0
θ (C ₃ –C ₄ –C ₅)	112.0	113.9	113.8	112.9	114.1
ϕ (H–O ₁ –C ₃ –C ₄)	180.0	173.9	175.6	173.2	172.3
ϕ (H–O ₂ –C ₅ –C ₄)	-46.0	-46.3	-46.7	-48.4	-41.4

Table 6.1: Structural parameters. ^a Bond lengths in Å, angles in degrees. ^b DFT and MP2 calculations have been performed with the GAMESS [SBB⁺93] suite of programs in conjunction with 6-311++G(d,p) basis set. ^c CPMD geometry optimisations have been performed with the same parameters reported for the MD simulations.

In order to use a larger time-step (δt of 5 a.u. \sim 0.12 femtoseconds),

hydrogen atoms have been replaced by deuterium atoms.

The fictitious electronic mass has been set at 700 a.u. to allow for a good decoupling between electronic and nuclear degrees of freedom. The systems has been thermalised by scaling the atomic velocities for 2 picoseconds in order to keep the temperature at the value of 300 ± 50 K and the trajectories for both systems have been collected for ~ 32 picoseconds.

The data collected from the two simulations of the glycols have been compared with corresponding CPMD simulations in vacuum. The computational parameters used in vacuum simulations, such as cell parameters, energy cut-off and time-step, are the same as those in water simulations. To further check the accuracy in reproducing the structural parameters of isolated diols with the computational approach adopted, geometry optimisations have been carried out with the GAMESS package [SBB⁺93] at the MP2 and density functional theory (DFT) level of theory (using the BLYP [Bec88, LYP88] and B3LYP [Bec93b, Bec93a, LYP88] exchange-correlation functionals) with 6-311++G(d,p) basis set.

The comparison of experimental and calculated data is reported in table 6.1, showing only minor differences between all electrons and pseudopotentials calculations. All calculated normal modes frequencies are real and positive, ensuring that the systems are in their respective equilibrium configurations.

6.1.2 Structural and electronic properties

Structural properties

PDO and EG interact with water mainly through H-bonds, acting as both acceptor and donor. The H-bond can be characterised structurally through the O–D···O distance r and the O–D···O angle θ . Figures 6.2 and 6.3 report the pair radial $g(r)$ and angular $g(\theta)$ distribution functions for the acceptor and donor H-bonds, respectively. It can be seen that radial distribution functions of the two glycols differ very slightly. The deep first minima are indicative of the slow mobility of the water molecules bound to the glycols. On the contrary, it can be noted that the $g(\theta)$ functions for PDO are narrower, probably due to the fact that the two OD sites are farther away than in EG, allowing a more independent solvation dynamics. This also implies that the angular parameters can be more sensitive to structural details. In the present case, the narrower $g(\theta)$ function is an evidence for a more stable H-bond in the PDO solution. The av-

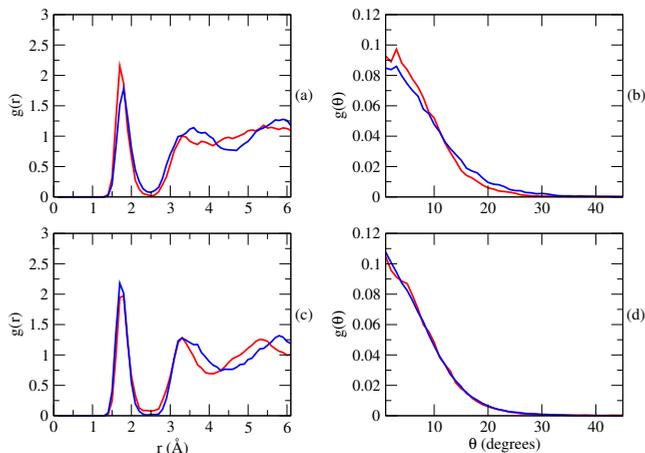


Figure 6.2: Acceptor $g(r)$ and $g(\theta)$: red for “site 1”, blue for “site 2”; panel (a) and (b) for EG, panel (c) and (d) for PDO.

average coordination number for the OD sites, extracted from the $g(r)$, is reported in table 6.2, and is always close to the expected value of 1 for the donor interactions, whereas the acceptor coordination number is larger and shows a significant spread as a consequence of weaker interactions. The differences occurring for site 1 and 2 of both diols arise from statistical uncertainty due to the finite temporal length of the simulations and the use of a single solute molecule for each system.

	site 1	site 2
EG donor	1.00	1.01
EG acceptor	1.70	1.42
PDO donor	0.98	1.00
PDO acceptor	1.56	1.94

Table 6.2: Average coordination numbers.

To better characterise the strength and structural features of the H-bond, the weighted joint radial and angular distribution function $g(r, \theta)$ can be calculated by adopting the weighting function F^{HB} [PCRS03, PCS05, FPCS06] already discussed in Section 3.1.3 (eqn. 3.8).

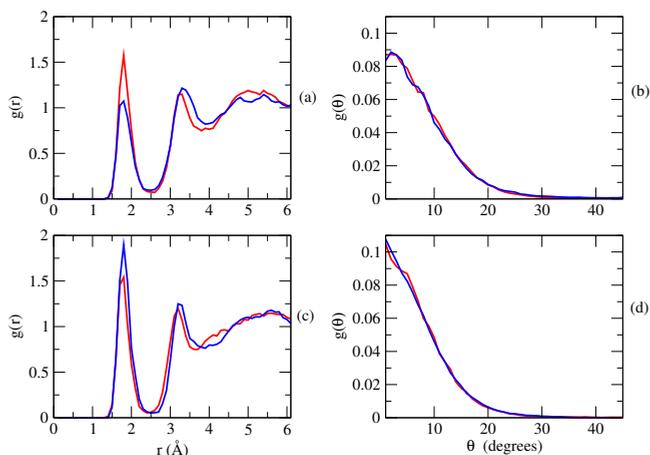


Figure 6.3: Donor $g(r)$ and $g(\theta)$: red for “site 1”, blue for “site 2”; panel (a) and (b) for EG, panel (c) and (d) for PDO.

The results shown in fig. 6.4 confirm that H-bonding in PDO is stronger, due to the lower spread of the joint distribution function along the angular parameters, since in the present simulations the angular conditions are a more stringent requirement for the hydrogen bond formation. A pictorial view related to the weighted $g(r, \theta)$ is shown in fig. 6.5, where the spatial distribution functions (SDF) of the OD groups of the water molecules are reported, displaying a three-dimensional description of the motion for the water molecules around the OD groups of the diols. The cyan-coloured isosurfaces (each point has to be visited 55 times) close to the glycols represent the probability to find the D atoms of water around the O-sites of the solutes, whereas the orange-coloured ones represent the probability to find the O atoms of water around the alcoholic D atoms of the diols. These combined isosurfaces give a clear representation of the global mobility of the solvation cage. The spread of the isosurfaces for PDO is smaller than for EG, in line with the previous observations from the $g(\theta)$ and $g(r, \theta)$ functions.

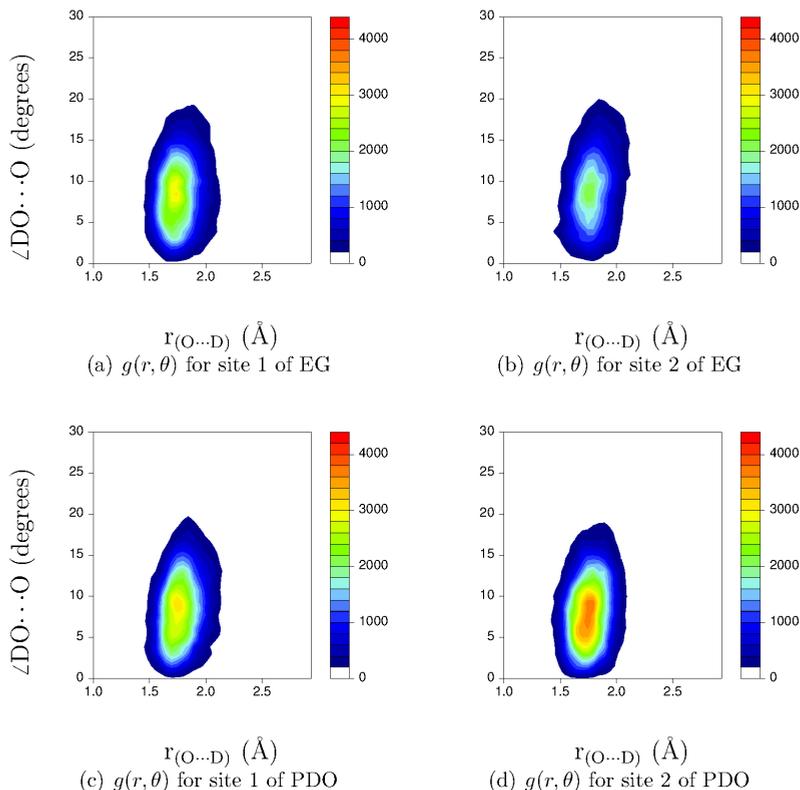


Figure 6.4: Weighted acceptor $g(r, \theta)$ for the two OD sites of the two diols.

Electronic structure analysis

The polarisation effects due to the interaction of the glycols with the solvent have been analysed in MLWF centres [MV97, SMVP98], briefly discussed in Section 3.2. The positions of the MLWF centres can be related to the electron pairs and give a direct picture of the electronic structure. In order to obtain useful insight on the electronic structure changes due to the interactions of the glycols with the solvent, the molecular dipole moment computed on the basis of the MLWF centres positions of PDO and EG has been monitored during the simulations. For this purpose the dipole moments of PDO and EG have been calculated on a series of time-equispaced configurations obtained by CPMD simulations (1 configuration every 10^4 time-step, corresponding to 1.2 ps). The polarisation enhancement of glycols in solution has been obtained by adopting the following computational strategy. For each configuration the MLWF centres

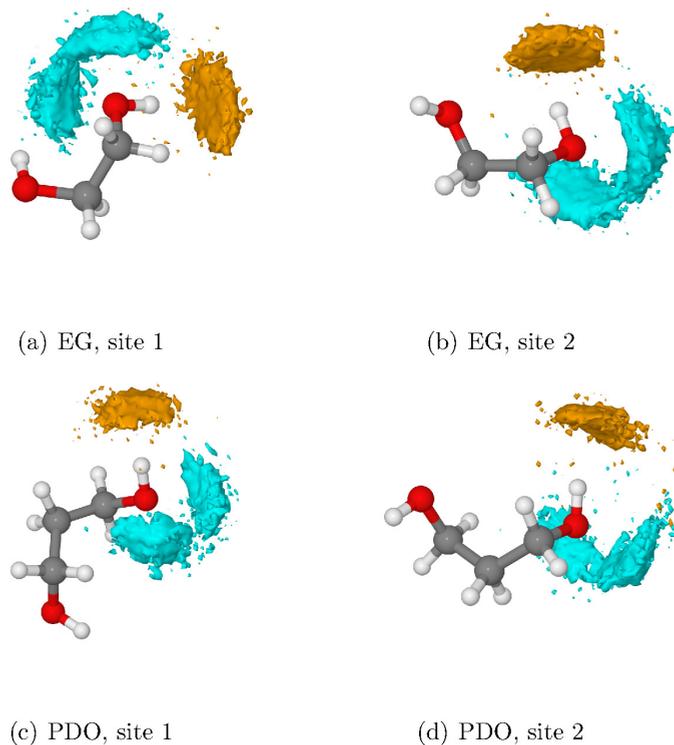


Figure 6.5: SDF of PDO and EG. Cyan and orange surfaces show the mobility of D and O atoms of H-bonded water molecules, respectively.

have been obtained for all the system, for the diols and the H-bonded water molecules and for diols alone without geometry optimisation.

In table 6.3 the dipole moments calculated by MLWF centres analysis by localising the valence shell electron doublets are reported. The dipole moment of the diols increases as reported in table 6.3, as a consequence of the interaction with the solvent. The time dependent data also show an evident correlation in the time evolution of the dipole moments for the three model systems. However, the solvent increases the dipole moment of the diols and larger fluctuations are observed for EG, which being smaller in size is more sensitive to polarisation effects by the water cage. Furthermore, the dipole moment increase, due to solvent interaction, is larger in EG than in PDO for the same reason.

	isolated diol	diol + H-bonded water molecules	diol in solution
EG	2.71 D	3.11 D	4.55 D
PDO	2.05 D	2.18 D	3.87 D

Table 6.3: average dipole moments for diols according to MLWF centres analysis on the three model systems.

6.1.3 Time-frequency analysis

To analyse PDO and EG vibrational response in aqueous environment, I have employed a combination of FT and WT methods. I have used the direct space implementation of WT (see Section 4.2.3), due to the fact that it was the first WT code I wrote.

In Ref. 6.6 the VDoS of the water solutions, of the diols with the two closest molecules and of the diols by themselves in the water solutions are shown, obtained by FFT of the velocity autocorrelation functions. The OD stretching region ($2100\text{--}2600\text{ cm}^{-1}$), which is the most interesting region of the spectrum, will be discussed in the following on the basis of WT analysis.

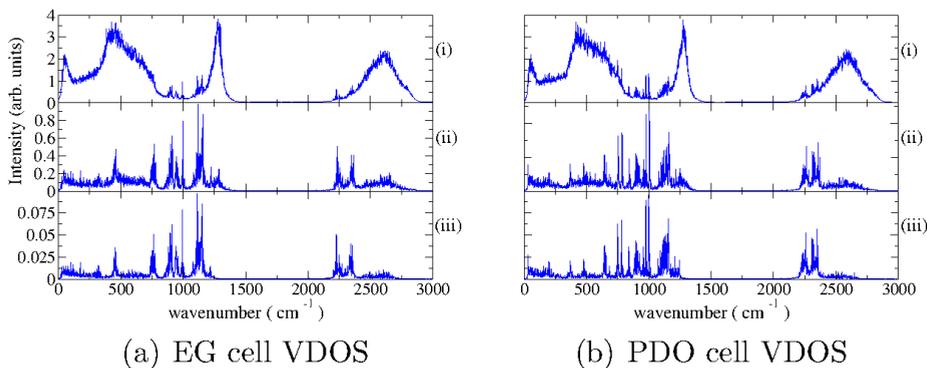


Figure 6.6: Vibrational densities of states: (i) VDoS of the complete simulation cells; (ii) VDoS of the diols and the two H-bonded water molecules; (iii) VDoS of the diols by themselves.

The structure of the OD stretching band results from the convolution of solvent-solvent and solute-solvent interactions and I focus on the structure, extracted from the simulation, of the VDoS of the solute molecule and of the two nearest water molecules bound to the diols. This can be

appreciated in more details analysing the VDoS extracted by FFT of the oscillations of the intra-molecular Δr_{O-D} functions, as reported in fig. 6.7, where the density of states obtained by simulations of the diols in solution (panel *a* and *b*) and in vacuum (panel *c* and *d*) are reported. The redshift due to H-bonding interactions with the water solvent is evident, along with the broadening of the stretching band. In fig. 6.8, I report

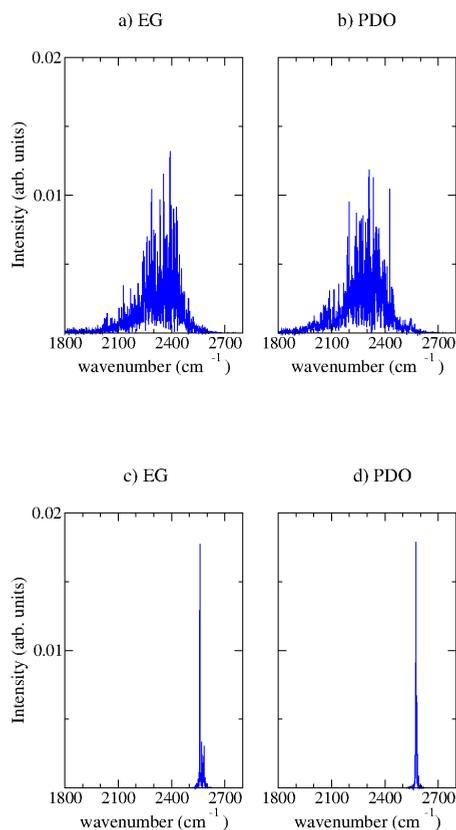


Figure 6.7: VDoS for hydroxyl group stretching mode in water (panels *a* and *b*) and in vacuum (panels *c* and *d*) simulations.

again fig. 4.8. While in Section 4.2.2 this figure was exploited to show how spectrum reconstruction changes with values of σ parameter of the MG mother wavelet, without knowledge of the meaning of the input signal, here I want to underline that the VDoS of a diol can be reproduced by both FFT and WT, using the inter-molecular $\Delta r_{O...D}(t)$ fluctuation as input function. The inhomogeneous broadening of the band can be attributed to time changes of the structure of the solvent cage around the hydroxyl

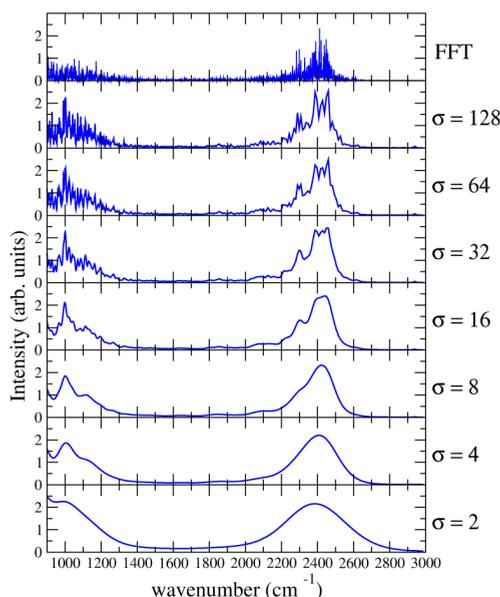


Figure 6.8: Fourier and wavelet power spectrum of the displacement of intermolecular $\Delta r_{O...D}(t)$ function of EG, site 2.

groups of the diols, which modulates the instantaneous vibrational frequency.

To better enlighten the origin of the various frequency components, the trajectory of site 1 of PDO has been analysed by WT at specific time intervals. Fig. 6.9 shows the details of the trajectory referring to site 1 of PDO. It can be seen that along the first 5 picoseconds of simulation site 1 is alternatively free or bound to water molecule # 22. Later, along the trajectory, the site is constantly bound to the same water molecule. The WT spectrum has been calculated at the simulation times of 1.2, 2.2, 3.8 and 4.8 picoseconds (dotted red lines in fig. 6.9) and the results, reported in fig. 6.10, clearly shows that the redshift nicely correlates with the H-bonding character. The same correlation has been observed at all time-steps probed for both sites of the two glycols.

An alternative and more immediate way of illustrating this behaviour is obtained by directly correlating the most intense O–D stretching frequency, obtained by WT of the corresponding $\Delta r_{O-D}(t)$ function, with the values of the F^{HB} donor function. This is displayed in fig. 6.11, showing how the wavelet-calculated VDOS for all OD sites of the diols changes with the value of the F^{HB} function (which is confined in the

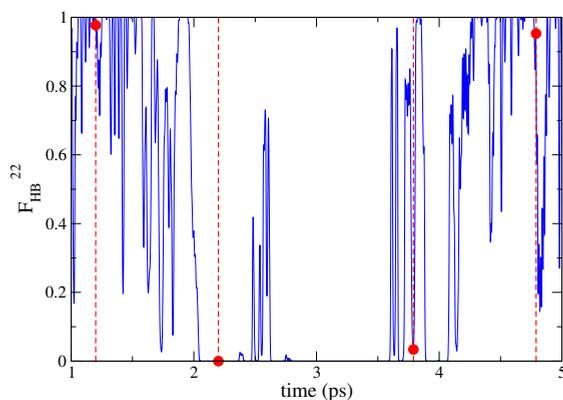


Figure 6.9: F_{HB}^{22} evolution with time for interaction of D_2O molecule # 22 with site 1 of PDO (full blue line). In dotted red lines the sampled time-steps for which the wavelet spectra are reported in fig. 6.10. The red circles represent the F_{22}^{HB} value sampled at those time-steps.

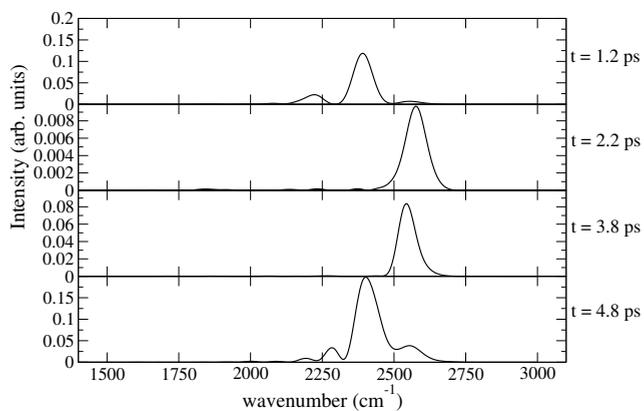


Figure 6.10: Wavelet spectra at four different time-steps calculated from intramolecular $\Delta r_{\text{O-D}}(t)$ function for site 1 of PDO (parameter $\sigma = 10$).

[0 – 1] interval due to the fact that only a single donor H-bond can be present at a time). The plotted quantities are the probability distributions

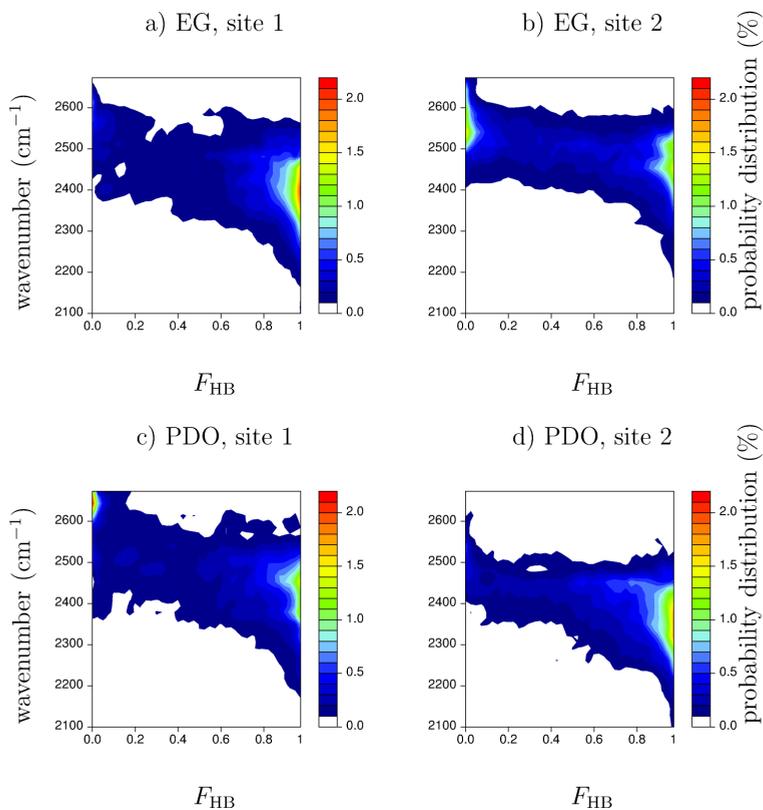


Figure 6.11: Correlation spectrograms between O – D stretching frequency and F^{HB} function.

of the VDOS, which thicken and are redshifted when F^{HB} approaches 1, meaning that the inter-molecular H-bond lowers the O–D intra-molecular stretching frequency. Consistently, the VDOS is peaked at higher frequencies when $F^{\text{HB}} \simeq 0$. As it can be seen, there is a sort of pathway between the two extremes in the frequency/configurational space explored by the simulations. For sites 2 of both glycols the pathway is continuous, due to a more oscillating character of the F^{HB} function with too fast oscillations to be precisely resolved even with WT.

As a whole, the present analysis shows how a structural quantity like F^{HB} , designed as a probe of the H-bonding, can be efficiently correlated with the stretching frequency of an intra-molecular vibration, because the

molecular oscillators are coupled to the water environment and are sensitive to small fluctuations in the solvation cage. Moreover, the advantage of a continuous function like F^{HB} to monitor on the fly the dynamics of the hydrogen bond and its effect on the vibrational spectrum is quite evident.

Wavelet analysis has also been used to sort out the time evolution of the frequency related to the most intense peak in the high resolution spectra (obtained like those of fig. 6.10) and its correlation with the intermolecular O...D bond length between nearest molecules along the simulation. The result is shown in fig. 6.12. The blue line in fig. 6.12, drawn as a guide for the eye, has been obtained by a 10^{th} -order polynomial fit of the raw data. A straight correlation between the time evolution of vibrational

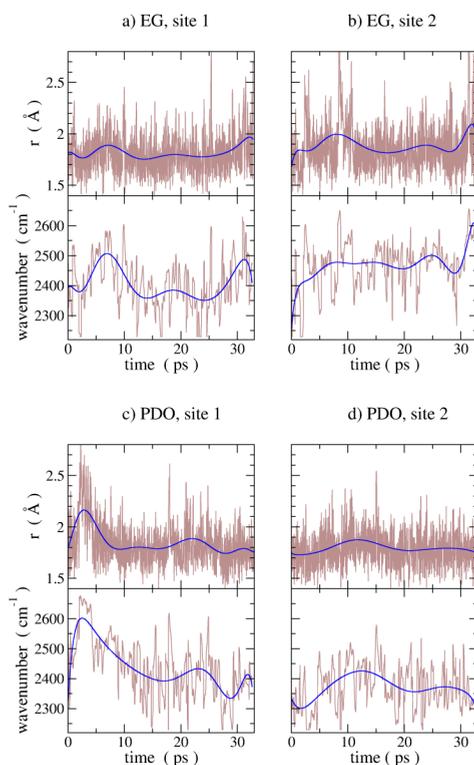


Figure 6.12: Comparison of the time evolutions of H-bond length and O–D stretching frequency. The brown lines correspond to the actual values found during the simulations, whereas the blue lines represent the smoothed trends of the former.

frequencies and the time evolution of intra-molecular O–D stretching distance can be appreciated.

A more straightforward way to show the correlation between stretching frequencies and inter-molecular $D_{diol} \cdots O_{water}$ distance is via the two-dimensional spectrogram plot of fig. 6.13, where distance and frequency are directly correlated at each time-step. The probability distribution reported in fig. 6.13 represents the maximum of the intra-molecular O–D band stretching obtained by WT correlated with the corresponding value of the inter-molecular H-bond length. These spectrograms show the same

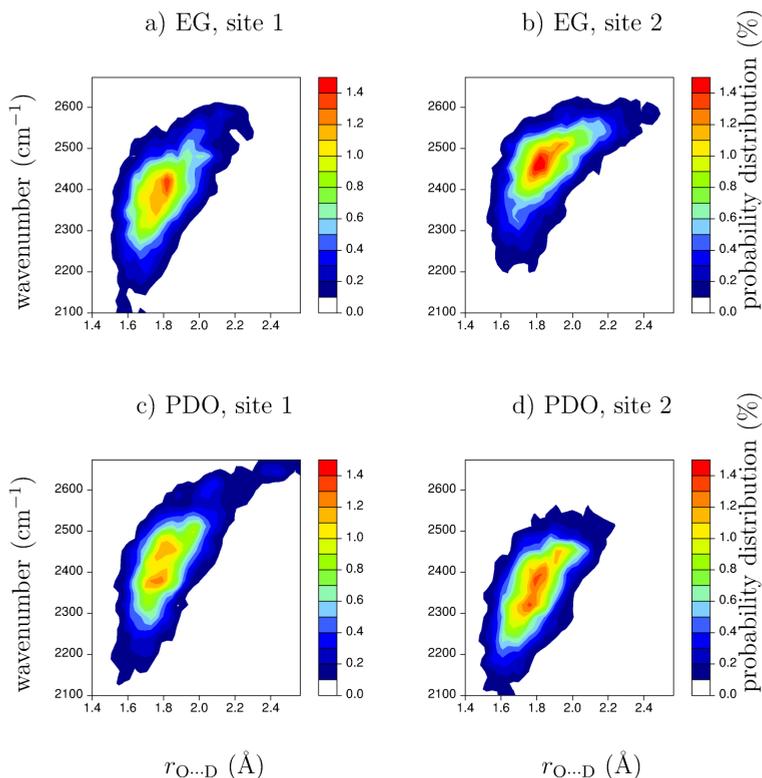


Figure 6.13: H-bond length – frequency spectrograms for both PDO and EG.

“banana shaped” distributions (as referred to in References [MSC08b, MSC08a]), obtained for other systems like heavy water and Cl^- anions in heavy water. Moreover, the same type of shape is obtained using inter-molecular H-bond distances or intra-molecular O–D bond lengths, since they both mainly probe the stretching frequency associated to the O–D stretching of the glycols. The VDOS thickens at bond lengths corresponding to the first peak of the $g(r)$ and $g(r, \theta)$ functions. The frequency shift is thus simply correlated to the strength and stability of the H-bond. A strong

H-bond gives a $g(r)$ function peaked at shorter distances and, within the harmonic approximation, it reduces the intra-molecular stretching force constant and, consequently, shifts the VDOS at smaller wavenumbers. The slightly different plot for EG, site 2, is due to the fact that the most strongly interacting water molecule for that site is not H-bonded for the first picoseconds of simulation, thus resulting in a final spectrogram showing a weaker H-bond, as expected.

It is not possible to establish an exact bijective relationship between bond length and frequency, but this is not surprising because the radial distance is only one of the structural parameters that characterise the H-bond, the other being the angular parameter. Hence, the distance is degenerate in the angle space. Moreover, the Heisenberg-like uncertainty principle between time and frequency, and therefore between distance and frequency, affects the spectrogram resolution (see Section 4.2). However, these plots clearly show that a reliable correlation between bond length and density of vibrational states exists, which can be understood on the basis of the H-bond stability. This is in agreement with previous results obtained using static *ab initio* calculations [Kle02].

Thus, it has been found that the complex pattern of the hydrogen bond stretching mode has an inhomogeneous origin and arises from the convolution of a number of differently shifted vibrational modes that can be sorted out by wavelet transform at different time intervals of the MD simulation. The time resolved vibrational modes obtained by WT can be correlated in a straightforward way with the structure of the solvent cage around the solute diols and with structural parameters like the O–D bond length or with the strength of the hydrogen bonding. Therefore, WT is not only a suitable tool to analyse the vibrational features in term of modulation of the local structure of the solvation cage but also offers unique opportunities to study time-resolved spectroscopic experiments.

6.2 Propanediol in Acetonitrile

6.2.1 Experimental Setup

The absorption spectra of PDO in ACN (1:60 volume ratio) at room temperature were recorded in the 3300–300 nanometers region with a spectrophotometer equipped with near-IR and visible light sources. Infrared spectra of PDO in ACN and as a pure liquid were recorded in the 2700–3700 cm^{-1} region. Raman spectra of ACN in various solvents and at various concentrations at room temperature were recorded using the 514.5 nanometers line of an Ar^+ laser. The overtone spectra of PDO in ACN reported in fig. 6.16 were obtained by subtracting the contribution of the pure solvent.

6.2.2 Computational Chemistry Details

Car-Parrinello Molecular Dynamics Simulation

I performed *ab initio* Molecular Dynamics simulation with the CPMD package [CPM] on a system made up of 32 ACN and 1 PDO molecules in the microcanonical (NVE) ensemble. I simulated the sample in a cubic box of edge length 14.6983 Å (to match the room temperature density of pure ACN), with periodic boundary conditions. BLYP exchange and correla-

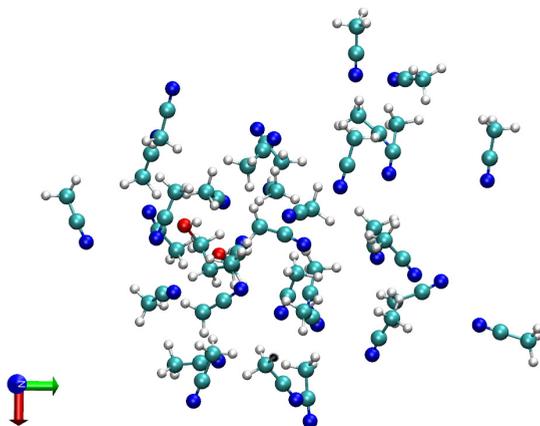


Figure 6.14: Snapshot extracted from the CPMD trajectory of PDO surrounded by 32 ACN molecules. The total number of atoms is equal to the simulation of PDO in water.

tion functional [Bec88, LYP88] was adopted, along with norm-conserving

Martins-Troullier pseudopotentials [TM91] and Kleinman-Bylander decomposition to describe core electrons. [KB82] Plane wave expansion of the electronic wavefunction was truncated at 85 Ry, while electronic fictitious mass was set to 400 atomic units (a.u.). The system was simulated for more than 26 picoseconds with a time-step of 4 a.u. (~ 0.096 femtoseconds) after a thermalisation run of about ~ 3 picoseconds at 300 K by velocity rescaling. The average temperature during the main production run was 295 ± 17 K.

Classical Molecular Dynamics Simulation

I performed classical molecular dynamics simulation with the FIST routine included into the CP2K suite of programs [CP2]. To model the interactions of PDO I employed the force field used by Chelli *et al.* [CPC⁺99] to simulate glycerol. For ACN I used the AMBER force field. [CCB⁺95] The simulated system is the same of the CPMD run reported in the previous paragraph. The classical simulation was 500 picoseconds long in the microcanonical (NVE) ensemble, with a time-step of 0.5 femtoseconds. Ewald summation was used to model electrostatic interactions. ESP charges were obtained with Gaussian 09 [FTS⁺b] at the BLYP/6-31G* level of theory. The production run was preceded by a 25 picoseconds long thermalisation with velocity rescaling.

Ab Initio Calculations

In order to validate the computational approach employed in the CPMD simulation, I performed static *ab initio* calculations with the Gaussian09 [FTS⁺b] package at various levels of theory (BLYP [Bec88, LYP88], B3LYP [Bec88, LYP88, Bec93b, Bec93a] and MP2) on isolated ACN and on a cluster made up of one PDO and one ACN molecule. I would have liked to use Gaussian03 [FTS⁺a] like in the calculations of PDO and EG in water, but, unfortunately, with Gaussian03 the geometry optimisation never reached convergence at the MP2 level of theory. All Gaussian09 calculations were performed with the 6-311++G(d,p) basis set, adopting very tight convergence criteria for both wavefunctions and geometrical optimisation, with an ultrafine integral grid. The structural results are summarised in table 6.4. All the computed harmonic vibrational frequencies are real and positive, ensuring that the systems are in their respective equilibrium configurations. The CPMD optimisation was carried out adopting the same

theory level and system	$r_{\text{CN}}/\text{\AA}$	$r_{\text{CC}}/\text{\AA}$	$r_{\text{CH}}/\text{\AA}$	$\alpha_{\text{N}\equiv\text{C}-\text{C}}/^\circ$	dip./D
ACN BLYP [†]	1.159	1.461	1.096	180.0	3.97
ACN BLYP [‡]	1.164	1.463	1.099	179.9	4.01
ACN B3LYP [‡]	1.165	1.463	1.099	180.0	4.05
ACN MP2 [‡]	1.174	1.463	1.092	180.0	4.30
Cluster (ACN+PDO) BLYP [†]	1.157	1.459	1.096	179.8	
Cluster (ACN+PDO) BLYP [‡]	1.163	1.461	1.099	179.8	
Cluster (ACN+PDO) B3LYP [‡]	1.151	1.454	1.092	179.8	
Cluster (ACN+PDO) MP2 [‡]	1.172	1.461	1.091	179.8	

Table 6.4: Structural parameters of ACN obtained after geometry optimisation at various levels of theory. [†] Optimisation performed with the CPMD code [CPM]; [‡] optimisation performed with the Gaussian 09 suite of programs [FTS[†]b]. The r_{CH} distance is the same for the 3 C–H bonds present in a ACN molecule.

cell dimension and parameters of the simulation run. As table 6.4 shows, there are only minor differences between BLYP CPMD and Gaussian09 calculations.

6.2.3 Intra-molecular H-Bond of Propanediol

The IR spectra of PDO as a pure liquid and in ACN solution are reported in fig. 6.15. The OH stretching band shows a more pronounced redshift in neat liquid PDO than in ACN solution, as a consequence of different hydrogen bond interactions. In ACN the band shows a well pronounced tail on the low frequency side. Fig. 6.16 shows that two components occur in both overtone bands. A deconvolution procedure employing pseudo-Voigt profile [OL77] (mixture of 90% Gaussian and 10% Lorentzian functions) points to the presence of two maxima located at ~ 6856 and ~ 7006 cm^{-1} for the first overtone and at ~ 10119 and ~ 10320 cm^{-1} for the second one. Adopting the usual empirical formula for overtones derived from the Morse diatomic oscillator model [KPTH96],

$$\nu/n = \omega - \omega x(n + 1) \quad (6.1)$$

where ν is the actual frequency, ω is the mechanical “harmonic” frequency, ωx the anharmonicity and n the vibrational quantum number (which is 2 for the first overtone and 3 for the second one), I obtain that the two fundamental bands should occur at ~ 3483 and ~ 3566 cm^{-1} .

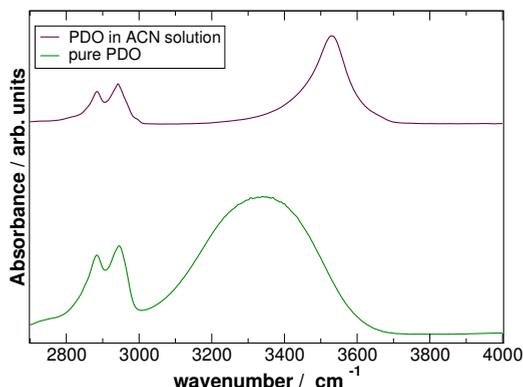


Figure 6.15: Experimental IR spectra of PDO in the CH and OH stretching region.

We extended the same approach to the analysis of the fundamental band around 3500 cm^{-1} ; a tentative fitting, adopting the same Voigt profiles previously described, gave two components at ~ 3485 and $\sim 3550\text{ cm}^{-1}$, as shown in fig. 6.17. The agreement with the data extrapolated from the overtone spectra is not complete ($\Delta\nu \sim 83\text{ cm}^{-1}$ from overtone calculation, $\Delta\nu \sim 65\text{ cm}^{-1}$ from fundamental band fitting); the different frequency separation in the two cases ($\sim 83\text{ cm}^{-1}$ and $\sim 65\text{ cm}^{-1}$) can be reasonably attributed to the different anharmonicity effects in fundamental and overtone vibrations.

A spectral behaviour similar to what observed here for PDO in ACN has been reported for gas phase PDO. In the latter case, Cheng *et al.* [CCT11b] reported and calculated the IR spectra for the second and third OH stretching overtones, finding a good agreement with the experimental data. They interpreted their results on the basis of intra-molecular H-bonding, associating the lower and the higher frequency bands to “bonded” and “free” conformations, respectively. Analogously, recent papers by Chen *et al.* [CCT11a] for 1,5-pentanediol, 1,6-hexanediol, and by Hazra *et al.* [HKS11] for peroxyacetic acid found evidences of intra-molecular H-bonds in the gas phase studying the overtone spectra.

Ref. [GTK12] reports on a computational investigation of the intra-molecular bonding in 9-hydroxy-9-fluorenicarboxylic acid. In Ref. [CCT11a] the reported frequency difference between the two components of the sec-

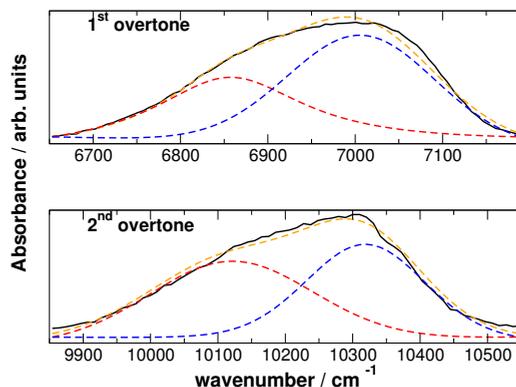


Figure 6.16: Overtone spectra of PDO in ACN. Black lines are the experimental spectra, red and blue lines are the fitting bands and orange lines represent the total fitted spectra.

ond overtone is $\Delta\nu_{0-3} = 310 \text{ cm}^{-1}$, of the same order of magnitude of the value (210 cm^{-1}) that I obtained from the fitting of the spectra in ACN solution. The occurrence of intra-molecular hydrogen bond was also recently reported in other similar systems, both from experimental results and from *ab initio* calculations [VFMC11,DPL11].

An intra-molecular H-bond is also present in our CPMD simulations of PDO in ACN, as well as an inter-molecular H-bond between one diol OH group and the nitrogen atom of a solvent molecule.

A relevant aspect of the simulation trajectory in the phase space has to be stressed. The two OH groups of PDO are obviously chemically equivalent and, in the ergodic limit of an infinitely long simulation, should exhibit the same behaviour. In fact, during the simulation run both groups, labelled OH(1) and OH(2), form and break a H-bond with the CN group of the solvent molecules (see fig. 6.18B). On the contrary, of the two possible intra-molecular H-bonded configurations, *i.e.* OH(1) donor to OH(2), and OH(2) donor to OH(1), only the first conformation happens to be realized in the 26 picoseconds run. As shown in fig. 6.18A, OH(1) forms a "bridge" with OH(2); the latter, in turn, may be involved as a donor in a H-bond to an ACN molecule. This "chained" H-bond configuration, illustrated in fig. 6.18A, is actually observed in the first ten picosecond of our simulation. This behaviour has to be ascribed to the limited length

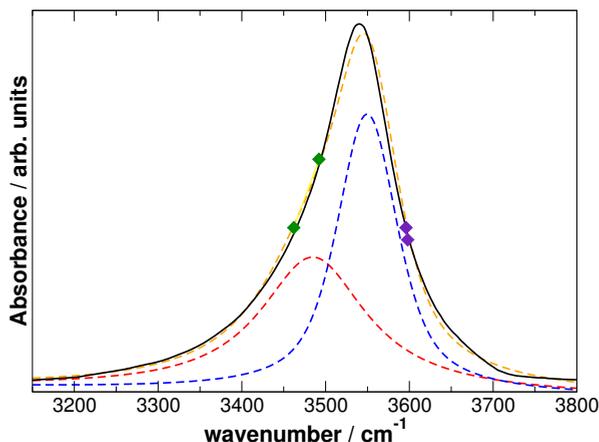


Figure 6.17: OH stretching fundamental band of PDO in ACN. Black line is the experimental spectrum, red and blue lines are the fitting bands and orange line represents the total fitted spectrum. The two green and two violet points are the harmonic frequencies obtained from normal mode calculations performed with Gaussian 09 at BLYP/6-311++G(d,p) level of theory on two model system made of one PDO and two ACN molecules forming one intra and one inter-molecular H-bond (green points) or two inter-molecular H-bond (violet points); intensities are not in scale and are reported on the experimental spectrum for better visibility.

of the simulation and to the fact that the sample contains only one PDO molecule. The resulting statistical limitation, indeed, does not allow us to make any quantitative estimate of the relative abundances of the different intra- and inter-molecular conformations. Nevertheless, it does not prevent us from pointing out the structure-frequency correlations that are the main objective of our work. Actually, the different roles played by OH(1) and OH(2) in the simulation allow treating the two groups as two different sites and to easily distinguish the contributions to the spectral features of the different structural realisations.

In order to obtain an estimate of the statistical significance of the conformational exchange resulting from the CPMD run, I also performed a classical MD simulation on the same system, extending the run to 500 ps (~ 20 times longer than the CPMD run). The classical trajectory shows a

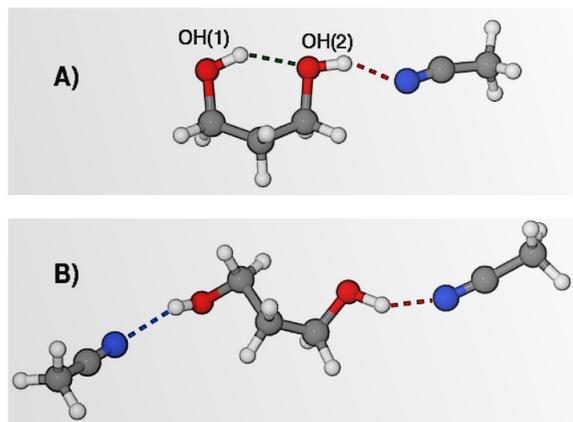


Figure 6.18: Two snapshots extracted from the CPMD trajectory; panel (A) shows an intra-molecular H-bond between the OH(1) \cdots OH(2) (green dotted line), and an inter-molecular H-bond between OH(2) and one acetonitrile molecule (red dotted line); panel (B) shows both OH groups engaged in inter-molecular H-bonds (red and blue dotted lines).

total of 34 “switching events” involving both OH groups,¹ corresponding to an average switching rate per hydroxyl group of ~ 0.9 events in 25 ps, in qualitative agreement with the CPMD simulation. This is displayed in fig. 6.19. However, I have to warn that adopting different choices for the force-field, different behaviours are found, suggesting, once again, that classical MD simulation results should be handled with much caution.

¹In the case of this classical MD run, for the sake of simplicity, I have considered the two hydroxyl groups H-bonded if at least one of the two possible OH \cdots OH distances is below the cutoff value of 2.5 Å.

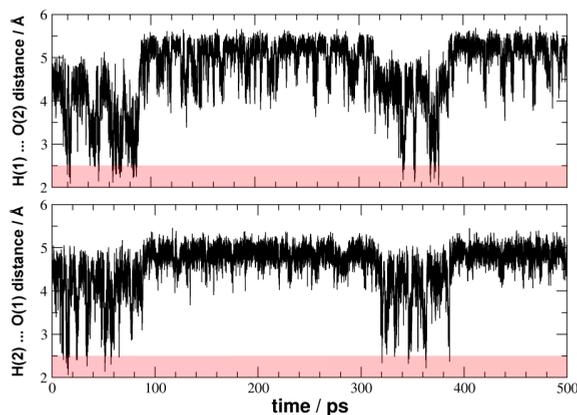


Figure 6.19: Intra-molecular H-bonds distances obtained from a classical MD run. There are ~ 34 H-bonding events. The region coloured in light red denotes the criteria adopted to consider PDO engaged into intra-molecular H-bonds.

Having assessed the plausibility of the intra-molecular H-bond, I now go on with the discussion of the CPMD simulation. Also the unnormalised and normalised pair radial distribution functions ($h(r)$ and $g(r)$, respectively) for the $\text{OH} \cdots \text{N}$ interactions involving the two OH groups (see fig. 6.20) show the signature of the different behaviours of the two moieties. It is evident that OH(1) forms H-bond with the solvent less frequently than OH(2). The average bond length of $\text{O}-\text{H}(1)$ ($\langle r_{\text{OH}(1)} \rangle$) throughout the simulation run is $\sim 0.983 \text{ \AA}$, whereas $\langle r_{\text{OH}(2)} \rangle$ is $\sim 0.994 \text{ \AA}$. This suggests that OH(2) covalent bond is weaker than OH(1), in qualitative agreement with the fact that OH(2) can be engaged simultaneously in H-bonding with two different partners (as an acceptor with OH(1) and as a donor with a ACN molecule). A more quantitative estimate can be extracted from the radial distribution functions of the $\text{OH} \cdots \text{OH}$ pairs of the PDO molecule, as reported in fig. 6.21. The maximum at about 2 \AA for the $\text{OH}(1) \cdots \text{OH}(2)$ interaction (blue line) and its absence for the $\text{OH}(2) \cdots \text{OH}(1)$ pair (red line) show that the intra-molecular H-bond actually occurs only in one direction, as shown in fig. 6.18A.

This explains the differences seen in fig. 6.20: in fact, summing up the values for all the unnormalised $h(r)$ functions for each OH sites, a very similar distribution is obtained. We also want to point out that a direct comparison between the radial distribution functions of fig. 6.20 and fig. 6.21 should only be restricted to the $h(r)$ functions, since the normalisation factors adopted for the two $g(r)$ are different. In fact, for the two $\text{OH} \cdots \text{N}$

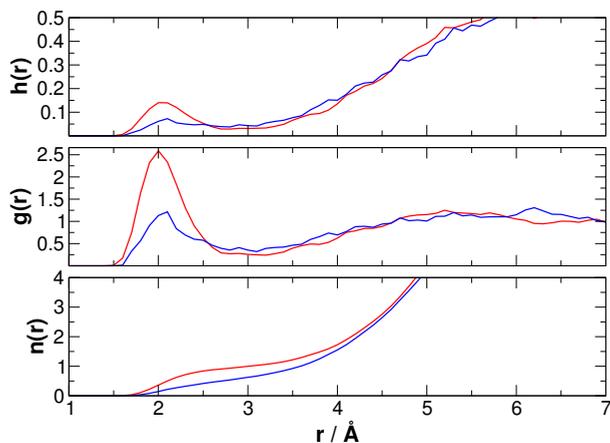


Figure 6.20: Radial distribution functions for the OH \cdots N pairs. Red and blue lines are associated to the two different OH groups of propanediol. OH(1) site in blue, OH(2) in red.

$g(r)$, the normalisation factor corresponds to all the possible OH(i) \cdots N pairs, which are 32, whereas for the OH \cdots OH $g(r)$ the number of all possible OH(i) \cdots OH(j) pairs is 1.

In order to extract the H-bond dynamics from MD simulations, I used the F^{HB} function [PCRS03,PCS05,FPCS06,CPLR09,PMMC⁺10,PMMC⁺11] already employed for PDO and EG in water. In the upper panel of fig. 6.22 each horizontal line corresponds to a different solvent molecule (the molecular index); the blue and red bars represent the occurrence of solute-solvent inter-molecular H-bond, involving OH(1) and OH(2). In the lower panel the value of F^{HB} for the intra-molecular bond is reported as a function of time. The inter-molecular H-bond involving OH(2) is clearly more stable, whereas that of OH(1) has a very short life time resulting in an intermittent behaviour. The intra-molecular H-bond is present for the first 10 picoseconds of the simulation time and is rather stable. The different solvation state of the two OH groups of PDO strongly affects the vibrational spectrum. The vibrational density of states (VDoS) is obtained from the CPMD trajectories as the Fourier transform of the atomic velocity autocorrelation function. In fig. 6.23, upper panel, I show the overall VDoS of the simulated system (upper panel) and the VDoS of the so-

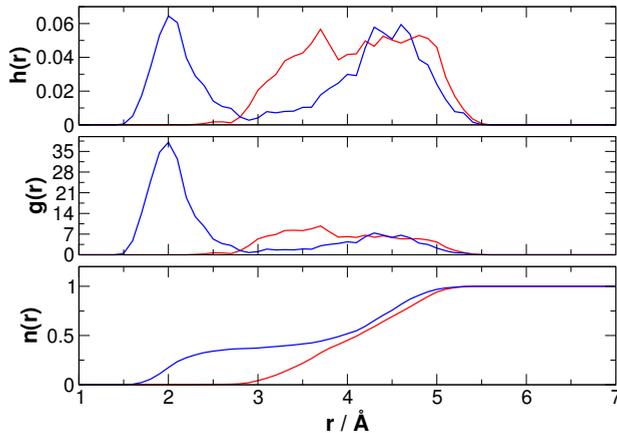


Figure 6.21: Radial distribution functions for the two $\text{OH}\cdots\text{HO}$ pairs. The $\text{OH}(1)\cdots\text{OH}(2)$ and $\text{OH}(2)\cdots\text{OH}(1)$ interactions are represented with blue and red lines, respectively.

lute only, obtained by restricting the Fourier transform to the atoms of PDO. A more detailed analysis can be done by restricting the VDoS calculation to the OH groups of PDO, which are the most affected by intra- and inter-molecular H-bonding. As shown in fig. 6.24, the VDoS of OH(2) is significantly redshifted with respect to that of OH(1). Quantifying this effect on the basis of Fourier analysis of the velocity autocorrelation function is rather difficult, as the resulting power spectra show a multitude of peaks in the OH stretching region (fig. 6.24, upper panel). WT analysis provides a convenient method of smoothing the calculated spectra [SLC03, Ehr02, SSZ⁺06], so that the maxima of the two VDoS can be located with better accuracy; fig. 6.24, lower panel, shows the resulting VDoS amplitude, consisting of the time-integrated spectrogram plotted along the frequency axis.

I have to point out that the wavelet code adopted to study the PDO/ACN mixture (and all other research studies exposed hereafter) is based on the Fourier space implementation (see Section 4.2.3).

The two main maxima are found at about ~ 3514 and ~ 3594 cm^{-1} , giving a frequency splitting $\Delta\nu \sim 80$ cm^{-1} , to be compared to the value (~ 83 cm^{-1}) obtained from the overtones using equation 6.1. Therefore,

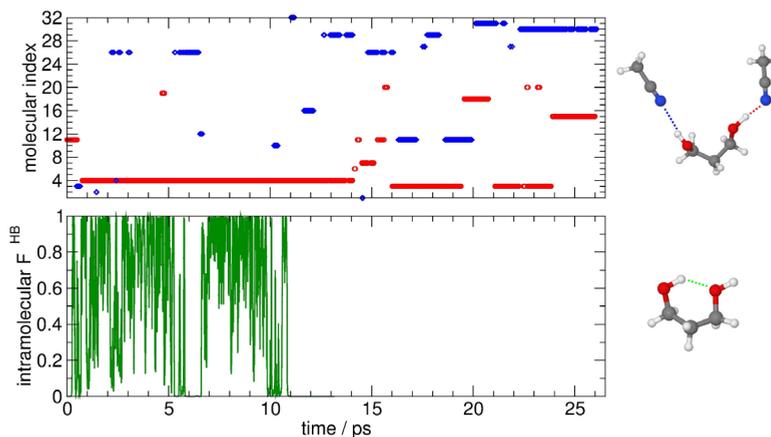


Figure 6.22: (upper panel) F^{HB} as a function of the ACN molecular index, a point is added if $F^{\text{HB}} \geq 10^{-3}$. Red points stand for OH(2) \cdots ACN interactions, whereas blue points stand for OH(1) \cdots ACN interactions; (lower panel) $F^{\text{HB}}(t)$ for the intra-molecular H-bond, depicted as green line and points.

CPMD simulation analysis confirms the results of the fitting of the experimental data and the extrapolated frequency splitting. Moreover, it suggests that the origin of the observed structured band in the OH stretching region is the different solvation dynamics and, in particular, the different H-bonding behaviour. This is analogous to what observed and calculated for PDO in the gas phase [CCT11b], with the remarkable difference that in the gas phase the system alternates between intra-molecular “bonded” and “free” conformations only, while in ACN solution there is, in addition, the possibility of an inter-molecular H-bond which, according to our simulation, involves at least one hydroxyl group for most of the time. Since only OH(1) acts as intra-molecular H-bond donor group during our 26 picoseconds long simulation (and, specifically, only for the first 10 picoseconds), it has been easy to separate the two “free” and “bonded” contributions.

The free energy difference between the intra-molecular “bonded” and “free” conformations can give an idea of the relative populations of the two conformations and of their contributions to the total spectrum. Table 6.5 reports the results of BLYP/6-311++G(d,p) calculations done with Gaussian09 [FTS⁺b] (very tight convergence criteria for wavefunction and geometric optimisation, ultrafine integral grid) on two model systems, consisting of a PDO molecule surrounded by two ACN molecules with

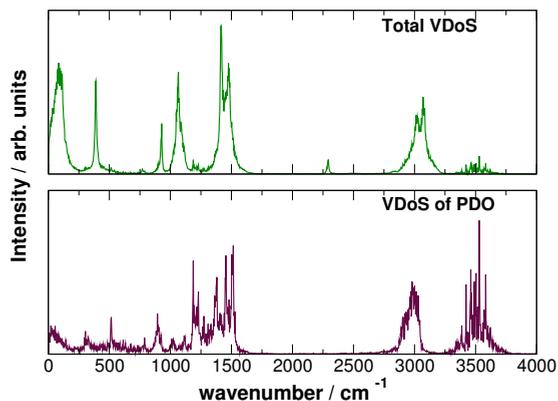


Figure 6.23: VDoS obtained from Fourier transform of the velocity autocorrelation functions for the simulated system (upper panel) and for the solute only (lower panel). Spectra are uniformly scaled by a factor of 1.055.

and without intra-molecular H-bonds, to estimate their relative difference in free energy, and thus their relative population in gas phase. The cal-

system	$r_{\text{O-H}(1)}/\text{\AA}$	$r_{\text{O-H}(2)}/\text{\AA}$	$\phi \text{ OH}(1)/^\circ$	$\phi \text{ OH}(2)/^\circ$	$\Delta G / \text{kJ mol}^{-1}$
“intra” cluster	0.985	0.984	-43.8	-84.9	0
“open” cluster	0.978	0.978	-81.0	-65.6	+1.02

Table 6.5: Comparison of structural and energetic features of a cluster of PDO and 2 ACN molecules. PDO in the cluster is engaged in two donor H-bonds with the two ACN molecules (“open cluster”) or with just one due to intra-molecular H-bonds (“intra cluster”). Structural parameters obtained after geometry optimisation Gaussian 09 suite of programs [FTS⁺b] adopting the BLYP/6-311++G(d,p) level of theory. Angles $\phi \text{ OH}(1)$ and $\phi \text{ OH}(2)$ are dihedral angles involving two carbon atoms of the PDO and OH(1) and OH(2), respectively.

culated geometric parameters of the two systems are collected in the first four columns of table 6.5. Within the limits of this simplified model, the relatively small free energy difference at 300 K (1.02 kJ/mole) suggests that the two conformers can both be present at room temperature. These two possible conformations correspond to specific normal modes whose frequencies, obtained from Gaussian09 [FTS⁺b] calculations, are reported in table 6.6 and displayed in fig. 6.17 as green (“intra” cluster) and violet

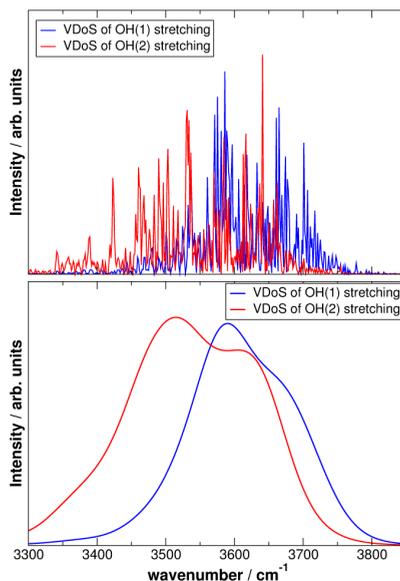


Figure 6.24: VDoS extracted from the trajectories of the hydroxyl groups only. Spectra are uniformly scaled by a factor of 1.055. Upper panel: spectrum calculated with Fourier Transform; lower panel: spectrum calculated by WT adopting the value of $\sigma = 2\pi$ for optimal smoothing.

(“open” cluster) points. It is worth noting that the frequency separation

system	ν_1/cm^{-1}	ν_2/cm^{-1}
“intra” cluster	3462	3490
“open” cluster	3596	3598

Table 6.6: Normal mode calculations performed with Gaussian09 [FTS⁺b] at BLYP/6-311++G(d,p) level of theory. Normal modes ν_1 and ν_2 stand for the two OH stretching normal mode.

between the two normal modes involving OH stretching is appreciable for the “intra” cluster, due to the different environments of the two hydroxyl groups. On the contrary, the frequencies for the “open” cluster are almost coincident and substantially higher than those of the intra-molecularly bonded cluster, since both OH are engaged in the same type of H-bond with ACN. The superposition of these two rather different spectral profiles, corresponding to the coexistence of the two conformers in the room temperature solution, is fully compatible with the broad and asymmetric

shape of the experimental spectrum (see fig. 6.17)

The correlation between H-bonding and O–H stretching frequency can be appreciated further exploiting the time-frequency analysis provided by wavelets, along with the F^{HB} function. For the analysis reported in fig. 6.25 I adopted a different choice for the σ_r and σ_θ parameters defining the F^{HB} function according to equations 3.9 and 3.10, in order to appreciate also minor changes in the solvation environment. In particular, I put σ_r equal to 2 and σ_θ to 10.0 degrees, twice its usual value (see Section 3.1.3 for details). The contribution of the intra-molecular H-bond function (see fig. 6.22, lower panel) is added to OH(1) and OH(2) inter-molecular F^{HB} functions, since both groups are involved in the bond. The correlation of the instantaneous frequency to the solvation state can be appreciated in fig. 6.25, where I report the opposite of F^{HB} and the H-bond function as a function of the simulation time.

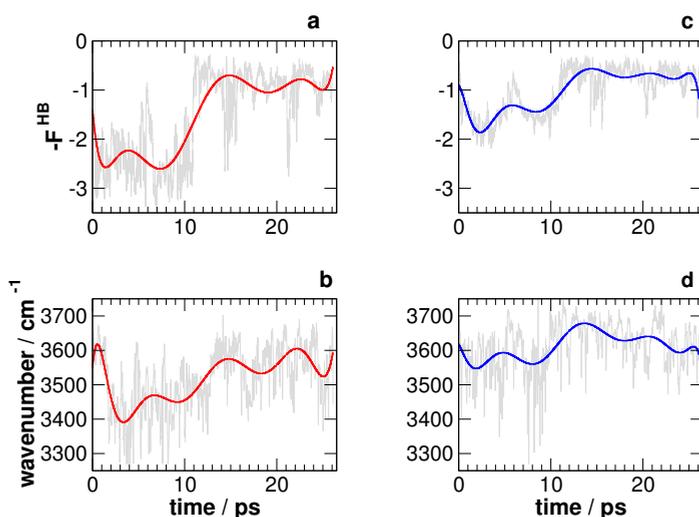


Figure 6.25: a) Hydrogen-bond function (changed in sign) $-F^{\text{HB}}$ as a function of time for OH(2) site; b) “instantaneous” O-H(2) stretching frequency obtained by WT with $\sigma = 2$; c) $-F^{\text{HB}}$ as a function of time for OH(1) site; d) “instantaneous” O-H(1) stretching frequency obtained by WT with $\sigma = 2$. Gray lines correspond to the actual values found during the simulation; the red and blue lines, drawn as a guide for the eye, were obtained by a 10th order polynomial fit of the raw datasets. The frequencies are uniformly scaled by a factor of 1.055.

The results summarised in fig. 6.25 are generally consistent with those shown in fig. 6.22 and 6.24: the broader frequency range spanned by VDoS of OH(2) and the lower value of the mean frequency correspond to the larger values (≥ 2) attained by F^{HB} for this group, which is engaged simultaneously in intra- and inter-molecular H-bonds for the first 10 picoseconds of the simulation. It is of some interest to correlate the instantaneous OH frequencies to the shortest OH \cdots N distance at every simulation time-step as done for the diols in water. The results of this analysis performed for PDO in ACN are shown in fig. 6.26. The spectrograms

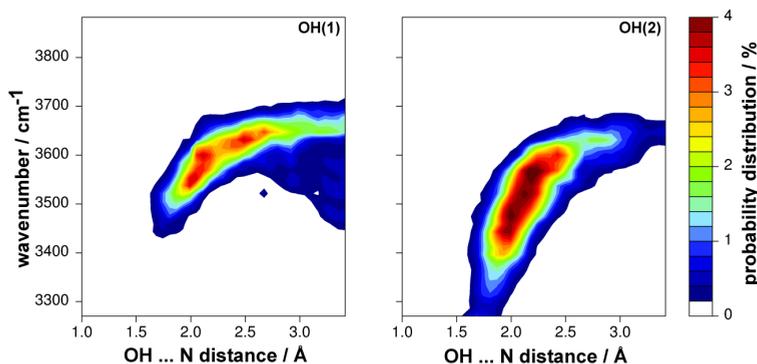


Figure 6.26: Correlation spectrograms between the OH stretching frequency and lowest OH \cdots N distance; a) OH(1); b) OH(2). Frequencies are uniformly scaled by a factor of 1.055. WT parameter σ has been put equal to 2.

have the usual banana shape; as expected, the OH(1) VDoS is distributed over a larger interval of OH \cdots N distance, and it thickens also out of the bonding region (*i.e.* for $r_{\text{OH}\cdots\text{N}} \geq 2.5$ Å). In contrast, the distribution of OH(2) frequencies is redshifted and it is practically all included in the bonding region.

In conclusion, the peculiar spectral features in the region of the OH stretching vibrations were interpreted as due to the presence of concurring intra- and inter-molecular hydrogen bonds, with the acceptor hydroxyl stretching frequency more redshifted than that of the donor. The dynamics of acetonitrile in the presence of the glycol solute can be seen as a case of “chemical exchange” widely studied by means of 2D IR spectroscopy.

6.2.4 ... and What About the Nitrile Stretch Blueshift?

The dynamical response of the $C\equiv N$ group of ACN to H-bond formation is completely opposite to that of the hydroxyl group of PDO. In general, the occurrence of H-bond causes a redshift of the OH stretching frequency. The presence of an intra-molecular H-bond (as in the case of PDO in ACN) can make the interpretation of the spectra more complicated, but the overall effect on the OH frequency is well established. The $C\equiv N$ stretching band, conversely, is blueshifted when nitrile is engaged as acceptor of H-bond. Experimental works by Kim and Hochstrasser reported this somewhat unusual feature for several different and heterogeneous nitrile mixtures (ACN in methanol [KH05, KH09]; benzonitrile, cinnamonnitrile and HIV-1 RT/TMC278 complex, which contains a cyanovinyl nitrile group, in aqueous solution [KH09]) by 2D IR experiments. In particular, for ACN in methanol they observed a $\Delta\nu$ separation of $\sim 8\text{ cm}^{-1}$ between “free” and H-bonded CN stretching frequency [KH05]. This behaviour can also be ascertained by measuring the Raman spectra of ACN in various solvents, as reported in fig. 6.27 (the small peak at about $\sim 2300\text{ cm}^{-1}$ is just a combination band of CC stretch and CH bend of ACN [KH05]).

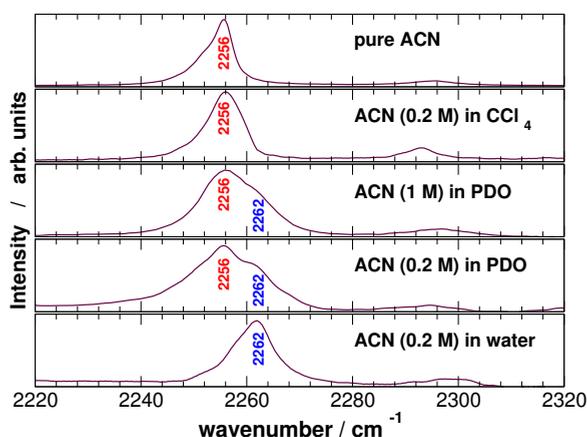


Figure 6.27: Experimental Raman spectra of ACN in various solvent.

In non H-bonding solvent (CCl_4) the $C\equiv N$ stretching frequency is about $\sim 6\text{ cm}^{-1}$ lower than in water. The same could be said for pure ACN, whose spectral response is overall similar to that of PDO in CCl_4 .

Even more interesting, using PDO as a solvent I observe a somewhat intermediate behaviour, with a main maximum at 2256 cm^{-1} but at the same time a clear shoulder at about $\sim 2262\text{ cm}^{-1}$. The shoulder becomes more evident with the increase of the $\frac{[\text{PDO}]}{[\text{ACN}]}$ molar ratio. This experimental result suggests that, in the presence of PDO, the ACN molecule gives rise to a sort of “chemical exchange” like that observed by Kim and Hochstrasser in References [KH05] and [KH09] for ACN in methanol. The experimental conditions of these measurements (ACN in PDO solvent) are different from those of our simulations (PDO in ACN solvent), but the dynamic response of the $\text{C}\equiv\text{N}$ stretching when involved in the H-bond can be assumed to be comparable. The frequency calculations for free and H-bonded ACN, performed with the DFT and MP2 approach (see Section 6.2.2), provide a reliable support to the interpretation of the shoulder clearly visible at 2262 cm^{-1} in the $\text{C}\equiv\text{N}$ stretching band of ACN peaked at 2256 cm^{-1} .

The BLYP calculation predicts for the $\text{C}\equiv\text{N}$ stretching mode the frequency of 2262 cm^{-1} for unbound ACN, and of 2276 cm^{-1} for ACN bound to PDO, with a frequency shift of $\Delta\nu = 14\text{ cm}^{-1}$. Very similar results are obtained from B3LYP and MP2 calculations, as shown in table 6.7. Although the frequency difference is larger than the experimental value, it is remarkable that all the calculations predict, in agreement with the experiment, an upshift of the $\text{C}\equiv\text{N}$ frequency when the group is engaged in H-bonding.

In other words, both experiments and calculations point to an increase of the strength of the $\text{C}\equiv\text{N}$ bond when ACN is H-bonded.

	BLYP	B3LYP	MP2
$\Delta\nu_{\text{C}\equiv\text{N}}$	14 cm^{-1}	13 cm^{-1}	15 cm^{-1}
$\Delta r_{\text{C}\equiv\text{N}}$	0.0018 \AA	0.0015 \AA	0.0014 \AA

Table 6.7: Difference between the $\text{C}\equiv\text{N}$ stretch frequency of H-bonded and free ACN at various levels of theory, performed with the Gaussian09 suite of programs. [FTS⁺b]

A confirmation of this picture comes from the analysis of the electron density obtained by calculating the Maximally Localised Wannier Functions (MLWF) centres [MV97, SMVP98] (see Section 3.2). In particular, I extracted two configurations from the simulated trajectory and performed

static *ab initio* calculations of the MLWF using the CPMD code [BMR⁺00]. The positions of these centres represent the most probable locations where to find electron pairs. The distance between the nitrogen electron lone pair and the three bonding electron pairs of the triple C≡N bond increases by $\Delta r_{e-e} \sim 0.01 \text{ \AA}$ when nitrogen is engaged as an acceptor in H-bond with the hydroxyl groups of PDO. [Wea79] This difference is certainly small in absolute terms, but the frequency splitting of the vibrational band is small too. We notice that this difference is consistent with the results of the Gaussian 09 calculations (irrespective of the level of theory) shown in table 6.4: the distance between C and N atoms is slightly reduced when ACN acts as H-bond acceptor from PDO. To check the reliability of our calculation, I have also computed, using CPMD, the dipole moment of bonded and free ACN: I found a value of 3.97 D for isolated ACN and of 5.01 D for hydrogen-bonded ACN: the isolated ACN value is close to the experimental value of 3.91 D measured in the gas phase. We used the CPMD program to calculate also the Electron Localisation Functions [BE90] (ELFs) (see Section 3.2), collected in fig. 6.28, to characterise the electron density in the region of the C≡N bond. The figure shows that a change in the shape of the electronic probability density decreases on the nitrogen atom upon formation of the PDO-ACN hydrogen bond.

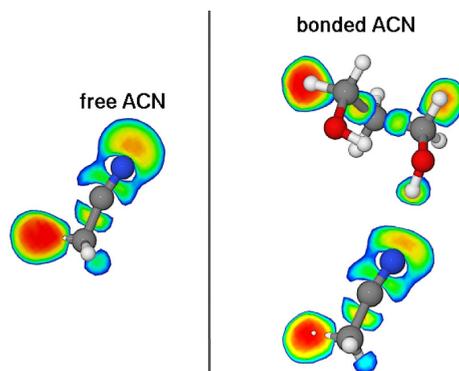


Figure 6.28: Contour lines of the ELF along the plane passing through the N, C, C atoms of ACN, for free and isolated ACN (left) and with the environment and PDO (right).

Both MLWF centres and ELFs suggest that some charge spreading actually occurs along the N \cdots OH line, mainly due to the electron density displacement from the nitrogen lone pair toward the PDO molecule. Consequently, a decreased electronic repulsion results between the three

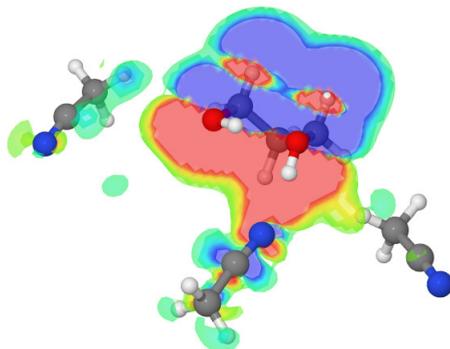


Figure 6.29: Difference between the electron density distributions of free and hydrogen bonded ACN, along the plane passing through the C_{ACN} , N_{ACN} , $H_{PDO}^{hydroxyl}$ atoms.

bonding pairs of the $C\equiv N$ triple bond. A support to our interpretation is obtained by taking the difference between the electron density distributions calculated for free and hydrogen bonded ACN, reported in fig.6.29. The blue regions denote an increase in the electron density, whereas the red ones correspond to a decrease. A charge transfer from the N atom toward the hydroxyl group evidently takes place, and the electron density along the nitrile triple bond increases due to the reduced repulsive interaction. This effect corresponds to the strengthening of the $C\equiv N$ bond force constant, resulting in the upshift of the corresponding vibrational frequency.

Recovering the $C\equiv N$ Stretch Blueshift with CPMD

My CPMD simulation does not reproduce the small frequency blueshift of the $C\equiv N$ band: as a matter of fact, weighting the VDoS by $F^{HB}(t)$ function and its complementary as we have done in Ref. [PMMC⁺10] for methyl acetate in methanol (see following Chapter 7), “free” and H-bonded contributions to the VDoS appears at about the same frequency of ~ 2260 cm^{-1} . To find the reasons of this inability to reproduce the spectral features of ACN, many possibilities have been taken into account.

In Ref. [AL07] and [TJM⁺05], effects due to the formation of “head-to-tail” complexes (mainly due to improper H-bonds) of ACN molecules are included in the $C\equiv N$ stretching elucidation. However, the formation of such complexes in my CPMD simulation is rather sporadic and apparently uncorrelated to the instantaneous $C\equiv N$ stretching frequency as calculated

by WT.

To check the influence of the exchange-correlation functional adopted, I performed another simulation (hereafter called PBE^I), adopting the same procedure and parameters described in Section 6.2.2, albeit employing PBE exchange-correlation functional [PBE96], because recently Thanthiriwatte *et al.* [THBS11] found that this choice appears to lead to better results in H-bonded systems. Simulation PBE^I reproduces the same structural and spectroscopic results discussed above: the intra-molecular HB is formed and the spectral shift for OH(2) is recovered, but a blueshift for C≡N stretch is still not observed.

Having asserted that this inconsistency probably is not functional-specific for PBE nor BLYP,² I performed a third CPMD simulation (simulation PBE^{II}) adopting the same procedure and parameters of simulation PBE^I, but this time thermalised at a temperature of 180 ± 15 K. At this much lower temperature, H-bonds no longer are broken nor formed. PDO only interacts with the same two ACN molecules and no intra-molecular H-bond is formed (the starting configuration that we have adopted did not have intra-molecular H-bond as for the other two simulations). The two inter-molecular H-bonds are stable for the entire simulation run, as summarily described by fig. 6.30. As a result, the VDoS of the two hydroxyl groups are completely overlapped, because they display the same H-bonding behaviour. Comparing the VDoS of the 30 “free” ACN molecules against the two H-bonded ones, a blueshift of about $\sim 8 \text{ cm}^{-1}$ is appreciable, as reported in fig. 6.31, while the main features of the three Car-Parrinello simulations are summarised in table 6.8. I performed fur-

	T	intra-molecular HB of PDO	blueshift of H-bonded CN groups
BLYP	300K	yes	no
PBE ^I	300K	yes	no
PBE ^{II}	180K	no	yes

Table 6.8: Summary of the three CPMD simulations performed.

ther analysis with WT, adopting the same procedure discussed for PDO, and the resulting spectrogram for the two H-bonded ACN molecules is

²This could also be inferred from the static *ab initio* calculation reported in the previous section, which were able to reproduce the blueshift using the BLYP functional

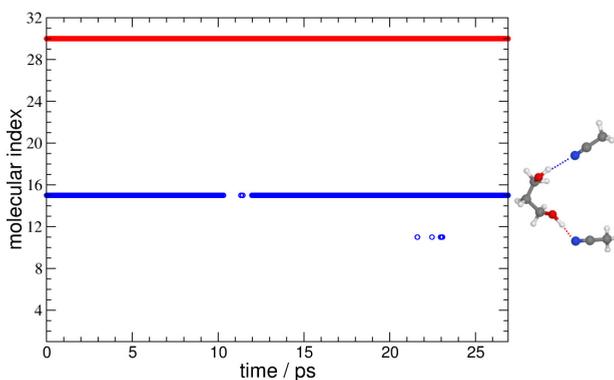


Figure 6.30: F^{HB} as a function of the ACN molecular index for PBE^{II} simulation, a point is added if $F^{\text{HB}} \geq 10^{-3}$. Red and blue points stand for the two different OH groups.

displayed in fig. 6.32. It can be appreciated that a “banana shaped” correlation is present, but inverted with respect to that of fig. 6.26: the decrease of the $\text{N} \cdots \text{HO}$ length is associated to a shift toward higher frequencies, recovering the known experimental behaviour. Thus, our results is that CPMD can, from a theoretical point of view, reproduce this unusual spectral feature. On the other hand, longer H-bond lifetimes are required to appreciate difference in frequency of just $\sim 4 \text{ cm}^{-1}$. Since $\Delta\nu = 1/2Tc$, with $\Delta\nu$ the frequency separation (in cm^{-1}), T the H-bond lifetime (in seconds) and c the speed of light in vacuum (in meters/seconds), a ~ 16 picoseconds long lifetime is required to correctly separate two peaks due to different H-bonding environments.

This can explain why both simulations performed at 300 K (BLYP and PBE^{I}) failed to recover the ACN blueshift, whereas simulation PBE^{II} at 180 K gives a frequency-distance correlation that is opposite to that of the OH stretching, as expected. Moreover, stochastic structural fluctuations in the solvation cage can induce frequency fluctuations large enough to hide (or compensate) the small blueshift induced by hydrogen bonding: a more stable H-bond ensures that these fluctuations have only minor impact onto the vibrational spectra.

I need to point out that the thermodynamic conditions of simulation PBE^{II} are obviously different from those of the experimental measure-

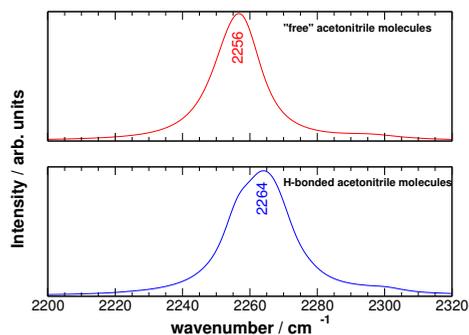


Figure 6.31: Calculated VDoS of the 30 free (upper panel) and the 2 H-bonded (lower panel) ACN molecules for the PBE^{II} simulation. Frequencies are uniformly scaled by a factor of 1.033. A Lorentzian smoothing has been adopted.

ments. In particular, at 180 K, acetonitrile should be a solid and therefore our conclusion regarding the blueshift recovery with CPMD have to be carefully considered: we have elucidated the causes of the incorrect vibrational response at room temperature performing a simulation at a lower temperature and in a different phase. Anyway, the idea of “freezing” the system is not so unusual in the field of MD simulations, as shown by recent works of Nicolini *et al.* [NC09,NFC11]. In these two latter references, the “freezing” of (part of) the molecular system was exploited to better sample the conformational space of a large biomolecule, whereas in my case I resorted to a low temperature simulation to slow the dynamics for a better sorting out of the effects affecting the C≡N stretching.

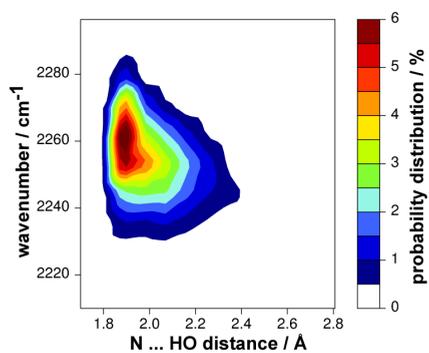


Figure 6.32: Correlation spectrogram between the CN stretching frequency and lowest N...HO distance. Frequencies are uniformly scaled by a factor of 1.033. WT parameter σ has been put equal to 2.

7. Methyl Acetate in Protic Solvents

Methyl acetate (MA) has been subject to detailed study at LENS laboratories for ultrafast spectroscopy. MA is a molecule simple enough to work as a target for high accuracy experiments, and at the same time exhibits many features in common with biologically active molecules, such as amino acids.

The solvation dynamics of MA in heavy water was probed by two-dimensional IR experiments by Dr. M. Candelaresi (then at LENS, Univ. of Florence, now at the Univ. of Glasgow, UK) and by CPMD simulations performed by Dr. M. Pagliai (Chemistry Dept., Univ. of Florence) during year 2009 [CPLR09]. The C=O stretching IR band of MA splits into a doublet in water as a consequence of the H-bond interaction with the solvent, which leads to the equilibrium between two solvated species, consisting of one MA molecule bonded to one and two water molecules [CPLR09]. The combined approach consisting of 2D IR experiments and CPMD simulation proved useful to understand what happens to MA in water and how the environment affects the C=O stretching mode, just falling short to reconstruct the IR spectrum shape. In fact, on year 2009, wavelets were not yet in use at LENS nor at the Chemistry Dept. of Florence to perform time-frequency analysis.

I applied wavelets to the trajectories of MA in water and methanol provided to me by Dr. Pagliai to perform this last step of interpretation, allowing to immediately associate VDoS to a specific solute-environment interaction.

As mentioned above, the vibrational spectrum of MA changes significantly due to the type of solvent, in particular in the region associated to the C=O stretching [CPLR09], indicatively at $\sim 1700\text{ cm}^{-1}$. For example, in non-protic solvents like CCl_4 and ACN, the C=O stretching mode gives rise to just a single band in the IR spectrum, and the frequency in acetonitrile is lower than in CCl_4 . The interactions in protic solvents are much more complex, because the solvents can interact with MA through

H-bonds. The number of solvent molecules interacting through H-bonds in methanol and water is different, and more caution to approach the water case is required. Thus, since MA in water poses more computational challenges than MA in methanol, the latter shall be presented before the former.

7.1 Methyl Acetate in Methanol

Performing sub-picoseconds IR experiments, Banno *et al.* [BOY⁺09,BOT08] observed that the C=O stretching band of MA in methanol splits into a doublet, like Candelaresi *et al.* found in aqueous solvent [CPLR09], with an average stretching frequency for this vibration higher than in water. Banno *et al.* [BOY⁺09,BOT08] explained this behaviour by assuming an equilibrium between two molecular forms, corresponding to MA involved alternatively in zero and one H-bond with the solvent, interpreting the presence of a shoulder centred at 1706 cm^{-1} as a proof of MA interacting with two methanol molecules.

7.1.1 Computational Chemistry Details

The simulations were performed on a system made up by 1 MA and 64 methanol molecules in a periodic cubic box with sides 16.3095 \AA . BLYP exchange and correlation functional [Bec88, LYP88] was employed along with norm-conserving Martins-Troullier [TM91] pseudopotentials within the Kleinman-Bylander [KB82] decomposition. The plane wave expansion was truncated at 85 Ry, while the fictitious electronic mass was 600 a.u.. The system has been thermalised at 300 K for ~ 2.6 picoseconds by velocity scaling. The simulation (in the NVE ensemble) was ~ 20 picoseconds long, saving the coordinates at each integration time-step of 5 a.u. ($\sim 0.12\text{ fs}$). The average temperature was $315 \pm 10\text{ K}$.

7.1.2 Structural Data and Spectrum Interpretation

To elucidate the H-bond interaction, the most important structural parameters for the CPMD simulation of MA in methanol are summarised in fig 7.2. In fig. 7.2 we reported the pair radial and angular distribution functions for MA in methanol (panels a and b) and in water (panels c and d), taken as a reference (the simulation of MA in water has been performed with the same computational approach [CPLR09]). In particular,

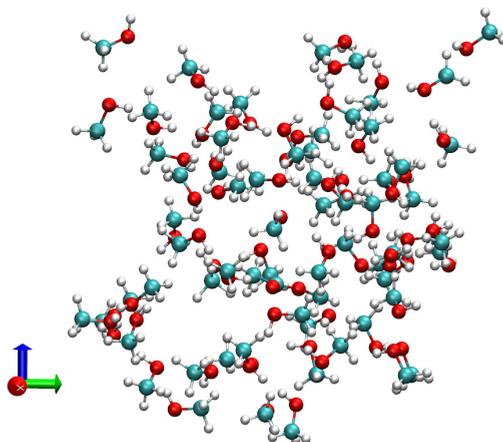


Figure 7.1: Snapshot of the simulation box of MA in CD₃OD.

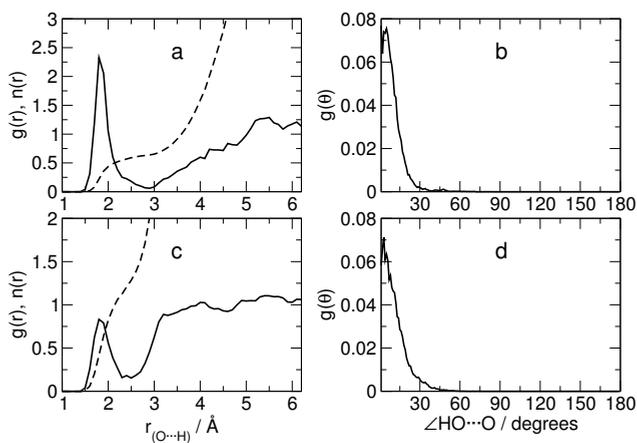


Figure 7.2: Panel a: pair radial distribution function (full line) and running integration number (dashed line) for MA in methanol; panel c: the same for MA in water. Panel b: angular distribution function for MA in methanol; panel d: the same for MA in water.

the a and c panels show the radial distribution functions for the O...H distance, whereas panels b and d show the distribution function of the HO...O angles limited to the H-bond interactions.

By comparison of the $g(r)$ functions (fig. 7.2 a and c) for MA in the two solvents, some peculiar differences in the H-bond region can be noted, related to the different number of solvent molecules that directly interact with the MA molecule (the coordination number, $n(r)$ in fig. 7.2 a and c). Analysing the $g(r)$ up to the first minimum (fig. 7.2 a and c), it is possible to correlate the structural parameters with the possibility of one solvent molecule to leave or enter the first solvation shell.

Since the accuracy of the $g(r)$ was challenged by the referees, we checked it performing subaverages, without observing differences in the positions of the first maximum and minimum. The H-bond evolution differs in time in the two subaverages, inducing slightly changes in the first peak height and in the value of the minimum deep in the $g(r)$. This is a check that basically none does as long as distribution functions are reasonable enough, and was asked to us to verify that our simulation reached an ergodic enough behaviour to have statistical meaning.

For the methanol solution the minimum of $g(r)$ at about 2.8 Å is much lower than in the water solution and, correspondently, the height of the first peak is more than twice higher. These two observations indicate that the solvent molecules in the first solvation shell are more localised in methanol than in water. In other words, the exchange rate of solvent molecules between the first and second solvation shell is definitely lower in the alcoholic solution. Thus, the H-bond formed by MA with methanol is then more stable and probably stronger than with water. As a consequence of the higher stability of the H-bond interaction, the angular distribution function is sharper in methanol than in water, as can be observed in fig. 7.2 b and d.

To further clarify this, Dr. Pagliai considered two simple models, consisting of a MA and 1 or 2 solvent molecules. The geometry of the clusters were optimised with DFT calculations at the BLYP/6-311++G(d,p) level of theory, adopting the Gaussian03 suite of programs [FTS⁺a]. Fig. 7.4 shows the calculated IR spectra of the two clusters in the C=O stretching region and in fig. 7.3 the IR spectra of MA in methanol and water are reported for comparison. A comparison with the calculated spectra of analogous clusters with water molecules is presented. It can be seen that the redshift of the IR band is directly correlated with the number of hydrogen bond

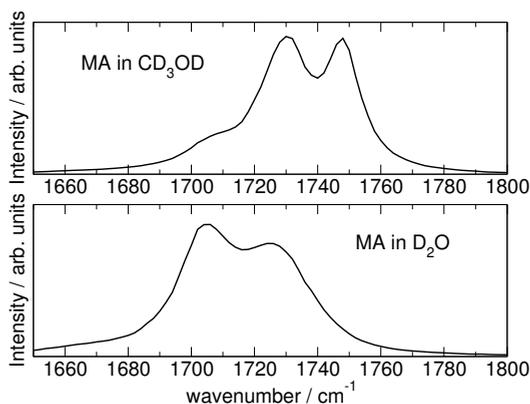


Figure 7.3: Experimental IR spectra of MA in methanol and water.

formed. In the case of MA forming one H-bond two bands are present at slightly different frequencies, corresponding to two distinct minimum structures, with the methanol or water molecule located on opposite sides with respect to the C=O bond.

For an accurate description of the H-bond dynamics of MA in methanol, we adopted for the analysis of our CPMD simulation the hydrogen bond function F^{HB} (see Section 3.1.3). The F_j^{HB} function assumes values close to 1 for strong H-bond interactions, whereas goes rapidly to 0 when the H-bond becomes weaker.

The combined analysis of the F^{HB} function with the instantaneous frequency computed by means of the WT provide a more descriptive visualisation of the H-bond effect on the vibrational frequencies. The time series analysed is represented by the variation of the C=O bond length during the CPMD simulation, thus allowing us to compute the frequency of the C=O stretching vibration at each time-step of the simulation. Please notice that this is an intra-molecular parameter, whereas WT has been adopted in the previous Chapter to correlate frequency with an inter-molecular distance.

The wavelet analysis of the C=O stretching vibrational mode allows obtaining more information compared to the standard Fourier analysis, and can be of valuable help in the spectroscopic and structural characterisation of the H-bond of MA in methanol solution. In fact, as can be appreciated in fig. 7.5, FFT reconstructed VDoS of the C=O stretch presents

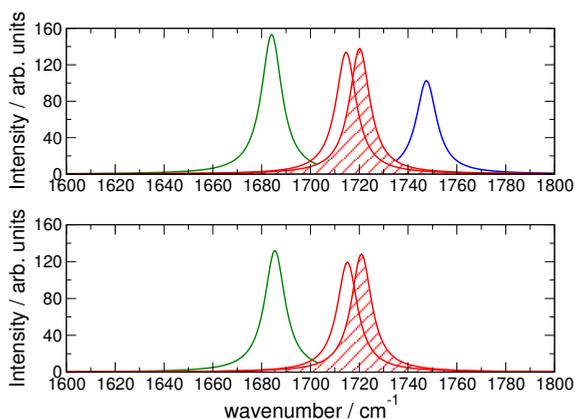


Figure 7.4: Calculated IR spectra for MA and clusters of MA with 1 or 2 methanol (upper panel) and water (lower panel) molecules in the C=O vibrational stretching region. The frequencies have been uniformly scaled by a factor 1.021. Blue, red and green line correspond to MA involved in 0, 1 and 2 H-bond, respectively. The two red bands refer to solvent/MA clusters of different geometry. The calculated spectra have been reported by assigning to each C=O stretching mode a Lorentzian shape with 10 cm^{-1} full width at half-maximum.

a lot of maxima and minima and is of very little help to understand the H-bond dynamics.

Fig. 7.6 shows the analysis performed with wavelet transforms in conjunction with the F^{HB} function. The result gives a clear description of the response of the C=O stretching mode to the variation of the H-bond strength. The presence of the H-bond interaction can be appreciated from the F^{HB} function reported in panel a of fig. 7.6. It can be seen that a hydrogen bond is present only in the first ~ 4 picoseconds of the simulation with molecule 58 of the solvent and after ~ 12 picoseconds with molecule 2. The behaviour of the F^{HB} function (F_2^{HB} and F_{58}^{HB}) shows the importance of a continuous function for characterising the H-bond, like F^{HB} , that accounts for the fast oscillation in and out of the geometrical constraints usually taken as a probe of the formation of a H-bond. In fact (see panel b of fig. 7.6), as long as the H-bonded molecule resides in the first solvation shell, the fast modulation of the C=O stretching frequency takes place around 1719 cm^{-1} (red dashed horizontal line). Instead, when no methanol molecule is H-bonded with MA, the C=O frequency oc-

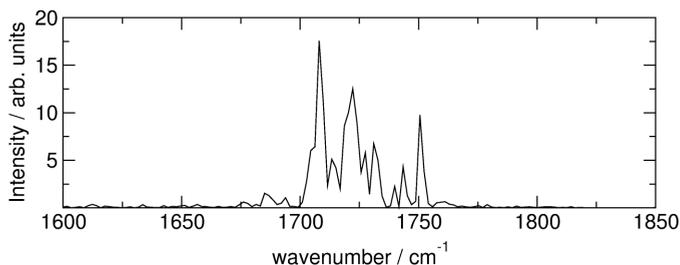


Figure 7.5: Power spectrum of MA in methanol obtained by FFT. Frequencies are uniformly scaled by a factor of 1.095.

curs at higher wavenumber 1750 cm^{-1} (blue dashed line). The interaction with the solvent molecules induces a redshift in the C=O stretching frequency of about 30 cm^{-1} , a value slightly higher than experimentally observed [BOY⁺09, BOT08].

WT allows to visualise this feature extracting “time resolved spectra” from the time-series of the C=O distances, as reported in fig. 7.7. Although no double H-bond configurations have been observed during the simulation, in some cases a second molecule approaches MA entering the first solvation shell region. This corresponds to the low frequency shoulder observed in the experimental spectrum. The C=O stretching frequency distribution can be appreciated in panel c of fig. 7.6, showing a bandwidth close to that measured in infrared spectra [BOY⁺09, BOT08]. An attempt to calculate the H-bond life time has been performed by computing the correlation function of F^{HB} obtaining an estimate of the order of 1 picoseconds which compares with the experimental relaxation time. [BOT08, BOY⁺09] We have applied other methods to compute the H-bond life time according to the definition given in Ref. [PCRS03]; the order of magnitude of the life-time does not change ranging in the values between ~ 0.5 and ~ 2 picoseconds.

An alternative approach to correlate the results obtained from the F^{HB} function with the vibrational frequencies of the C=O stretching mode of MA can be accomplished taking advantage of the time localisation capability of wavelet transforms. Actually, by this technique it is possible to extract at each time-step of the simulation the C=O bond length and the corresponding vibrational frequency.

Thus, we can draw a two-dimensional plot in which the distribution of the instantaneous frequencies is expressed in terms of a specific struc-

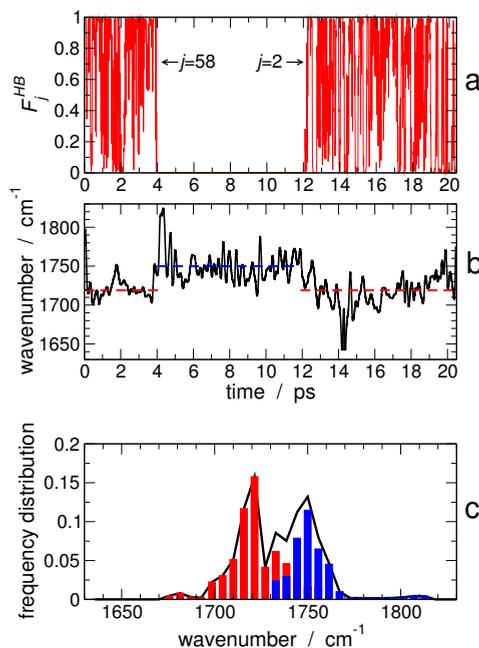


Figure 7.6: Panel a: Evolution of F_j^{HB} for the molecule with $j=2$ and 58. Molecule with $j=58$ forms an H-bond in the first part of the simulation, whereas molecule with $j=2$ interacts with MA in the last part of the simulation. Panel b: Time evolution of the maximum of the C=O stretching band during the simulation. The dashed lines represent the average frequency of C=O stretching mode in presence (red) or absence (blue) of H-bond. Panel c: C=O stretching frequency distribution for MA during the simulation. The red and blue bars indicate the frequency distribution of MA involved and not involved in H-bond interaction, respectively. The frequencies have been uniformly scaled by a factor 1.095

tural parameter. Fig. 7.8a shows the probability distribution of the C=O stretching frequency as a function of C=O bond length. As expected, the surface is characterised by the presence of two main regions, with the maxima separated by $\sim 30 \text{ cm}^{-1}$, corresponding to MA engaged or not engaged in the H-bond interaction. It is interesting to observe that the shape of the two regions is quite different: this allows to pinpoint the correspondence between structural changes and frequency values. The presence of the H-bond not only leads to a redshift of the C=O stretch frequency, but also makes the C=O bond to span a larger length range. This behaviour can be better appreciated in figures 7.8b and 7.8c, where a

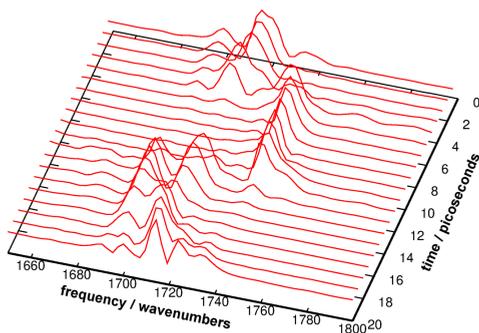


Figure 7.7: Wavelet spectra of MA in methanol as a function of time.

clear separation of the two spectral regions is obtained by weighting the probability distribution by F^{HB} and $(1-F^{HB})$, respectively.

The possibility to localise the vibrational modes in time and frequency domain by wavelet transform makes molecular dynamics simulations particularly effective in the study of H-bond dynamics and helpful in the comprehension of the experimental results obtained with time-resolved spectroscopies.

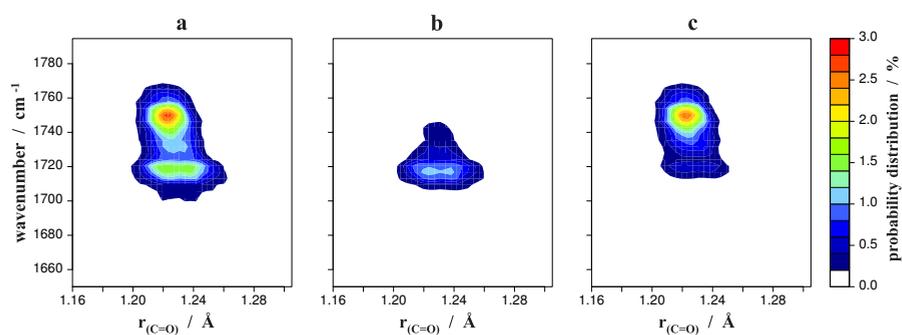


Figure 7.8: a) Probability distribution of the C=O stretching mode as a function of the C=O bond length. The b) and c) surfaces represent the probability distribution given in a), weighted by F^{HB} and $(1 - F^{HB})$ factor, respectively. The frequencies have been uniformly scaled by a factor 1.095

7.2 Methyl Acetate in Water

Simulation Details

The computational approach adopted is completely analogous to that employed in Ref. [CPLR09] and for studying MA in methanol, except for the simulation length (increased to 20 picoseconds, as for MA in CD_3OD), the plane waves cutoff (increased to 85 Ry), and the simulation cell size (the side is smaller, 12.6694 Å).

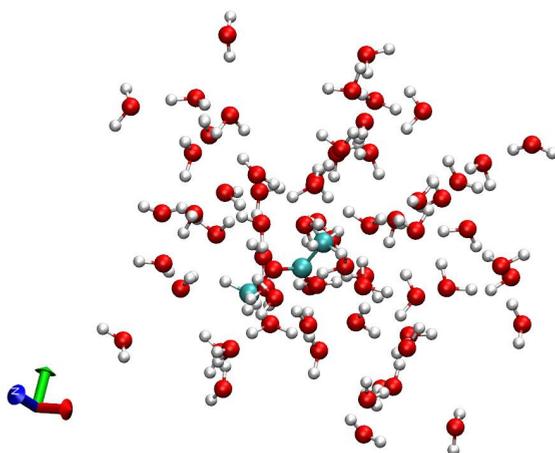


Figure 7.9: Snapshot of the simulation box of MA in D_2O .

7.2.1 Spectral Interpretation

Interpreting the vibrational spectrum of the $\text{C}=\text{O}$ stretch band of MA in water is actually more difficult than in methanol. In methanol, there is a “chemical exchange” between MA and an H-bonded cluster $\text{MA} \cdots \text{CD}_3\text{OD}$. In water, the “chemical exchange” is between the H-bonded cluster $\text{MA} \cdots \text{D}_2\text{O}$ and the double H-bonded cluster $\text{MA} \cdots 2 \times \text{D}_2\text{O}$. In fact, the O atom of the $\text{C}=\text{O}$ group of MA can act as H-bond acceptor with two H-bond donor molecules, and, while we have not seen this occurring in methanol, it happens in water.

This has deep consequences onto the vibrational spectrum, as can be appreciated in fig. 7.10. It can be seen that a doublet is prominent in both solvents. Deconvolution of the experimental spectra shows that in both cases a third, much weaker component is present; the three bands have

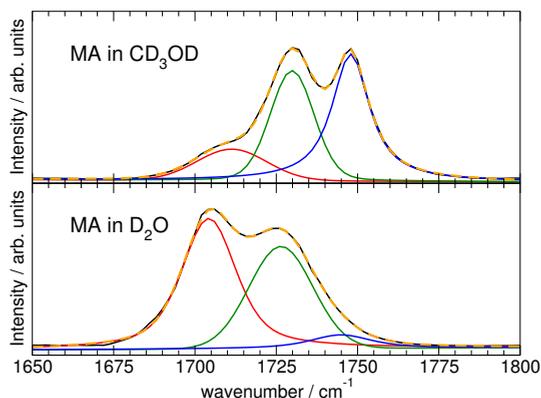


Figure 7.10: Experimental IR spectrum (black line) of MA in methanol (top) and water (bottom). The blue, green and red lines refer to MA H-bonded to 0, 1, and 2 solvent molecules, respectively. The orange dashed line represents to the resulting fitted spectrum.

rather similar frequencies in water and methanol solutions, while their relative intensities are remarkably different. As widely discussed in Refs. [BOT08, BOY⁺09, CPLR09], the three bands are attributed, in decreasing frequency order, to MA H-bonded to zero, one and two solvent molecules. The relative intensities of the bands show that MA in water is, in large prevalence, bonded to one or two solvent molecules, while in methanol most MA molecules form zero or one H-bond. This is summarised in table 7.1.

number of H-bonds	CD ₃ OD	D ₂ O
0	1748 cm ⁻¹	1745 cm ⁻¹ (weak)
1	1730 cm ⁻¹	1727 cm ⁻¹
2	1711 cm ⁻¹ (weak)	1704 cm ⁻¹

Table 7.1: Observed frequencies for MA in methanol and water, respectively. I have obtained the frequencies by a fitting procedure adopting a combined Gaussian (90%) and Lorentzian (10%) functions.

The different solvation number also affects the distance/frequency spectrogram for the C=O stretching, reported in fig. 7.11.

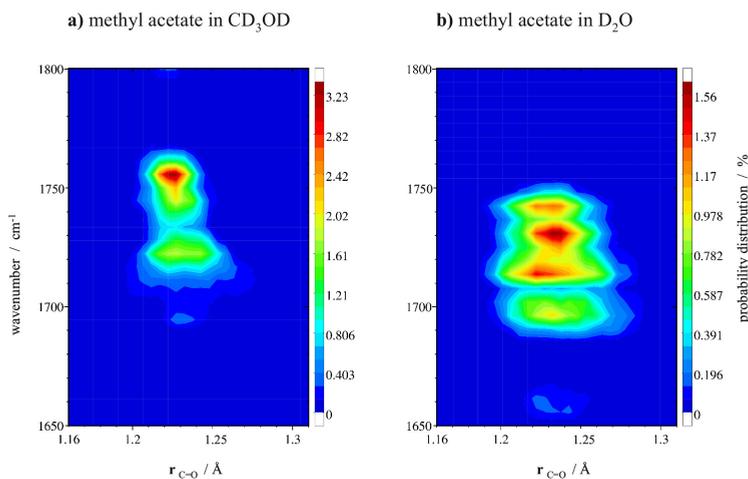


Figure 7.11: Probability distribution of the C=O stretching mode for MA in (a) methanol and (b) water, as a function of C=O bond length. A σ value of 4π has been used.

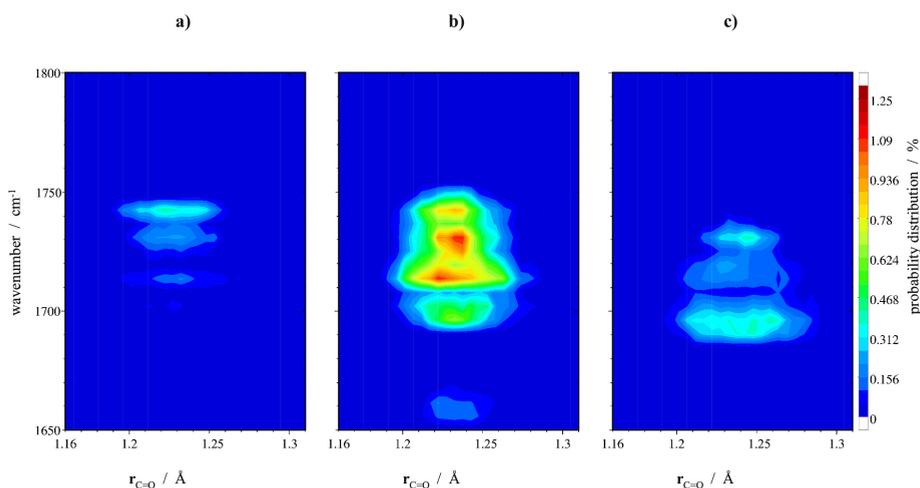


Figure 7.12: Probability distribution of the C=O stretching mode for MA in water interacting with (a) 0, (b) 1, and (c) 2 water molecules, as a function of C=O bond length. A σ value of 4π has been used.

The frequency distribution is significantly more complicated in water than in methanol, and in order to separate the various contribution the F^{HB} function has to be once again adopted.

A detail of the probability distribution for three different solute-solvent

interactions of MA, H-bonded with 0, 1 and 2 water molecules, is shown in fig. 7.12. These isosurfaces have been obtained starting from fig. 7.11, panel b, counting the configurations with values of F^{HB} lower than 10^{-4} (fig. 7.12, panel a) or with one or two values of F^{HB} greater than 10^{-4} (fig. 7.12, panels b and c, respectively); the overall pattern is in agreement with the spectral interpretation given above.

7.3 Discussion

This research showed that wavelets can be used to correlate the stretch frequency of a localised normal mode (in these cases, the C=O stretch) with an inter-molecular structural property (the bond length). The WT allowed to split the frequency doublets into components due to specific H-bonding arrangements, whereas standard Fourier analysis was not even able to show us a clear doublet (see fig. 7.10).

The case of MA in methanol is particularly lucky, because in that solvent MA forms alternatively just 1 or 0 H-bonds: it is a sort of binary choice, and wavelets completely elucidated the solvation behaviour. In water it is more difficult to sort the VDoS into the various H-bonding contribution, and the resulting spectrograms (fig. 7.12) require more caution to be interpreted, but the agreement is still qualitatively very good.

Remembering that MA was originally chosen as a target for 2D IR experiments due to its similarity with amino acids, this wavelet-based approach has great potential to be extended to biomolecules.

8. Water Solvation of Thiazole

The adsorption of molecules on metal surfaces is useful to understand reactions of heterogeneous catalysis or electroactivation in electrochemical cells. These processes can be probed and studied by surface-enhanced Raman scattering (SERS), owing to the giant enhancement of the Raman signal of adsorbates on high-reflectivity metals like silver, gold and copper.

Among SERS substrates, silver surfaces are the most efficient for enhancing the Raman signals of adsorbates, since for this metal high electromagnetic enhancements are expected by laser excitation in the overall visible region. In particular, water dispersions of silver nanoparticles offer important advantages with respect to other SERS platforms, such as thin films, island crystalline surfaces or other substrates roughened by chemical or electrochemical treatments.

However, chemisorption in metal hydrosols is a quite complex process, which involves the solvation of molecules in the aqueous environment, the ligand affinity to the metal, and the presence of active-sites (adatoms or adclusters) at the surface of the metal nanoparticles. Hence, a good understanding of the interaction of the adsorbates with water is relevant to understand the competition between solvation and adsorption on a metal surface.

It has been found that thiazole adsorbed on aggregated Ag particles gives a strong SERS response mainly due to chemisorption [MMPMMS11], thus the dynamics of thiazole in water has been approached by Dr. Pagliai using CPMD calculations, while I have analysed the structural and vibrational features of the simulated trajectory.

8.1 *Ab Initio* Molecular Dynamics Details

Ab initio MD simulation have been performed with the CPMD package [CPM] on a system made up of 64 heavy water and 1 thiazole molecules in

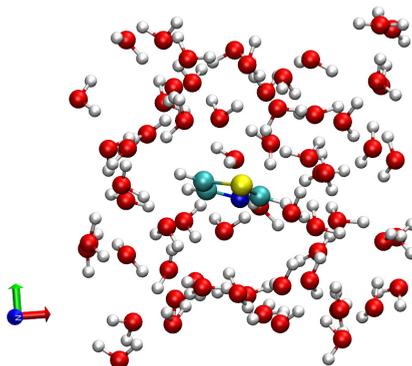


Figure 8.1: Snapshot of the simulation box adopted to simulate thiazole in water.

the NVE ensemble. The sample has been simulated in a cubic box of edge length 14.3683 Å, with periodic boundary conditions. BLYP exchange and correlation functional [Bec88,LYP88] has been adopted, along with norm-conserving Martins-Troullier pseudopotentials [TM91] and Kleinman-Bylander decomposition. [KB82] Plane wave expansion has been truncated at 85 Ry while a 400 atomic units (a.u.) electronic fictitious mass has been set. The system has been simulated for about 26 picoseconds with a time-step of 4 a.u. (~ 0.096 femtoseconds). The average temperature was 302 ± 16 K.

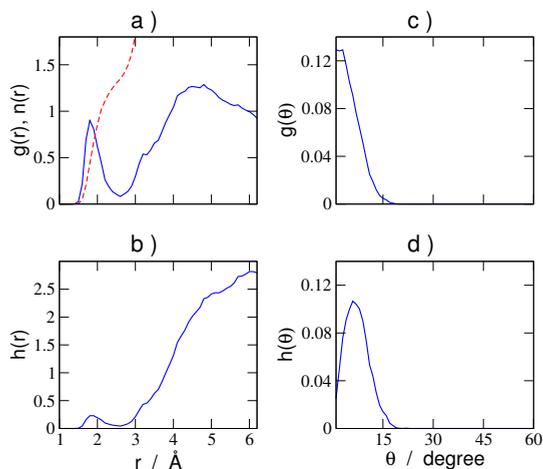


Figure 8.2: (blue lines) $g(r)$, $h(r)$, $g(\theta)$, $h(\theta)$, (red line) $n(r)$ for the $\text{N} \cdots \text{D}$ interactions; r is the $\text{N} \cdots \text{D}$ distance whereas θ is the $\text{N} \cdots \text{D} - \text{O}$ angle.

8.2 Simulation Analysis

Structural and Electronic Properties

Instead, in fig. 8.2 a and c are reported the pair radial and angular distribution functions $g(r)$ and $g(\theta)$, along with the integration number, $n(r)$, for the $N \cdots$ water interaction. As expected, the nitrogen atom of thiazole acts as H-bond acceptor, whereas no such feature has been observed for the S atom. This different behaviour of the two heteroatoms is further

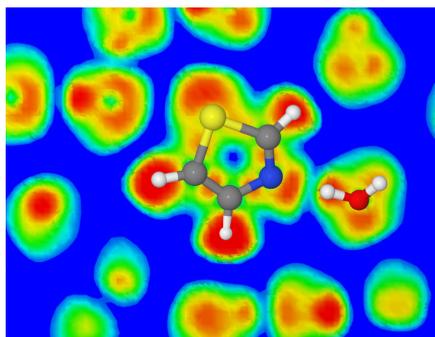


Figure 8.3: Contour plots of the ELF of thiazole interacting with water molecules from a snapshot of the CPMD simulation.

illustrated by fig. 8.3, which shows ELF contour plots (see Section 3.2) near the solute molecule. The ELF isosurfaces show that a charge transfer along the $N \cdots D$ direction occurs, whereas no such thing is observable for the S atom. On the contrary, around the S atom there is a systematic void to the water molecules that to move away from it.

Vibrational Properties

It would be interesting to probe the normal mode associated to the N atom motions inside thiazole. Unfortunately, N atom is involved in several non-localised normal modes, thus to extract them from a simulation would be a very complex task [Key97].

Thus, to probe the system reaction to H-bonding, I choose to study the VDoS of the water molecules involved in H-bonds with thiazole. In fig. 8.4 the reconstructed VDoS of the simulation cell is reported in the range associated to OD stretches. As can be seen, the agreement with the experimental Raman spectra is very good even adopting FT. Nevertheless,

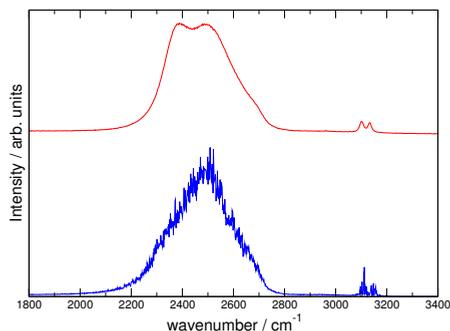


Figure 8.4: Comparison of experimental Raman spectrum of thiazole in D_2O (red upper line) with vibrational density of states (lower blue line) calculated by FFT of the velocity autocorrelation function (obtained by the CPMD simulation) corrected by a uniform scaling factor of 1.05.

to correlate vibrations and structural properties, I once again employed wavelets as shown in fig. 8.5. Water, like the diols of Chapter 6, shows the

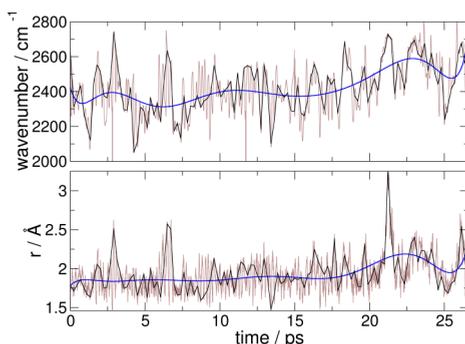


Figure 8.5: (upper panel) Maximum of the OD stretching band time-resolved frequencies of the water molecule H-bonded to thiazole; (lower panel) time distribution of the H-bond distance. In brown, the raw datasets; in black, the averaged quantities; and in blue, the 10th degree polynomial fitting are reported.

same correlation between H-bond length and vibrational frequency. This could be presented also in the form of the “banana shaped” spectrograms already seen in Chapter 6, as reported in fig. 8.6. However, in this case, the plot has been constructed weighting the VDoS of the O–D stretching mode by the F^{HB} H-bond function (see Section 3.8), in order to sort out

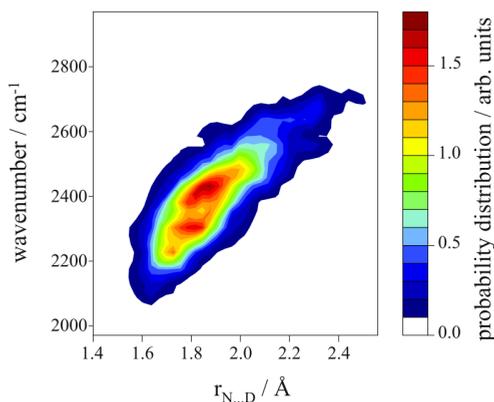


Figure 8.6: Weighted probability distribution of the OD stretching mode as a function of the N \cdots D inter-molecular length.

the contribution due to water molecules interacting with each other. This trick was not required for diols.

8.2.1 Discussion

In this research the CPMD simulation analysis showed that thiazole mainly interacts with water through its nitrogen atom and not via the sulphur atom. I have probed the VDoS of water performing Fourier and wavelet transforms of the O–D time series, but this approach is slightly undermined by the normal mode coupling between the two different OD groups present in a water molecule. In the next Chapter, a study of the dynamics and vibrational properties of water shall be again presented. Partially deuterated water (HOD) has been employed for the research presented in the next Chapter, where the O–D and O–H normal modes are significantly decoupled occurring in two very distinct spectral regions.

9. Bifurcated Hydrogen-Bond of Confined Water

H-bond occurring in water is closely related to the most peculiar properties of this solvent, and it has been the subject of many experimental [ISS⁺06,Sti80,CS05,CBH⁺05,LRT06a,KCP⁺08,Cho08,PHMM08,RRT09,BS10] and computational studies [OT93,CIKP03,KMM⁺04,PXV09,ZDG10]. As a matter of fact, the network of water molecules evolves in time exploring different types of H-bonds during their formation and breaking [LS03,LRT06b,SL10]. In particular, Loparo *et al.* [LRT06b] speculated about the mechanism of H-bond creation and breaking and proposed two alternative models to describe this phenomenon: in one model the switching of H-bonding partners goes through a stable intermediate, otherwise through a transition state.

This evolution, whatever the mechanism is, can be viewed as succession of different transient structures occurring in the dynamics of water molecules; recently Laage and Hynes [LH06,LH08] proposed a bifurcated model to describe the transition state of H-bonded species as they switch from one H-bond acceptor to the other.

This type of bond is deeply involved in the formation of biologically relevant structures [ZM92], but, actually, in pure water bifurcated or strained H-bonds have extremely short lifetime (~ 100 femtoseconds as an order of magnitude), as shown by 2D IR studies [FEL⁺03,ELF⁺05,RRN⁺11]. Thus, it is impractical to study bifurcated H-bonds with *ab initio* MD simulations in liquid water.

However, Pandelov *et al.* [PPWI09] recently showed that the geometry and dynamics of those peculiar structures are, instead, well accessible for water molecules confined in crystalline ionic hydrates. In fig. 9.1, Pandelov *et al.* [PPWI09] reported the IR spectra of various hydrated salts, two of them (hydrated NaCl and LiCl) showing multiple structured bands in the OH stretching region. Werhahn *et al.* [WPXI11] reported a similar spectrum shape for lithium nitrate trihydrate, as shown in fig. 9.2. The structured bands of the OH stretching modes are a clue that water

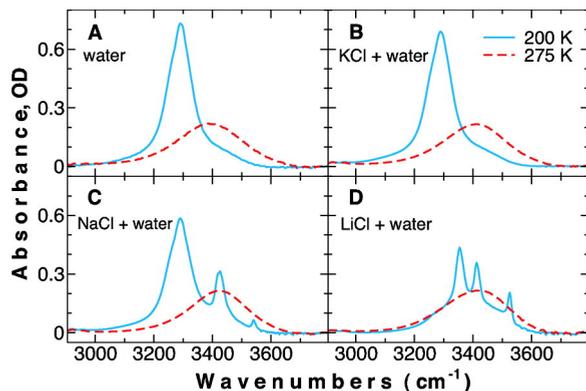


Figure 9.1: FTIR spectra of various aqueous solutions (see inset). The samples are measured in the liquid (275 K, red dashed lines) and in the solid phase (200 K, blue solid lines). The figure has been taken from Ref. [PPWI09].

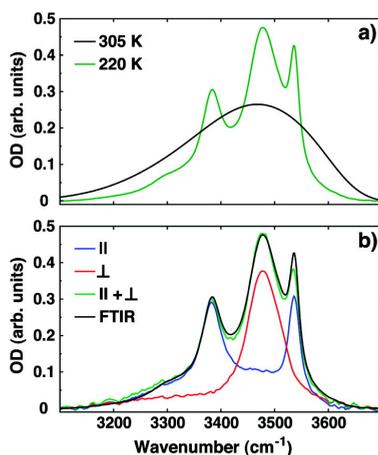


Figure 9.2: IR spectra of $\text{LiNO}_3 \cdot (\text{HDO})(\text{D}_2\text{O})_2$. (a) The absorption spectrum of the liquid sample at 305 K (black line) is compared to that of the crystalline sample at 220 K (green line). Three distinct peaks appear upon crystallisation, pointing toward discrete, distinguishable H-bonding environments of the OH groups. (b) Comparison of the polarisation-resolved and IR spectra. The polarisation-resolved measurements resulted in parallel (blue line) and perpendicular (red line) signals. For the definition of the parallel and perpendicular planes of polarisation, refer to [WPXI11] and the description of the used polarisation-resolving laser setup. The sum of these signals (green line) reproduces the IR spectrum quite accurately. The figure has been taken from Ref. [WPXI11].

molecules are engaged in H-bonds of different strength and stability, thus hydrated NaCl, LiCl and LiNO₃ salts are good candidates to contain water involved in this elusive bifurcated H-bond.

In fact, lithium nitrate trihydrate is an inorganic salt where the water molecules are forced by the lattice forces in a characteristic crystalline arrangement [HTO77,HTO80]. In this crystal, water is involved in three different types of H-bonds [WPXI11], a strong one between water molecules, a weak one between water and nitrate anions, and a bifurcated one where the hydroxyl group of water is directed halfway between two N–O bonds of nitrate (see figures 9.3 and 9.4). Due to the crystalline nature of this

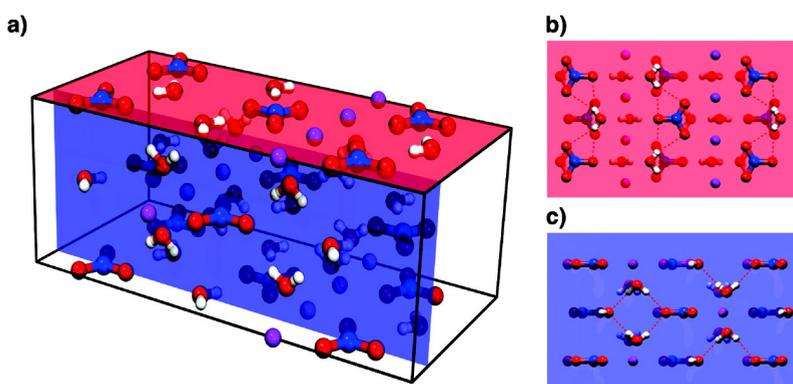


Figure 9.3: (a) Crystal structure of LiNO₃·(H₂O)₂. The crystal structure belongs to the orthorhombic group *Cmcm* with the OH groups of the water molecules all lying in planes perpendicular to each other. These planes are projected in panels b and c for a better view. (b) Plane showing the bifurcated H-bond giving rise to the absorption line indicated in red in fig. 9.2, panel b. (c) Plane showing two types of hydrogen bonds: a weak bond to the oxygen atom of a nitrate anion, and a strong, ice-like bond between two water molecules, which results in the blue absorption spectrum in fig. 9.2, panel b. This figure has been taken from Ref. [WPXI11].

salt, about a third of water molecules is engaged in bifurcated H-bonds that persist for quite long time, in contrast to what happens in pure liquid water. Werhahn *et al.* [WPXI11] have recently characterised crystalline LiNO₃ · (HOD)(D₂O)₂ by means of IR spectroscopy and interpreted the OH stretching region of the IR spectrum as due to the superposition of the stretching bands of the hydroxyl groups engaged in strong, weak, and bifurcated H-bonds.

As the bifurcated H-bonds lie in crystallographic planes orthogonal to

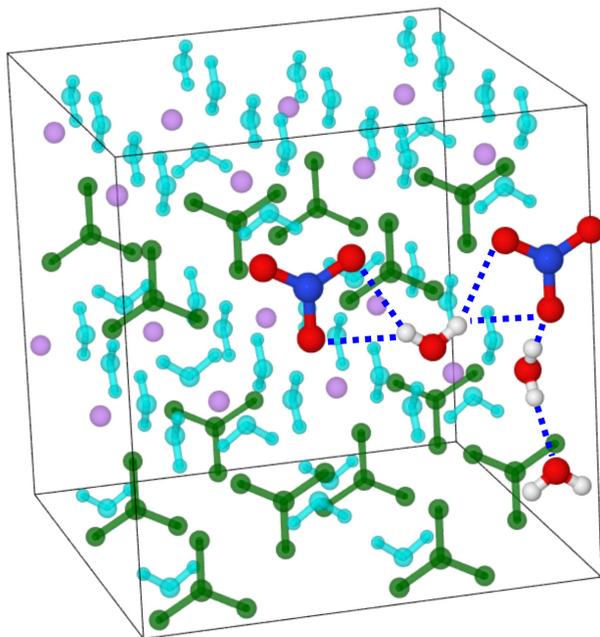


Figure 9.4: Simulation cell. Some selected H-bond interactions are represented with dotted blue lines. Water molecules and nitrate ions, whose interactions are not explicitly shown, are represented with translucent cyan and green colours, respectively.

those where the water molecules form weak and strong H-bonds, Werhahn *et al.* [WPXI11] were able to discriminate the three strong peaks occurring in the OH stretching region from polarised IR spectra. Moreover, they measured slightly different lifetimes for these types of H-bonds, ranging from ~ 1 to ~ 2 picoseconds. Fortunately, these lifetimes are accessible to *ab initio* MD.

9.1 Computational Details

9.1.1 *Ab Initio* Molecular Dynamics Simulation

In this study, my aim was to give a molecular level interpretation of the experimental findings very recently obtained by Werhahn *et al.* [WPXI11], adopting a computational approach based on Car-Parrinello simulations and WT. Making use of the CPMD code [CPM], I have simulated a $\text{LiNO}_3 \cdot (\text{HOD})_3$ crystal composed by 48 water molecules, 16 Li^+ cations, and 16 NO_3^- anions in an orthorhombic box of $13.6036 \times 12.7132 \times 11.998$

\AA^3 volume, with periodic boundary conditions. I chose the structural parameters reported in the supporting information of Ref. [WPXI11] as starting configuration for the simulation. In this lattice, two “families” of HOD molecules are present: 16 of them are engaged in two bifurcated H-bonds, whereas 32 molecules form simultaneously one weak and one strong H-bond. I had care to dispose the HOD molecules so to have an equal number of H-bonds formed along the O–H and the O–D directions for every type of H-bonds (*i.e.* 16 weak, 16 strong, and 8 bifurcated H-bonds along either the O–H and the O–D directions). Moreover, HOD molecules are disposed to break the crystal symmetry, in order to avoid spectral contribution from higher K-points for a better comparison with the experimental data. Thus, I have 8 bifurcated H-bonds along O–H direction, 8 bifurcated H-bonds along O–D, 16 weak H-bonds along O–H, 16 weak H-bonds along O–D, 16 strong H-bonds along O–H and 16 strong H-bonds along O–D. HOD molecules were disposed in the crystal lattice so to break the crystal symmetry, in order to avoid spectral contribution from higher \mathbf{k} -points for a better comparison with the experimental data [WPXI11]. The BLYP [Bec88, LYP88] exchange-correlation functional was employed along with GTH pseudopotentials [GTH96], the wavefunctions were expanded in a plane wave basis truncated at energies of 85 Ry, while the fictitious electronic mass was 400 a.u. The equations of motion were integrated with a time-step of 4 a.u. (~ 0.096 femtoseconds) over 17.6 picoseconds at temperature of 224 ± 9 K in the microcanonical ensemble (NVE), after a thermalisation of ~ 3 picoseconds by velocity rescaling.

9.1.2 Dipole Moment Calculation

MLWF centres [MV97, SMVP98] were calculated every 128 a.u. (~ 3.072 femtoseconds) to obtain molecular dipole moments. Within the linear response theory (see Section 3.5.2), I have calculated the contribution to the IR absorption spectrum due to water molecules by Fourier transforming the molecular dipole time series according to the equation 3.24.

Saving the MLWF centres at ~ 3 femtoseconds intervals for a ~ 17.6 pico-second-long trajectory, I have a resolution of $\sim 1.9 \text{ cm}^{-1}$ and a maximum accessible frequency of $\sim 5500 \text{ cm}^{-1}$ for the calculated spectra. The dipole moment has been projected along the unit vectors in order to enhance the contribution to the spectrum due to OH/OD stretching mode.

9.2 Analysis of the Trajectory

I calculated the IR spectrum of $\text{LiNO}_3 \cdot (\text{HOD})_3$ by FT of the dipole moment of the whole simulation cell. The result is presented in fig. 9.5 (upper panel); in the lower panel I report a magnification of the OD and OH stretching region, in order to sort out the high intensity peaks due to lattice modes and those at $1100\text{--}1500\text{ cm}^{-1}$, which are due to the vibrations of nitrate anions [WS61, SCWJ66, WSK01]. These latter modes are clearly

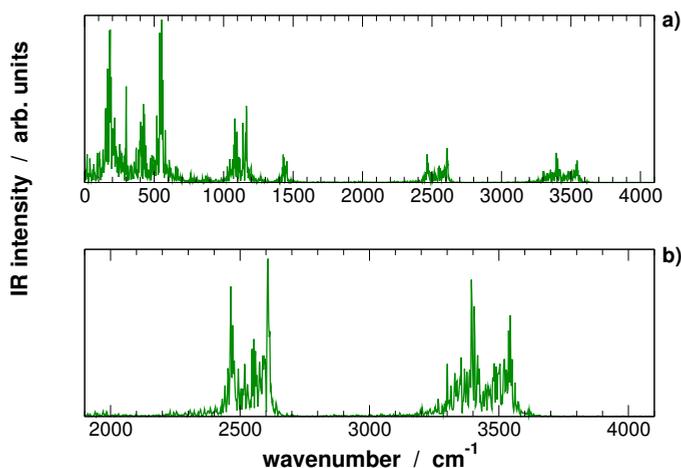


Figure 9.5: Panel a: IR spectrum of the simulation cell ; panel b: magnification of the $2000\text{--}4000\text{ cm}^{-1}$ region. Frequencies are uniformly scaled by a factor of ~ 1.027 .

separated from the rest of the spectrum, but their analysis is difficult due to the congestion of spikes and peaks that make it difficult to extract the actual band shapes. Therefore, I calculated the IR spectrum for O–H and O–D stretching adopting MLWF centres, as explained in the previous section: projecting the dipole moment along the O–H and O–D unit vectors is equivalent to the well established one-phonon approximation (see for example Cardini *et al.* [COK85]), which is adopted here to discriminate and highlight the contributions due to a specific normal mode (the OH or OD stretching mode). The experimental spectrum [WPXI11] for $\text{LiNO}_3 \cdot (\text{HOD})(\text{D}_2\text{O})_2$ is reported in panel a of fig. 9.6: Werhahn *et al.* assigned the peak centered at $\sim 3380\text{ cm}^{-1}$ to the OH stretching mode of

water molecules forming strong H-bonds, the intermediate one at ~ 3480 cm^{-1} to the OH stretching of molecules forming bifurcated H-bonds and that at ~ 3535 cm^{-1} to the OH stretching of molecules that form weak H-bonds. As the weakly and strongly interacting OH groups lie on the same crystallographic plane, orthogonal to that of the bifurcated H-bonds, they measured the contributions of weakly and strongly bonded water molecules in parallel polarised IR spectra, while that due to bifurcated hydrogen-bonded water molecules appears only in the perpendicular polarisation.

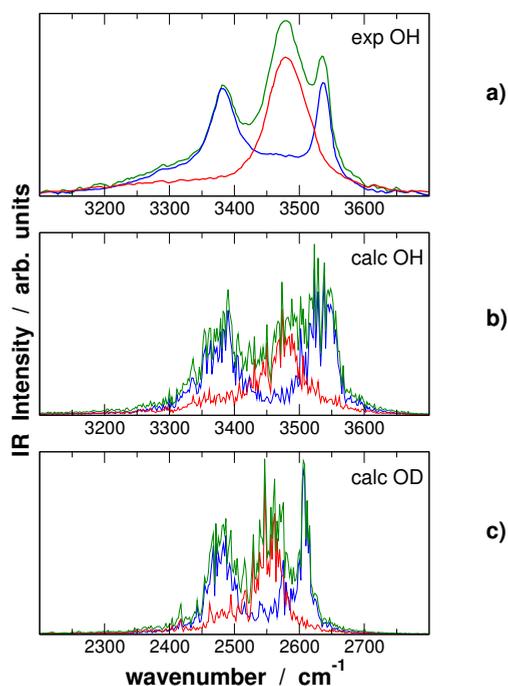


Figure 9.6: Panel a: experimental IR spectra [WPXI11] of $\text{LiNO}_3 \cdot (\text{HOD})(\text{D}_2\text{O})_2$ for perpendicular (blue line) and parallel (red line) polarisation, and their sum (green line); panel b and c: calculated IR spectra of $\text{LiNO}_3 \cdot (\text{HOD})_3$ for water molecules interacting with weak and strong H-bonds (blue lines), bifurcated bonds (red lines) and their sum (green lines) in the spectral region of both OH (panel b) and OD (panel c) stretching modes. Frequencies are uniformly scaled by a factor of ~ 1.027 .

The same separation of the peaks is obtained in our calculated spectra (shown in panels b and c of fig. 9.6 for O–H and O–D stretchings, respectively): in fact, we have calculated the dipole moment time series for each water molecule, and grouped the spectral contribution according to the strength of the H-bond: every molecule forming a strong H-bond with one of its hydroxyl groups has the other OH group engaged in a weak H-bond. These two contributions then appear together in the calculated spectrum (blue lines in panels b and c of fig. 9.6), whereas the contribution of the water molecules forming bifurcated H-bond appears separated (red line in panels b and c of fig. 9.6). Our results fully confirm the assignment [WPXI11] of the three spectral components for both OH and OD stretching modes, in order of increasing frequency, to strong, bifurcated and weak H-bondings. Also the relative bandwidths agree with the experiment, the central peak being substantially broader (both for OH and OD stretching modes) than the lateral ones. Werhahn *et al.* [WPXI11] tentatively attributed the large bandwidth (80 cm^{-1}) of the central peak to librations of bifurcated H-bonded water molecules in the plane defined by the two nearest NO_3^- anions. From static *ab initio* calculations they obtained a potential energy profile along the librational coordinate with 1 kcal/mol deep well centred at 0 degree. I extracted the distribution of this angle from our simulated CPMD trajectory: the result is summarised in fig. 9.7. The main feature reported by Werhahn *et al.* [WPXI11] is recov-

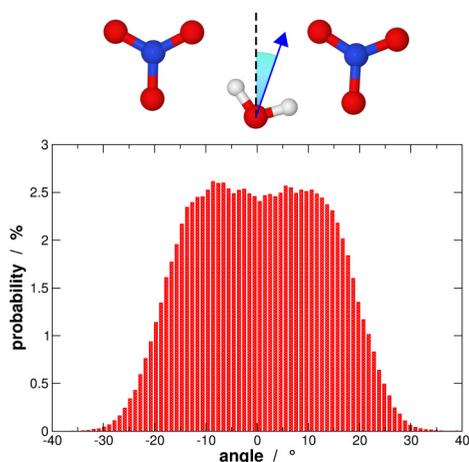


Figure 9.7: Libration angle distribution for every water molecule interacting with bifurcated H-bonds along the simulation run.

ered, but the maxima of the probability distribution are centered at ± 8 degree. I noticed that this value agrees with that obtained by Ramasesha *et al.* [RRN⁺11] from simulations and 2D IR experiments on pure water. This could suggest that water molecules involved in bifurcated H-bonds at finite temperature oscillate between two slightly displaced positions, symmetrically closer to one of the two nitrate groups.

ELFs (see Section 3.2) have been calculated and projected onto two normal planes of the lithium nitrate trihydrate lattice, as reported in fig. 9.8 for the optimised structure [WPXI11]: weak, strong (A panel) and bifurcated (B panel) hydrogen-bonds are displayed. For the strong interaction, a larger charge transfer occurs, as can be appreciated from the fact that the ELF isosurfaces of strongly bonded water pairs overlap slightly, whereas for weak and bifurcated interactions there is a clear separation.

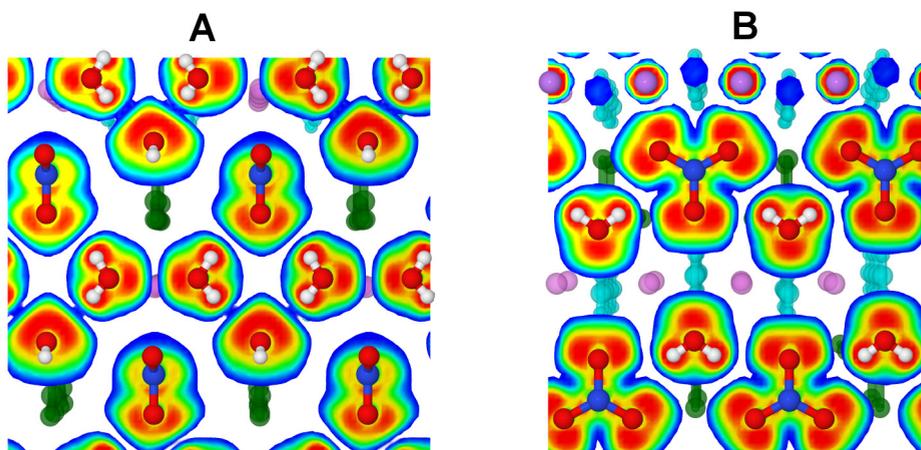


Figure 9.8: (A) ELF projected onto the plane containing weak and strong H-bonds. (B) ELF projected onto the plane containing bifurcated H-bonds.

The analysis of the dipole moment dynamics by means of WT is of help in the interpretation of the strong and broad band attributed to bifurcated H-bond. Exploiting the wavelet ability to localise a signal both in time and in frequency, I can correlate in time, with steps of ~ 3 femtoseconds, the most intense frequency of the IR wavelet spectrum (calculated replacing FT with WT in equation 6.13) with the $\text{H} \cdots \text{O}$ inter-molecular distance of the two nearest neighbours of each water molecule, as illustrated in fig. 9.9. The result of this analysis is reported in panels a and b of fig. 9.10, where the correlation is displayed as two-dimensional contour plot.

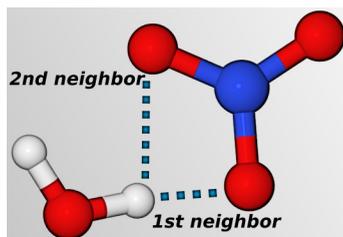


Figure 9.9: In the wavelet spectrograms, the OH stretching frequency has been correlated with the instantaneous inter-molecular distances between the OH group and its two nearest oxygen atoms.

In previous works, wavelets have been employed to recover from MD simulation the inter-molecular structure-frequency correlation of the O–D/O–H stretching band [MRH04, LS03, RRT09] for water [MSC08a, MSC08b] and glycols, leading to “banana-like” plots, whose global shape is not very sensitive to the specific molecular system.

Conversely, since I included both first and second neighbours (in view of the ambiguity of such definition for the water molecules interacting with bifurcated H-bonds) in the present analysis major differences appear between weak and strong (panel a) and bifurcated (panel b) hydrogen-bonding interactions. The Wavelet plot of the formers clearly shows two separated spatial regions: hydrogen-bonded molecules are within the first 2.4 \AA , where the IR activity thickens at frequencies of around 3540 and 3380 cm^{-1} due to weak and strong H-bonds, respectively; a second region at distances larger than 2.4 \AA contains slightly broader peaks at the same frequencies, but no overlap between the IR activities of the two zones occurs. On the other hand, Wavelet analysis shows a definite continuous pathway between the two regions for the bifurcated-bonded water molecules, corroborating the assumption that the IR profile of those molecules is modulated in time not by one of the two $\text{H} \cdots \text{O}_{\text{NO}_3^-}$ distances, but by both of them. These graphs can be seen as an evident “signature” of bifurcated hydrogen-bonding.

Experiments of Loparo et al. [LRT06b] are consistent with the assumption that bifurcated H-bond configurations are visited only transiently during the switch of bonding acceptor. Our calculated angle distribution (see fig. 9.7) supports this finding. Moreover, with WT I calculated the distribution of the OH stretching frequency as a function of the H-bond length. Our calculation (see fig. 9.10) confirms the hypothesis of Loparo

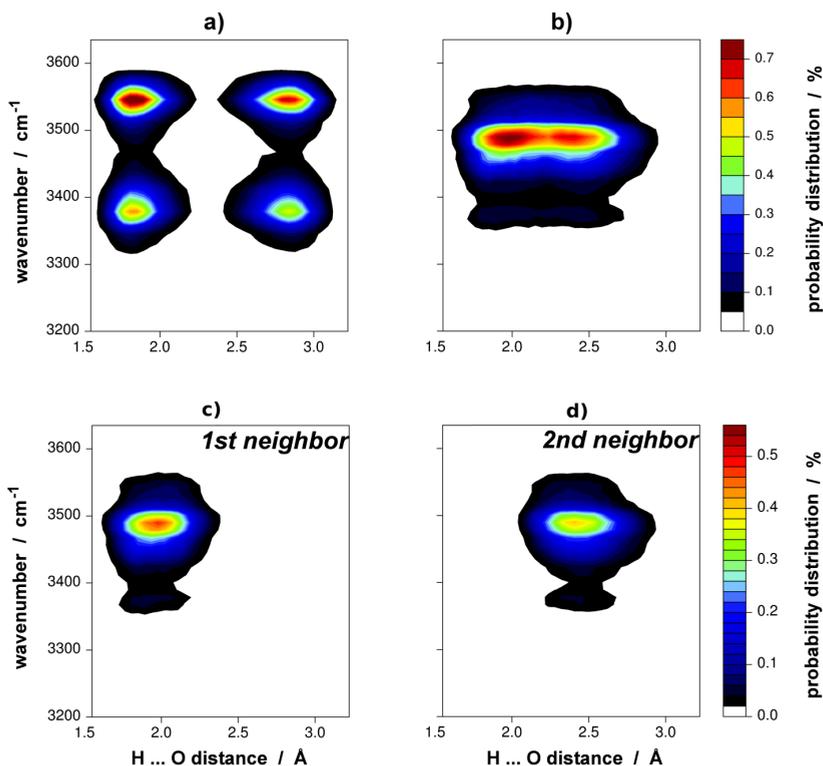


Figure 9.10: Distribution of the most intense frequency of the IR wavelet spectrum with the change of the two shortest H \cdots O inter-molecular distances; panel a: weakly and strongly hydrogen-bonded water molecules; panel b: bifurcated hydrogen-bonded water molecules. Panels c and d show the same distribution reported in panel b splitted between the first and second neighbours. Wavelets plots are calculated with $\sigma = 24\pi$. Frequencies are uniformly scaled by a factor of ~ 1.027 .

et al. [LRT06b], if the probability distribution of frequencies is adopted in place of the free-energy profile and if the H \cdots O distance replaces the fictitious “H-bond coordinate”. In panels c and d of fig. 9.10, the same distribution of probability is reported for the nearest (panel c) and second nearest neighbour (panel d) separately, in order to avoid cancellation effects in the plot relative to the bifurcated interaction due to the two nearest neighbours simultaneously. In the present case the shape of the two distributions is less extended along the H \cdots O coordinate, since the mobility of molecules in a crystalline salt is much less than in a liquid. Moreover, the value adopted for the σ parameter (24π) in fig. 9.10 is op-

timal for frequency localisation, but has poor time-localisation accuracy due to a Heisenberg-like uncertainty [Chu92]. Therefore, I set σ at a much lower value (π) to improve time-accuracy, as reported in fig. 9.11: with this choice the relationship between the time-dependent distances and the time-dependent frequency is more apparent, albeit with larger spread along the frequency axis due to time-frequency uncertainty. In particular, the “banana-shape” of the distribution [MSC08b, MSC08a] is largely recovered in bifurcated-bonded molecules for the interaction with the first neighbour (panel c), while for the interaction with the second neighbour (panel b) the change of frequency with the distance appears to be the opposite. In fact, when the second neighbour moves away from the OH group, the first neighbour comes closer and the hydroxyl stretching mode shifts to lower frequencies. This proves that the two structure-frequency distributions are actually anti-correlated, an effect that in previous studies in liquid solutions [MSC08b, MSC08b] was not observable, due to the mobility of the first and (in particular) of the second solvation shells. Loparo *et al.* [LRT06b] proposed two different schematic free-energy surfaces that describe the changes of OH stretching frequency in liquid water during the switch of H-bond partners along a fictitious “H-bond coordinate”, as reported in fig. 9.12. However, their plot reported in fig. 9.12 are pure speculation, it is just an artifact figure to illustrate some general concept, whereas my contour plot of fig. 9.11 comes from wavelet analysis of an *ab initio* MD trajectory.

The lifetimes of the three types of H-bond have been obtained calculating the autocorrelation functions adopting the procedure of Pagliai *et al.* [PCRS03]. These correlation functions do not show a single exponential decay. The integral lifetimes, albeit with a significant uncertainty, range in the 1.6 – 2.1 picoseconds interval, which is of the same order of magnitude of those reported by Werhahn *et al.* [WPXI11].

Differences between weak/strong hydrogen-bonded molecules and bifurcated ones do not occur only in structural and vibrational properties, but also in the electronic structure, as analysed by MLWF centres. As a matter of fact, dipole moment distribution for the two families of water molecules lead to two different profiles, as reported in fig. 9.13: the distribution of the 32 weakly and strongly hydrogen-bonded molecules display a Gaussian-shaped profile centred at ~ 3 D, a value coincident with that previously calculated by Silvestrelli and Parrinello [SP99] and by Gubskaya and Kusalik [GK02] in liquid water. On the other hand, the

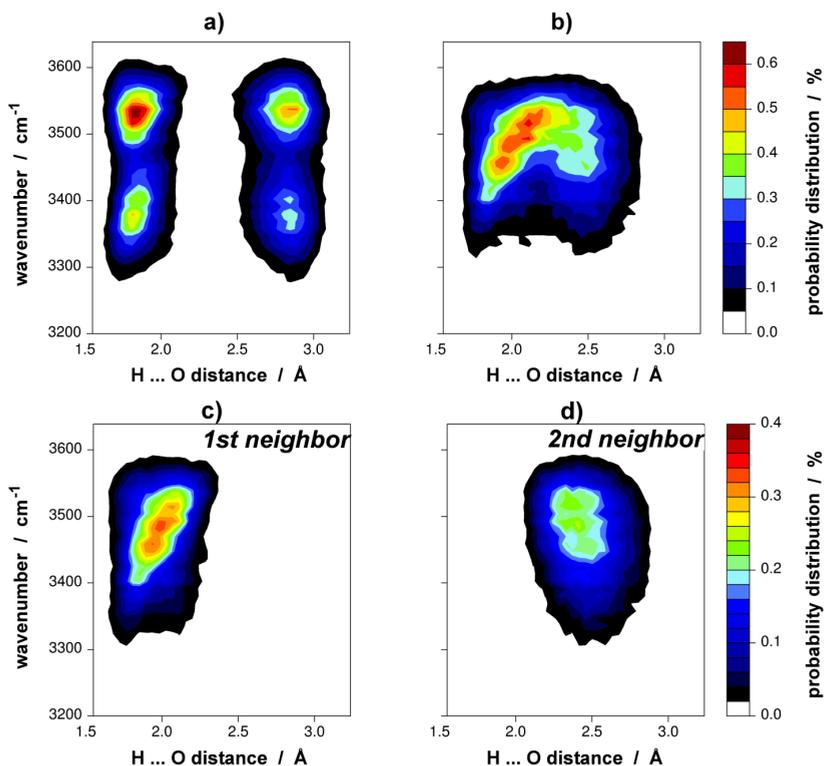


Figure 9.11: Distribution of the most intense frequency of the IR Wavelet spectrum with the change of the two shortest H \cdots O inter-molecular distances; panel a: weakly and strongly hydrogen-bonded water molecules; panel b: bifurcated hydrogen-bonded water molecules. Panels c and d show the same distribution reported in panel b splitted between the first and second neighbours. Wavelets plots are calculated with $\sigma = \pi$. Frequencies are uniformly scaled by a factor of ~ 1.027 .

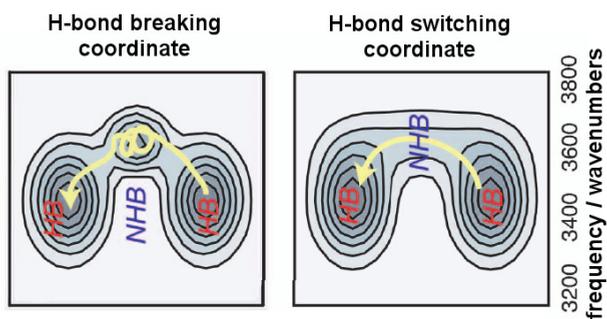


Figure 9.12: OH stretching frequency distribution spreaded along the fictitious “H-bond switching/breaking coordinate”. Figure is taken from Ref. [LRT06b].

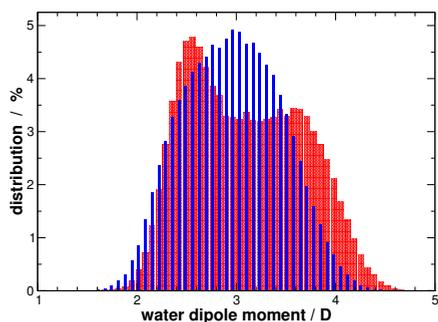


Figure 9.13: Dipole moments distribution for the water molecules. Blue and red histograms correspond to the distribution of the weakly/strongly and bifurcated hydrogen-bonded molecules, respectively. Both histograms are normalised to 100% .

population of the remaining 16 bifurcated hydrogen-bonded molecules shows a bimodal profile of the molecular dipole moments distribution.

I tentatively associated the two maxima occurring in the red distribution of fig. 9.13 to the instantaneous presence of distorted bifurcated H-bonds, two of the four $\text{H} \cdots \text{O}_{\text{NO}_3^-}$ distances being significantly longer (*i.e.* $> 0.5 \text{ \AA}$) than the others, and we attribute the central minimum to a perfectly bifurcated arrangement.

III

Concluding Remarks

The main leitmotif of this dissertation is hydrogen-bonding, which I have studied in some molecular systems adopting molecular dynamics and wavelet analysis.

In the case of the dipeptides in a phospholipid bilayer (see Chapter 5), the analysis of the hydrogen-bonding interaction between C=O groups and water provided relevant insights about the position of the dipeptides into the membrane, filling some interpretative voids left unanswered even by 2D infrared experiments. In fact, the information obtained by means of 2D infrared experiments often needs complement from calculations. In particular, the plausibility of some structural deductions usually requires molecular dynamics simulations to be substantiated. While the study of systems made of tens of thousands atoms is challenging, the accuracy in spectrum reconstruction offered by classical molecular dynamics does not allow elucidating many vibrational features. Moreover, a lot of modern implementations of molecular dynamics, in order to better sample the phase-space, is nowadays completely disconnected from the very idea of a true time-resolved trajectory (*e.g.* the “replica exchange method”), which, on the other hand, is pivotal in vibrational analysis of molecular dynamics datasets. As a matter of fact, a frequency analysis of the simulation cannot be directly performed without a true trajectory, and the only other viable option left is to extract some configuration and then perform harmonic static *ab initio* calculations on these configurations.

Therefore, I mainly adopted *ab initio* molecular dynamics to study smaller systems at higher level of accuracy.

I have employed this approach to study diols in water (see Chapter 6) and their vibrational response changing in solvation. A correlation between the instantaneous vibrational frequency and change of the environment was found, allowing to explain the inhomogeneous broadening

of the O–D stretching mode directly from simulated data. This correlation was previously assessed only adopting mixed techniques at different levels of theory (*i.e.* classical MD simulation followed by *ab initio* static computations on selected configurations, as hinted before) and for much simpler systems, made up of relatively few atoms.

While these results were consistent and encouraging, I then searched for a more direct comparison with experimental data, simulating propanediol in acetonitrile. In fact, with this choice of solvent the vibrational signals of solute (the O–D stretching mode of propanediol) and solvent (acetonitrile) do not overlap anymore. Adopting wavelets and *ab initio* molecular dynamics I found an intriguing structural feature, an intra-molecular hydrogen-bond of propanediol, which was only observed and studied before in the gas phase with IR experiments and some static *ab initio* calculations. It was surprising that this feature could manifest itself in a condensed phase simulation, also because I did not observe it in the aqueous-phase simulation of propanediol, thus suggesting that, while in water the intra-molecular hydrogen-bond is not competitive compared with solvation, this could be true in acetonitrile. Static *ab initio* calculations subsequently proved this, giving comparable results for the free energy of clusters of propanediol and acetonitrile with and without this intra-molecular feature. This predicted hydrogen-bond was also tested by new infrared experiments in condensed phase, which actually supported the computational finding, hence strengthening the common view of computational methods as reliable “*in silico* experiments”.

In this system I have also tried to recover another and much more subtle vibrational effect, the blueshift of the $C\equiv N$ stretching band, which is due to hydrogen-bonding. This was a challenging task, because the usual effect of hydrogen-bonding is to redshift the vibrational frequencies of the involved groups. It was not taken for granted that a density functional theory approach, as implemented in *ab initio* molecular dynamics, was able to reproduce this effects. And, to tell the truth, first results seemed to highlight a minor shortcoming (after all, the blueshift is of the order of few wavenumbers) of this computational method. In fact, static *ab initio* calculations were able to recover the blueshift, suggesting that this issue with the *ab initio* molecular dynamics simulation was not due to some “theoretical” shortcomings (*e.g.* the adopted functional), but probably caused

by statistical limitations. And this was indeed the case, since, performing a new simulation at a lower temperature, the effect was largely recovered. Albeit this approach is slightly unorthodox, a similar method has been recently exploited by Nicolini *et al.* [NC09,NFC11] to perform “smarter” simulations. In the specific case of propanediol in acetonitrile, I resorted to this “trick” in order to “slow down” the dynamics, thus isolating the effect of hydrogen-bonding onto the $C\equiv N$ stretch and sorting out this latter from all other environment rearrangements that could induce frequency shifts.

Wavelets were also pivotal into elucidating the behaviour of methyl acetate in methanol and water (see Chapter 7), allowing the extraction of new interesting information even from old *ab initio* molecular dynamics calculations. In fact, while Fourier transforms could not even reproduce the frequency doublet of the $C=O$ stretching band, the vibrational spectrum as obtained by wavelet analysis showed it very clearly. This represents a relevant increase in the ability of simulations to reproduce spectroscopic features with wavenumber accuracy. Moreover, wavelets allowed us to associate each frequency component of the doublet to a specific intermolecular situation. In this way, a complete assignment of the two peaks of the vibrational doublet was possible, obtaining a computational understanding of the effects of the “chemical exchange” between different hydrogen-bonding states onto the vibrational spectrum, and assessing a one-to-one correspondence between peaks and structural realisations.

Finally, I have shown how a combined approach of computational methods provides new significant insights on a somewhat elusive topic such as bifurcated hydrogen-bonding (see Chapter 9) in water, which is relevant because it is seen as an intermediate-step of hydrogen-bond breaking and formation in aqueous environment. Therefore, it is also pivotal to understand how water switches between hydrogen-bonding partners. The infrared spectrum was reproduced with surprising accuracy, distinguishing the hydroxyl stretching of the water molecules engaged in bifurcated or “normal” hydrogen-bonds.

Furthermore, wavelets allowed to extract a contour-map that basically spreads the infrared activity along the hydrogen-bond switching coordinate, providing more solid computational basis to one of the two competitive models proposed by Loparo *et al.* [LRT06b]. In fact, these calculations anchored one of this two models to a real physical coordinate, at the same time reproducing the spectroscopic and structural features in excel-

lent agreement with the experimental data. Thus, even in this case (all other simulated systems of this dissertation were liquids) wavelet analysis makes possible to extract correlations between the structural and spectroscopic properties, providing a better understanding of the physical origins of the experimental findings.

Moreover, this approach also provides a peculiar “signature” of the bifurcated interaction, allowing (in principle) to recover it also in other molecular systems when and where it should arise. As a matter of fact, this computational approach based on time-frequency localisation can be easily generalised, and thus could be applied to pinpoint the contribution of bifurcated interactions to the vibrational spectra of different molecular systems, even when characterised by faster hydrogen-bonding dynamics, as in liquid water.

In conclusion, I would like to point out the effectiveness of the combined use of high-level-of-theory molecular dynamics and wavelet transforms to monitor vibrational properties, in order to understand how these latter are affected and modulated by the environment. I have focused my research mainly to hydrogen-bonded systems because hydrogen-bond is the strongest of the inter-molecular interactions, but it is my hope that in future even weaker interactions could be realistically modelled and probed by this type of calculations. The continuous research in extending the density functional formalism to Van der Waals forces (see Ref. [SC11]), as well as the recent corrections due to Grimme (see Appendix C.4 and Ref. [Gri06]) for *ab initio* molecular dynamics, suggest optimism.

Besides, the ever-expanding size of the molecular systems that can be studied by *ab initio* molecular dynamics, leads to hope that in future also the vibrational properties of large biomolecules could be computationally reproduced and fully elucidated. The combined approaches described and adopted in this dissertation represent a suitable mean to achieve this goal.

Besides, while this dissertation is mainly devoted to hydrogen-bonding (due to its ability to shift vibrational frequencies up to hundreds of wavenumbers), the possibility to reproduce the doublet in the vibrational spectrum of the C=O stretching of methyl acetate (with a splitting of just $\sim 20 \text{ cm}^{-1}$) suggests that weaker interactions could be probed with wavelet analysis too. To corroborate this hope, the finding that even the small C \equiv N blueshift in acetonitrile (less than 10 cm^{-1}) can in principle be recovered, which hints that a complete picture of the molecular behaviour and vibrations shall be at hand.

Bibliography

- [ACAR96] H. J. Anderson, J. E. Coleman, R. J. Andersen, and M. Roberge. Cytotoxic peptides hemiasterlin, hemiasterlin a and hemiasterlin b induce mitotic arrest and abnormal spindle formation. *Cancer Chemotherapy and Pharmacology*, 39:223–226, 1996.
- [ACR96] A. Askar, A. E. Cetin, and H. Rabitz. Wavelet transform for analysis of molecular dynamics. *J. Phys. Chem.*, 100:19165–19173, 1996.
- [AL07] D.S. Ahn and J.S. Lee. Density functional theory study of acetonitrile-water clusters: Structures and infrared frequency shifts. *Bull. Korean Chem. Soc.*, 28:725–729, 2007.
- [AWW⁺97] B. K. Alsberg, A. M. Woodward, M. K. Winson, J. Rowland, and D. B. Kell. Wavelet denoising of infrared spectra. *Analyst*, 122:645–652, 1997.
- [Bad91] R. F. W. Bader. A quantum theory of molecular structure and its applications. *Chem. Rev.*, 91:893–928, 1991.
- [BE90] A.D. Becke and K.E. Edgecombe. A simple measure of electron localization in atomic and molecular systems. *J. Chem. Phys.*, 92:5397–5403, 1990.
- [Bec88] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behaviour. *Phys. Rev. A*, 38(6):3098–3100, 1988.
- [Bec93a] A. D. Becke. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.*, 98:1372–1377, 1993.
- [Bec93b] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648–5652, 1993.
- [Bec14] Axel D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics*, 140(18), 2014.
- [BMR⁺00] G. Berghold, C. J. Mundy, A. H. Romero, J. Hutter, and M. Parrinello. General and efficient algorithms for obtaining maximally localized wannier functions. *Phys. Rev. B*, 61:10040–10048, 2000.
- [BOT08] M. Banno, K. Ohta, and K. Tominaga. Ultrafast Vibrational Dynamics and Solvation Complexes of Methyl Acetate in Methanol Studied by Sub-Picosecond Infrared Spectroscopy. *J. Raman Spectrosc.*, 39:1531–1537, 2008.
- [BOY⁺09] M. Banno, K. Ohta, S. Yamaguchi, S. Hirai, and K. Tominaga. Vibrational Dynamics of Hydrogen-Bonded Complexes in Solutions Studied with Ultrafast Infrared Pump-Probe Spectroscopy. *Acc. Chem. Res.*, 42:1259–1269, 2009.

Modelling of spectroscopic and structural properties using molecular dynamics

- [BP76] B. J. Berne and R. J. Pecora. *Dynamic Light Scattering*. John Wiley and Sons Ltd., 1976.
- [BS98] F. A. Bornemann and C. Schütte. A mathematical investigation of the Car-Parrinello method. *Numer. Math.*, 78:359–376, 1998.
- [BS10] H. J. Bakker and J. L. Skinner. Vibrational spectroscopy as a probe of structure and dynamics in liquid water. *Chem. Rev.*, 110:1498–1517, 2010.
- [C. 92] C. K. Chui. *An introduction to wavelet*. Academic Press Inc., San Diego, California (USA), 1992.
- [CAJL93] K. Cho, T. A. Arias, J. D. Joannopoulos, and Pui K. Lam. Wavelets in electronic structure calculations. *Phys. Rev. Lett.*, 71:1808–1811, 1993.
- [CBH⁺05] M. L. Cowan, B. D. Bruner, N. Huse, J. R. Dwyer, B. Chugh, E. T. J. Nibbering, T. Elsaesser, and R. J. D. Miller. Ultrafast memory loss and energy redistribution in the hydrogen bond network of liquid H₂O. *Nature*, 434:199–202, 2005.
- [CCB⁺95] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.
- [CCT11a] H. Y. Chen, Y. L. Cheng, and K. Takahashi. Theoretical calculation of the oh vibrational overtone spectra of 1,5-pentanediol and 1,6-hexanediol. *J. Phys. Chem. A*, 115:14315–14324, 2011.
- [CCT11b] Y. L. Cheng, H. Y. Chen, and K. Takahashi. Theoretical calculation of the oh vibrational overtone spectra of 1-n alkane diols (n = 2-4): Origin of disappearing hydrogen-bonded oh peak. *J. Phys. Chem. A*, 115:5641–5653, 2011.
- [Cha87] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, Inc., 1987.
- [CHC03] J. H. Choi, S. Ham, and M. Cho. Local amide i mode frequencies and coupling constants in polypeptides. *J. Phys. Chem. B*, 107(34):9132–9138, 2003.
- [Cho08] M. Cho. Coherent two-dimensional optical spectroscopy. *Chem. Rev.*, 109:1331–1418, 2008.
- [CHT98] R. Carmona, W. Hwang, and B. Torresani. *Practical Time-Frequency Analysis: Gabor and Wavelet Transforms with an implementation*. Academic Press: New York, 1998.
- [Chu92] C. K. Chui. *An introduction to wavelet*. Wavelet analysis and its applications. Academic Press, Inc., 1992.
- [CIKP03] B. Chen, I. Ivanov, M. Klein, and M. Parrinello. Hydrogen bonding in water. *Phys. Rev. Lett.*, 91:215503–215507, 2003.
- [CLS04] S. A. Corcelli, C. P. Lawrence, and J. L. Skinner. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H₂O and D₂O. *J. Chem. Phys.*, 120(17):8107–8117, 2004.

- [COK85] G. Cardini, S. F. O'Shea, and M. L. Klein. Gas-surface potentials and the dynamics of overlayers. *Faraday Discuss. Chem. Soc.*, 80:227–238, 1985.
- [CP85] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, 1985.
- [CP2] CP2K, version 2.3, available from <http://www.cp2k.org/>. by the CP2K developers group (2012).
- [CPC⁺99] R. Chelli, P. Procacci, G. Cardini, R. G. Della Valle, and S. Califano. Glycerol condensed phases part i. a molecular dynamics study. *Phys. Chem. Chem. Phys.*, 1:871–877, 1999.
- [CPLR09] M. Candelaresi, M. Pagliai, M. Lima, and R. Righini. Chemical equilibrium probed by two-dimensional ir spectroscopy: Hydrogen bond dynamics of methyl acetate in water. *J. Phys. Chem. A*, 113:12783–12790, 2009.
- [CPM] CPMD, available from <http://www.cpmc.org/>. Copyright MPI für Festkörperforschung Stuttgart 1997-2001, Copyright IBM Corp 1990-2008.
- [CS05] S. A. Corcelli and J. L Skinner. Infrared and Raman Line Shapes of Dilute HOD in Liquid H₂O and D₂O from 10 to 90 °C. *J. Phys. Chem. A*, 109:7725–7729, 2005.
- [D. 00] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, 2nd edition edition, 2000.
- [DAE03] I. P. Daykov, T. A. Arias, and Torkel D. Engeness. Robust *Ab Initio* calculation of condensed matter: Transparent convergence through semiclassical multiresolution analysis. *Phys. Rev. Lett.*, 90:216402, 2003.
- [dip06] *IDM Pharma Submits New Drug Application to the FDA for Junovan (mifamurtide) in the Treatment of Osteosarcoma*, New Drug Applications Archive, IRVINE, Calif., USA, October 26, 2006.
- [DPL11] P. D. Dopieralski, C. L. Perrin, and Z. Latajka. On the intramolecular hydrogen bond in solution: Car-parrinello and path integral molecular dynamics perspective. *J. Chem. Theory and Comput.*, page DOI: 10.1021/ct200580c, 2011.
- [ea01] W. Koch et al. *A Chemist's Guide to Density Functional Theory*. Wiley VCH, 2001.
- [Ege] S. N. Ege. *Organic Chemistry: Structure and Reactivity*. Houghton Mifflin Harcourt.
- [Ehr02] F. Ehrentreich. Wavelet transform applications in analytical chemistry. *Anal. Bioanal. Chem.*, 372:115–121, 2002.
- [ELF⁺05] J. D. Eaves, J. J. Loparo, C. J. Fecko, A. Tokmakoff, and P. Geissler. Hydrogen-bonds in liquid water are broken only fleetingly. *Proc. Natl. Acad. Sci. U.S.A.*, 102:13019–13022, 2005.
- [EPB⁺95] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.
- [FEL⁺03] C. J. Fecko, J. D. Eaves, J. J. Loparo, A. Tokmakoff, and P. Geissler. Ultrafast hydrogen-bond dynamics in the infrared spectroscopy of water. *Science*, 301:1698–1702, 2003.

Modelling of spectroscopic and structural properties using molecular dynamics

- [FJ97] M. Frigo and S. G. Johnson. The fastest Fourier transform in the west. Technical Report MIT-LCS-TR-728, Massachusetts Institute of Technology, September 1997.
- [FJ98] M. Frigo and S. G. Johnson. FFTW: An adaptive software architecture for the FFT. In *Proc. 1998 IEEE Intl. Conf. Acoustics Speech and Signal Processing*, volume 3, pages 1381–1384. IEEE, 1998.
- [FJ00] S.E. Feller and A.D. MacKerell Jr. An improved empirical potential energy function for molecular simulations of phospholipids. *J. Phys. Chem. B*, 104:7510–7515, 2000.
- [FPCS06] C. Faralli, M. Pagliai, G. Cardini, and V. Schettino. Structure and Dynamics of Br⁻ Ion in Liquid Methanol. *J. Phys. Chem. B*, 110:14923–14928, 2006.
- [FS02] D. Frenkel and B. Smit. *Understanding Molecular Simulation*. Academic Press, 2002.
- [FTS⁺a] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [FTS⁺b] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, , and D. J. Fox. Gaussian 09, revision c.01. Gaussian, Inc., Wallingford CT, 2010.

- [FWT02] H. Fukunishi, O. Watanabe, and S. Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Phys. Chem.*, 116:9058, 2002.
- [GI99] S. Goedecker and O. V. Ivanov. Frequency localization properties of the density matrix and its resulting hypersparsity in a wavelet representation. *Phys. Rev. B*, 59:7270–7273, 1999.
- [GK02] A. V. Gubskaya and P. G. Kusalik. The total molecular dipole moment for liquid water. *J. Chem. Phys.*, 11:5290–5302, 2002.
- [GLS97] W. Gropp, E. Lusk, and A. Skjellum. *Using MPI: Portable Parallel Programming with Message-Passing Interface*. The MIT Press, 1997.
- [Gri06] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27:1787–99, 2006.
- [GTH96] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, 1996.
- [GTK12] Q. Gu, C. Trindle, and J. L. Knee. Communication: Frequency shifts of an intramolecular hydrogen bond as a measure of intermolecular hydrogen bond strengths. *J. Chem. Phys.*, 137(9):091101, 2012.
- [Haa10] A. Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.
- [HAO⁺06] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.*, 65:712–725, 2006.
- [HJK05] D. L. Howard, P. Jorgensen, and H. G. Kjaergaard. Weak intramolecular interactions in ethylene glycol identified by vapor phase oh stretching overtone spectroscopy. *J. Am. Chem. Soc.*, 127(48):17096–17103, 2005.
- [HKS11] M. K. Hazra, X. Kuang, and A. Sinha. Influence of intramolecular hydrogen bonding on oh-stretching overtone intensities and band positions in peroxyacetic acid. *J. Phys. Chem. A*, 116:5784–5795, 2011.
- [HLDH99] P. Hamm, M. Lim, W. F. DeGrado, and R. M. Hochstrasser. The two-dimensional ir nonlinear spectroscopy of a cyclic penta-peptide in relation to its three-dimensional structure. *Proc. Natl. Acad. Sci.*, 96(5):2036–2041, 1999.
- [HLH98] P. Hamm, M. Lim, and R. M. Hochstrasser. Structure of the amide i band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B*, 102(31):6123–6138, 1998.
- [Hoo85] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.
- [HTO77] K. Hermansson, J. O. Thomas, and I. Olovsson. Hydrogen bond studies. cxx. an x-ray determination of the crystal structure of $\text{lino}_3 \cdot 3\text{h}_2\text{o}$. *Acta Crystallogr. B*, 33:2875–2861, 1977.
- [HTO80] K. Hermansson, J. O. Thomas, and I. Olovsson. Hydrogen bond studies. cxxxviii. neutron diffraction studies of $\text{lino}_3 \cdot 3\text{h}_2\text{o}$ at 120 and 295 k. *Acta Crystallogr. B*, 36:1032–1040, 1980.

Modelling of spectroscopic and structural properties using molecular dynamics

- [HTP95] M. Herrenbauer, D. P. Tieleman, and C. Posten. Molecular modelling of diffusional motion and transfer of pyrene in lipid membranes. *Computer Applications in Biotechnology*, 8:317–321, 1995.
- [ISP⁺99] E. D. Isaacs, A. Shukla, P. M. Platzman, D. R. Hamann, B. Barbiellini, and C. A. Tulk. Covalency of the hydrogen bond in ice: A direct x-ray measurement. *Phys. Rev. Lett.*, 82:600–603, 1999.
- [ISS⁺06] H. Iglev, M. Schmeisser, K. Simeonidis, A. Thaller, and A. Laubereau. Ultrafast superheating and melting of bulk ice. *Nature*, 439:183–186, 2006.
- [JCM⁺83] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. *J. Chem. Phys.*, 79:926–935, 1983.
- [JDMN00] K. Jetter, U. Depczynski, K. Molt, and A. Niemoller. Principles and applications of wavelet transformation to chemometrics. *Analytica Chimica Acta*, 420:169–180, 2000.
- [Jef] G. A. Jeffrey. *An Introduction to Hydrogen Bonding*. Oxford University Press.
- [JF08] S. G. Johnson and M. Frigo. Implementing FFTs in practice. In C. Sidney Burrus, editor, *Fast Fourier Transforms*, chapter 11. Connexions, Rice University, Houston TX, September 2008.
- [JM11] W. Jones and N.H. March. *Theoretical Solid State Physics: Perfect Lattices in Equilibrium*. Dover, 2011.
- [JO90] G. James and E. N. Olson. Fatty acylated proteins as components of intracellular signaling pathways. *Biochemistry*, 29(11):2623–2634, 1990.
- [JW89] R. E. Jacobs and S. H. White. The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. *Biochemistry*, 28(8):3421–3437, 1989.
- [KB82] L. Kleinman and D. M. Bylander. Efficacious form for model pseudopotentials. *Phys. Rev. Lett.*, 48(20):1425–1428, 1982.
- [KCP⁺08] D. Kraemer, M. L. Cowan, A. Paarmann, N. Huse, E. T. J. Nibbering, T. Elsaesser, and R. J. D. Miller. Temperature dependence of the two-dimensional infrared spectrum of liquid H_2O . *Proc. Natl. Acad. Sci. U.S.A.*, 105:437–442, 2008.
- [Key97] T. Keyes. Instantaneous normal mode approach to liquid state dynamics. *J. Phys. Chem. A*, 101(16):2921–2930, 1997.
- [KH05] Y.S. Kim and R.M. Hochstrasser. Chemical exchange 2d ir of hydrogen-bond making and breaking. *Proc. Natl. Acad. Sci.*, 102:11185–11190, 2005.
- [KH09] Y.S. Kim and R.M. Hochstrasser. Applications of 2d ir spectroscopy to peptides, proteins, and hydrogen-bond dynamics. *J. Phys. Chem. B*, 113:8231–8251, 2009.
- [Kir05] J. F. Kirby. Which wavelet best reproduces the fourier power spectrum? *Comput. Geosci.*, 31:846–864, 2005.
- [KJ09] I. Kopriva and I. Jerić. Blind decomposition of infrared spectra using flexible component analysis. *Chemometrics and Intelligent Laboratory System*, 97:170–178, 2009.

- [Kle02] R. A. Klein. Ab initio conformational studies on diols and binary diol-water systems using dft methods. intramolecular hydrogen bonding and 1:1 complex formation with water. *J. Comput. Chem.*, 23(6):585–599, 2002.
- [KMM⁺04] I. F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello. Liquid water from first principles: Investigation of different sampling approaches. *J. Phys. Chem. B*, 108:12990–12998, 2004.
- [KMMR82] A.J. Kinning, V. Mom, F.C. Mijlhoff, and G.H. Renes. The molecular structure of 1,3-propanediol in the gas phase, an electron diffraction study. *J. Mol. Struct.*, 82:271 – 275, 1982.
- [KPTH96] H. G. Kjaergaard, R. J. Proos, D. M. Turnbull, and B. R. Henry. Ch stretching overtone investigation of relative ch bond lengths in pyridine. *J. Phys. Chem.*, 100:19273–19279, 1996.
- [KS65] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A*, 140:A1133–1138, 1965.
- [LH06] D. Laage and J. T. Hynes. A molecular jump mechanism of water reorientation. *Science*, 299:1367–1372, 2006.
- [LH08] D. Laage and J. T. Hynes. On the molecular mechanism of water reorientation. *J. Phys. Chem. B*, 112:14230–14242, 2008.
- [LRT06a] J. J. Loparo, S. T. Roberts, and A. Tokmakoff. Multidimensional infrared spectroscopy of water. i. vibrational dynamics in two-dimensional ir line shapes. *J. Chem. Phys.*, 125:194521, 2006.
- [LRT06b] J. J. Loparo, S. T. Roberts, and A. Tokmakoff. Multidimensional infrared spectroscopy of water. ii. hydrogen bond switching dynamics. *J. Chem. Phys.*, 125:194522, 2006.
- [LS03] C. P. Lawrence and J. L. Skinner. Vibrational spectroscopy of hod in liquid d₂o. iii. spectral diffusion, and hydrogen-bonding and rotational dynamics. *J. Chem. Phys.*, 118:264–272, 2003.
- [LYP88] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the correlation energy. *Phys. Rev. B*, 37(2):785–789, 1988.
- [Mar04] R. M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004.
- [MH09] D. Marx and J. Hütter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.
- [MKO93] S. D. Meyers, B. G. Kelly, and J. J. O’Brien. An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of yanai waves. *Monthly Weather Rev.*, 121:2858–2865, 1993.
- [MMH04] I. De. Moortel, S. A. Munday, and A. W. Hood. Wavelet analysis: the effect of varying basic wavelet parameters. *Solar Phys.*, 222:203–228, 2004.
- [MMPCR12a] F. Muniz-Miranda, M. Pagliai, G. Cardini, and R. Righini. Bifurcated hydrogen bond in lithium nitrate trihydrate probed by ab initio molecular dynamics. *J. Phys. Chem. A*, 116:2147–2153, 2012.

Modelling of spectroscopic and structural properties using molecular dynamics

- [MMPCR12b] F. Muniz-Miranda, M. Pagliai, G. Cardini, and R. Righini. Hydrogen bond effects in the vibrational spectra of 1,3-propanediol in acetonitrile: ab initio and experimental study. *J. Chem. Phys.*, 137:244501, 2012.
- [MMPCS11] F. Muniz-Miranda, M. Pagliai, G. Cardini, and V. Schettino. Wavelet transform for spectroscopic analysis: Application to diols in water. *J. Chem. Theory Comput.*, 7:1109–1118, 2011.
- [MMPMMS11] M. Muniz-Miranda, M. Pagliai, F. Muniz-Miranda, and V. Schettino. Raman and computational study of solvation and chemisorption of thiazole in silver hydrosol. *Chem. Commun.*, 47:3138–3140, 2011.
- [MP98] M. Marchi and P. Procacci. Coordinates scaling and multiple time step algorithms for simulation of solvated proteins in the npt ensemble. *J. Chem. Phys.*, 109(13):5194–5202, 1998.
- [MRH04] K. B. Møller, R. Rey, and J. T. Hynes. Hydrogen bond dynamics in water and ultrafast infrared spectroscopy: A theoretical study. *J. Phys. Chem. A*, 108:1275–1289, 2004.
- [MSC08a] B. S. Mallik, A. Semparathi, and A. Chandra. A first principles theoretical study of vibrational spectral diffusion and hydrogen bond dynamics in aqueous ionic solutions: D₂O in hydration shells of Cl⁻ ions. *J. Chem. Phys.*, 129:194512–194527, 2008.
- [MSC08b] B. S. Mallik, A. Semparathi, and A. Chandra. Vibrational spectral diffusion and hydrogen bond dynamics in heavy water from first principles. *J. Phys. Chem. A*, 112:5104–5112, 2008.
- [MSC⁺10] S. Marsili, G.F. Signorini, R. Chelli, M. Marchi, and P. Procacci. Orac: a molecular dynamics simulation program to explore free energy surfaces in biomolecular systems at the atomistic level. *J. Comput. Chem.*, 31:1106–16, 2010.
- [Mul55] R. S. Mulliken. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.*, 23:1833–1831, 1955.
- [MV97] N. Marzari and D. Vanderbilt. Maximally-localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56:12847–12862, 1997.
- [NC09] P. Nicolini and R. Chelli. Improving fast-switching free energy estimates by dynamical freezing. *Phys. Rev. E*, 80:041124, 2009.
- [NFC11] P. Nicolini, D. Frezzato, and R. Chelli. Exploiting configurational freezing in nonequilibrium monte carlo simulations. *J. Chem. Theory and Comput.*, 7:582–593, 2011.
- [NG06] A.I. Neelov and S. Goedecker. An efficient numerical quadrature for the calculation of the potential energy of wavefunctions expressed in the daubechies wavelet basis. *J. Comput. Phys.*, 217(2):312 – 339, 2006.
- [OHS08] W. J. Orts, K. M. Holtman, and J. N. Seiber. Agricultural chemistry in bioenergy. *J. Agr. Food Chem.*, 56:3892–3899, 2008.
- [OL77] J.J. Olivero and R.L. Longbothuma. Empirical fits to the voigt line width: A brief review. *J. Quant. Spectrosc. Radiat. Transfer*, 17:233–236, 1977.

- [OT93] I. Ohmine and H. Tanaka. Fluctuation, relaxations, and hydration in liquid water. hydrogen-bond rearrangement dynamics. *Chem. Rev.*, 93:2545–2566, 1993.
- [PBE96] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, 1996.
- [PBMM⁺] M. Pagliai, L. Bellucci, M. Muniz-Miranda, G. Cardini, and V. Schettino. A combined raman, dft and md study of the solvation dynamics and the adsorption process of pyridine in silver hydrosols. *Phys. Chem. Chem. Phys.*
- [PCC⁺08] M. Pagliai, C. Cavazzoni, G. Cardini, G. Erbacci, M. Parrinello, and V. Schettino. Anharmonic infrared and raman spectra in car-parrinello molecular dynamics simulations. *J. Chem. Phys.*, 128:224514, 2008.
- [PCRS03] M. Pagliai, G. Cardini, R. Righini, and V. Schettino. Hydrogen bond dynamic in liquid methanol. *J. Chem. Phys.*, 119:6655–6662, 2003.
- [PCS05] M. Pagliai, G. Cardini, and V. Schettino. Solvation dynamics of li^+ and cl^- ions in liquid methanol. *J. Phys. Chem. B*, 109:14923–14928, 2005.
- [PHMM08] A. Paarmann, T. Hayashi, S. Mukamel, and R. J. D. Miller. Probing intermolecular couplings in liquid water with two-dimensional infrared photon echo spectroscopy. *J. Chem. Phys.*, 128:191103, 2008.
- [PMMC⁺10] M. Pagliai, F. Muniz-Miranda, G. Cardini, R. Righini, and V. Schettino. Hydrogen bond dynamics of methyl acetate in methanol. *J. Phys. Chem. Lett.*, 1:2951–2955, 2010.
- [PMMC⁺11] M. Pagliai, F. Muniz-Miranda, G. Cardini, R. Righini, and V. Schettino. Spectroscopic properties with a combined approach of ab initio molecular dynamics and wavelet analysis. *J. Mol. Struc.*, 993:438–442, 2011.
- [PMM⁺df] P. Procacci, M. Marchi, S. Marsili, G.F. Signorini, R. Chelli, and M. Souaille. *ORAC: A Molecular Dynamics Program to Simulate Complex Molecular Systems at the Atomistic Level*. ORAC 5 User Manual, <http://www.chim.unifi.it/orac/orac-manual.pdf>.
- [PPWI09] S. Pandelov, B. M. Pilles, J. C. Werhahn, and H. Iglev. Time-resolved dynamics of the oh stretching vibration in aqueous nacl hydrate. *J. Phys. Chem. A*, 113:10184–10188, 2009.
- [PR80] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, 1980.
- [PSB91] G. Pastore, E. Smargiassi, and F. Buda. Theory of *ab initio* molecular-dynamics calculations. *Phys. Rev. A*, 44:6334–6347, 1991.
- [PXV09] F. Paesani, S. S. Xantheas, and G. A. Voth. Infrared spectroscopy and hydrogen-bond dynamics of liquid water from centroid molecular dynamics. *J. Phys. Chem. B*, 113:13118–13130, 2009.
- [PY89] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford University Press, Oxford, 1989.
- [RCB77] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327, 1977.

Modelling of spectroscopic and structural properties using molecular dynamics

- [RP95] G. Rauhut and P. Pulay. Transferable scaling factors for density functional derived vibrational force fields. *J. Phys. Chem.*, 99:3093–3100, 1995.
- [RRN⁺11] K. Ramasesha, S. T. Roberts, R. A. Nicodemus, A. Mandal, and A. Tokmakoff. Ultrafast 2d ir anisotropy of water reveals reorientation during hydrogen-bond switching. *J. Chem. Phys.*, 135:54509, 2011.
- [RRT09] S. T. Roberts, K. Ramasesha, and A. Tokmakoff. Structural rearrangements in water viewed through two-dimensional infrared spectroscopy. *Acc. Chem. Res.*, 42:1239–1249, 2009.
- [RW05] A. Rahaman and R. A. Wheeler. Wavelet transforms for determining time-dependent vibrational frequencies. *J. Chem. Theory Comput.*, 1:769–771, 2005.
- [SBB⁺93] M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery. *J. Comput. Chem.*, 14:1347–1363, 1993.
- [SC11] S. N. Steinmann and C. Corminboeuf. Comprehensive benchmarking of a density-dependent dispersion correction. *J. Chem. Theory and Comput.*, 7:3567–3577, 2011.
- [SCWJ66] A. T. Ward S. C. Wait and G. J. Janz. Anion-cation interactions in molten inorganic nitrates: Vibrational analysis. *J. Chem. Phys.*, 45:133–137, 1966.
- [SL10] G. Stirnemann and D. Laage. Direct evidence of angular jumps during water reorientation through two-dimensional infrared anisotropy. *J. Phys. Chem. Lett.*, 1:1511–1516, 2010.
- [SLC03] X. G. Shao, A. K. M. Leung, and F. T. Chau. Wavelet: a new trend in chemistry. *Acc. Chem. Res.*, 36:276–283, 2003.
- [SMVP98] P. L. Silvestrelli, N. Marzari, D. Vanderbilt, and M. Parrinello. Maximally-localized generalized wannier functions for disordered systems: application to amorphous silicon. *Solid State Comm.*, 107(1):7–11, 1998.
- [SO99] Y. Sugitaa and Y. Okamotoa. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.
- [SP99] P. L. Silvestrelli and M. Parrinello. Water molecule dipole in the gas and in the liquid phase. *Phys. Rev. Lett.*, 82:3308–3311, 1999.
- [SS94] B. Silvi and A. Savin. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature*, 371:683–686, 1994.
- [SSZ⁺06] M. Sundling, N. Sukumar, H. Zhang, M. J. Embrechts, and C. M. Breneman. Wavelets in chemistry and cheminformatics. *Reviews in Computational Chemistry*, 22:295–322, 2006.
- [Sti80] F. H. Stillinger. Water revised. *Science*, 209:451–457, 1980.
- [TBM92] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, 97(3):1990–2001, 1992.
- [TC98] C. Torrence and G. P. Compo. A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.*, 79:61–78, 1998.

- [THBS11] K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns, and C. D. Sherrill. Assessment of the performance of dft and dft-d methods for describing distance dependence of hydrogen-bonded interactions. *J. Chem. Theory Comput.*, 7:88–96, 2011.
- [TJM⁺05] F. H. Tukhvatullin, A. Jumabaev, G. Muradov, H. A. Hushvaktov, and A. A. Absanov. Raman spectra of cn vibrations of acetonitrile in aqueous and other solutions. experimental results and ab initio calculations. *J. Raman Spectrosc.*, 36:932–937, 2005.
- [TM91] N. Troullier and J. L. Martins. Efficient pseudopotential for plane-wave calculations. *Phys. Rev. B*, 43(3):1993–2006, 1991.
- [Tro59] H. F. Trotter. On the product of semi-groups operators. *Proc. Am. Math. Soc.*, 10:545, 1959.
- [TT98] H. Torii and M. Tasumi. Ab initio molecular orbital study of the amide i vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. *J. Raman Spect.*, 29(1):81–86, 1998.
- [Tuc10a] M. E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 2010.
- [Tuc10b] M. E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts. Oxford University Press Inc., 2010.
- [UP06] H. Uktu and S. Piravadili. Wavelets as a powerful method to filter ir spectra. *Infrared Physics and Technology*, 49:167–172, 2006.
- [VCMMR11] V. V. Volkov, R. Chelli, F. Muniz-Miranda, and R. Righini. Structural properties of a membrane associated anchor dipeptide. *J. Phys. Chem. B*, 115(18):5294–5303, 2011.
- [VCZ⁺07] V. V. Volkov, R. Chelli, W. Zhuang, F. Nuti, Y. Takaoka, A. M. Papini, S. Mukamel, and R. Righini. Electrostatic interactions in phospholipid membranes revealed by coherent 2d ir spectroscopy. *Proc. Natl. Acad. Sci.*, 104(39):15323–15327, 2007.
- [VFMC11] B. Velino, L. B. Favero, A. Maris, and W. Caminati. Conformational equilibria in diols: The rotational spectrum of chiral 1,3-butandiol. *J. Phys. Chem. A*, 115:9585–9589, 2011.
- [VKM⁺05] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Comp.Phys.Comm.*, 167:103–128, 2005.
- [VR09] V. V. Volkov and R. Righini. Partitioning of an anchor dipeptide in a phospholipid membrane. *J. Phys. Chem. B*, 113(50):16246–16250, 2009.
- [Wea79] R.C. Weast, editor. *CRC Handbook of Chemistry and Physics*. CRC Press, Inc., Boca Raton, Florida 33431, 1979.
- [WH00] S. Woutersen and P. Hamm. Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy. *J. Phys. Chem. B*, 104(47):11316–11320, 2000.

Modelling of spectroscopic and structural properties using molecular dynamics

- [WPH⁺02] S. Woutersen, R. Pfister, P. Hamm, Y. Mu, D. S. Kosov, and G. Stock. Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations. *J. Chem. Phys.*, 117(14):6833–6840, 2002.
- [WPXI11] J. C. Werhahn, S. Pandelov, S. S. Xantheas, and H. Iglev. Dynamics of weak, bifurcated, and strong hydrogen bonds in lithium nitrate trihydrate. *J. Phys. Chem. Lett.*, 2:1633–1638, 2011.
- [WS61] J. K. Wilmshurst and S. Senderoff. Vibrational spectra of inorganic molecules. ii. infrared reflection spectra of liquid lithium, sodium, potassium, and silver nitrates. *J. Chem. Phys.*, 35:1078–1084, 1961.
- [WSK01] M. R. Waterland, David Stockwell, and A. M. Kelley. Symmetry breaking effects in no_3^- : Raman spectra of nitrate salts and ab initio resonance raman spectra of nitrate-water complexes. *J. Chem. Phys.*, 114:6249–6258, 2001.
- [ZDG10] C. Zhang, D. Donadio, and G. Galli. First-principle analysis of the ir stretching band of liquid water. *J. Phys. Chem. Lett.*, 1:1398–1402, 2010.
- [ZM92] Y. J. Zheng and K. M. Merz. Study of hydrogen-bonding interactions relevant to biomolecular structures and function. *J. Comput. Chem.*, 13:1151–1169, 1992.

Appendix

A. Some DFT Details

A.1 Proof of Hohenberg-Kohn Theorems

The aim of this Appendix is to demonstrate that $\rho(\mathbf{r})$ determines the nucleus-electron potential V^{Ne} save for an additive constant. Demonstrating this, is equivalent to demonstrating that there is a bijective correspondence between $\rho(\mathbf{r})$ and V^{Ne} .

The demonstration will go on by *ab absurdo* reasoning, so let's just assume that to two different potentials V^{Ne} and V'^{Ne} correspond the same density $\rho(\mathbf{r})$ for the non-degenerate ground state of the system. If this stands true, then we should have two different electronic Hamiltonians \hat{H}^e and \hat{H}'^e with different eigenstates $|\varphi\rangle$ and $|\varphi'\rangle$, such that they give the same total density ρ . If we apply the variational principle (eqn. 2.7) adopting $|\varphi'\rangle$ as a test-ket for the \hat{H} Hamiltonian, we obtain that

$$E_0 \leq \langle \varphi' | \hat{H} | \varphi' \rangle = \langle \varphi' | \hat{H}' | \varphi' \rangle + \langle \varphi' | \hat{H} - \hat{H}' | \varphi' \rangle = E'_0 + \langle \varphi' | \hat{H} - \hat{H}' | \varphi' \rangle \quad . \quad (\text{A.1})$$

With the same reasoning and adopting the test-ket $|\varphi\rangle$ and the Hamiltonian \hat{H}' we obtain

$$E'_0 \leq \langle \varphi | \hat{H}' | \varphi \rangle = \langle \varphi | \hat{H} | \varphi \rangle + \langle \varphi | \hat{H}' - \hat{H} | \varphi \rangle = E_0 + \langle \varphi | \hat{H}' - \hat{H} | \varphi \rangle \quad . \quad (\text{A.2})$$

Moreover, calculating $\langle \varphi' | \hat{H} - \hat{H}' | \varphi' \rangle = \langle \varphi' | V^{Ne} - V'^{Ne} | \varphi' \rangle$ we get

$$\begin{aligned} \langle \varphi' | \hat{H} - \hat{H}' | \varphi' \rangle &= \langle \varphi' | V^{Ne} - V'^{Ne} | \varphi' \rangle = \int \varphi'^* [V^{Ne} - V'^{Ne}] \varphi' \mathbf{d}\mathbf{r} = \\ &= \int \varphi'^* \varphi' [V^{Ne} - V'^{Ne}] \mathbf{d}\mathbf{r} = \int \rho(\mathbf{r}) [V^{Ne} - V'^{Ne}] \mathbf{d}\mathbf{r} = -\langle \varphi | \hat{H}' - \hat{H} | \varphi \rangle \quad . \end{aligned}$$

Thus, the two equations A.1 and A.2 become

$$E_0 \leq E'_0 + \int \rho(\mathbf{r}) [V^{Ne} - V'^{Ne}] \mathbf{d}\mathbf{r} \quad (\text{A.3})$$

and

$$E'_0 \leq E_0 - \int \rho(\mathbf{r}) [V^{Ne} - V'^{Ne}] \mathbf{d}\mathbf{r} \quad . \quad (\text{A.4})$$

If we add eqn. A.4 to eqn. A.3 we obtain the following inequality

$$E_0 + E'_0 \leq E'_0 + E_0 \quad . \quad (\text{A.5})$$

But, if $E_0 + E'_0 = E'_0 + E_0$ was true, then the two states $|\varphi\rangle$ and $|\varphi'\rangle$ were energy degenerate, violating one of the hypothesis. Besides, inequality $E_0 + E'_0 < E'_0 + E_0$ is a clear algebraic contradiction. This demonstrates that there are not two different potential V^{Ne} and V'^{Ne} able to produce the same ground-state density $\rho(\mathbf{r})$: from this follows that the correspondence between $\rho(\mathbf{r})$ and V^{Ne} has to be bijective. This is the first Hohenberg-Kohn theorem.

Modelling of spectroscopic and structural properties using molecular dynamics

The second Hohenberg-Kohn theorem is just an application of the variational theorem to the energy of the system written as a functional of the electronic density of the ground-state:

$$E = \langle \varphi | \hat{H} | \varphi \rangle = E[\rho] = T^e[\rho] + V^{ee}[\rho] + V^{Ne}[\rho] = F_{HK}[\rho] + \int \rho(\mathbf{r}) V^{Ne} d\mathbf{r} ;$$

here $F^{HK} = T^e[\rho] + V^{ee}[\rho]$ is called ‘‘Hohenberg-Kohn functional’’ and does not depend on the specific electronic system, since it is just the sum of electronic kinetic energy ($T^e[\rho]$) and electronic potential energy ($V^{ee}[\rho]$). It is often said that F^{HK} is a ‘‘universal functional’’, because it does not depend on the molecular geometry (*i.e.* the nuclear coordinates $\{\mathbf{R}_N\}$). The second Hohenberg-Kohn theorem states that

With a test electronic density $\tilde{\rho}(\mathbf{r})$, provided that $\tilde{\rho}(\mathbf{r}) \geq 0$ and $\int \tilde{\rho}(\mathbf{r}) d\mathbf{r} = n$, we have that

$$E_0 \leq E[\tilde{\rho}] \quad (\text{A.6})$$

This theorem follows immediately noticing that

$$E[\tilde{\rho}] = \langle \tilde{\varphi} | \hat{H} | \tilde{\varphi} \rangle = \int \tilde{\rho}(\mathbf{r}) V^{Ne} d\mathbf{r} + F^{HK}[\tilde{\rho}] \geq E[\rho] = E_0 .$$

A.2 Discussion about Kohn-Sham Equations

In this Appendix a way to obtain the Kohn-Sham Equations (eqn. 2.21) is presented. following the treatment of Parr and Yang [PY89]. Equation 2.18 can be written in terms of single-electron wavefunctions as

$$E[\rho] = \sum_i \int \phi_i^*(\mathbf{r}) \left(-\frac{1}{2} \nabla_{(i)}^2 \right) \phi_i(\mathbf{r}) d\mathbf{r} + \int d\mathbf{r} \rho(\mathbf{r}) V^{Ne} + \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho]$$

$$\rho(\mathbf{r}) = n \int \dots \int |\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)|^2 ds_1 dx_2 \dots dx_N = \sum_j |\phi^j(\mathbf{x})|^2$$

$$\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{1}{\sqrt{n!}} \det [\phi^1 \phi^2 \dots \phi^n]$$

Single-electron states $|\phi^i\rangle$ are, by a hypothesis, a complete basis, *i.e.* they should span the entire Hilbert space accessible to the system. Thus, if $|\phi^i\rangle$ are a complete basis, then they can always be arranged as an orthonormal basis with some unitary transformation, to obtain the orthonormality condition

$$\langle \phi_i | \phi^j \rangle = \int \phi_i^*(\mathbf{r}) \phi^j(\mathbf{r}) d\mathbf{r} = \delta_{ij}$$

Now a new functional Y of single-electron states $\{|\phi^i\rangle\}$ can be defined:

$$Y[|\phi^i\rangle] = E[\rho] - \sum_{ij} \epsilon_{ij} \langle \phi_i | \phi^j \rangle$$

To minimise the $E[\rho]$ functional with respect to the $|\phi^i\rangle$ single-electron states, then the following condition must hold

$$\delta Y[|\phi^i\rangle] = 0$$

which is the analogous of eqn. 2.14. Adopting the variational calculus, these equations lead to

$$\hat{h}_{(i)}^{\text{eff}}|\phi^i\rangle = \left[-\frac{1}{2}\nabla_{(i)}^2 + V^{\text{eff}}\right]|\phi^i\rangle = \sum_j \epsilon_{ij}|\phi^j\rangle$$

Here $\hat{h}_{(i)}^{\text{eff}} = \left[-\frac{1}{2}\nabla_{(i)}^2 + V^{\text{eff}}\right]$ is the single-electron Kohn-Sham Hamiltonian, but the equation still does not appear like eqn. 2.21. However, since the Kohn-Sham Hamiltonian is a self-adjunct operator, the matrix of its eigenvalues can always be made diagonal by a unitary transformation that leads us to eqn. 2.21.

A.3 Exchange-Correlation Functionals

The simpler functional: LDA

The most crude approximation done in DFT is the so-called local density approximation (LDA). This approximation is rather inadequate to treat complex molecular systems, but it easily shows how the DFT equations work in practice. Within LDA one assumes that near \mathbf{r} the function $\rho(\mathbf{r})$ does not change significantly, so that one could write down $\nabla\rho(\mathbf{r}) \approx 0$ and

$$E_{xc}^{LDA} = \int \rho(\mathbf{r})\epsilon_{xc}^{LDA}[\rho]\mathbf{d}\mathbf{r} \quad ,$$

with $\epsilon_{xc}^{LDA}[\rho] = \epsilon_x^{LDA}[\rho] + \epsilon_c^{LDA}[\rho]$ being the exchange-correlation energy of a single electron in a constant-density gas. The energy ϵ_{xc}^{LDA} is assumed to be the sum of an exchange energy derived from Thomas-Fermi-Dirac theory

$$\epsilon_x^{LDA}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}\rho(\mathbf{r})^{\frac{2}{3}} \quad ,$$

and a correlation energy that is obtained by previous computer simulations (mainly Quantum Monte-Carlo). The resulting exchange-correlation potential is

$$V_{xc}^{LDA} = \frac{\delta E_{xc}^{LDA}}{\delta\rho(\mathbf{r})} = \epsilon_{xc}^{LDA}[\rho] + \rho(\mathbf{r})\frac{\partial\epsilon_{xc}^{LDA}}{\partial\rho} \quad ,$$

which, when put into the equation 2.21, gives the LDA Kohn-Sham equation

$$\left\{-\frac{1}{2}\nabla_i^2 + V^{Ne} + \sum_i \int \frac{\rho(\mathbf{r}_j)}{\mathbf{r}_{ij}}\mathbf{d}\mathbf{r}_j + \epsilon_{xc}^{LDA}[\rho] + \rho(\mathbf{r})\frac{\partial\epsilon_{xc}^{LDA}}{\partial\rho}\right\}\phi^i = \epsilon_i\phi^i \quad . \quad (\text{A.7})$$

It has to be noticed that the LDA functional, albeit simple and inaccurate for molecular systems, nevertheless should be considered fully *ab initio*. In fact, the more complex functional discussed in the following are both more complex and accurate, yet they contain some part of empiricism.

BLYP, a GGA Functional

To deal with real quantum chemical problems, nowadays LDA is discarded and new DFT methods are employed. DFT methods that take into account the first derivative of the electron density (*i.e.*, such that $V_{xc} = V_{xc}[\rho(\mathbf{r}), \nabla\rho(\mathbf{r})]$) are called Generalized Gradient

Approximation (GGA) methods. The first corrective term to the exchange energy is usually of the type:

$$-\frac{5}{216}(3\pi^5)^{-1/3} \int \frac{|\nabla\rho(r)|^2}{\rho(r)^{4/3}} dr \quad ,$$

where the gradient of the density is present. An analogous formula exists for the correlation energy. One of the first and most widely used functional of this type was the B88 functional proposed by Becke [Bec88] as exchange correction to the LDA functional:

$$E_x^{B88} = E_x^{LDA} + \Delta E_x^{B88}$$

$$\Delta E_x^{B88} = -\beta\rho^{1/3} \frac{x^2}{1 + 6\beta x \sinh^{-1} x}$$

$$x = \frac{|\nabla\rho|}{\rho^{4/3}} \quad ;$$

here the β parameter is obtained by fitting of rare gas data. Just this simple correction improves the LDA results of two order of magnitude. A widely adopted correlation correction is included into the LYP functional, proposed by Lee, Yang and Parr [LYP88], containing four parameters (a, b, c and d) extrapolated from Helium atom data fitting:

$$E^{LYP} = -4a \frac{\rho^+ \rho^-}{\rho^2(1+d\rho^{1/3})} - ab\omega \left\{ \frac{\rho^+ \rho^-}{18} \left[144(2^{2/3})C_F((\rho^+)^{8/3} + (\rho^-)^{8/3}) \right. \right.$$

$$+ (47 - 7\delta) |\nabla\rho|^2 - (45 - \delta)(|\nabla\rho^+|^2 + |\nabla\rho^-|^2)$$

$$\left. \left. + 2\rho^{-1}(11 - \delta)(\rho^+ |\nabla\rho^+|^2 + \rho^- |\nabla\rho^-|^2) \right] \right\} \quad (A.8)$$

$$\omega = \frac{\exp[-c\rho^{1/3}]}{\rho^{4/3}(1+d\rho^{-1/3})} \quad (A.9)$$

$$\delta = c\rho^{-1/3} + \frac{d\rho^{-1/3}}{1+d\rho^{-1/3}} \quad , \quad (A.10)$$

with ρ^+ and ρ^- densities of the spin states. In fact, there is no correlation if all spin vectors are parallel to each others. The B88 and LYP functionals are often used together and, when this happens, it is usual to refer to this pair as "BLYP" functional. The BLYP functional is the more adopted functional in this dissertation, for a number of reasons, the main ones being the good accuracy in reproducing static and dynamic properties, and the fact that it is implemented inside the CPMD suite of programs.

An hybrid Functional: B3LYP

Basically, hybrid functionals (sometimes called Hyper-GGA) are GGA functional where the exact Hartree-Fock exchange term is added as a corrective term. The HF exchange integral has the form:

$$K_{ij} = \iint \phi_j(\mathbf{x}_1)\phi_i^*(\mathbf{x}_1) \frac{1}{r_{12}} \phi_j^*(\mathbf{x}_2)\phi^i(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \quad . \quad (A.11)$$

Since hybrid functionals mix together HF and DFT features, the reason of their name is obvious. One of the most employed hybrid functionals is the so-called B3 exchange [Bec93b, Bec93a], which is dependent from three parameters a, b and c extrapolated by

data-fitting, and which is often combined with the LYP correlation to obtain the B3LYP functional:

$$E_{xc}^{B3LYP} = (1 - a)E_x^{LDA} + aE_x^{HF} + b\Delta E_x^{B88} + (1 - c)E_c^{LDA} + cE_c^{LYP} \quad ;$$

here with the symbol E_x^{HF} I mean the total exchange energy of the HF method, which is defined as a sum of the n double-electron integrals K_{ij} :

$$E_x^{HF} = - \sum_i \sum_{j \leq i} K_{ij} \quad .$$

The exact HF exchange energy is scaled by a factor a : this is due to the fact that an exact exchange long range energy (*i.e.* E_x^{HF}) is not compensated by an approximated long range correlation (*i.e.* E^{LYP}) as the theory requires (see [ea01]).

Unfortunately, the B3LYP is not currently integrated into the CPMD package, nor other Hybrids, because it is difficult (both at the practical and theoretical level) to implement HF exchange with plane waves as basis set. Anyway, the B3LYP functional, due to its effectiveness in reproducing many molecular properties, has risen to a position of "reference DFT calculation" and in the present dissertation is often used in static calculations in order to check the quality of BLYP calculations.

I have also to pinpoint that calculations with LDA or a GGA functional are much faster than analogous calculations carried out with a hybrid one. This is due to the fact that the HF exchange (eqn. A.11) is non-local, because cannot be written as an integral of the type

$$\iint \frac{\rho(\mathbf{x}_1)\rho(\mathbf{x}_2)}{r_{12}} d\mathbf{x}_1 d\mathbf{x}_2 \quad (\text{A.12})$$

due to the presence of the mixed terms $\phi_j(\mathbf{x}_1)\phi_i^*(\mathbf{x}_1)$ and $\phi_j^*(\mathbf{x}_2)\phi^i(\mathbf{x}_2)$ that prevent a simple reduction to single-electron densities $\rho(\mathbf{x}_1)$ and $\rho(\mathbf{x}_2)$. The integral in eqn.A.12 is the Coulomb integral (see for example ref. [Bec14], eqn. (16)).

B. Some MD Details

B.1 Liouville Operator and Integration Algorithms

The time-rate of a function $A = A(\mathbf{q}(t), \mathbf{p}(t))$ can be put as

$$\frac{dA}{dt} = \sum_i^{3N} \dot{q}^i \frac{\partial A}{\partial q^i} + \dot{p}_i \frac{\partial A}{\partial p_i} = \sum_i^{3N} \frac{\partial A}{\partial q^i} \frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial H}{\partial q^i} . \quad (\text{B.1})$$

It can be created a Liouville operator $i\mathbb{L}$ defined as (see [PMM⁺df]):

$$i\mathbb{L} = \sum_i^{3N} \dot{q}^i \frac{\partial}{\partial q^i} + \dot{p}_i \frac{\partial}{\partial p_i} = \sum_i^{3N} \dot{q}^i \nabla_{q^i} + \dot{p}_i \nabla_{p_i} = \dot{\mathbf{q}}(t) \nabla_{\mathbf{q}} + \dot{\mathbf{p}}(t) \nabla_{\mathbf{p}} ; \quad (\text{B.2})$$

the operator is self-adjunct and symbols $\nabla_{\mathbf{q}}$ and $\nabla_{\mathbf{p}}$ refer to the coordinate and momenta divergence, respectively. With the help of this operator, eqn. B.1 can be rewritten as

$$\frac{dA}{dt} = i\mathbb{L}A . \quad (\text{B.3})$$

This first order differential equation can be easily integrated to obtain

$$A(\mathbf{q}, \mathbf{p}) = e^{i\mathbb{L}t} A(\mathbf{q}(0), \mathbf{p}(0)) . \quad (\text{B.4})$$

It can be defined a new operator

$$e^{i\mathbb{L}t} = \hat{U}(t) , \quad (\text{B.5})$$

which is called “propagator” of the dynamics, because when acts on $A(\mathbf{q}(0), \mathbf{p}(0))$ it propagates the A function in time to obtain $A(\mathbf{q}(t), \mathbf{p}(t))$ due to eqn. B.4:

$$\hat{U}(t) \cdot \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} = e^{i\mathbb{L}t} \cdot \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} = \begin{pmatrix} \mathbf{q}(t) \\ \mathbf{p}(t) \end{pmatrix} .$$

Furthermore, it is a unitary operator, since we have

$$\hat{U}(t) \hat{U}(t)^\dagger = e^{i\mathbb{L}t} e^{-i\mathbb{L}t} = e^{i\mathbb{L}(t-t)} = e^0 = \mathbb{I} ,$$

with \mathbb{I} representing the identity operator. This is an important property that provides that the equations of motion solved with such propagator are time-reversible as required by theory. In turn, time reversible integrators allow the conservation of energy, which usually drifts when approached with other algorithms. The Liouville operator can be thought as composed of two sub-operators, one acting on the coordinates, $i\mathbb{L}_q(t) = \dot{\mathbf{q}}(t) \nabla_{\mathbf{q}}$, and the other on the momenta, $i\mathbb{L}_p(t) = \dot{\mathbf{p}}(t) \nabla_{\mathbf{p}}$:

$$i\mathbb{L}(t) = i\mathbb{L}_q(t) + i\mathbb{L}_p(t) .$$

Modelling of spectroscopic and structural properties using molecular dynamics

using eqn. B.4, the action of $i\mathbb{L}_q(t)$ can be underlined:

$$\begin{aligned} A(\mathbf{q}(t), \mathbf{p}(t)) &= A(\mathbf{q}(0), \mathbf{p}(0)) + i\mathbb{L}_q t A(\mathbf{q}(0), \mathbf{p}(0)) + \frac{(i\mathbb{L}_q t)^2}{2!} A(\mathbf{q}(0), \mathbf{p}(0)) + \dots \\ &= e^{i\mathbb{L}_q(0)t} A(\mathbf{q}(0), \mathbf{p}(0)) = \sum_{k=0}^{\infty} \frac{(\dot{\mathbf{q}}(0)t)^k}{k!} \nabla_{\mathbf{q}}^k A(\mathbf{q}(0), \mathbf{p}(0)) = \\ &= A(\mathbf{q}(0) + t\dot{\mathbf{q}}(0), \mathbf{p}(0)) . \end{aligned}$$

It can be appreciated that the $e^{i\mathbb{L}_q t}$ operator has caused a time-shift of the coordinates. In the same way it can be demonstrated that operator $e^{i\mathbb{L}_p t}$ can induce a time-shift on momenta. Thus, the combined effect of these operators is a time-shift of the entire state of the system and they can be applied to generate a continuous dynamics. However, the true propagator is $e^{i\mathbb{L}t}$ given in eqn. B.5, and this propagator has both the addends $i\mathbb{L}_q(0)$ and $i\mathbb{L}_p(0)$. Unfortunately, $i\mathbb{L}_q(0)$ and $i\mathbb{L}_p(0)$ do not commute with each other

$$e^{i\mathbb{L}_q(0) + i\mathbb{L}_p(0)} \neq e^{i\mathbb{L}_q(0)} e^{i\mathbb{L}_p(0)} .$$

This problem has been tackled by Trotter [Tro59] who proved the following theorem for two non-commuting operators α and β :

$$e^{\alpha + \beta} = \lim_{\tau \rightarrow +\infty} (e^{\frac{\alpha}{2\tau}} e^{\frac{\beta}{\tau}} e^{\frac{\alpha}{2\tau}}) = \lim_{\tau \rightarrow +\infty} (e^{\frac{\beta}{2\tau}} e^{\frac{\alpha}{\tau}} e^{\frac{\beta}{2\tau}}) . \quad (\text{B.6})$$

Therefore, when $\tau \rightarrow +\infty$ the propagator can be partitioned in two independent operators whose application is relatively easy. Obviously the limit $\tau \rightarrow +\infty$ cannot be reached and this is a spring of numerical error, but it is unavoidable in the context of numerical simulations. Defining

$$\begin{aligned} \tau &\equiv \frac{t}{\Delta t} \\ \frac{\alpha}{\tau} &\equiv \frac{i\mathbb{L}_p t}{\tau} = \Delta t \dot{\mathbf{p}}(0) \nabla_{\mathbf{p}} \end{aligned}$$

and

$$\frac{\beta}{\tau} \equiv \frac{i\mathbb{L}_q t}{\tau} = \Delta t \dot{\mathbf{q}}(0) \nabla_{\mathbf{q}}$$

and exploiting Trotter's formula, the propagator becomes

$$\hat{U}(t) = e^{i\mathbb{L}t} = \lim_{\Delta t \rightarrow 0, t \rightarrow +\infty} e^{i\mathbb{L}_p \frac{\Delta t}{2}} e^{i\mathbb{L}_q \Delta t} e^{i\mathbb{L}_p \frac{\Delta t}{2}} , \quad (\text{B.7})$$

where Δt is the finite time-step adopted to integrate the equations of motion. Obviously, analytically solving the molecular classical equations of motion is not possible, and the time range is discretised in time-steps; during each time-step dynamical variables are assumed to be constant. To finally deduce the integration algorithm is useful to analyse how acts

$$e^{i\mathbb{L}_p \frac{\Delta t}{2}} e^{i\mathbb{L}_q \Delta t} e^{i\mathbb{L}_p \frac{\Delta t}{2}}$$

on the (classical) initial state vector

$$\begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} .$$

We find that:

$$e^{i\mathbb{L}_p \frac{\Delta t}{2}} e^{i\mathbb{L}_q \Delta t} e^{i\mathbb{L}_p \frac{\Delta t}{2}} \cdot \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} = \begin{pmatrix} \mathbf{q}(0) + \Delta t \dot{\mathbf{q}}(\frac{\Delta t}{2}) \\ \mathbf{p}(0) + \frac{\Delta t}{2} [\dot{\mathbf{p}}(0) + \dot{\mathbf{p}}(\Delta t)] \end{pmatrix} =$$

$$= \begin{pmatrix} \mathbf{q}(0) + \Delta t \dot{\mathbf{q}}(0) + \dot{\mathbf{p}}(0) \frac{(\Delta t)^2}{2} \\ \mathbf{p}(0) + \frac{\Delta t}{2} [\dot{\mathbf{p}}(0) + \dot{\mathbf{p}}(\Delta t)] \end{pmatrix}$$

Thus, the so-called “velocity verlet” algorithm at every time-step Δt updates the state vector as

$$p_i(t_n) \longrightarrow p_i(t_{n+1}) = p_i(t_n + \Delta t) = p_i(t_n) + \frac{\Delta t}{2} [\dot{p}_i(t_n) + \dot{p}_i(t_{n+1})] \quad (\text{B.8})$$

$$q^i(t_n) \longrightarrow q^i(t_{n+1}) = q^i(t_n + \Delta t) = q^i(t_n) + \dot{q}^i(t_n) \Delta t + \dot{p}_i(t_n) \frac{(\Delta t)^2}{2} \quad (\text{B.9})$$

adopting the approximated propagator

$$e^{i\mathbb{L}_p \frac{\Delta t}{2}} e^{i\mathbb{L}_q \Delta t} e^{i\mathbb{L}_p \frac{\Delta t}{2}} .$$

It is immediate to check than equations B.8 and B.9 correspond to equations 2.34 and 2.36.

Verlet-based algorithms are not the best ones to guarantee that energy and other relevant constant of motion are conserved from one time-step to the subsequent: as a matter of fact, “predictor-corrector” algorithms can succeed much better in this specific task and generate a trajectory that is much closer to the real one. The point is that verlet-type integration schemes display a better behaviour during the overall simulation, with small drifts in conserved quantities: this quality is often made explicit by saying that “verlets are less accurate but more robust”. Moreover, as pinpointed in Ref. [Mar04], pag. 372, “the errors do not accumulate” since they are prone to erase each others.

B.2 Replica Exchange

In MD the mean quantity $\langle \theta \rangle$ is calculated as a time-average. On the contrary, in Monte-Carlo simulations (MC) the systems explores different geometric configurations not interconnected by a true time-dependent trajectory, so the mean quantity $\langle \theta \rangle$ is computed as a true state-average. The replica-exchange (REM) [SO99] method aims to retain the best of both worlds and to speed up the exploration of the phase space, which is a critical requirement to investigate large molecular systems. Basically, with REM many different simulations are performed concurrently on the same system, but at different temperatures: they are called replicas. The dynamics evolves according to normal MD in all replicas, but from time to time a MC move is attempted to switch configurations between replicas according to the Metropolis acceptance rule.

Let’s follow the dynamics of the 300 K replica, for example, and imagine that after some MD steps configuration are switched with the 500 K replica. When the switching occurs, the velocities of the 500 K replica are rescaled to match the 300 K requirement, and *vice versa*. If we want to calculate properties at 300 K, we have to focus just onto the 300 K replica event after the switching with the 500 K replica.

It has been demonstrated that this approach is able to speed up computations, albeit forcing to perform many different simulations. A slightly different REM scheme, called “hamiltonian-REM” [FWT02], requires that the various replicas do not differ in temperature but in the Hamiltonian: within this approach, the classical potentials of the different replicas are scaled by a factor that weaken or strengthen interactions. This is similar to “normal” REM, since the dynamical effects of rising temperatures (*i.e.* kinetic energy) or decreasing potential energy is about the same. Anyway, hamiltonian-REM allows turning off specific interactions selectively (*e.g.* intra-molecular ones, solute-solvent ...) that REM does not.

Modelling of spectroscopic and structural properties using molecular dynamics

The drawback of this technique is that REM-MD cannot be used to study time-dependent phenomena, since the MC-moves disrupt the time-sequence trajectory.

C. Some *Ab Initio* MD Details

C.1 Hellman-Feynmann Theorem

In quantum mechanics every observable variable is represented by a self-adjunct operator acting on a suitable Hilbert space. So, it appears that there is no space for the classical notion of force. However, the classical concept of “force” can be introduced into a quantum theoretical framework by the Hellman-Feynmann theorem that will be reported in this Appendix.

I call $\hat{H}(\lambda)$ the Hamiltonian operator of the system that happens to depend on the λ parameter. $|\psi_E\rangle$ is autoket of $\hat{H}(\lambda)$ so that $\hat{H}(\lambda)|\psi_E\rangle = E(\lambda)|\psi_E\rangle$. Then

$$\langle\psi_E|\hat{H}(\lambda)|\psi_E\rangle = E(\lambda)\langle\psi_E|\psi_E\rangle = E(\lambda) \quad .$$

Taking the derivative of $E(\lambda)$ with respect of λ we obtain:

$$\begin{aligned} \frac{\partial E(\lambda)}{\partial\lambda} &= \frac{\partial}{\partial\lambda}\langle\psi_E|\hat{H}(\lambda)|\psi_E\rangle = \\ &= \left(\frac{\partial}{\partial\lambda}\langle\psi_E|\right)\hat{H}(\lambda)|\psi_E\rangle + \langle\psi_E|\left(\frac{\partial}{\partial\lambda}\hat{H}(\lambda)\right)|\psi_E\rangle + \langle\psi_E|\hat{H}(\lambda)\left(\frac{\partial}{\partial\lambda}|\psi_E\rangle\right) = \\ &= \left\langle\frac{\partial\hat{H}(\lambda)}{\partial\lambda}\right\rangle + E\left(\frac{\partial}{\partial\lambda}\langle\psi_E|\right)|\psi_E\rangle + E\langle\psi_E|\left(\frac{\partial}{\partial\lambda}|\psi_E\rangle\right) = \\ &= \left\langle\frac{\partial\hat{H}(\lambda)}{\partial\lambda}\right\rangle + E\frac{\partial}{\partial\lambda}\langle\psi_E|\psi_E\rangle = \left\langle\frac{\partial\hat{H}(\lambda)}{\partial\lambda}\right\rangle \quad , \end{aligned} \quad (\text{C.1})$$

where we have exploited the fact that $\frac{\partial}{\partial\lambda}\langle\psi_E|\psi_E\rangle = 0$ since the state $|\psi_E\rangle$ does not depend on λ . The equality $\frac{\partial}{\partial\lambda}E(\lambda) = \left\langle\frac{\partial}{\partial\lambda}\hat{H}(\lambda)\right\rangle$ can be related to classical “forces” if we adopt nuclear coordinates $\{\mathbf{R}_N\}$ as parameter for the Hamiltonian in eqn. 2.5, obtaining the following relation

$$\nabla_{\mathbf{R}_N}E(\mathbf{R}_N) = \langle\nabla_{\mathbf{R}_N}\hat{H}^e(\mathbf{R}_N)\rangle = \langle\varphi_E|\nabla_{\mathbf{R}_N}\hat{H}^e(\mathbf{R}_N)|\varphi_E\rangle \quad .$$

But from the classical mechanics, we know that for conservative forces we have (eqn. 2.32)

$$F_i = \dot{p}_i = -\frac{\partial\mathcal{H}}{\partial q^i} \quad ,$$

so that we can define the forces acting on the atomic nuclei as

$$F_N = \nabla_{\mathbf{R}_N}E(\mathbf{R}_N) = \langle\varphi_E|\nabla_{\mathbf{R}_N}\hat{H}^e(\mathbf{R}_N)|\varphi_E\rangle \quad , \quad (\text{C.2})$$

is the Hellmann-Feynman theorem.

C.2 Pulay Forces and BSSE

Eqn. C.2 is the theoretical pillar of much *ab initio* MD and requires correct eigenfunctions to calculate reliable semi-classical forces. Nevertheless, SCF methods (like DFT) are more accurate in calculating eigenvalues than eigenfunctions. Moreover, force calculations based on eqn. C.2 are affected by another problem: we assumed that electronic eigenfunctions of \hat{H}^e were independent from $\{\mathbf{R}_N\}$ parameters, but this fact is only superficially true.

In fact, electronic autokets are built as linear combinations of a vector basis $|k\rangle$ that, in real computer calculations, often depend on the nuclear coordinates. The Gaussian bases (see Appendix 2.1.5) in static *ab initio* calculations are usually constructed around nuclear position to better describe the static features of molecules. They are called localised bases for this reason. In static calculations localised bases do not create serious problems: the error induced by localised bases, called “basis set superposition error” (BSSE), can be effectively tackled with some computational tricks, for example adopting the “counterpoise” procedure included into the Gaussian03 and Gaussian09 packages [FTS⁺a, FTS⁺b].

However, adopting these Gaussian localised basis to describe molecules that actually move and interact with others, like in *ab initio* MD, the position changes introduce serious errors in the force calculations at every time-step. Thus, performing *ab initio* MD with localised bases, the forces are cursed by so-called “Pulay forces” [RP95] that are spurious. There is no currently available way to erase these spurious forces.

This is one of the reason why CPMD employ non-localised plane waves that spans all the available space, which moreover are orthogonal by construction.

C.3 Pseudopotentials

Pseudopotentials are employed in CPMD in order to minimise the number of plane waves used to describe electrons. Since core electrons would require a huge amount of plane waves to correctly reproduce fast oscillations in their corresponding wavefunctions, pseudopotentials effectively replace core-electrons eigenvalues. Pseudopotentials have to correctly simulate the limit behaviour of the core electrons wavefunctions where these approach a certain cut-off radius. Inside this cut-off, the pseudopotential and the wavefunctions should be slow oscillating to reduce the number of plane waves to employ. Moreover, it is an important requirement that the pseudopotential have to be transferable, which means that pseudopotentials could be adopted in calculations for different chemical environments producing results of the same accuracy.

Usually pseudopotentials are categorised as “norm conserving” or “ultrasoft”. A norm conserving pseudopotential (*e.g.* see Ref. [TM91] and [GTH96]) is constructed to reproduce the valence properties of all electron-calculation and produce norm conserving pseudo-wavefunctions that satisfy the usual orthonormality condition. To achieve this requirement, usually norm conserving pseudo-wavefunctions tend to retain a certain amount of oscillating behaviour. Ultrasoft pseudopotentials, on the other hand, create pseudo-wavefunctions that are as “smooth” as possible to minimise the range in Fourier space needed to accurately describe valence properties. Ultrasoft pseudopotentials have not been used in this dissertation since they do not allow to calculate with the desired accuracy some structural and spectroscopic properties (such as $g(r)$, IR intensities and MLWF centres).

C.4 Van der Waals Forces

As shown in Section 2.2.5, classical MD uses force fields containing Lennard-Jones type potentials to describe weak interactions such as van der Waals forces. Thus, in the standard CPMD approach, van der Waals forces are almost entirely lost due to the shortcomings of actual implemented exchange-correlation functionals (see Appendix A.3). Stefan Grimme [Gri06] proposed a semi-empirical correction E_{Grimme} to the DFT energy in order to retain dispersive effects:

$$E_{tot} = E_{DFT} + E_{Grimme} \quad .$$

For a pair of atoms j and i , their interaction energy calculated within the Grimme approach takes the form

$$E_{Grimme} \sim \sum_{j=1}^{N-1} \sum_{i=1+j}^N \frac{c^{ji} f_{dump}(\mathbf{R}_{ji})}{\mathbf{R}_{ji}^6} \quad ,$$

where c^{ji} is the dispersion coefficient for the ji pair of atoms calculated as

$$c^{ji} = \sqrt{c^j c^i} \quad ,$$

N is the total number of atoms and f_{dump} is a dumping function used to avoid singularity when the interatomic distance \mathbf{R}_{ji} is small, defined as

$$f_{dump}(\mathbf{R}_{ji}) = \frac{1}{1 + e^{d(\mathbf{R}_{ji}/\mathbf{R}_{ji}^r) - 1}}$$

with \mathbf{R}_{ji}^r being the sum of the van der Waals radii for atoms i and j .

Acknowledgements

This work was possible only through fruitful and productive collaborations, providing opportunities to meet technical and scientific challenges.

First and foremost, I thank my supervisor, Prof. Vincenzo Schettino, for advising me and giving inspiration for the research work. I also have to thank Prof. Gianni Cardini, whose invaluable computational experience saved the day in many occasions. I thank Prof. Roberto Righini, whose point of view, much more experimental-oriented, always suggested wise alternative approaches to problems. Last but not least, I have to thank Dr. Marco Pagliai, the post-doctoral fellow with whom I worked the most during these three years.

I want to thank my college friends too, who shared with me the burdens of the Ph.D. *course*, which, too often, are enough to change it into a Ph.D. *curse*. In particular, my deepest thanks go to Laura, Elvy (no, I still haven't watched TBBT), and Samuele: from the beginning of our chemistry degree, till the last days of the doctorate, you have always stood by my side, and I'm deeply in debt with you all. Thank you so much.

My deepest thanks also go to all my other college friends, which unfortunately now I meet too rarely that I'd want: Lorenzo & Stefano (both of them essential fellows of many "nerd nights"), Irene (who I also thank for our long international calls via Skype®), Jacopo, Marta, Daniele, Nicola, Luisa (who was the first to experience all the horrors of the Italian "3+2" university reforms) and Riccardo (who was about to lose its train to Lausanne and remain in Florence!)

I thank Giulia, who I had the luck to meet at an international conference.

I want to thank all the dwellers of Room 313 of the Chemistry Dept. that I had the pleasure to know: in particular Simona, Nuno (and its awesome lighter!), Federico, Enrica, Giulia (from Pisa), and Paolo (whose odyssey in the Pyrenees is still remembered and celebrated).

I thank the other organisers of the 3rd Ph.D.-Day here in Sesto Fiorentino, in particular Antonio (a.k.a. Priscilla) and Lucio (along with Laura and Elvy, already mentioned above, and Andrea, below).

I also absolutely want to thank my amazing comics fandom friends (which, by the way, are all geologists): Andrea (a man I've the honour to call a friend since we were both in the middle school), Emanuele, and Pietro.

I sincerely wish you all the best.

And, last but not least, I wish to thank my parents, who supported me even during the most awful days of this 3-years work and encouraged me to go on.¹

The good news is that, after the bachelor and the master theses, this is the last dissertation I'll ever write. Or so I do hope.

Francesco Muniz

Friday, December 21st, 2012, Sesto Fiorentino

¹Here a citation of something like "*Spartans never surrender, Spartans never retreat*" would be marvellous. Unfortunately, my parents aren't used to expressing themselves by words like these.

Modelling of spectroscopic and structural properties using molecular dynamics

Note added in August, 2014: I have to thank Dr. Alfonso Pedone for allowing me to carry on new research at his laboratory at UniMoRE as a post-doctoral fellow, also leaving me time to make some corrections to the present text. Differently from what I wrote in December, 2012, it seems I've once again to (re)write my dissertation for publication. I'm grateful to the Firenze University Press (FUP) for that. My academic career (who knows how long it will last...) can be followed at

<https://sites.google.com/site/compmaterchem/people/francesco-muniz-miranda>

https://scholar.google.it/citations?user=puaMVC_EAAAAAJ&hl=en

https://www.researchgate.net/profile/Francesco_Muniz-Miranda?ev=hdr_xprf

<https://personale.unimore.it/rubrica/dettaglio/fmunizmi>

PREMIO TESI DI DOTTORATO

ANNO 2007

- Bracardi M., *La Materia e lo Spirito. Mario Ridolfi nel paesaggio umbro*
Coppi E., *Purines as Transmitter Molecules. Electrophysiological Studies on Purinergic Signalling in Different Cell Systems*
Mannini M., *Molecular Magnetic Materials on Solid Surfaces*
Natali I., *The Ur-Portrait. Stephen Hero ed il processo di creazione artistica in A Portrait of the Artist as a Young Man*
Petretto L., *Imprenditore ed Università nello start-up di impresa. Ruoli e relazioni critiche*

ANNO 2008

- Bemporad F., *Folding and Aggregation Studies in the Acylphosphatase-Like Family*
Buono A., *Esercizio, istituzioni, territorio. Alloggiamenti militari e «case Herme» nello Stato di Milano (secoli XVI e XVII)*
Castenasi S., *La finanza di progetto tra interesse pubblico e interessi privati*
Colica G., *Use of Microorganisms in the Removal of Pollutants from the Wastewater*
Gabbiani C., *Proteins as Possible Targets for Antitumor Metal Complexes: Biophysical Studies of their Interactions*

ANNO 2009

- Decorosi F., *Studio di ceppi batterici per il biorisanamento di suoli contaminati da Cr(VI)*
Di Carlo P., *I Kalasha del Hindu Kush: ricerche linguistiche e antropologiche*
Di Patti F., *Finite-Size Effects in Stochastic Models of Population Dynamics: Applications to Biomedicine and Biology*
Inzitari M., *Determinants of Mobility Disability in Older Adults: Evidence from Population-Based Epidemiologic Studies*
Macrì F., *Verso un nuovo diritto penale sessuale. Diritto vivente, diritto comparato e prospettive di riforma della disciplina dei reati sessuali in Italia*
Pace R., *Identità e diritti delle donne. Per una cittadinanza di genere nella formazione*
Vignolini S., *Sub-Wavelength Probing and Modification of Complex Photonic Structures*

ANNO 2010

- Fedi M., *«Tuo lumine». L'accademia dei Risvegliati e lo spettacolo a Pistoia tra Sei e Settecento*
Fondi M., *Bioinformatics of genome evolution: from ancestral to modern metabolism. Phylogenomics and comparative genomics to understand microbial evolution*
Marino E., *An Integrated Nonlinear Wind-Waves Model for Offshore Wind Turbines*
Orsi V., *Crisi e Rigenerazione nella valle dell'Alto Khabur (Siria). La produzione ceramica nel passaggio dal Bronzo Antico al Bronzo Medio*
Polito C., *Molecular imaging in Parkinson's disease*
Romano R., *Smart Skin Envelope. Integrazione architettonica di tecnologie dinamiche e innovative per il risparmio energetico*

ANNO 2011

- Acciaioli S., *Il trompe-l'œil letterario, ovvero il sorriso ironico nell'opera di Wilhelm Hauff*
Bernacchioni C., *Sfingolipidi bioattivi e loro ruolo nell'azione biologica di fattori di crescita e citochine*
Fabbri N., *Bragg spectroscopy of quantum gases: Exploring physics in one dimension*
Gordillo Hervás R., *La construcción religiosa de la Hélade imperial: El Panhelenion*
Mugelli C., *Indipendenza e professionalità del giudice in Cina*
Pollastri S., *Il ruolo di TAF12B e UVR3 nel ciclo circadiano dei vegetali*
Salizzoni E., *Paesaggi Protetti. Laboratori di sperimentazione per il paesaggio costiero euro-mediterraneo*

ANNO 2012

- Adessi A., *Hydrogen production using Purple Non-Sulfur Bacteria (PNSB) cultivated under natural or artificial light conditions with synthetic or fermentation derived substrates*
- Alberti M., *La 'scoperta' dei disoccupati. Alle origini dell'indagine statistica sulla disoccupazione nell'Italia liberale (1893-1915)*
- Bondi D., *Filosofia e storiografia nel dibattito anglo-americano sulla svolta linguistica*
- Evangelisti E., *Structural and functional aspects of membranes: the involvement of lipid rafts in Alzheimer's disease pathogenesis. The interplay between protein oligomers and plasma membrane physicochemical features in determining cytotoxicity*
- Gualdani R., *Using the Patch-Clamp technique to shed light on ion channels structure, function and pharmacology*
- Petrucci F., *Petri Candidi Decembrii Epistolarum iuveniliium libri octo. A cura di Federico Petrucci*
- Ramalli A., *Development of novel ultrasound techniques for imaging and elastography. From simulation to real-time implementation*

ANNO 2013

- Bacci M., *Coarse-grained molecular dynamics and continuum models for the transport of protein molecules*
- Brancasi I., *Architettura e Illuminismo. Filosofia e progetti di città nel tardo Settecento francese*
- Cucinotta E., *Produzione poetica e storia nella prassi e nella teoria greca di età classica*
- Locatelli M., *Mid infrared digital holography and terahertz imaging*
- Lunghi C., *Early cross-modal interactions and adult human visual cortical plasticity revealed by binocular rivalry*
- Martelli R., *Characteristics of raw and cooked fillets in species of actual and potential interest for Italian aquaculture. Rainbow trout (*Oncorhynchus mykiss*) and meagre (*Argyrosomus regius*)*
- Muniz Miranda F., *Modelling of spectroscopic and structural properties using molecular dynamics*
- Pellegrini L., *Circostanze del reato: trasformazioni in atto e prospettive di riforma*

