

Università degli Studi di Bologna

---

Dipartimento di Chimica Fisica e Inorganica

MOLECULAR DYNAMICS SIMULATIONS  
FOR LIQUID CRYSTALS  
AND PHOTORESPONSIVE SYSTEMS  
XIX Ciclo  
Settore Scientifico Disciplinare: CHIM02

Philosophy Doctor Thesis

Presented by:  
GIUSTINIANO TIBERIO

Thesis Coordinator  
Prof. VINCENZO BALZANI

Supervisor:  
Prof. CLAUDIO ZANNONI

Co-supervisor:  
Prof. ALBERTO ARCIONI  
Dr. ROBERTO BERARDI  
Dr. LUCA MUCCIOLI

---

2004–2006



*Al mio, ormai ex, gruppo di ricerca che mi ha sempre aiutato  
nei momenti difficili!! Grazie a tutti!!*



# Contents

<b>Table of content</b>	<b>4</b>
<b>1 Preface</b>	<b>5</b>
<b>2 Introduction</b>	<b>7</b>
2.1 Azobenzene . . . . .	7
2.2 Liquid crystals: <i>n</i> CB . . . . .	10
<b>3 Molecular Dynamics</b>	<b>11</b>
3.1 Computer simulations . . . . .	11
3.1.1 Physics . . . . .	11
3.1.2 Climate . . . . .	12
3.1.3 Ground water . . . . .	12
3.1.4 Fusion energy . . . . .	13
3.1.5 Life science . . . . .	14
3.1.6 Materials and Chemistry . . . . .	14
3.2 Atomistic simulations . . . . .	16
3.2.1 Introduction . . . . .	16
3.2.2 Hierarchical modelling of materials phenomena . . . . .	17
3.3 Molecular Dynamics technique . . . . .	23
3.3.1 Symplectic and Reversible Integrators . . . . .	28
3.3.2 Multiple Time Steps Algorithms for the Isothermal– Isobaric Ensemble . . . . .	44
3.3.3 Multiple Time Steps Algorithms For Large Size Flexi- ble Systems with Strong Electrostatic Interactions . . . . .	61
3.3.4 The smooth particle mesh Ewald method . . . . .	67
3.3.5 Subdivision the Non Bonded Potential . . . . .	72

<b>4</b>	<b>Azobenzene in organic solvents</b>	<b>81</b>
4.1	Introduction . . . . .	81
4.2	Models and Simulations . . . . .	83
4.2.1	Azobenzene ground and excited states . . . . .	83
4.2.2	Azobenzene excitation and decay . . . . .	85
4.2.3	Solvents and solutions . . . . .	87
4.2.4	Simulation conditions . . . . .	88
4.3	Results and discussion . . . . .	89
4.3.1	Photoisomerization quantum yield . . . . .	90
4.3.2	Kinetic model for isomerization . . . . .	91
4.3.3	Geometry modifications during the isomerization process . . . . .	94
4.3.4	Permanence time in the excited state . . . . .	96
4.3.5	Isomerization mechanism . . . . .	96
4.3.6	Geometrical species identification . . . . .	98
4.4	Conclusions . . . . .	100
<b>5</b>	<b>4,n-alkyl,4'-cyano-biphenyls simulations</b>	<b>105</b>
5.1	Introduction . . . . .	105
5.2	Simulation models . . . . .	106
5.2.1	Force field . . . . .	106
5.2.2	Tuning the molecular force field . . . . .	108
5.3	Phase transition determination . . . . .	110
5.4	Simulation details . . . . .	111
5.5	Results . . . . .	112
5.5.1	Phase transitions . . . . .	112
5.5.2	Radial distributions . . . . .	113
5.5.3	Length analysis . . . . .	122
5.5.4	Smectic phase . . . . .	126
5.6	Conclusions . . . . .	128
<b>6</b>	<b>Azobenzene photo-isomerization in 5CB</b>	<b>129</b>
6.1	Introduction . . . . .	129
6.2	Model . . . . .	129
6.3	Results . . . . .	130
6.3.1	Order parameter $P_2$ . . . . .	130
6.3.2	Photo-isomerization quantum yield and permanence time in $S_1$ . . . . .	131
6.3.3	Geometrical analysis . . . . .	132

6.3.4	Isomerization mechanism . . . . .	134
6.4	Conclusions . . . . .	134
<b>7</b>	<b>Azobenzene in liquid crystal: ESR study</b>	<b>135</b>
7.1	Introduction . . . . .	135
7.2	Experimental . . . . .	136
7.3	Results . . . . .	139
7.4	Discussion . . . . .	144
7.5	Conclusions . . . . .	149
<b>8</b>	<b>Orientation of rigid solutes in 5CB: MD study</b>	<b>151</b>
8.1	Dipolar couplings . . . . .	152
8.2	Mean Field models . . . . .	153
8.2.1	Maier Saupe model . . . . .	154
8.2.2	Surface tensor model . . . . .	154
8.3	Samples preparation . . . . .	155
8.4	Simulation details . . . . .	156
8.5	Results and discussion . . . . .	156
8.5.1	Order parameters . . . . .	157
8.5.2	Dipolar coupling . . . . .	160
8.6	Comparison with mean field models . . . . .	161
8.6.1	Maier–Saupe . . . . .	162
8.6.2	Surface tensor . . . . .	164
<b>9</b>	<b>Avian prion hexarepeat</b>	<b>169</b>
9.1	Introduction . . . . .	169
9.2	Materials and methods NMR Measurements. . . . .	172
9.2.1	Structure Calculations . . . . .	172
9.2.2	Molecular Dynamics . . . . .	173
9.3	Results and Discussion . . . . .	174
9.3.1	NMR Analysis and Solution Structure Determination . . . . .	174
9.3.2	Molecular Dynamics simulations . . . . .	179
9.3.3	Ramachandran analysis . . . . .	180
9.3.4	Intrapeptide interactions . . . . .	183
9.4	Conclusions . . . . .	185
	<b>List of Tables</b>	<b>193</b>

<b>List of Figures</b>	<b>200</b>
<b>Glossary</b>	<b>203</b>
<b>Bibliography</b>	<b>222</b>
<b>Appendix</b>	<b>225</b>
<b>A Azobenzene force field</b>	<b>225</b>
A.1 Structures and atom labels . . . . .	225
A.2 Atomic charges in atomic units . . . . .	226
A.3 Bonded force field . . . . .	227
A.3.1 Bond and bendings . . . . .	227
A.3.2 Proper torsion . . . . .	227
A.4 Nonbonded mixrule . . . . .	228
<b>B <math>n</math>CB force field</b>	<b>229</b>
B.1 Structures and atom labels . . . . .	229
<b>C Solutes force fields</b>	<b>233</b>
<b>D Radial distribution functions</b>	<b>239</b>
<b>E Mean Field models</b>	<b>241</b>
E.0.1 Maier Saupe Model . . . . .	241
E.1 MF models including the solute shape . . . . .	242
E.1.1 C model . . . . .	243
E.1.2 CZ model . . . . .	243
E.1.3 I model . . . . .	244
E.1.4 CI model . . . . .	244
E.1.5 Surface tensor model . . . . .	245
E.1.6 Terzis-Photinos's model . . . . .	246
<b>F Substitution algorithm</b>	<b>247</b>



# Chapter 1

## Preface

Not very far in time from the birth of the first mechanic Turing's calculator, today is possible to obtain at relatively low cost a powerful machine able to execute some billions of operation per second. At this technological evolution has been followed a cultural revolution in each field of our life. All the principal daily action are related to the use of some informatics system that help all of us to reduce the amount of time spent for boring and recursive actions, that until few years ago must be done by hand. In the world of science the computational devices are not only a useful way to reduce our waste time, but in a lot of cases are become a necessity and can be considered as a research instrument (we can simply think about an computational analysis done for each single spectra from a Fourier transform instrument to convert experimental signal from time to frequency domain). The increase of computational power of modern calculator is really incredible and is very difficult to imagine what will be at the end of this century (or simply at the next year!!). All this power give us the opportunity to investigate the universe around us in a virtual way. Is possible predict the next supernova explosion, the progress of a planetary viral infection, the trend of macro economy, the oxygen photosynthesis of chlorophyll, etc... We can extract information about natural or socio-economic event in the present, in the past and in the future. The starting consideration are inside a more or less long list of initial assumptions and a deterministic series of laws able to describe each specific phenomena. From this point of view a superficial analysis could consider the end of human in favour to a hyper-technological world governed by computer. I do not think so. The extraordinary capabilities of calculator are inside the velocity execution of a code that include all the instruction

about “how?”, “what?”, “where?” and “when?” to do something, but they don’t know “why?”. They are not able to sense the reason of their work, they have not a target. They simply reflect out our ideas, our passions our dreams.

# Chapter 2

## Introduction

The aims of this PhD. thesis are to investigate the molecular behaviour of functional material trying to understand the correlation between macro and microscopic properties proper of these systems. My approach will be to try to explain, using atomistic computer simulations, important aspect of molecular properties related to a better material comprehension.

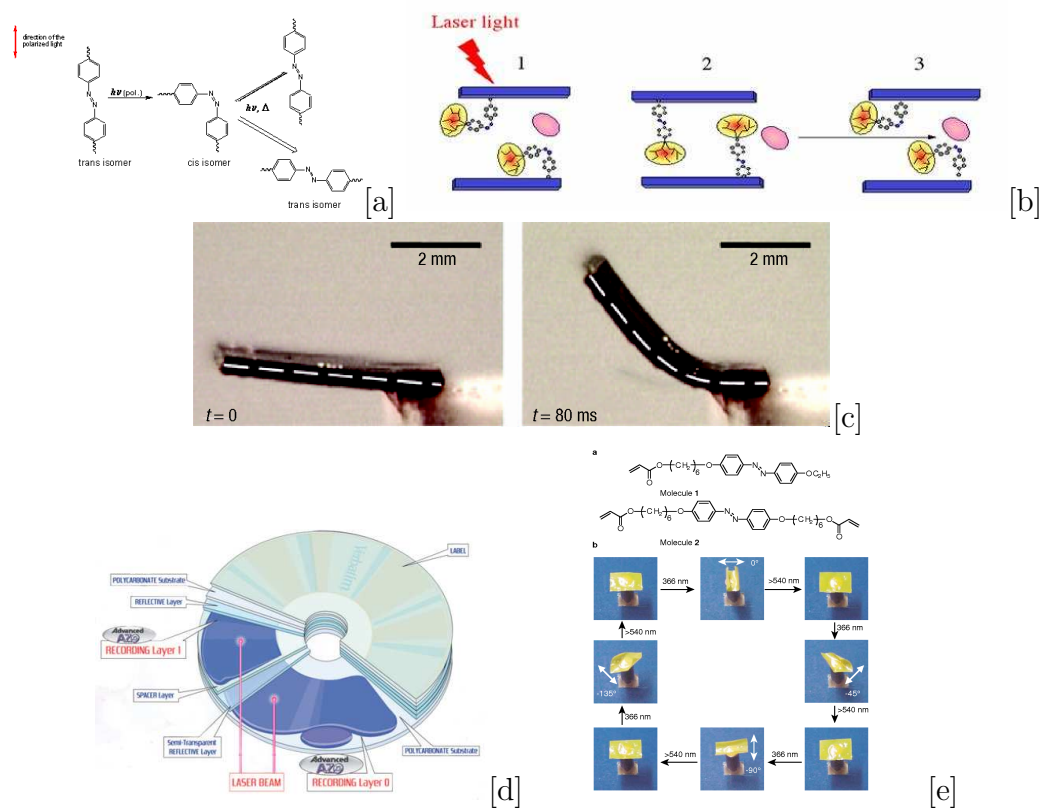
I will show a number of cases where the molecular interactions are strictly related to the material properties. This will permit us to predict or confirm an experimental result and add new information about the inner structure of complex materials. These studies are developed on two different line of works: the first one about the photoisomerization of azobenzene in condensed phase and the second one on the development of a correct parametrization able to reproduce the chemical-physical properties of a series on nematogenic molecules (n-alkyl,4-cyano biphenyls). The final target will be to focus our study on azobenzene like molecules inside ordered liquid crystals and liquid crystal polymers joining the knowledge obtained from previous separated studies.

### 2.1 Azobenzene

At the basis of this study there is the photoisomerization of the azo-derivatives molecules, that due a strong geometrical, dipolar, light-induced properties modifications at molecular level, able to modify macroscopic material properties so to create a new generation of functional materials. A very large number of applications are created using photoisomerization properties of

azo-group (Fig. 3.1) inside a liquid crystal matrix, bonded on a surface or in a polymer structure. The effect of the geometrical change of azo-groups inside a complex matrix can be observed macroscopically as a length shortening of an elastomer piece, a change of optical properties (order–disorder transition) of a liquid crystal, or a darkening on a polymer sheet. The application fields are large and interesting for new responsive material engineering developments in macroscopic and nanometric scale for the production of devices able to reply to external stimuli (light irradiation) and return to the starting state after a second impulse (light at different wavelength) or more by thermal influence.

Figure 2.1: Application of azobenzene-like molecules. [a] **Optical alignment**. [b] **Molecular pumps**. [c] **Artificial muscles**. [d] **Storage disk**. [e] Deformable sheet [Yu et al. 2003].



The basic mechanism for the photoisomerization using a continuous light source can be summarised as follow:

1. Selective excitation of target isomer with visible or near ultra violet radiation
2. Relaxation process
3. Increasing of other isomer concentration by selective excitation of remaining starting isomer

The advantage of this mechanism is that is generally reversible so is possible convert from a configuration to another simply by changing the irradiation wavelenght.

In spite of the large number of experimentals works and application on this class of compound, the microscopic molecular mechanism of isomerization process, is still not completely clarified at the moment, especially in condensed phases. There are many theoretical papers that try to address the nature of atomic reorganisations around the azoic bond during isomerization, and there is strong controversy about the pathway followed by molecules in the lower excited state. The principal mechanisms involved are three:

- Torsional pathway along  $N = N$  torsional degree of freedom
- Inversion pathway around one pyramidal  $N$  atom
- Coupled double inversion involved the two  $N$  atoms together

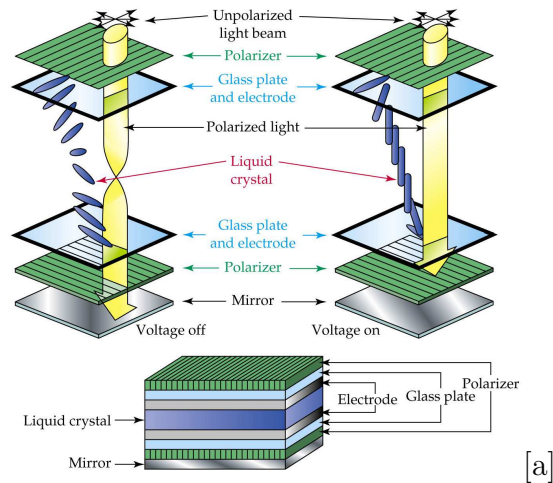
In this content we have developed a simulation method (see Chap. 4) to investigate the open theoretical question about the isomerization process, able to reproduce the isomerization of an azoic bond inside a condensed phase and adapt to study properties into a complex real matrix like liquid crystals or polymers in order to give a theoretical contribution to understand and improve the properties of these materials.

After the setting of isomerization process model, we have tried to study the isomerization effect the liquid crystal matrix (Chap. 7) compared with experimental measures (Chap. 7) and on the azo-containing side-chain polymer (Chap. ??) under unpolarized and linearly polarised excitation light parallel or perpendicular to the molecular director or main chain orientations.

## 2.2 Liquid crystals: $n$ CB

In the last years, the number of application of liquid crystals are became incredibly important in different technological fields. The reason reside inside the extraordinary capabilities of these materials to auto-assemble in an ordered system by an external stimulus. Generally a temperature reduction, an electric or magnetic field, light irradiation or pressure. The studies on this system are also focused on the possibility to build auto-assembling molecular system, part of nanometric machines able to connect themselves without human interactions.

Figure 2.2: **Scheme of a twisted Nematic display**



The developments of computational resources give us the possibility to perform a realistic simulation, taking into account the chemical structure and relative molecular properties to reproduce the macroscopic effect that can be simply observed in a common laboratory such as the nematic–isotropic transition temperature of a liquid crystal. Setting up the simulation parameters to reproduce the chemical–physics parameters we have the intrinsic possibility to study the exact motions of a molecule inside the matrix and so the complete knowledge of the molecular driving forces that govern the properties of a material and we can suggest the improves to obtain better desired characteristics.

# Chapter 3

## Molecular Dynamics

### 3.1 Computer simulations

The computer simulations begin to become a very important field for science applications. In principle each theoretical model could be used in a simulation. The simulation can extend our comprehension about process that are very difficult or impossible to observe by an experimental technique (imagine the future evolution of a geological process like mountain formation or the exact position of a molecule inside a solution for a specific short time–window). The simulation models strongly depends on the phenomena to be studied especially on the time and dimension–scale and on the finding properties to be studied or predicted. Here we want to show some examples about the available applications now for the scientists in different knowledge fields.

#### 3.1.1 Physics

Nuclear physics research provides new insights to advance our knowledge on the nature of matter and energy. At its heart, nuclear physics attempts to understand the composition, structure, and properties of atomic nuclei. High energy physics strives to understand the universe at a more basic level by investigating the elementary particles that are the fundamental constituents of matter and the forces between them. The study of physics has advanced with the technology to study ever–higher energies and very rare phenomena that probe the smallest dimensions we can see and tell us about the very early history of our universe. While physics has revolutionised our understanding of how the universe works, elements of physics technology have helped transform other fields of science, medicine, and even everyday life. Physics

research and its impacts will be remembered as one of the highlights of the history of the late 20<sup>th</sup> century.

### 3.1.2 Climate

Fossil fuel combustion for energy production is altering the chemical makeup of the Earth's atmosphere. The consequences for climate change and the potential for significant, even catastrophic, nonlinear feedbacks through the Earth system are topics of significant international debate. These global changes, principally driven by human activities at regional scales, require us to acquire an unprecedented understanding of potential regional and global changes in our environment, economy, and society. Climate scientists today face challenging uncertainties about how climate systems will respond to future environmental changes. There are many compelling questions that must be answered if we are to predict future climate and gain the understanding needed to assess and ameliorate the potential impacts of energy use. The key science question that drives researcher is complex and complicated. Simply put, it is: *"How will the Earth's climate respond to physical, chemical, and biological changes produced by global alterations of the atmosphere, ocean, and land?"*. The science charge to understand the impacts of energy production and use on the environment continues to lead the evolution of the climate modelling and simulation research. The research on this field will advance the development of future climate models based on theoretical foundations and improved computational methods that dramatically increase both the accuracy and throughput of computer model-based predictions of future climate system response to the increased atmospheric concentrations of greenhouse gases.

### 3.1.3 Ground water

One of the most challenging problems in environmental remediation involves hazardous materials which have leached into the subsurface and are at risk of being more widely dispersed by the flow of groundwater through contaminated areas. The result can be that contaminants located in a remote area may be carried by groundwater to more sensitive water resource areas such as rivers, lakes or wells. Research efforts to contain and resume contaminated sites challenges the state of the science in many areas. Scientifically rigorous models of subsurface reactive transport that accurately simulate the move-



ment of contaminants across multiple length scales remain elusive. This will benefit environmental cleanup efforts as well as improve the monitoring of contaminants in groundwater around existing and future radionuclide waste disposal and storage sites. These efforts will also assist the research on using deep geological formations to store carbon dioxide taken from the atmosphere.

### 3.1.4 Fusion energy

Fusion has the potential to provide a long-term, environmentally-acceptable source of energy for the future. While research during the past 20 years indicates that it will likely be possible to design and build a fusion power plant, the major challenge of making fusion energy economical remains. Improved simulation and modelling of fusion systems using terascale computers is essential to achieving the predictive scientific understanding needed to make fusion practical. Answers to several long-standing questions could give the planners of reactors a useful edge in the design of future fusion power plants. Magnetised fusion plasmas contain electrons and the fusion fuel-ions of deuterium and tritium. Plasma contained within a fusion device behaves very differently depending on the shape of the magnetic field and distribution of the electric current. Because no material can withstand the 100 million degree temperature of the plasma, it is the magnetic field that actually contains the plasma. Being able to control the plasma is critical to the success of fusion as a source of energy. Integrated simulation of magnetic fusion systems involves the simultaneous modelling of the core plasma, the edge plasma, and the plasma-wall interactions. In each region of the plasma, there is anomalous transport driven by turbulence, there are abrupt rearrangements of the plasma caused by large-scale instabilities, and there are interactions with neutral atoms and electromagnetic waves. Many of these processes must be computed on short time and space scales, while the results of integrated modelling are needed for the whole device on long time scales. The mix of complexity and widely differing scales in integrated modelling results in a unique computational challenge. At present the understanding of the small-scale ("micro") instabilities that degrade plasma confinement by causing the turbulent transport of energy and particles and the large-scale ("macro") instabilities that can produce rapid topological changes in the confining magnetic field are too incomplete to begin developing integrated models. Similarly the understanding of plasma-material interactions

and the propagation of electromagnetic waves are also too primitive to begin to develop integrated models. Thus, the first phase in fusion energy sciences focuses on the development of improved physics models of each of these elements.

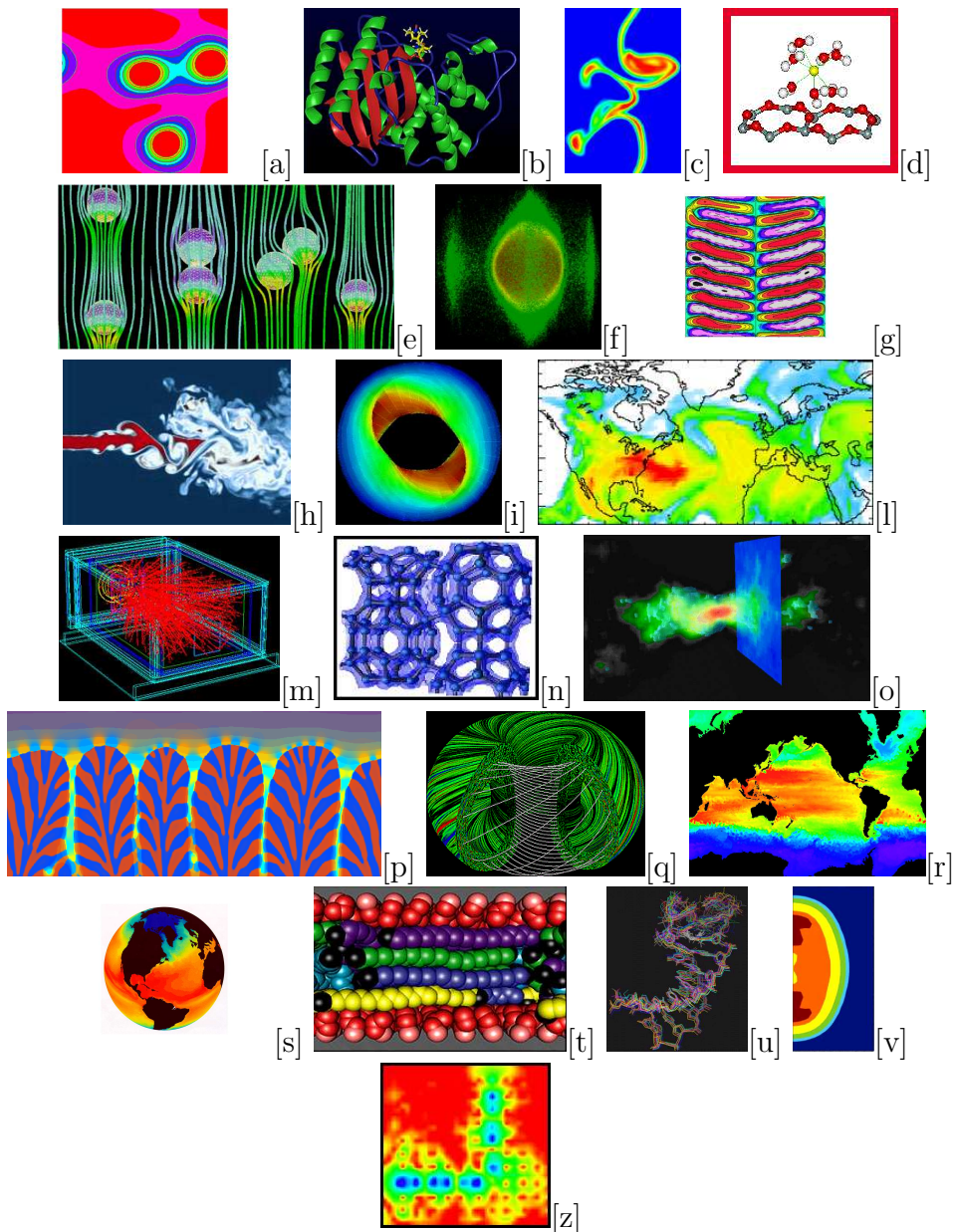
### 3.1.5 Life science

This field is focused on developing new methods for modelling complex biological systems, including molecular complexes, metabolic and signalling pathways, individual cells and, ultimately, interacting organisms and ecosystems. Such systems act on time scales ranging from microseconds to thousands of years and the systems must couple to huge databases created by an ever-increasing number of high-throughput experiments.

### 3.1.6 Materials and Chemistry

Great progress has been made in the past half century in bringing molecular theory and modelling from a purely interpretive science to an accurate, predictive tool for describing molecular energetics and chemical reactions. Predictions that rival experimental accuracy are now possible for molecules comprised of two to six atoms from the first two rows of the periodic table. However, the analysis and optimisation of many processes of importance, such as combustion, phase transitions, etc... require expansion of current modelling capabilities to more complex molecules and to molecules interacting with extended structures such as clusters or surfaces. These include quantum simulations of materials and nanostructures; stress corrosion cracking; multi-scale simulations of strongly correlated materials. New efforts will be coordinated with existing, off-cycle, efforts to improve understanding and accurate modelling of material properties, reactions and interactions, on length scales that are 10 orders of magnitude or more.

Figure 3.1: Example of computer simulation applications. Data from <http://www.science.doe.gov/scidac/>. [a] Vortices in a superfluid. [b] Protein dynamics. [c] Turbulent methane flame. [d] Clay–mineral geochemistry. [e] Two spheres mixing in a stream. [f] HEP particle beam halo. [g] Transport barrier dynamics. [h] Combustion turbulence modelling. [i] Fusion magnetic field. [l] Perturbation in clear–sky and cloud albedo. [m] Au–Au collision. [n] Crystal structure for  $C_{36}$  solid. [o] Lattice quantum chromodynamics. [p] Binary alloy solidification. [q] Perturbed plasma density. [r] DOE Parallel Climate Model. [s] Sea surface temperature. [t] Molecular simulation of complex fluids. [u] Structural biology. [v] Nuclear theory. [z] Waveguide optics.



## 3.2 Atomistic simulations

### 3.2.1 Introduction

The physical and chemical characterisation of materials is made possible by a huge array of sophisticated techniques, with often atomic-scale resolution and extreme sensitivity. The modelling and simulation of materials are based on the development of theories and computational methods, and range from the quantum physics of atomic-scale phenomena to the continuum descriptions of macroscopic behaviour. The realm of materials is vast, ranging from the inorganic world of metallic, semiconducting and insulating materials to organic polymers as well as biologically relevant and bio-mimetic materials and structures. Their physical, chemical and biological properties vary enormously, from ultra-hard solids to soft tissues and DNA strands, from ceramic superconductors to organic semiconductors, from ferromagnetic liquids to amorphous insulators. The structural variety is unbounded, and can be accessed with sophisticated manipulation and processing techniques. The infinite number of possibilities in materials research underscores the importance of predictive theory and modelling. The community of materials theorists has traditionally been divided into sub-communities largely defined by the length and timescales of interest to them. In the macroscopic spatial regime from millimetres to metres, modelling has typically developed around continuum equations solved by finite-element (FE) and finite-difference techniques. In the mesoscopic regime from micrometres to millimetres, phenomenological approaches have developed around stochastic methods for the material's micro-structure such as grain boundaries and dislocations. In the microscopic, Ångström-scale regime physicists and quantum chemists have based their work on the Schrödinger equation and other expressions of quantum mechanics of interacting electrons and atomic nuclei. The rapid increase in computational capabilities, in terms of both raw computing power and new algorithms, has enabled spectacular developments in each of these length scales. The challenge now faced by theorists is to bridge the different length (and time) scales to a more general framework, which has been coined as multi-scale modelling. One should be able to move, as seamlessly as possible, from one scale to another so that the calculated parameters, properties and other numerical information calculated can be efficiently transferred across scales. Although all the information is in principle available at the finest (microscopic) level on description, it would be mindless and totally

impractical to use it for macroscopic phenomena. At each length scale there are emergent phenomena which are best described by new, coarse-grained equations, which eliminate the unnecessary detail and emphasise the emergent properties. For example, the energy barriers, vibrational frequencies and entropies for atomic motion can be calculated at the microscopic level and then transferred to a kinetic simulation as discrete jump probabilities or as coefficients to reaction-diffusion-type equations. The properties of materials and structures made thereof should ultimately be explainable in terms of their constituent atoms and their mutual interactions, and their motion at finite temperatures and under external forces. Below, the basic steps from the microscopic towards the macroscopic are briefly described.

### 3.2.2 Hierarchical modelling of materials phenomena

#### Atomistic calculations: density-functional theory

Density-functional theory (DFT) [Nieminen 1999, Dreizler and Gross 1998] is the undisputed workhorse for quantum mechanical atomistic calculations. DFT transforms the complex many-body problem of interacting electrons and nuclei into a coupled set of one-particle (KohnSham) equations, which are computationally much more manageable. The theory allows parameter-free calculations of all ground-state physical observables, such as charge and spin densities, total energies and many related quantities, such as bonding distances, elastic moduli, vibrational frequencies, defect and surface energies, migration barriers, reaction energies, magnetic moments etc. There are numerous popular implementations of DFT to large-scale calculations of materials properties, for example those using plane-wave basis functions and pseudo-potentials for valence electrons or those using linearised methods for all-electron calculations. System sizes of up to several hundred atoms are feasible. While the simple local-(spin-) density approximation (L(S)DA) is robust and provides often surprising quantitative accuracy, there are also workable methods beyond L(S)DA which recover aspects of the non local nature of electronic exchange and correlation interactions. These methods can lead to a numerical accuracy similar that of full quantum calculations (configuration interactions, coupled cluster, quantum Monte Carlo (MC)) for small systems. The latter methods are currently computationally too heavy for large system sizes. However, one should note the important recent development towards a linear-scaling ('order-N') formulation of the quan-

tum MC method [Williamson et al. 2001]. The static formulation of DFT enables also ‘first-principles’ molecular dynamics simulations, where the interatomic forces are calculated for the adiabatic motion of nuclei from the electronic degrees of freedom, utilising the HellmannFeynman theorem. The density-functional approach can be generalised to the time-dependent case [Onida et al. 2002] as well. In the linear-response regime, such properties as photo-absorption spectra or frequency dependent electromagnetic susceptibilities can be attacked. While the construction of the exchange-correlation functional for time-dependent calculations is still a challenge, the simple adiabatic local-density approximation (TDLDA), which implies an instantaneous response dependent only on the local electron density, has turned out to be surprisingly accurate for many purposes. The time-dependent formulation also allows an approach to address the question of excitation energies, for which the ground-state KohnSham eigenvalues often give a poor estimate. For finite systems such an approach has been shown to be quite accurate [MacKerrel et al. 1995], comparable to more laborious quasi particle ‘self-energy’ methods. In the case of strong external perturbations, such as laser dissociation of molecules, DFT provides a working scheme to simulate the full nonlinear response of the combined system of electrons and nuclei. The numerical solution of the KohnSham equations using real-space (RS) methods [Beck 2000, Chelikowsky et al. 2000, Heiskanen et al. 2001] instead of basis sets is currently an active research topic. Real-space grids can offer several benefits. Firstly, RS methods can in principle be used in both pseudo potential and all-electron calculations. Secondly, and more importantly in the present context, systems containing different length scales (e.g. nanostructures, surfaces etc) can be treated economically as one need not waste many grid points in empty regions. Also different types of boundary condition (free or periodic) are easily implemented. Finally, RS methods can be efficiently adapted to parallel computing through domain decomposition. Multi-grid techniques enable substantial speedups in the convergence of RS methods, which are now becoming competitive with more conventional basis-set approaches such as Fourier series (plane-wave methods).

### **Moving atoms: molecular dynamics simulations**

Within the BornOppenheimer (BO) approximation, the total ground-state energy associated with the electronic degrees of freedom defines the potential energy hyper-surface for the atomic motion, which is classical and controlled

by the nuclear masses. Exceptions are the lightest elements, hydrogen and helium, for which quantum mechanical tunnelling plays a role. The interatomic forces can be obtained as derivatives of the potential energy hyper-surface with respect to nuclear coordinates. The low-lying excitations of the nuclear subsystem are quantised phonons, which can be accurately addressed by density-functional perturbation theory [Baroni et al. 2001]. At high temperatures, the atomic motion becomes entirely classical. The atomic excursions from their equilibrium positions increase with increasing temperature. The motion becomes anharmonic and eventually leads to diffusive motion characterised by hops over barriers separating equivalent positions in the potential energy landscape. The atomic motion can be modelled by solving Newton's equations of motion. This approach goes under the generic name of molecular dynamics simulations [Frenkel and Smit 1996, Rapaport 1995], and has become a popular and powerful way to investigate the complicated, collective processes associated with atomic motion. Various thermodynamical ensembles (constant total energy–microcanonical, constant temperature–canonical, constant volume, constant pressure) can be simulated using, if necessary, auxiliary variables for atomic velocity scaling and unit-cell dimensions. As molecular dynamics simulation works in ‘real time’, i.e. the computational timestep (typically less than 10–15 s) is dictated by the physical constants, following a particular physical event may be painfully slow and require a huge number of computational timesteps. This is the problem of rare events, such as a thermally activated jump over a migration barrier at low temperatures, which makes the brute force application of molecular dynamics simulation unpractical for such cases. Ingenious schemes have been proposed [Sorensen and Voter 2000] to overcome the rare event bottleneck while preserving the deterministic nature of molecular dynamics simulation. The most direct, parameter-free approach to molecular dynamics simulation is to calculate the forces from first principles, i.e. by evaluating the electronic total energy and interatomic forces at each timestep. This can be done either ‘on the fly’ through the CarParrinello algorithm [Car and Parrinello 1985] (which updates the electronic and ionic degrees of freedom in unison in the vicinity of the BO surface) or through direct minimisation of the electronic total energy on the BO surface (adiabatic molecular dynamics). First-principles molecular dynamics simulations are computationally demanding and still limited to rather modest system sizes and short time sequences. The accurate evaluation of the forces also puts stringent requirements on the computational techniques used. Most of the practical implementations of first-principles

molecular dynamics methods use Fourier techniques (plane-wave basis sets) for the electronic degrees of freedom in conjunction with pseudo potentials to fold out the inert core-electron states. Plane-wave methods have the advantage that the spatial resolution is uniform, i.e. independent of the nuclear positions, which enables accurate force evaluations. Even with the fastest (parallel) algorithms and computers, the calculations are typically limited to a few hundred electrons and atoms in size and to a few nanoseconds in length. The molecular dynamics simulations can be speeded up at the expense of ‘first-principles’ accuracy of the force evaluations. Several schemes utilising minimal basis sets in the form of local orbitals in the spirit of the tight-binding approximation have been introduced [Bowler and Gillan 1998, Frauenheim et al. 2000, Ordejon 1998], with varying degrees of accuracy and portability. The force evaluations in such schemes are usually faster than the first-principles methods by at least one order of magnitude. Schemes utilising parametrised classical interactions in analytic form are naturally fastest. Such schemes range from simple pairwise interatomic potentials of insulating solids to angle-dependent models for covalent solids and to sophisticated many-atom force fields designed for organic molecules [Brenner 2000].

### **Moving atoms: kinetic Monte Carlo methods**

Another possible approach to simulating atomic motion is provided by stochastic simulation by the MC method [Landau and Binder 2000]. The basic idea is very simple. The potential energy hypersurface is first evaluated for all relevant atomic configurations. These include not only (meta)stable configurations but also those associated with saddle points. A typical saddle point configuration is one where an atom is at the transition state, at the barrier separating two nearby valleys in the hypersurface. A complete mapping of the hypersurface is of course impossible. Thus physical intuition is needed to divide the atoms into those actively moving during various processes of interest and those that are mere ‘spectators’ during the events. Identification of the barriers is a nontrivial task, and algorithms such as the ‘nudged elastic band’ (NEB) [Jonsson et al. 1998], the ‘locally activated Monte Carlo’ (LAMC) [Kaukonen et al. 1998] or the ‘tossing ropes’ [Geissler et al. 2001] are needed. The relative probabilities of different atomic movements are proportional to Boltzmann factors involving the energy differences between the transition state and the initial state. In the kinetic Monte Carlo (KMC) simulation, one replaces the short-time dynamics of the system by discrete



hops in a network. This is done by mapping the potential energy surface to a grid, where the grid points are associated with minima in the potential energy landscape. The possible initial and final states for atomic configurations are classified, typically in terms of ‘atomic neighbourhoods’, which largely determine the energy of a given configuration. The number, position and chemical identity of neighbouring atoms are used to define the class of the configuration. Given the classified configurations, one defines the transition probability  $W_{fi}$  from an initial state  $i$  to a final state  $f$  for all possible transitions between the two. These can include not only single-atom events but also concerted events, i.e. simultaneous movements of several atoms. For an event class labelled by  $k$ , the transition rate is:

$$\Gamma^k = W(f, i) = \Gamma_0^{fi} \exp\left(-\frac{E_{TS}^{fi} - E_i}{k_B T}\right) \quad (3.2.1)$$

where  $\Gamma_0^{fi}$  is a prefactor,  $E_{TS}^{fi}$  the total energy in the transition state for process  $k$  and  $E_i$  the energy of the initial state. The prefactor depends on entropic factors associated with the possible vibrational modes, and can be explicitly evaluated within the classical and harmonic approximation. Thermal equilibrium can be ensured by invoking the detailed balance condition. The KMC simulation is efficiently executed with the BortzKalosLebowitz [Bortz et al. 1975] algorithm. One first evaluates the total rate  $R$  by summing  $\Gamma^k$  over all possible processes. Secondly, three random numbers  $\rho_1$ ,  $\rho_2$  and  $\rho_3$  are drawn from the interval  $[0, 1]$ . One of them, say  $\rho_1$ , is used to determine the class  $k$  for the event chosen:

$$\sum_{j=0}^k \Gamma^j \leq \rho_1 R \leq \sum_{j=0}^{k-1} \Gamma^j \quad (3.2.2)$$

The second random number  $\rho_2$  is used to choose a particular event from the class  $k$ , and the third random number  $\rho_3$  is used to advance the ‘clock’ through  $t = t - \ln(\rho_3)/R$ .

The procedure can obviously be generalised for non-equilibrium dynamics controlled by arbitrary rate constants. KMC simulations are closely related to lattice-gas and cellular automaton models [Chopard and Droz 1998], where the updating of the local variables (configurations) follows rules written in terms of the neighbourhoods of each atom. The updating can take place either sequentially or in parallel. In the continuous cellular automaton

scheme, the condition of discrete occupation numbers (e.g. that an atom either occupies a lattice site (occupation = 1) or the site is empty (occupation = 0)) is released. A continuous variable ranging from zero to unity is used to characterise the ‘mass’ or ‘occupation state’ of a lattice site. Starting from a site initially occupied by an atom, this continuous variable is reduced by an amount corresponding to its net removal probability. When the continuous variable reaches the value zero, the atom is definitely removed. This mode of KMC methods is particularly useful in simulations of surface etching,

### **From molecular dynamics to continuum equations**

The coupling of the atomic-scale description of molecular dynamics to mesoscale and eventually macroscale models is absolutely essential to successful realisation of true multi-scale modelling. Let us consider the case of mechanical properties of solids as an example, including fracture and crack propagation under external load. Even though molecular dynamics simulations can currently run with hundreds of millions of atoms (with classical potentials), these calculations cannot properly represent environment of a dynamical system in the meso and macroscales. The important issue is the flow of mechanical energy into and out of the system, rather than the statistical physics of the atomistic assembly (which can be modelled by a relatively small sample). Much of the important physics is in long-range interactions, such as the elastic field. Thus it is natural to envisage embedding the molecular-dynamics region in a continuum mechanical description. Far away from the central region (e.g. the tip of the propagating crack) the atoms are displaced only slightly from their equilibrium positions, and (linear) elasticity theory is expected to work well. The FE method [Dhatt and Touzot 1984] is the method of choice for this region. The FE algorithms for continuum equations are of course much more efficient computationally than molecular dynamics: only a small number of degrees of freedom are necessary. In crack propagation, much of the action is focused near the crack faces and the emitted dislocations. Far away from the crack, little is happening as atoms mainly vibrate around their equilibrium positions: a mean-field-type description is totally adequate. However, one cannot dispense with the distant region entirely. Molecular dynamics simulations show that long-wavelength pressure waves emanate from the crack tip. These waves must be properly propagated into the surrounding medium, which can only be accomplished by an appropriate embedding continuum. The FE equations of motion can be derived straight-

forwardly from the Hamiltonian for linear elasticity theory, with the physical parameters contained in the elastic constant tensor  $\mathbf{C}$  and the mass matrix  $\mathbf{M}$ . The crucial issue then is the ‘handshaking’ between the molecular dynamics region and the embedding FE region. This remains an active area of study, and several schemes have been proposed [Rudd and Broughton 2000, B.I. Lundqvist and Wahnström 2002].

### From kinetic Monte Carlo methods to rate equations

KMC schemes discussed above carry the spatial information defined in terms of discrete configurations of atoms through which the system evolves as a generalised random-walk process. This information can easily become overwhelming and difficult to handle and the number of possible configurations can become unmanageable. Instead of solving the full stochastic differential equation (the ‘master equation’), as the KMC methods do, one can define density fields for a set of relevant configurations, for example spatial densities  $n_k(r)$  for atom clusters of size  $k$ . These density fields can then be linked with macroscopic mass currents, as for example in the case of simple chemical diffusion driven by density gradients. The diffusion current is

$$J_1 = -(Dn_1(\vec{r})/k_B T)\Delta\mu \quad (3.2.3)$$

where  $D$  is the phenomenological diffusion constant and  $\mu$  the chemical potential. In the dilute limit

$$\mu = k_B T \ln(n_1(\vec{r})) \quad (3.2.4)$$

which leads to the Schmoluchowski equation and Fick’s law. Combined with the constraint of mass conservation, the familiar diffusion equation is obtained, and a link between the microscopic random walk and the macroscopic diffusion model is established. Similarly, simple rate-reaction equations can be derived for spatially averaged quantities such as average concentrations of clusters with a given size. A famous example of such rate equations is Waite’s formulation [Waite 1957] for diffusion-limited reactions.

## 3.3 Molecular Dynamics technique

<sup>1</sup>In atomistic models the coordinates of all atomic are treated explicitly and interactions between distant atoms are represented by a pairwise additive

---

<sup>1</sup>Materials from Ref. [Marchi and Procacci 1997]

dispersive–repulsive potential and a Coulomb contribution due to the atomic charges. Furthermore, nearby atoms interact through special two body, three body and four body functions representing the valence bonds, bending and torsional interaction energies surfaces. The validity of such an approach as well as the reliability of the various potential models proposed in the literature [Wiener et al. 1986, Cornell et al. 1995, Brooks et al. 1983, van Gunsteren and Berendsen 1987] is not the object of the present book. For reading on this topic, we refer to the extensive and ever growing literature [Cornell et al. 1995, MacKerrel et al. 1995, Pavelites et al. 1997, and D. Bashford et al. 1998]. Here, we want only to stress the general concept that atomistic simulations usually have more predictive power than simplified models, but are also very expensive with respect to the latter from a computational standpoint. This predictive power stems from the fact that, in principle, simulations at the atomistic level do not introduce any uncontrolled approximation besides the obvious assumptions inherent in the definition of the potential model and do not assume any *a priori* knowledge of the system, except of course its chemical composition and topology. Therefore, the failure in predicting specific properties of the system for an atomistic simulation is due only to the inadequacy of the adopted interaction potential. We may define this statement as *the brute force postulate*. In practise, however, in order to reduce the computational burden, severe and essentially uncontrolled approximations such as neglect of long range interactions, suppression of degrees of freedom, dubious thermostatting or constant pressure schemes are often undertaken. These approximations, however, lessen the predictive power of the atomistic approach and incorrect results may follow due to the inadequacy in the potential model, baseless approximations or combinations of the two. Also, due to their cost, the predictive capability of atomistic level simulations might often only be on paper, since in practise only a very limited phase space region can be accessed in an affordable way, thereby providing only biased and unreliable statistics for determining the macroscopic and microscopic behaviour of the system. It is therefore of paramount importance in atomistic simulations to use computational techniques that do not introduced uncontrolled approximations and at the same time are efficient.

As stated above, simulations of complex systems at the atomistic level, unlike simplified models, have the advantage of representing with realistic detail the full flexibility of the system and the potential energy surface according to which the atoms move. Both these physically significant ingredients of the atomistic approach unfortunately pose severe computational

problems: on one hand the inclusion of full flexibility necessarily implies the selection of small step size thereby reducing in a MD simulation the sampling power of the phase space. On the other hand, especially the evaluation of inter-particle long range interactions is an extremely expensive task using conventional methods, its computational cost scaling typically like  $N^2$  (with  $N$  being the number of particles) quickly exceeding any reasonable limit. In this book we shall devote our attention to the methods, within the framework of classical MD simulations, that partially overcome the difficulties related to the presence of flexibility and long range interactions when simulating complex systems at the atomistic level.

Complex systems experience different kind of motions with different time scales: Intramolecular vibrations have a period not exceeding few hundreds of femtoseconds while reorientational motions and conformational changes have much longer time scales ranging from few picoseconds to hundreds of nanoseconds. In the intra-molecular dynamics one may also distinguish between fast stretching involving hydrogens with period smaller than 10 fs, stretching between heavy atoms and bending involving hydrogens with double period or more, slow bending and torsional movements and so on. In a similar manner in the diffusional regime many different contributions can be identified.

In a standard integration of Newtonian equations, all these motions, irrespectively of their time scales, are advanced using the same time step whose size is inversely proportional to the frequency of the fastest degree of freedom present in the system, therefore on the order of the femtosecond or even less. This constraint on the step size severely limits the accessible simulation time. One common way to alleviate the problem of the small step size is to freeze some supposedly irrelevant and fast degrees of freedom in the system. This procedure relies on the the so-called SHAKE algorithm [Ryckaert et al. 1977, Ciccotti and Ryckaert 1986a, Allen and Tildesley 1989] that implements holonomic constraints while advancing the Cartesian coordinates. Typically, bonds and/or bending are kept rigid thus removing most of the high frequency density of states and allowing a moderate increase of the step size. The SHAKE procedure changes the input potential and therefore the output density of the states. Freezing degrees of freedom, therefore, requires in principle an a priori knowledge of the dynamical behaviour of the system. SHAKE is in fact fully justified when the suppressed degrees of freedom do not mix with the "relevant" degrees of freedom. This might be "almost" true for fast stretching involving hydrogens which approximately defines an

independent subspace of internal coordinates in almost all complex molecular systems [Procacci et al. 1996] but may turn to be wrong in certain cases even for fast stretching between heavy atoms. In any case the degree of mixing of the various degrees of freedom of a complex system is not known a priori and should be on the contrary considered one of the targets of atomistic simulations. The SHAKE algorithm allows only a moderate increase of the step size while introducing, if used without caution, essentially uncontrolled approximations. In other words indiscriminate use of constraints violates the brute-force postulate. A more fruitful approach to the multiple time scale problem is to devise a more efficient "multiple time step" integration of the equation of motion. Multiple time step integration in MD is a relatively old idea [Street et al. 1978, Teleman and Joensonn 1986, Tuckerman et al. 1991, Tuckerman and Berne 1991, Tuckerman et al. 1990, Grubmuller et al. 1991] but only in recent times, due to the work of Tuckerman, Martyna and Berne and coworkers [Tuckerman et al. 1992; 1993, Humphreys et al. 1994, Procacci and Berne 1994a, Procacci and Marchi 1996, Martyna et al. 1996] is finding widespread application. These authors introduced a very effective formalism for devising multilevel integrators based on the symmetric factorisation of the Liouvillean classical time propagator. The multiple time step approach allows integration of all degrees of freedom at an affordable computational cost. In the simulation of complex systems, for a well tuned multilevel integrator, the speed up can be sensibly larger than that obtained imposing bond constraints. Besides its efficiency, the multiple time steps approach has the advantage of not introducing any a priori assumption that may modify part of the density of the state of the system.

The Liouvillean approach to multiple time steps integrator lends itself to the straightforward, albeit tedious, application to extended Lagrangian systems for the simulation in the canonical and isobaric ensembles: once the equations of motions are known, the Liouvillean and hence the scheme, is *de facto* available. Many efficient multilevel schemes for constant pressure or constant temperature simulation are available in the literature [Procacci and Berne 1994b, Martyna et al. 1996, Marchi and Procacci 1998].

As already outlined, long range interactions are the other stumbling block in the atomistic MD simulation of complex systems. The problem is particularly acute in biopolymers where the presence of distributed net charges makes the local potential oscillate wildly while summing, e.g. onto spherical non neutral shells. The conditionally convergent nature of the electrostatic energy series for a periodic system such as the MD box in periodic boundary

conditions (PBC) makes any straightforward truncation method essentially unreliable [Saito 1994, Allen and Tildesley 1989, Lee et al. 1995].

The reaction field [Barker and Watts 1973, Barker 1980] is in principle a physically appealing method that correctly accounts for long range effects and requires only limited computational effort. The technique assumes explicit electrostatic interactions within a cutoff sphere surrounded by a dielectric medium which exerts back in the sphere a "polarisation" or reaction field. The dielectric medium has a dielectric constant that matches that of the inner sphere. The technique has been proved to give results identical to those obtained with the exact Ewald method in Monte Carlo simulation of dipolar spherocylinders where the dielectric constant that enters in the reaction field is updated periodically according to the value found in the sphere. The reaction field method does however suffer of two major disadvantages that strongly limits its use in MD simulations of complex systems at the atomistic level: during time integration the system may experience instabilities related to the circulating dielectric constant of the reaction field and to the jumps into the dielectric of molecules in the sphere with explicit interactions. The other problem, maybe more serious, is that again the method requires an a priori knowledge of the system, that is the dielectric constant. In pure phases this might not be a great problem but in inhomogeneous systems such as solvated protein, the knowledge of the dielectric constant might be not easily available. Even with the circulating technique, an initial unreliable guess of the unknown dielectric constant, can strongly affect the dielectric behaviour of the system and in turn its dynamical and structural state.

The electrostatic series can be computed in principle exactly using the Ewald re-summation technique [Ewald 1921, deLeeuw et al. 1980]. The Ewald method rewrites the electrostatic sum for the periodic system in terms of two absolutely convergent series, one in the direct lattice and the other in reciprocal lattice. This method, in its standard implementation, is extremely CPU demanding and scales like  $N^2$  with  $N$  being the number of charges with the unfortunate consequence that even moderately large size simulations of inhomogeneous biological systems are not within its reach. The rigorous Ewald method, which does not suffers of none of the inconveniences experienced by the reaction field approach, has however regained resurgent interest in the last decade after publication by Darden, York and Pedersen [Darden et al. 1993] of the Particle Mesh technique and later on by Essmann, Darden et al. [Essmann et al. 1995] of the variant termed Smooth Particle Mesh Ewald (SPME). SPME is based on older idea idea of Hockney [Hockney 1989] and

is essentially an interpolation technique with a charge smearing onto a regular grid and evaluation via fast Fourier Transform (FFT) of the interpolated reciprocal lattice energy sums. The performances of this technique, both in accuracy and efficiency, are astonishing. Most important, the computational cost scales like  $N \log N$ , that is essentially linearly for any practical application. Other algorithm like the Fast Multipole Method (FMM) [Petersen et al. 1994, Greengard and Rokhlin 1987, Shimada et al. 1994, Zhou and Berne 1996] scales exactly like  $N$ , even better than SPME. However FMM has a very large prefactor and the break even point with SPME is on the order of several hundred thousand of particles, that is, as up to now, beyond any reasonable practical limit.

The combination of the multiple time step algorithm and of the SPME [Procacci et al. 1996] makes the simulation of large size complex molecular systems such as biopolymers, polar mesogenic molecules, organic molecules in solution etc., extremely efficient and therefore affordable even for long time spans. Further, this technique do not involve any uncontrolled approximation and is perfectly consistent with standard PBC. Of course SPME is itself an approximation of the true electrostatic energy.

### 3.3.1 Symplectic and Reversible Integrators

In a Hamiltonian problem, the symplectic condition and microscopic reversibility are inherent properties of the true time trajectories which, in turn, are the exact solution of Hamilton's equation. A stepwise integration defines a  $t$ -flow mapping which may or may not retain these properties. Non symplectic and/or non reversible integrators are generally believed [Sanz-Serna 1992, Grey et al. 1994, Biesiadecki and Skeel 1993, Channel and Scovel 1990] to be less stable in the long-time integration of Hamiltonian systems. In this section we shall illustrate the concept of reversible and symplectic mapping in relation to the numerical integration of the equations of motion.

#### Canonical Transformation and Symplectic Conditions

Given a system with  $n$  generalised coordinates  $q$ ,  $n$  conjugated momenta  $p$  and Hamiltonian  $H$ , the corresponding Hamilton's equations of motion are:

$$\dot{q}_i = \frac{\delta H}{p_i}$$



$$\dot{p}_i = \frac{\delta H}{q_i} \quad i = 1, 2, \dots, n \quad (3.3.5)$$

These equations can be written in a more compact form by defining a column matrix with  $2n$  elements such that:

$$\mathbf{x} = \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} \quad (3.3.6)$$

In this notation the Hamilton's equations 3.3.5 can be compactly written as:

$$\dot{\mathbf{x}} = \mathbf{J} \frac{\delta H}{\mathbf{x}} \quad \mathbf{J} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (3.3.7)$$

where  $\mathbf{J}$  is a  $2n \times 2n$  matrix,  $1$  is an  $n \times n$  identity matrix and  $0$  is a  $n \times n$  matrix of zeroes. Eq. 3.3.7 is the so-called symplectic notation for the Hamilton's equations.

Using the same notation we now may define a transformation of variables from  $\mathbf{x} \equiv \{q, p\}$  to  $\mathbf{y} \equiv \{\mathbf{Q}, \mathbf{P}\}$  as:

$$\mathbf{y} = \mathbf{y}(\mathbf{x}) \quad (3.3.8)$$

For a restricted canonical transformation [Goldstein 1980, Arnold 1989] we know that the function  $H(x)$  expressed in the new coordinates  $\mathbf{y}$  serves as the Hamiltonian function for the new coordinates  $\mathbf{y}$ , that is the Hamilton's equations of motion in the  $\mathbf{y}$  basis have exactly the same form as in Eq. 3.3.7:

$$\dot{\mathbf{y}} = \mathbf{J} \frac{\delta H}{\delta \mathbf{y}} \quad (3.3.9)$$

If we now take the time derivative of Eq. 3.3.8, use the chain rule relating  $\mathbf{x}$  and  $\mathbf{y}$  derivatives and use Eq. 3.3.9, we arrive at:

$$\dot{\mathbf{y}} = \mathbf{M} \mathbf{J} \mathbf{M}^t \frac{\delta H}{\delta \mathbf{y}}. \quad (3.3.10)$$

Here  $\mathbf{M}$  is the Jacobian matrix with elements:

$$M_{ij} = \delta y_i / \delta x_j, \quad (3.3.11)$$

and  $\mathbf{M}^t$  is its transpose. By comparing Eqs. 3.3.9 and 3.3.10, we arrive at the conclusion that a transformation is canonical if, and only if, the Jacobian matrix  $\mathbf{M}$  of the transformation Eq. 3.3.8 satisfies the condition:

$$\mathbf{M}\mathbf{J}\mathbf{M}^t = \mathbf{J}. \quad (3.3.12)$$

Eq. 3.3.12 is known as the symplectic condition for canonical transformations and represents an effective tool to test whether a generic transformation is canonical. Canonical transformations play a key role in Hamiltonian dynamics. It may also be proved that there exist a canonical transformation, parametrically dependent on time, changing the coordinates and momenta  $\mathbf{y}(t) \equiv \{p, q\}$  at any given time  $t$  to their initial values  $\mathbf{x} \equiv \{p, q\}$  [Goldstein 1980]. Obtaining such a transformation is of course equivalent to finding the solution for Hamilton's equations of motion.<sup>2</sup> The transformation  $\phi$

$$\mathbf{z}(t) = \phi(t, \mathbf{z}(0)) \quad (3.3.13)$$

from the initial values  $\{p_0, q_0\} \equiv \mathbf{z}(0)$  to the generalised coordinates  $\{\mathbf{P}, \mathbf{Q}\} \equiv \mathbf{z}(t)$  at a generic time  $t$  defines the  $t$ -flow mapping of the systems and, being canonical, its Jacobian matrix obeys the symplectic condition 3.3.12. An important consequence of the symplectic condition, is the invariance under canonical (or symplectic) transformations of many properties of the phase space. These invariant properties are known as "Poincare invariants" or canonical invariants. For example transformations or  $t$ -flow's mapping obeying Eq. 3.3.12 preserve the phase space volume. This is easy to see, since the infinitesimal volume elements in the  $\mathbf{y}$  and  $\mathbf{x}$  bases are related by:

$$d\mathbf{y} = |\det\mathbf{M}|d\mathbf{x} \quad (3.3.14)$$

where  $|\det\mathbf{M}|$  is the Jacobian of the transformation. Taking the determinant of the symplectic condition Eq. 3.3.12 we see that  $|\det\mathbf{M}| = 1$  and therefore

$$d\mathbf{y} = d\mathbf{x} \quad (3.3.15)$$

---

<sup>2</sup>This concept may be effectively understood using the generating function approach [Goldstein 1980, Sanz-Serna 1992] for canonical transformation applied to the particular case of an infinitesimal canonical transformation (*ict*) dependent on a "trajectory" parameter  $\alpha$ . If the generator is the Hamiltonian itself and the parameter is time  $t$ , then the canonical transformation made up of a succession of infinitesimal *ict*'s is the solution of Hamilton's equation [Goldstein 1980].

For a canonical or symplectic  $t$ -flow mapping this means that the phase total space volume is invariant and therefore Liouville theorem is automatically satisfied.

A stepwise numerical integration scheme defines a  $\Delta t$ -flow mapping or equivalently a coordinates transformation, that is

$$\begin{aligned} \mathbf{Q}(\Delta t) &= \mathbf{Q}(q(0), p(0), \Delta t) \\ \mathbf{P}(\Delta t) &= \mathbf{P}(q(0), p(0), \Delta t) \end{aligned} \quad y(\Delta t) = y(x(0)). \quad (3.3.16)$$

We have seen that exact solution of the Hamilton equations has  $t$ -flow mapping satisfying the symplectic conditions 3.3.12. If the Jacobian matrix of the transformation 3.3.16 satisfies the symplectic condition then the integrator is termed to be symplectic. The resulting integrator, therefore, exhibits properties identical to those of the exact solution, in particular it satisfies Eq. 3.3.15. Symplectic algorithms have also been proved to be robust, i.e resistant to time step increase, and generate stable long time trajectory, i.e. they do not show drifts of the total energy. Popular MD algorithms like Verlet, leap frog and velocity Verlet are all symplectic and their robustness is now understood to be due in part to this property [Biesiadecki and Skeel 1993, Tuckerman et al. 1992, Procacci and Berne 1994a, Channel and Scovel 1990].

### Liouville Formalism: a Tool for Building Symplectic and Reversible Integrators

In the previous paragraphs we have seen that it is highly beneficial for an integrator to be symplectic. We may now wonder if there exists a general way for obtaining symplectic and possibly, reversible integrators from "first principles". To this end, we start by noting that for any property which depends on time implicitly through  $p, q \equiv \mathbf{x}$  we have

$$\begin{aligned} \frac{dA(p, q)}{dt} &= \sum_{q,p} \left( \dot{q} \frac{\delta A}{\delta q} + \dot{p} \frac{\delta A}{\delta p} \right) = \sum_{q,p} \left( \frac{\delta H}{\delta p} \frac{\delta A}{\delta q} - \frac{\delta H}{\delta q} \frac{\delta A}{\delta p} \right) \\ &= iLA \end{aligned} \quad (3.3.17)$$

where the sum is extended to all  $n$  degrees of freedom in the system.  $L$  is the Liouvillean operator defined by

$$iL \equiv \sum_{q,p} \left( \dot{q} \frac{\delta}{\delta q} + \dot{p} \frac{\delta}{\delta p} \right) = \sum_{q,p} \left( \frac{\delta H}{\delta p} \frac{\delta}{\delta q} - \frac{\delta H}{\delta q} \frac{\delta}{\delta p} \right). \quad (3.3.18)$$

Eq. 3.3.17 can be integrated to yield

$$A(t) = e^{iLt} A(0) : \quad (3.3.19)$$

If  $A$  is the state vector itself we can use Eq. 3.3.19 to integrate Hamilton's equations:

$$\begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = e^{iLt} \begin{bmatrix} q(0) \\ p(0) \end{bmatrix}. \quad (3.3.20)$$

The above equation is a formal solution of Hamilton's equations of motion. The exponential operator  $e^{iLt}$  times the state vector defines the  $t$ -flow of the Hamiltonian system which brings the system phase space point from the initial state  $q_0, p_0$  to the state  $p(t)q(t)$  at a later time  $t$ . We already know that this transformation obeys Eq. 3.3.12. We may also note that the adjoint of the exponential operator corresponds to the inverse, that is  $e^{iLt}$  is unitary. This implies that the trajectory is exactly time reversible. In order to build our integrator, we now define the discrete time propagator  $e^{iL\Delta t}$  as

$$e^{iLt} = \left[ e^{iLt/ni} \right]^n, \quad \Delta t = t/n \quad (3.3.21)$$

$$e^{iL\Delta t} = e \sum_{q,p} \left( \dot{q} \frac{\delta}{\delta q} + \dot{p} \frac{\delta}{\delta p} \right) \Delta t. \quad (3.3.22)$$

In principle, to evaluate the action of  $e^{iL\Delta t}$  on the state vector  $p, q$  one should know the derivatives of all orders of the potential  $V$ . This can be easily seen by Taylor expanding the discrete time propagator  $e^{iL\Delta t}$  and noting that the operator  $\dot{q}/q$  does not commute with  $-\delta V/\delta q(\delta/\delta p)$  when the coordinates and momenta refer to same degree of freedom. We seek therefore approximate expressions of the discrete time propagator that retain both the symplectic and the reversibility property. For any two linear operators  $A, B$  the Trotter formula [Trotter 1959] holds:

$$e^{(A+B)t} = \lim_{n \rightarrow \infty} \left( e^{At/n} e^{Bt/n} \right)^n \quad (3.3.23)$$

We recognise that the propagator Eq. 3.3.22 has the same structure as the left hand side of Eq. 3.3.23; hence, using Eq. 3.3.23, we may write for  $\Delta t$  sufficiently small

$$e^{iL\Delta t} = e^{(\dot{q} \frac{\delta}{\delta q} + \dot{p} \frac{\delta}{\delta p}) \Delta t} \simeq e^{\dot{q} \frac{\delta}{\delta q} \Delta t} e^{\dot{p} \frac{\delta}{\delta p} \Delta t} + \mathcal{O}(\Delta t^2). \quad (3.3.24)$$

Where, for simplicity of discussion, we have omitted the sum over  $q$  and  $p$  in the exponential. Eq. 3.3.24 is exact in the limit that  $\Delta t \rightarrow 0$  and is first order for finite step size. Using Eq. 3.3.12 it is easy to show that the  $t$ -flow defined in Eq. 3.3.24 is symplectic, being the product of two successive symplectic transformations. Unfortunately, the propagator Eq. 3.3.24 is not unitary and therefore the corresponding algorithm is not time reversible. Again the non unitarity is due to the fact that the two factorised exponential operators are non commuting. We can overcome this problem by halving the time step and using the approximant:

$$e^{(A+B)t} \simeq e^{At/2} e^{Bt/2} e^{Bt/2} e^{At/2} = e^{At/2} e^{Bt} e^{At/2} \quad (3.3.25)$$

The resulting propagator is clearly unitary, therefore time reversible, and is also correct to the second order [de Raedt and Raedt 1983]. Thus, requiring that the product of the exponential operator be unitary, automatically leads to more accurate approximations of the true discrete time propagator [Yoshida 1990, de Raedt and Raedt 1983]. Applying the same argument to the propagator 3.3.22 we have

$$e^{iL\Delta t} = e^{(\dot{q}\frac{\delta}{\delta q} + \dot{p}\frac{\delta}{\delta p})\Delta t} \simeq e^{\dot{p}\frac{\delta}{\delta p}\Delta t/2} e^{\dot{q}\frac{\delta}{\delta q}\Delta t} e^{\dot{p}\frac{\delta}{\delta p}\Delta t/2} + \mathcal{O}(\Delta t^3). \quad (3.3.26)$$

The action of an exponential operator  $e^{(a\delta/\delta x)}$  on a generic function  $f(x)$  trivially corresponds to the Taylor expansion of  $f(x)$  around the point  $x$  at the point  $x + a$ , that is

$$e^{(a\delta/\delta x)} f(x) = f(x + a). \quad (3.3.27)$$

Using Eq. 3.3.27, the time reversible and symplectic integration algorithm can now be derived by acting with our Hermitian operator Eq. 3.3.26 onto the state vector at  $t = 0$  to produce updated coordinate and momenta at a later time  $\Delta t$ . The resulting algorithm is completely equivalent to the well known velocity Verlet:

$$\begin{aligned} p(\Delta t/2) &= p(0) + F(0)\Delta t/2 \\ q(\Delta t) &= q(0) + \left( \frac{p(\Delta t/2)}{m} \right) \Delta t \\ p(\Delta t) &= p(\Delta t/2) + F(\Delta t)\Delta t/2. \end{aligned} \quad (3.3.28)$$

We first notice that each of the three transformations obeys the symplectic condition Eq. 3.3.12 and has a Jacobian determinant equal to one. The product of the three transformation is also symplectic and, thus, phase volume preserving. Finally, since the discrete time propagator 3.3.26 is unitary, the algorithm is time reversible.

One may wonder what it is obtained if the operators  $\dot{q}\delta/\delta q$  and  $-\delta V/\delta q(\delta/\delta p)$  are exchanged in the definition of the discrete time propagator 3.3.26. If we do so, the new integrator is

$$\begin{aligned} q(\Delta t/2) &= q(0) + \frac{p(0)}{m}\Delta t/2 \\ p(\Delta t) &= p(0) + F[q(\Delta t/2)]\Delta t \\ q(\Delta t) &= q(\Delta t/2) + \frac{p(\Delta t)}{m}\Delta t/2. \end{aligned} \quad (3.3.29)$$

This algorithm has been proved to be equivalent to the so-called Leap-frog algorithm [Toxvaerd 1987]. Tuckerman et al. [Tuckerman et al. 1992] called this algorithm position Verlet which is certainly a more appropriate name in the light of the exchanged role of positions and velocities with respect to the velocity Verlet. Also, Eq. 3.3.25 clearly shows that the position Verlet is essentially identical to the Velocity Verlet. A shift of a time origin by  $\Delta t/2$  of either Eq. 3.3.29 or Eq. 3.3.28 would actually make both integrator perfectly equivalent. However, as pointed out [Tuckerman et al. 1993], half time steps are not formally defined, being the right hand side of Eq. 3.3.25 an approximation of the discrete time propagator for the full step  $\Delta t$ . Velocity Verlet and Position Verlet, therefore, do not generate numerically identical trajectories although of course the trajectories are similar.

We conclude this section by saying that is indeed noticeable that using the same Liouville formalism different long-time known schemes can be derived. The Liouville approach represent therefore a unifying treatment for understanding the properties and relationships between stepwise integrators.

### Potential Subdivision and Multiple Time Steps Integrators for *NVE* Simulations

The ideas developed in the preceding sections can be used to build multiple time step integrators. Multiple time step integration is based on the concept of reference system. Let us now assume that the system potential  $V$  be

subdivided in  $n$  terms such that

$$V = V_0 + V_1 + \dots + V_n. \quad (3.3.30)$$

Additionally, we suppose that the corresponding average values of the square modulus of the forces  $F_k = |\delta V_k / \delta x|$  and of their time derivatives  $\dot{F}_k = |d = dt(\delta V_k / \delta x)|$  satisfy the following condition:

$$\begin{aligned} F_0^2 &\gg F_1^2 \gg \dots \gg F_n^2 \\ \dot{F}_0^2 &\gg \dot{F}_1^2 \gg \dots \gg \dot{F}_n^2. \end{aligned} \quad (3.3.31)$$

These equations express the situation where different time scales of the system correspond to different pieces of the potential. Thus, the Hamiltonian of the  $k$ -th reference system is defined as

$$H = T + V_0 + \dots + V_k, \quad (3.3.32)$$

with a perturbation given by:

$$P = V_{(k+1)} + V_{(k+2)} + \dots + V_n. \quad (3.3.33)$$

For a general subdivision of the kind given in Eq. 3.3.30 there exist  $n$  reference nested systems. In the general case of a flexible molecular systems, we have fast degrees of freedom which are governed by the stretching, bending and torsional potentials and by slow intermolecular motions driven by the intermolecular potential. As we shall discuss with greater detail in Sec. 3.3.3, in real systems there is no clearcut condition between intra and intermolecular motions since their time scales may well overlap in many cases. The conditions Eq. 3.3.31 are, hence, never fully met for any of all possible potential subdivisions.

Given a potential subdivision Eq. 3.3.30, we now show how a multi-step scheme can be built with the methods described in section 3.3.1. For the sake of simplicity, we subdivide the interaction potential of a molecular system into two components only: one intra molecular,  $V_0$ , generating mostly "fast" motions and the other intermolecular,  $V_1$ , driving slower degrees of freedom. Generalisation of the forthcoming discussion to a  $n$ -fold subdivision, Eq. 3.3.30, is then straightforward.

For the 2-fold inter/intra subdivision, the system with Hamiltonian  $H = T + V_0$  is called the intra-molecular reference system whereas  $V_1$  is the inter-

molecular perturbation to the reference system. Correspondingly, the Liouvillean may be split as

$$\begin{aligned} iL_0 &= \dot{q} \frac{\delta}{\delta q} - \frac{\delta V_0}{\delta q} \frac{\delta}{\delta p} \\ iL_1 &= -\frac{\delta V_1}{\delta q} \frac{\delta}{\delta p} \end{aligned} \quad (3.3.34)$$

Here  $L_0$  is the Liouvillean of the 0-th reference system with Hamiltonian  $T + V_0$ , while  $L_1$  is a perturbation Liouvillean. Let us now suppose now that  $\Delta t_1$  is a good time discretization for the time evolution of the perturbation, that is for the slowly varying intermolecular potential. The discrete  $e^{iL\Delta t_1} \equiv e^{(iL_0+iL_1)\Delta t_1}$  time propagator can be factorised as

$$\begin{aligned} e^{iL\Delta t_1} &= e^{iL_1\Delta t_1/2} (e^{iL_0\Delta t_1/n})^n e^{iL_1\Delta t_1/2} \\ &= e^{iL_1\Delta t_1/2} (e^{iL_0\Delta t_0})^n e^{iL_1\Delta t_1/2}, \end{aligned} \quad (3.3.35)$$

where we have used Eq. 3.3.25 and we have defined

$$\Delta t_0 = \frac{\Delta t_1}{n} \quad (3.3.36)$$

as the time step for the "fast" reference system with Hamiltonian  $T + V_0$ . The propagator 3.3.35 is unitary and hence time reversible. The external propagators depending on the Liouvillean  $L_1$  acting on the state vectors define a symplectic mapping, as it can be easily proved by using Eq. 3.3.12. The full factorised propagator is therefore symplectic as long as the inner propagator is symplectic. The Liouvillean  $iL_0 \equiv \dot{q}/q - \delta V_0/\delta q/\delta p$  can be factorised according to the Verlet symplectic and reversible breakup described in the preceding section, but with an Hamiltonian  $T + V_0$ . Inserting the result into Eq. 3.3.35 and using the definition 3.3.34, the resulting double time step propagator is then

$$e^{iL\Delta t_1} = e^{-\frac{\delta V_1}{\delta q} \frac{\delta}{\delta p} \Delta t_1/2} \left( e^{-\frac{\delta V_0}{\delta q} \frac{\delta}{\delta p} \Delta t_0/2} e^{\dot{q} \frac{\delta}{\delta q} \Delta t_0} e^{-\frac{\delta V_0}{\delta q} \frac{\delta}{\delta p} \Delta t_0/2} \right)^n e^{-\frac{\delta V_1}{\delta q} \frac{\delta}{\delta p} \Delta t_1/2} \quad (3.3.37)$$

This propagator is unfolded straightforwardly using the rule 3.3.27 generating the following symplectic and reversible integrator from step  $t = 0$  to  $t = \Delta t_1$ :



$$\begin{aligned}
p\left(\frac{\Delta t_1}{2}\right) &= p(0) + F_1(0)\frac{\Delta t_1}{2} \\
&\quad DO \ i = 1, n \\
p\left(\frac{\Delta t_1}{2} + i\frac{\Delta t_0}{2}\right) &= p\left(\frac{\Delta t_1}{2} + [i-1]\frac{\Delta t_0}{2}\right) + F_0\left([i-1]\frac{\Delta t_0}{2}\right)\frac{\Delta t_0}{2} \\
q(i\Delta t_0) &= q([i-1]\Delta t_0) + p\left(\frac{\Delta t_1}{2} + i\frac{\Delta t_0}{2}\right)\frac{\Delta t_0}{m} \\
p\left(\frac{\Delta t_1}{2} + i\Delta t_0\right) &= p\left(\frac{\Delta t_1}{2} + i\frac{\Delta t_0}{2}\right) + F_0(i\Delta t_0)\frac{\Delta t_0}{2} \\
&\quad ENDDO \\
p(\Delta t_1) &= p'\left(\frac{\Delta t_1}{2}\right) + F_1(n\Delta t_0)\frac{\Delta t_1}{2}
\end{aligned} \tag{3.3.38}$$

Note that the slowly varying forces  $F_1$  are felt only at the beginning and the end of the macrostep  $\Delta t_1$ . In the inner  $n$  steps loop the system moves only according to the Hamiltonian of the reference system  $H = T + V_0$ . When using the potential breakup, the inner reference system is rigorously conservative and the total energy of the reference system (i.e.  $T + V_0 + \dots + V_k$ ) is conserved during the  $P$  micro-steps.

The integration algorithm given an arbitrary subdivision of the interaction potential is now straightforward. For the general subdivision 3.3.30 the corresponding Liouvillean split is

$$iL_0 = \dot{q}\frac{\delta}{\delta q} - \frac{\delta V_i}{\delta q}\frac{\delta}{\delta p}, \quad iL_1 = -\frac{\delta V_1}{\delta q}\frac{\delta}{\delta p} \dots, \quad iL_n = -\frac{\delta V_k}{\delta q}\frac{\delta}{\delta p}. \tag{3.3.39}$$

We write the discrete time operator for the Liouville operator  $iL = L_0 + \dots + L_n$  and use repeatedly the Hermitian approximant and Trotter formula to get a hierarchy of nested reference systems propagator, viz.

$$e^{i(\sum_{i=0}^n L_i)\Delta t_n} = e^{iL_n\frac{\Delta t_n}{2}} \left( e^{i(\sum_{i=0}^{n-1} L_i)\Delta t_{n-1}} \right)^{P_{n-1}} e^{iL_n\frac{\Delta t_n}{2}} \tag{3.3.40}$$

$$\begin{aligned}
\Delta t_n &= \Delta t_{n-1} P_{n-1} \\
e^{i(\sum_{i=0}^{n-1} L_i)\Delta t_{n-1}} &= e^{iL_{n-1}\frac{\Delta t_{n-1}}{2}} \left( e^{i(\sum_{i=0}^{n-2} L_i)\Delta t_{n-2}} \right)^{P_{n-2}} e^{iL_{n-1}\frac{\Delta t_{n-1}}{2}}
\end{aligned} \tag{3.3.41}$$

$$\begin{aligned} \Delta t_{n-1} &= \Delta t_{n-2} P_{n-2} \\ e^{i(L_0+L_1+L_2)\Delta t_2} &= \overset{\dots\dots\dots}{e^{iL_2\frac{\Delta t_2}{2}}} \left( e^{i(L_1+L_0)\Delta t_1} \right)^{P_1} e^{iL_2\frac{\Delta t_2}{2}} \end{aligned} \quad (3.3.42)$$

$$\begin{aligned} \Delta t_2 &= \Delta t_1 P_1 \\ e^{i(L_0+L_1)\Delta t_1} &= e^{iL_1\frac{\Delta t_1}{2}} \left( e^{iL_0\Delta t_0} \right)^{P_0} e^{iL_1\frac{\Delta t_1}{2}} \\ \Delta t_1 &= \Delta t_0 P_0 \end{aligned} \quad (3.3.43)$$

where  $\Delta t_i$  is the generic integration time steps selected according to the time scale of the  $i$ -th force  $F_i$ . We now substitute Eq. 3.3.44 into Eq. 3.3.43 and so on climbing the whole hierarchy until Eq. 3.3.41. The resulting multiple time steps symplectic and reversible propagator is then

$$\begin{aligned} e^{-iL\Delta t_n} &= e^{F_n\frac{\delta}{\delta p}\Delta t_n} \left( e^{F_{n-1}\frac{\delta}{\delta p}\frac{\Delta t_{n-1}}{2}} \dots \right. \\ &\dots \left. \left( \left( e^{F_0\frac{\delta}{\delta p}\frac{\Delta t_0}{2}} e^{i\frac{\delta}{\delta q}\Delta t_0} e^{F_0\frac{\delta}{\delta p}\frac{\Delta t_0}{2}} \right)^{P_0} \dots e^{F_{n-1}\frac{\delta}{\delta p}\frac{\Delta t_{n-1}}{2}} \right)^{P_{n-1}} e^{F_n\frac{\delta}{\delta p}\Delta t_n} \right) \end{aligned} \quad (3.3.44)$$

The integration algorithm that can be derived from the above propagator was first proposed by Tuckerman, Martyna and Berne and called r-RESPA, reversible reference system propagation algorithm [Tuckerman et al. 1992]

### Constraints and r-RESPA

The r-RESPA approach makes unnecessary to resort to the SHAKE procedure [Ryckaert et al. 1977, Ciccotti and Ryckaert 1986a] to freeze some fast degrees of freedom. However the SHAKE and r-RESPA algorithms are not mutually exclusive and sometimes it might be convenient to freeze some degrees of freedom while simultaneously using a multi-step integration for all other freely evolving degrees of freedom. Since r-RESPA consists in a series of nested velocity Verlet like algorithms, the constraint technique RATTLE [Andersen 1983] used in the past for single time step velocity Verlet integrator can be straightforwardly applied. In RATTLE both the constraint conditions on the coordinates and their time derivatives must be satisfied. The resulting coordinate constraints is upheld by a SHAKE iterative procedure which corrects the positions exactly as in a standard Verlet integration,

while a similar iterative procedure is applied to the velocities at the half time step.

In a multi time step integration, whenever velocities are updated, using part of the overall forces (e.g. the intermolecular forces), they must also be corrected for the corresponding constraints forces with a call to RATTLE. This combined RATTLE-r-RESPA procedure has been described for the first time by Tuckerman and Parrinello [Tuckerman and Parrinello 1994] in the framework of the Car-Parrinello simulation method. To illustrate the combined RATTLE-r-RESPA technique in a multi-step integration, we assume a separation of the potential into two components deriving from intramolecular and intermolecular interactions. In addition, some of the covalent bonds are supposed rigid, i.e.

$$d_a = d_a^{(0)} \quad (3.3.45)$$

$$\dot{d}_a = 0 \quad (3.3.46)$$

where  $a$  runs over all constrained bonds and  $d^{(0)}_a$  are constants. In the double time integration 3.3.1, velocities are updated four times, i.e. two times in the inner loop and two times in the outer loop. To satisfy 3.3.46, SHAKE must be called to correct the position in the inner loop. To satisfy 3.3.46, RATTLE must be called twice, once in the inner loop and the second time in the outer loop according to the following scheme

$$\begin{aligned}
 p' \left( \frac{\Delta t_1}{2} \right) &= p(0) + F_1(0) \frac{\Delta t_1}{2} \\
 p(\Delta t_1) &= RATTLE_p \left\{ p' \left( \frac{\Delta t_1}{2} \right) \right\} \\
 DO & \quad i = 1, n \\
 & \quad p' \left( \frac{\Delta t_1}{2} + i \frac{\Delta t_0}{2} \right) = p \left( \frac{\Delta t_1}{2} + [i - 1] \frac{\Delta t_0}{2} \right) + F_0 \left( [i - 1] \frac{\Delta t_0}{2} \right) \frac{\Delta t_0}{2} \\
 & \quad p \left( \frac{\Delta t_1}{2} + i \frac{\Delta t_0}{2} \right) = RATTLE_p \left\{ p' \left( \frac{\Delta t_1}{2} + i \frac{\Delta t_0}{2} \right) \right\} \\
 & \quad q' (i \Delta t_0) = q \left( [i - 1] \Delta t_0 \right) + p \left( \frac{\Delta t_1}{2} + i \frac{\Delta t_0}{2} \right) \frac{\Delta t_0}{m} \\
 & \quad q (i \Delta t_0) = RATTLE_q \left\{ q' (i \Delta t_0) \right\} \\
 & \quad p \left( \frac{\Delta t_1}{2} + i \Delta t_0 \right) = p \left( \frac{\Delta t_1}{2} + i \frac{\Delta t_0}{2} \right) + F_0 (i \Delta t_0) \frac{\Delta t_0}{2} \\
 ENDDO & \\
 p(\Delta t_1) &= p' \left( \frac{\Delta t_1}{2} \right) + F_1 (n \Delta t_0) \frac{\Delta t_1}{2}
 \end{aligned} \quad (3.3.47)$$

Where  $RATTLE_p$  and  $RATTLE_q$  represent the constraint procedure on velocity and coordinates, respectively.

### Applications

As a first simple example we apply the double time integrator to the *NVE* simulation of flexible nitrogen at 100 K.

Table 3.1: Energy conservation ratio  $R$  for various integrators (see text). The last three entries refer to a velocity Verlet with bond constraints.  $\langle V_i \rangle$  and  $\langle V_m \rangle$  are the average value of the intra-molecular and intermolecular energies (in KJ/mole), respectively. CPU is given in seconds per picoseconds of simulation and  $\Delta t$  in fs. Single time step velocity Verlet with  $\Delta t = 4.5$  fs is unstable.

$\Delta t$	$n$	$R$	CPU	$\langle V_i \rangle$	$\langle V_m \rangle$
<b>0.3</b>	<b>1</b>	<b>0.005</b>	<b>119</b>	<b>0.1912</b>	<b>-4.75</b>
0.6	1	0.018	62	0.1937	-4.75
1.5	1	0.121	26	0.2142	-4.75
4.5	1	-	-	-	-
0.6	2	0.004	59	0.1912	-4.75
1.5	5	0.004	28	0.1912	-4.75
3.0	10	0.005	18	0.1912	-4.75
4.5	15	0.006	15	0.1912	-4.75
6.0	20	0.008	12	0.1912	-4.74
9.0	30	0.012	10	0.1911	-4.74
3.0	-	0.001	14	-	-4.74
6.0	-	0.004	8	-	-4.75
9.0	-	0.008	6	-	-4.74

The overall interaction potential is given by

$$V = V_{intra} + V_{inter}$$

Where  $V_{inter}$  is the intermolecular potential described by a Lennard-Jones model between all nitrogen atoms on different molecules [Nosé and Klein 1983].  $V_{intra}$  is instead the intramolecular stretching potential holding together the two nitrogen atoms of each given molecule. We use here a simple harmonic spring depending on the molecular bond length  $r_m$ , namely:

$$V_{intra} = \frac{1}{2} \sum_m k (r - r_0)^2, \quad (3.3.48)$$

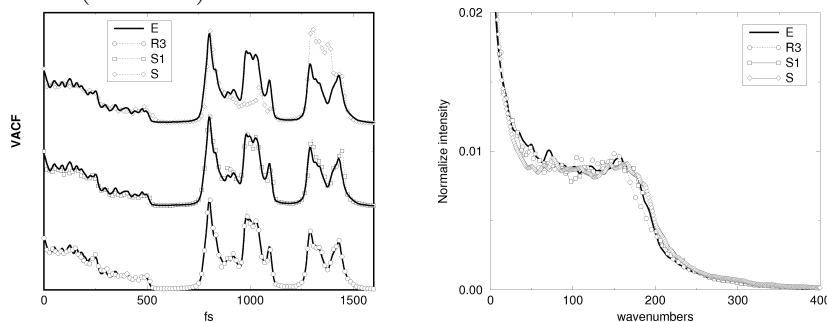
with  $r_0$  and  $r$  the equilibrium and instantaneous distance between the nitrogen atoms, and  $k$  the force constant tuned to reproduce the experimental gas-phase stretching frequency [Herzberg 1950].

As a measure of the accuracy of the numerical integration we use the adimensional energy conservation ratio [Procacci and Berne 1994a, an G. Dolling 1975, Procacci and Marchi 1996, Procacci et al. 1996]

$$R = \frac{\langle E^2 \rangle - \langle E \rangle^2}{\langle K^2 \rangle - \langle K \rangle^2} \quad (3.3.49)$$

where  $E$  and  $K$  are the total and kinetic energy of the system, respectively. In Tab. 3.1 we show the energy conservation ratio  $R$  and CPU timings on a IBM-43P/160MH/RS6000 obtained for flexible nitrogen at 100 K with the r-RESPA integrator as a function of  $n$  and  $\Delta t_1$  in Eq. 3.3.1 and also for single time step integrators. Results of integrators for rigid nitrogen using SHAKE are also shown for comparison. The data in Tab. 3.1 refer to a 3.0 ps run without velocity rescaling. They were obtained starting all runs from coordinates corresponding to the experimental  $P_{a3}$  structure [an G. Dolling 1975, Medina and Daniels 1976] of solid nitrogen and from velocities taken randomly according to the Boltzmann distribution at 100 K.

Figure 3.2: Time record of the torsional potential energy at about 300 K for a cluster of eight molecules of  $C_{24}H_{50}$  obtained using three integrators: solid line integrator  $E$ ; circles integrator  $R3$ ; squares integrator  $S1$ ; diamonds integrator  $S$  (see text)

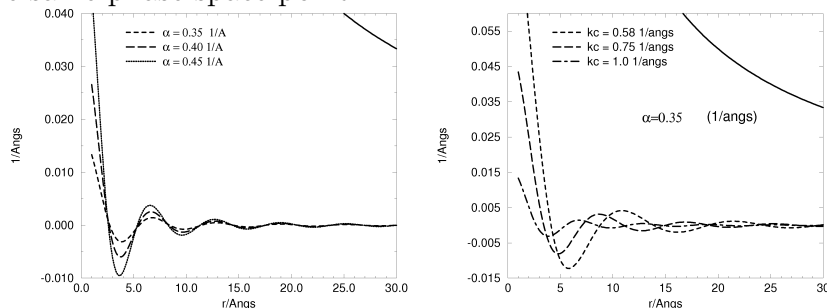


The entry in bold refers to the "exact" result, obtained with a single time step integrator with a very small step size of 0.3 fs. Note that  $R$  increases quadratically with the time step for single time step integrators whereas r-RESPA is remarkably resistant to outer time step size increase. For example

r-RESPA with  $\Delta t_1 = 9.0fs$  and  $P = 30$  (i.e.  $\Delta t_0 = 0.3fs$ ) yields better accuracy on energy conservation than single time step Velocity Verlet with  $\Delta t = 0.6fs$  does, while being more than six times faster. Moreover, r-RESPA integrates all degrees of freedom of the systems and is almost as efficient as Velocity Verlet with constraints on bonds. It is also worth pointing out that energy averages for all r-RESPA integrators is equal to the exact value, while at single time step even a moderate step size increase results in sensibly different averages intra-molecular energies. As a more complex example we now study a cluster of eight single chain alkanes  $C_{24}H_{50}$ . In this case the potential contains stretching, bending and torsional contributions plus the intermolecular Van-der-Waals interactions between non bonded atoms. The parameter are chosen according to the AMBER protocol [Cornell et al. 1995] by assigning the carbon and hydrogen atoms to the AMBER types *ct* and *hc*, respectively. For various dynamical and structural properties we compare three integrators, namely a triple time step r-RESPA (R3) a single time step integrator with bond constraints on X - H (S1) and a single time step integrator with all bonds kept rigid (S). These three integrators are tested, starting from the same phase space point, against a single time step integrator (E) with a very small time step generating the "exact" trajectory. In Fig. 3.2 we show the time record of the torsional potential energy. The R3 integrator generates a trajectory practically coincident with the "exact" trajectory for as long as 1.5 ps. The single time step with rigid X - H bonds also produces a fairly accurate trajectory, whereas the trajectory generated by S quickly drifts away from the exact time record. In Fig. 3.3 we show the power spectrum of the velocity autocorrelation function obtained with R3, S1 and S. The spectra are compared to the exact spectrum computed using the trajectories generated by the accurate integrator E. We see that R3 and S1 generates the same spectral profile within statistical error. In contrast, especially in the region above 800 wavenumbers, S generates a spectrum which differs appreciably from the exact one. This does not mean, of course, that S is unreliable for the "relevant" torsional degrees of freedom. Simply, we cannot a priori exclude that keeping all bonds rigid will not have an impact on the equilibrium structure of the alkanes molecules and on torsional dynamics. Actually, in the present case, as long as torsional motions are concerned all three integrators produce essentially identical results. In 20 picoseconds of simulation, R3 S1 and S predicted 60, 61, 60 torsional jumps, respectively, against the 59 jumps obtained with the exact integrator E. According to prescription of Ref. [Cardini and Schettino 1990], in order to avoid

period doubling, we compute the power spectrum of torsional motion from the autocorrelation function of the vector product of two normalised vector perpendicular to the dihedral planes. Rare events such as torsional jumps produce large amplitudes long time scale oscillations in the time autocorrelation function and therefore their contribution overwhelms the spectrum which appears as a single broaden peak around zero frequency. For this reason all torsions that did undergo a barrier crossing were discarded in the computation of the power spectrum. The power spectrum of the torsional motions is identical for all integrators within statistical error when evaluated over 20 ps of simulations.

Figure 3.3: Power spectra of the velocity autocorrelation function (left) and of the torsional internal coordinates (right) at 300 K for a cluster of 8  $C_{24}H_{50}$  molecules calculated with integrators  $E$ ,  $R3$ ,  $S1$  and  $S$  (see text) starting from the same phase space point



From these results it can be concluded that  $S1$  and  $R3$  are very likely to produce essentially the same dynamics for all "relevant" degrees of freedom. We are forced to state that also the integrator  $S$  appears to accurately predict the structure and overall dynamics of the torsional degrees of freedom at least for the 20 ps time span of this specific system.

Since torsions are not normal coordinates and couple to higher frequency internal coordinates such as bending and stretching, the ability of the efficient  $S$  integrator of correctly predicting low frequency dynamics and structural properties cannot be assumed a priori and must be, in principle, verified for each specific case. We also do not know how the individual eigenvectors are affected by the integrators and, although the overall density for  $S$  and  $S1$  appears to be the same, there might be considerable changes in the torsional dynamics.  $R3$  does not require any assumption, is accurate everywhere in the spectrum (see Fig. 3.2) and is as efficient as  $S$ . For these reasons  $R3$ , or

a multi-step version of the equally accurate S1, must be the natural choice for the simulation of complex systems using all-atoms models

### 3.3.2 Multiple Time Steps Algorithms for the Isothermal–Isobaric Ensemble

The integrators developed in the previous section generates dynamics in the microcanonical ensemble where total energy, number of particles and volume are conserved. The derivation based on the Liouvillean and the corresponding propagator, however lends itself to a straightforward generalisation to non microcanonical ensembles. Simulations of this kind are based on the concept of extended system and generate trajectories that sample the phase space according to a target distribution function. The extended system method is reviewed in many excellent textbooks and papers [Allen and Tildesley 1989, Frenkel and Smit 1996, Rapaport 1995, Nosé 1991a;b, Ferrario 1993, Martyna et al. 1994; 1996] to which we refer for a complete and detailed description. Here it suffices to say that the technique relies on the clever definition of a modified or extended Lagrangian which includes extra degrees of freedom related to the intensive properties (e.g. pressure or temperature) one wishes to sample with a well defined distribution function. The dynamics of the extended system is generated in the microcanonical ensemble with the true  $n$  degrees of freedom and, additionally, the extra degrees of freedom related to the macroscopic thermodynamic variables. With an appropriate choice, the equations of motion of the extended system will produce trajectories in the extended phase space generating the desired equilibrium distribution function upon integration over the extra (extended) variables. There are several extended system techniques corresponding to various ensembles, e.g. constant pressure in the  $NPH$  ensemble simulation with isotropic [Andersen 1980] and anisotropic [Parrinello and Rahman 1980] stress, constant temperature simulation [Nosé 1984] in the  $NVT$  ensemble and isothermal–isobaric simulation [Ferrario and Ryckaert 1985] in the  $NPT$  ensemble. As we shall see, the dynamic of the real system generated by the extended system method is never Hamiltonian. Hence, symplecticness is no longer an inherent property of the equations of motion. Nonetheless, the Liouvillean formalism developed in the preceding section, turns out to be very useful for the derivation of multiple time step reversible integrators for a general isothermal–isobaric



ensemble with anisotropic stress, or *NPT*<sup>3</sup>. This extended system is the most general among all non microcanonical simulations: The *NPT*, *NPH* the *NVT* and even *NVE* ensemble may be derived from this Lagrangian by imposing special constraints and/or choosing appropriate parameters [Martyna et al. 1996, Marchi and Procacci 1998]

### The Parrinello–Rahman–Nosé Extended Lagrangian

The starting point of our derivation of the multilevel integrator for the *NPT* ensemble is the Parrinello–Rahman–Nosé Lagrangian for a molecular system with  $N$  molecules or groups<sup>4</sup> each containing  $n_i$  atoms and subject to a potential  $V$ . In order to construct the Lagrangian we define a coordinate scaling and a velocity scaling, i.e.

$$r_{ik\alpha} = R_{i\alpha} + l_{ik\alpha} = \sum_{\beta} h_{\alpha\beta} S_{i\beta} + l_{ik\alpha}; \quad (3.3.50)$$

$$\begin{aligned} \dot{R}'_{i\alpha} &= \dot{R}_{i\alpha} s \\ \dot{l}'_{ik\alpha} &= \dot{l}_{ik\alpha} s \end{aligned} \quad (3.3.51)$$

Here, the indices  $i$  and  $k$  refer to molecules and atoms, respectively, while Greek letters are used to label the Cartesian components.  $r_{ik\alpha}$  is the  $\alpha$  component of the coordinates of the  $k$ -th atom belonging to the  $i$ -th molecule;  $R_{i\alpha}$  is the center of mass coordinates;  $S_{i\beta}$  is the scaled coordinate of the  $i$ -th molecular center of mass.  $l_{ik\alpha}$  is the coordinate of the  $k$ -th atom belonging to the  $i$ -th molecule expressed in a frame parallel at any instant to the fixed laboratory frame, but with origin on the instantaneous molecular center of mass. The set of  $l_{ik\alpha}$  coordinates satisfies  $3N$  constraints of the type  $\sum_{k=1}^{n_i} l_{ik\alpha} = 0$ .

The matrix  $\mathbf{h}$  and the variable  $s$  control the pressure and temperature of the extended system, respectively. The columns of the matrix  $\mathbf{h}$  are the Cartesian components of the cell edges with respect to a fixed frame. The elements of this matrix allow the simulation cell to change shape and size and are sometimes called the "barostat" coordinates. The volume of the MD cell is related to  $\mathbf{h}$  through the relation

<sup>3</sup>When  $P$  is not in boldface, we imply that the stress is isotropic

<sup>4</sup>For large molecules it may be convenient to further subdivide the molecule into groups. A group, therefore encompasses a conveniently chosen subset of the atoms of the molecule

$$\Omega = \det(\mathbf{h}). \quad (3.3.52)$$

$s$  is the coordinates of the so-called "Nos'e thermostat" and is coupled to the intramolecular and center of mass velocities.

We define the "potentials" depending on the thermodynamic variables  $P$  and  $T$

$$\begin{aligned} V_P &= P \det(\mathbf{h}) \\ V_T &= \frac{g}{\beta} \ln s. \end{aligned} \quad (3.3.53)$$

Where  $P$  is the external pressure of the system,  $\beta = k_B T$ , and  $g$  is a constant related to total the number of degrees of freedom in the system. This constant is chosen to correctly sample the  $NPT$  distribution function.

The extended  $NPT$  Lagrangian is then defined as

$$L = \frac{1}{2} \sum_i^N M_i s^2 \dot{\mathbf{S}}_i^t \mathbf{h}^t \mathbf{h} \dot{\mathbf{S}}_i + \frac{1}{2} \sum_{ik} m_{ik} s^2 \dot{\mathbf{i}}_{ik}^t \dot{\mathbf{i}}_{ik} + \frac{1}{2} W s^2 \text{tr}(\dot{\mathbf{h}}^t \dot{\mathbf{h}}) \quad (3.3.54)$$

$$+ \frac{1}{2} Q \dot{s}^2 - V - P_{ext} \Omega - \frac{g}{\beta} \ln s \quad (3.3.55)$$

The arbitrary parameters  $W$  and  $Q$  are the "masses" of the barostat and of the thermostats, respectively<sup>5</sup>. They do not affect the sampled distribution function but only the sampling efficiency [Procacci and Berne 1994b, Nosé 1984, Ryckaert and Ciccotti 1983]. For a detailed discussion of the sampling properties of this Lagrangian the reader is referred to Refs. [Martyna et al. 1994, Marchi and Procacci 1998].

### The Parrinello–Rahman–Nos'e Hamiltonian and the Equations of Motion

In order to derive the multiple time step integration algorithm using the Liouville formalism described in the preceding sections we must switch to the Hamiltonian formalism. Thus, we evaluate the conjugate momenta of the

---

<sup>5</sup> $W$  has actually the dimension of a mass, while  $Q$  has the dimension of a mass time a length squared

coordinates  $S_{i\alpha}$ ,  $l_{ik\alpha}$ ,  $h\alpha\beta$  and  $s$  by taking the derivatives of the Lagrangian in Eq. 3.3.55 with respect to corresponding velocities, i.e.

$$\mathbf{T}_i = M_i \mathbf{G} s^2 \dot{\mathbf{S}}_i \quad (3.3.56)$$

$$\mathbf{p}_{ik} = m_{ik} s^2 \dot{\mathbf{l}}_{ik} \quad (3.3.57)$$

$$\mathbf{P}_h = s^2 W \dot{\mathbf{h}} \quad (3.3.58)$$

$$p_s = Q \dot{s}. \quad (3.3.59)$$

Where we have defined the symmetric matrix

$$\mathbf{G} = \mathbf{h}^t \mathbf{h} \quad (3.3.60)$$

The Hamiltonian of the system is obtained using the usual Legendre transformation [Goldstein 1980]

$$H(p, q) = \sum \dot{q}p - L(q, \dot{q}). \quad (3.3.61)$$

One obtains

$$\begin{aligned} H = & \frac{1}{2} \sum_i^N \frac{\dot{\mathbf{T}}_i \mathbf{G}^{-1} \dot{\mathbf{T}}_i}{M s^2} + \frac{1}{2} \sum_{ik} \frac{\mathbf{p}_{ik}^t \mathbf{p}_{ik}}{m_{ik} s^2} + \frac{1}{2} \frac{\text{tr}(\mathbf{P}_h^t \mathbf{P}_h)}{s^2 W} + \frac{p_s^2}{2Q} \\ & + V + P\Omega + \frac{g \ln s}{\beta} \end{aligned} \quad (3.3.62)$$

In the extended systems formulation we always deal with real and virtual variables. The virtual variables in the Hamiltonian 3.3.62 are the scaled coordinates and momenta while the unscaled variables (e.g  $\mathbf{R}_i = \mathbf{h}\mathbf{S}_i$  or  $p'_{ik\alpha} = p_{ik\alpha}/s$  are the real counterpart. The variable  $s$  in the Nosé formulation plays the role of a time scaling [Frenkel and Smit 1996, Nosé and Klein 1983, Nosé 1984]. The above Hamiltonian is given in terms of virtual variables and in term of a virtual time and is indeed a true Hamiltonian function and has corresponding equation of motions that can be obtained applying Eq. 3.3.7 with  $x \equiv S_{i\alpha}, l_{ik\alpha}, h\alpha\beta, s, T_{i\alpha}, p_{ik\alpha}, \pi_{\alpha\beta}, p_s$  in a standard fashion. Nonetheless, the equations of motions in terms of these virtual variable are inadequate for several reasons since for example one would deal with a fluctuating time step [Frenkel and Smit 1996, Nosé 1984]. It is therefore convenient to work in terms of real momenta and real time. The real momenta are related to the virtual counterpart through the relations

$$T_{i\alpha} \rightarrow T_{i\alpha}/s \quad (3.3.63)$$

$$p_{ik\alpha} \rightarrow p_{ik\alpha}/s \quad (3.3.64)$$

$$(\mathbf{P}_h)_{\alpha\beta} \rightarrow (\mathbf{P}_h)_{\alpha\beta}/s \quad (3.3.65)$$

$$p_s \rightarrow p_s/s \quad (3.3.66)$$

It is also convenient [Procacci and Berne 1994b] to introduce new center of mass momenta as

$$\mathbf{P}_i \equiv \mathbf{G}^{-1}\mathbf{T}_i. \quad (3.3.67)$$

such that the corresponding velocities may be obtained directly without the knowledge of the "coordinates"  $\mathbf{h}$  in  $\mathbf{G}$

$$\dot{\mathbf{S}}_i = \frac{\mathbf{P}_i}{M}. \quad (3.3.68)$$

Finally, a real time formulation and a new dynamical variable  $j$  are adopted:

$$t \rightarrow t/s\eta \equiv ln \ s \quad (3.3.69)$$

The equations of motions for the newly adopted set of dynamical variables are easily obtained from the true Hamiltonian in Eq. 3.3.62 and then using Eqs. 3.3.64-3.3.69 to rewrite the resulting equations in terms of the new momenta. In so doing, we obtain:

$$\dot{\mathbf{i}}_{ik} = \frac{\mathbf{p}_i k}{m_{ik}}, \quad \dot{\mathbf{S}}_i = \frac{\mathbf{P}_i}{M_i}, \quad \dot{\mathbf{h}} = \frac{\mathbf{P}_h}{W}, \quad \dot{\eta} = \frac{p_\eta}{Q} \quad (3.3.70)$$

$$\dot{\mathbf{p}}_{ik} = f_{ik}^c - \frac{p_\eta}{Q}\mathbf{p}_{ik}, \quad (3.3.71)$$

$$\dot{\mathbf{P}}_i = \mathbf{h}^{-1}\mathbf{F}_i - \mathbf{G}^{-1}\dot{\mathbf{G}}\mathbf{P}_i - \frac{p_\eta}{Q}\mathbf{P}_i, \quad (3.3.72)$$

$$\dot{\mathbf{P}}_h = \left( \mathcal{V} + \mathcal{K} - \mathbf{h}^{-1}P_{ext} \det \mathbf{h} \right) - \frac{p_\eta}{Q}\mathbf{P}_h, \quad (3.3.73)$$

$$\dot{p}_\eta = \mathcal{F}_\eta. \quad (3.3.74)$$

It can be verified that the conserved quantity  $\mathcal{H}$  is associated with the above equations of motion, namely

$$\begin{aligned}
\mathcal{H} &= \frac{1}{2} \sum_{i=1}^N \frac{\mathbf{P}_i^t \mathbf{G} \mathbf{P}_i}{M_i} + \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^{n_i} \frac{\mathbf{P}_{ik}^t \mathbf{P}_{ik}}{m_{ik}} + \frac{1}{2} \frac{\text{tr}(\mathbf{P}_h^t \mathbf{P}_h)}{\mathbf{W}} \\
&+ \frac{1}{2} \frac{p_\eta p_\eta}{Q} + V + P_{ext} \det \mathbf{h} + g k_B T \eta.
\end{aligned} \tag{3.3.75}$$

The atomic force  $\mathbf{f}_{ik}^c = \frac{\delta V}{\delta r_{ik\alpha}} - \frac{m_{ik}}{M_i} \mathbf{F}_i$  includes a constraint force contribution which guarantees that the center of mass in the intramolecular frame of the  $l_{ik\alpha}$  coordinates remains at the origin.  $\mathcal{V}$  and  $\mathcal{K}$  are the virial and ideal gas contribution to the internal pressure tensor  $P_{int} = \mathcal{V} + \mathcal{K}$  and they are defined as <sup>6</sup>

$$\begin{aligned}
\mathcal{V} &= \sum_{i=1}^N \mathbf{F}_i \mathbf{S}_i^t \\
\mathcal{K} &= \sum_{i=1}^N M_i (\mathbf{h} \dot{\mathbf{S}}_i) \dot{\mathbf{S}}_i^t.
\end{aligned} \tag{3.3.76}$$

Finally  $\mathbf{F}_i$  is the force driving the Nos'e thermostat

$$\mathcal{F}_\eta = \frac{1}{2} \sum_{i=1}^N M_i \dot{\mathbf{S}}_i^t \mathbf{G} \dot{\mathbf{S}}_i + \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^{n_i} \frac{\mathbf{P}_{ik}^t \mathbf{P}_{ik}}{m_{ik}} - -g k_B T \tag{3.3.77}$$

with  $g$  equal to the number of all degrees of freedom  $N_f$  including those of the barostat<sup>7</sup>.

Eqs. 3.3.64–3.3.69 define a generalised coordinates transformation of the kind of Eq. 3.3.8. This transformation is non canonical, i.e. the Jacobian matrix of the transformation from the virtual coordinates does not obey Eq. 3.3.12. This means that  $\mathcal{H}$  in terms of the new coordinates Eq. 3.3.75 is "only" a constant of motion, but is no longer a true Hamiltonian: application

---

<sup>6</sup>In presence of bond constraints and if the scaling is group-based instead of molecular based, these expression should contain a contribution from the constraints forces. Complications due to the constraints can be avoided altogether by defining groups so that no two groups are connected through a constrained bond [Marchi and Procacci 1998]. In that case  $V$  does not include any constraint contribution.

<sup>7</sup>The thermostat degree of freedom must be included [Frenkel and Smit 1996, Martyna et al. 1994] in the count when working in virtual coordinates. Indeed in Eq. 3.3.62 we have  $g = N_f + 1$

of Eq. 3.3.5 does not lead to Eqs. 3.3.71–3.3.74. Simulations using the real variables are not Hamiltonian in nature in the sense that the phase space of the real variables is compressible [Tuckerman et al. 1997] and that Liouville theorem is not satisfied [Martyna et al. 1994]. This "strangeness" in the dynamics of the real variables in the extended systems does not of course imply that the sampling of the configurational real space is incorrect. To show this, it suffices to evaluate the partition function for a microcanonical distribution of the kind  $\delta(\mathcal{H} - -\mathcal{E})$ , with  $\mathcal{H}$  being given by Eq. 3.3.75. The Jacobian of the transformation of Eqs. 3.3.64–3.3.69 must be included in the integration with respect to the real coordinates when evaluating the partition function for the extended system. If the equations of motion in terms of the transformed coordinates are known, this Jacobian,  $\mathcal{J}$ , can be readily computed from the relation [Arnold 1989]:

$$\frac{d\mathcal{J}}{dt} = -\mathcal{J} \left( \frac{\delta}{\delta \mathbf{y}} \cdot \dot{\mathbf{y}} \right). \quad (3.3.78)$$

Where  $\mathbf{y}$  has the usual meaning of phase space vector containing all independent coordinates and momenta of the systems. Inserting the equations of motion of Eq. 3.3.74 into Eq. 3.3.78 and integrating by separation of variables yields

$$J = e^{Nf\eta} [\det \mathbf{h}] 6N : \quad (3.3.79)$$

Using 3.3.79 and integrating out the thermostat degrees of freedom, the partition function can be easily shown [Martyna et al. 1994, Melchionna et al. 1993] to be equivalent to that that of  $NPT$  ensemble, i.e.

$$\Delta_{NPT} \propto \int d\mathbf{h} e^{-\beta P_{\text{ext}} \det(\mathbf{h})} Q(\mathbf{h}) \quad (3.3.80)$$

with  $Q(\mathbf{h})$  being the canonical distribution of a system with cell of shape and size define by the columns of  $\mathbf{h}$ .<sup>8</sup>

---

<sup>8</sup>Actually in ref. [Martyna et al. 1994, Marchi and Procacci 1998] is pointed out that the virial theorem implied by the distribution is slightly different from the exact virial in the  $NPT$  ensemble. Martyna et al. [Martyna et al. 1994] proposed an improved set of equations of motion that generates a distribution satisfying exactly the virial theorem.

### Equivalence of Atomic and Molecular Pressure

The volume scaling defined in Eq. 3.3.51 is not unique. Note that only the equation of motion for the center of mass momentum, Eq. 3.3.73, has a velocity dependent term that depends on the coordinates of the barostat through the matrix  $\mathbf{G}$  defined in Eq. 3.3.60. The atomic momenta, Eq. 3.3.72, on the contrary, are not coupled to the barostat. This fact is also reflected in the equations of motion for the barostat momenta, Eq. 3.3.74, which is driven by the internal pressure due only to the molecular or group center of masses. In defining the extended Lagrangian one could as well have defined an atomic scaling of the form

$$r_{ik\alpha} = \sum_{\beta} h_{\alpha\beta} s_{i\alpha k}. \quad (3.3.81)$$

Atomic scaling might be trivially implemented by eliminating the kinetic energy, which depends on the  $l_i k \alpha$  velocities, from the starting Lagrangian 3.3.55 and replacing the term  $\frac{1}{2} \sum_i^N M_i s^2 \dot{\mathbf{S}}_i^t \mathbf{h}^t \mathbf{h} \dot{\mathbf{S}}_i$  with  $\frac{1}{2} \sum_{ik} m_{ik} s^2 \dot{\mathbf{s}}_{ik}^t \mathbf{h}^t \mathbf{h} \dot{\mathbf{s}}_{ik}$ . The corresponding equations of motions for atomic scaling are then

$$\dot{\mathbf{r}}_{ik} = \frac{\mathbf{p}_{ik}}{m_{ik}}, \quad \dot{\mathbf{h}} = \frac{\mathbf{P}_h}{W}, \quad \dot{\eta} = \frac{p_\eta}{Q} \quad (3.3.82)$$

$$\dot{\mathbf{p}}_{ik} = \mathbf{h}^{-1} \dot{\mathbf{p}}_{ik} - \mathbf{G}^{-1} \dot{\mathbf{G}} \dot{\mathbf{p}}_{ik} - \frac{p_\eta}{Q} \dot{\mathbf{p}}_{ik}, \quad (3.3.83)$$

$$\dot{\mathbf{P}}_h = \left( \mathcal{V} + \mathcal{K} - \mathbf{h}^{-1} P_{ext} \det \mathbf{h} \right) - \frac{p_\eta}{Q} \mathbf{P}_h, \quad (3.3.84)$$

$$\dot{p}_\eta = \mathcal{F}_\eta \quad (3.3.85)$$

where the quantities  $\mathcal{V}$ ,  $\mathcal{K}$ ,  $\mathcal{F}_\eta$  depend now on the atomic coordinates

$$\mathcal{V} = \sum_{i=1k}^N \mathbf{f}_{ik} \mathbf{s}_{ik}^t$$

$$\mathcal{K} = \sum_{i=1}^N M_i (\mathbf{h} \mathbf{s}_{ik}) \mathbf{s}_{ik}^t \quad (3.3.86)$$

$$\mathcal{F}_\eta = \frac{1}{2} \sum_{i=1}^N \sum_{k=1}^{n_i} \frac{\mathbf{p}_{ik}^t \mathbf{p}_{ik}}{m_{ik}} - g k_B T. \quad (3.3.87)$$

In case of atomic, Eq. 3.3.81, or molecular scaling, Eq. 3.3.51, the internal pressure entering in Eqs. 3.3.74,3.3.85 is then

$$P_{int} = \langle P_{atom} \rangle = \left\langle \frac{1}{3V} \sum_i \sum_k \left( \frac{\mathbf{p}_{ik}^2}{m_{ik}} + \mathbf{r}_{ik} \bullet \mathbf{f}_{ik} \right) \right\rangle \quad (3.3.88)$$

$$P_{int} = \langle P_{mol} \rangle = \left\langle \frac{1}{3V} \sum_i \left( \frac{\mathbf{P}_i^2}{M_i} + \mathbf{R}_i \bullet \mathbf{F}_i \right) \right\rangle \quad (3.3.89)$$

respectively. Where the molecular quantities can be written in term of the atomic counterpart according to:

$$\mathbf{R}_i = \frac{1}{M_i} \sum_k m_{ik} \mathbf{r}_{ik} \mathbf{P}_i = \sum_k \mathbf{p}_{ik} \mathbf{F}_i = \sum_k \mathbf{f}_{ik} \quad (3.3.90)$$

The equation of motion for the barostat in the two cases, Eqs.3.3.85,3.3.74, has the same form whether atomic or molecular scaling is adopted. The internal pressure in the former case is given by Eq. 3.3.89 and in the latter is given by Eq. 3.3.89. The two pressures, Eqs. 3.3.89,3.3.89, differ instantaneously. Should the difference persist after averaging, then it would be obvious that the equilibrium thermodynamic state in the *NPT* ensemble depends on the scaling method. The two formulae 3.3.89,3.3.89 are fortunately equivalent. To prove this statement, we closely follow the route proposed by H. Berendsen and reported by Ciccotti and Ryckaert [Ciccotti and Ryckaert 1986b] and use Eqs. 3.3.90–3.3.90 to rearrange Eq. 3.3.89. We obtain

$$\sum_i \langle \mathbf{R}_i \bullet \mathbf{F}_i \rangle = \sum_i \frac{1}{M_i} \sum_{kl} \langle m_{ik} \mathbf{r}_{ik} \bullet \mathbf{f}_{il} \rangle \quad (3.3.91)$$

Adding and subtracting  $m_{ik} \mathbf{r}_{il} \bullet \mathbf{f}_{il}$ , we get

$$= \sum_i \frac{1}{M_i} \sum_{kl} \langle m_{ik} (\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet \mathbf{f}_{il} + m_{ik} \mathbf{r}_{il} \bullet \mathbf{f}_{il} \rangle \quad (3.3.92)$$

which can be rearranged as

$$= \sum_i \frac{1}{M_i} \left\{ \sum_{kl} \left[ \frac{1}{2} \langle (\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet (m_i \mathbf{f}_{il} - m_j \mathbf{f}_{ik}) \rangle \right] + \sum_l \langle \mathbf{r}_{il} \bullet \mathbf{f}_{il} \rangle \right\} \quad (3.3.93)$$



using the newton law  $\mathbf{f}_{ik} = m_{ik}\mathbf{a}_{ik}$ , where  $\mathbf{a}_{ik}$  is the acceleration, we obtain

$$= \sum_i \frac{1}{M_i} \left\{ \sum_{kl} \left[ \frac{1}{2} \langle m_j m_i (\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet (\mathbf{a}_{il} - \mathbf{a}_{ik}) \rangle \right] + \sum_l \langle \mathbf{r}_{il} \bullet \mathbf{f}_{il} \rangle \right\}. \quad (3.3.94)$$

The first term in the above equation can be decomposed according to:

$$(\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet (\mathbf{a}_{il} - \mathbf{a}_{ik}) = \frac{d}{dt} [(\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet (\mathbf{v}_{il} - \mathbf{v}_{ik})] + (\mathbf{v}_{il} - \mathbf{v}_{ik})^2 \quad (3.3.95)$$

The first derivative term on the right hand side is zero rigorously for rigid molecules or rigid groups and is zero on average for flexible molecules or groups, assuming that the flexible molecules or groups do not dissociate. This can be readily seen in case of ergodic systems, by evaluating directly the average of this derivatives as

$$\left\langle \frac{d}{dt} [(\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet (\mathbf{v}_{il} - \mathbf{v}_{ik})] \right\rangle = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\infty \frac{d}{dt} [(\mathbf{r}_{ik} - \mathbf{r}_{il}) \bullet (\mathbf{v}_{il} - \mathbf{v}_{ik})] dt \quad (3.3.96)$$

$$= \lim_{\tau \rightarrow \infty} \frac{1}{\tau} [(\mathbf{r}_{ik}(\tau) - \mathbf{r}_{il}(\tau)) \bullet (\mathbf{v}_{il}(\tau) - \mathbf{v}_{ik}(\tau)) + C] \quad (3.3.97)$$

So if the quantity  $\mathbf{r}_{ikl}(\tau)\mathbf{v}_{ilk}(\tau)$  remains bounded (which is true if the potential is not dissociative, since  $k, l$  refers to the same molecule  $i$ ), the average in Eq. 3.3.97 is zero<sup>9</sup>. Thus, we can rewrite the average of Eq. 3.3.94 as

$$\left\langle \sum_i \mathbf{R}_i \bullet \mathbf{F}_i \right\rangle = \sum_i \frac{1}{2M_i} \sum_{kl} m_{ik} m_{il} \langle (\mathbf{v}_{il} - \mathbf{v}_{ik})^2 \rangle + \sum_{ik} \langle \mathbf{r}_{ik} \bullet \mathbf{f}_{ik} \rangle. \quad (3.3.98)$$

The first term on the right hand side of the above equation can be further developed obtaining the trivial identity:

$$\sum_{kl} m_{ik} m_{il} \langle (\mathbf{v}_{il} - \mathbf{v}_{ik})^2 \rangle = \sum_{kl} m_{ik} m_{il} \langle \mathbf{v}_{ik}^2 \rangle + \sum_{kl} m_{ik} m_{il} \langle \mathbf{v}_{il}^2 \rangle -$$

---

<sup>9</sup>The statement the molecule of group does not dissociate is even too restrictive. It is enough to say that the quantity 3.3.97 remains bound.

$$- \sum_{kl} 2m_{ik}m_{il} \langle \mathbf{v}_{il} \bullet \mathbf{v}_{ik} \rangle \quad (3.3.99)$$

$$= 2M_i \sum_k m_{ik} \langle \mathbf{v}_{ik}^2 \rangle - 2 \langle \mathbf{P}_i^2 \rangle \quad (3.3.100)$$

Substituting Eq. 3.3.100 in Eq. 3.3.98 we get

$$\left\langle \sum_i \mathbf{R}_i \bullet \mathbf{F}_i \right\rangle = \sum_{ik} m_{ik} \langle \mathbf{v}_{ik}^2 \rangle - \frac{1}{M_i} \langle \mathbf{P}_i^2 \rangle + \sum_{ik} \langle \mathbf{r}_{ik} \bullet \mathbf{f}_{ik} \rangle \quad (3.3.101)$$

Substituting Eq. 3.3.101 into Eq. 3.3.89 leads speedily to 3.3.89 which completes the proof. As a consequence of the above discussion, it seems likely that both the equilibrium and non equilibrium properties of the MD system are not affected by coordinate scaling. We shall see later that this is actually the case.

### Liouvillean Split and Multiple Time Step Algorithm for the *NPT* Ensemble

We have seen in section 3.3.1 that the knowledge of the Liouvillean allows us to straightforwardly derive a multi-step integration algorithm. Thus, for simulation in the *NPT* ensemble, the Liouvillean  $iL = \dot{\mathbf{y}} \nabla_{\mathbf{y}}$  is readily available from the equations of motion in 3.3.71–3.3.74. For sake of simplicity, to build our *NPT* multiple time step integrator we assume that the system potential contains only a fast intramolecular  $V_0$  term and a slow intermolecular term  $V_1$ , as discussed in Sec. 3.3.1. Generalisation to multiple intra and inter-molecular components is straightforward.

We define the following components of the *NPT* Liouvillean

$$iL_x = - \sum_i \mathbf{P}_i \frac{p_\eta}{Q} \nabla_{\mathbf{P}_i} - \sum_{ik} \mathbf{P}_{ik} \frac{p_\eta}{Q} \nabla_{\mathbf{P}_{ik}} - \sum_{\alpha\beta} (\mathbf{P}_h)_{\alpha\beta} \frac{p_\eta}{Q} (\nabla_{\mathbf{P}_h})_{\alpha\beta} \quad (3.3.102)$$

$$iL_y = \mathcal{F}_\eta \nabla_{p_\eta} \quad (3.3.103)$$

$$iL_z = \sum_i -\mathbf{G}^{-1} \dot{\mathbf{G}} \mathbf{P}_i \nabla_{\mathbf{P}_i} \quad (3.3.104)$$

$$iL_u = \sum_{\alpha\beta} \left( \mathcal{K} - \mathbf{h}^{-1} P_{ext} \det(\mathbf{h}) \right)_{\alpha\beta} (\nabla_{\mathbf{P}_h})_{\alpha\beta} \quad (3.3.105)$$

$$iL_s = \sum_i \mathbf{J}_i \nabla_{\mathbf{P}_i} + \sum_{ik} \mathbf{f}_{ik}^c \nabla_{\mathbf{p}_{ik}} + \sum_{\alpha\beta} (\mathcal{V})_{\alpha\beta} (\nabla_{\mathbf{P}_h})_{\alpha\beta} \quad (3.3.106)$$

$$\begin{aligned} iG_0 &= \sum_i \frac{\mathbf{P}_i}{M_i} \nabla_{\mathbf{s}_i} + \sum_{ik} \frac{\mathbf{p}_{ik}}{m_{ik}} \nabla_{\mathbf{l}_{ik}} + \sum_{\alpha\beta} \frac{(\mathbf{P}_h)_{\alpha\beta}}{W} (\nabla_{\mathbf{h}})_{\alpha\beta} + \\ &+ \frac{p_\eta}{Q} \nabla_\eta - \nabla_{\mathbf{l}_{ik}} V_0 \nabla_{\mathbf{p}_{ik}}, \end{aligned} \quad (3.3.107)$$

where in Eq. 3.3.107 the scaled forces  $\mathbf{F}_i$  have been replaced by its real space counterparts, i.e.  $\mathbf{J}_i = \mathbf{h}^{-1} \mathbf{F}_i$ .

The atomic scaling version of this Liouvillean breakup is derived on the basis of Eqs. 3.3.85. One obtains

$$iL_x = - \sum_{ik} \mathbf{p}_{ik} \frac{p_\eta}{Q} \nabla_{\mathbf{p}_{ik}} - \sum_{\alpha\beta} (\mathbf{P}_h)_{\alpha\beta} \frac{p_\eta}{Q} (\nabla_{\mathbf{P}_h})_{\alpha\beta} \quad (3.3.108)$$

$$iL_y = \mathcal{F}_\eta \nabla_{p_\eta} \quad (3.3.109)$$

$$iL_z = \sum_{ik} -\mathbf{G}^{-1} \dot{\mathbf{G}} \mathbf{p}_{ik} \nabla_{\mathbf{p}_{ik}} \quad (3.3.110)$$

$$iL_u = \sum_{\alpha\beta} (\mathcal{K} - \mathbf{h}^{-1} P_{ext} \det \mathbf{h})_{\alpha\beta} (\nabla_{\mathbf{P}_h})_{\alpha\beta} \quad (3.3.111)$$

$$iL_s = \sum_{ik} \mathbf{j}_{ik} \nabla_{\mathbf{p}_{ik}} + \sum_{\alpha\beta} (\mathcal{V})_{\alpha\beta} (\nabla_{\mathbf{P}_h})_{\alpha\beta} \quad (3.3.112)$$

$$\begin{aligned} iG_0 &= \sum_{ik} \frac{\mathbf{p}_{ik}}{m_{ik}} \nabla_{\mathbf{l}_{ik}} + \sum_{\alpha\beta} \frac{(\mathbf{P}_h)_{\alpha\beta}}{W} (\nabla_{\mathbf{h}})_{\alpha\beta} + \\ &+ \frac{p_\eta}{Q} \nabla_\eta - \nabla_{\mathbf{l}_{ik}} V_0 \nabla_{\mathbf{p}_{ik}}, \end{aligned} \quad (3.3.113)$$

where  $\mathbf{j}_{ik} = \mathbf{h}^{-1} \mathbf{f}_{ik}$  and  $\mathcal{V}$ ,  $\mathcal{K}$ ,  $\mathcal{F}_\eta$  are given in Eqs. 3.3.87, 3.3.87. For the time scale breakup in the *NPT* ensemble we have the complication of the extra degrees of freedom whose time scale dynamics can be controlled by varying the parameter  $Q$  and  $W$ . Large values of  $Q$  and  $W$  slow down the time dynamics of the barostat and thermostat coordinates. The potential  $V$  determines the time scale of the  $iG_0$  term (the fast component) and of the  $iL_s$  contribution (the slow component). All other sub-Liouvilleans either handle the coupling of the true coordinates to the extra degrees of freedom, or drive the evolution of the extra coordinates of the barostat and thermostat ( $iL_y$  and  $iL_u$ ). The time scale dynamics of these terms depends not only on the potential subdivision and on the parameters  $W$  and  $Q$ , but also on the type

of scaling [Marchi and Procacci 1998]. When the molecular scaling is adopted the dynamics of the virial term  $V$  contains contributions only from the intermolecular potential since the barostat is coupled only to the center of mass coordinates (see Eq. 3.3.76). Indeed, the net force acting on the molecular center of mass is independent on the intramolecular potential, since the latter is invariant under rigid translation of the molecules. When atomic scaling or group (i.e. sub-molecular) scaling is adopted, the virial  $V$  (see Eq. 3.3.87) depends also on the fast intramolecular such as stretching motions. In this case the time scale of the barostat coordinate is no longer slow, unless the parameter  $W$  is changed. For standard values of  $W$ , selected to obtain an efficient sampling of the  $NPT$  phase space [Paci and Marchi 1996, Nosé and Klein 1983], the barostat dependent Liouvilleans, Eqs. 3.3.106, 3.3.105, have time scale dynamics comparable to that of the intramolecular Liouvillean  $iG_0$  and therefore must be associated with this term.

Thus, the molecular split of the Liouvillean is hence given by

$$\begin{aligned} iL_1 &= iL_x + iL_y + iL_z + iL_u + iL_s \\ iL_0 &= iG_0 \end{aligned} \quad (3.3.114)$$

whereas the atomic split is

$$\begin{aligned} iL_1 &= iL_x + iL_y + iL_s \\ iL_0 &= iG_0 + iL_z + iL_u \end{aligned} \quad (3.3.115)$$

For both scaling, a simple Hermitian factorisation of the total time propagator  $e^{iLt}$  yields the double time discrete propagator

$$e^{iL_1+iL_0} = e^{iL_1\Delta t_1/2} \left( e^{iL_0\Delta t_0} \right)^n e^{iL_1\Delta t_1/2} \quad (3.3.116)$$

where  $\Delta t_0$ , the small time step, must be selected according to the intramolecular time scale whereas  $\Delta t_1$ , the large time step, must be selected according to the time scale of the intermolecular motions. We already know that the propagator 3.3.116 cannot generate a symplectic. In this case the symmetric form of the multiple time step propagator Eq. 3.3.116 does not imply necessarily time reversibility. Some operators appearing in the definition of  $L_1$  (e.g.  $iL_z$  and  $iL_s$ ) for the molecular scaling and in the definitions of  $iL_1$  and  $iL_0$  for the atomic scaling are in fact non commuting. We have

seen in Sec. 3.3.1 that first order approximation of non commuting propagators yields time irreversible algorithms. We can render the propagator in Eq. 3.3.116 time reversible by using second order symmetric approximant (i.e. Trotter approximation) for any two non commuting operators. For example in the case of the molecular scaling, when we propagate in Eq. 3.3.116 the slow propagator  $e^{iL_1\Delta t/2}$  for half time step, we may use the following second order  $\mathcal{O}(\Delta t^3)$  split

$$e^{iL_1\frac{\Delta t_1}{2}} \simeq e^{iL_y\frac{\Delta t_1}{4}} e^{iL_z\frac{\Delta t_1}{2}} e^{iL_y\frac{\Delta t_1}{4}} e^{iL_x\frac{\Delta t_1}{4}} e^{i(L_s+L_u)\frac{\Delta t_1}{2}} e^{iL_x\frac{\Delta t_1}{4}} \quad (3.3.117)$$

An alternative simpler and equally accurate approach when dealing with non commuting operators is simply to preserve the unitarity by reversing the order of the operators in the first order factorisation of the right and left operators of Eq. 3.3.116 without resorting to locally second order  $\mathcal{O}(\Delta t^3)$  approximation like in Eq. 3.3.117. Again for the molecular scaling, this is easily done by using the approximant

$$\left( e^{iL_1\frac{\Delta t_1}{2}} \right)_{left} = e^{iL_x\frac{\Delta t_1}{2}} e^{iL_y\frac{\Delta t_1}{2}} e^{iL_z\frac{\Delta t_1}{2}} e^{iL_u\frac{\Delta t_1}{2}} e^{iL_s\frac{\Delta t_1}{2}} \quad (3.3.118)$$

for the left propagator, and

$$\left( e^{iL_1\frac{\Delta t_1}{2}} \right)_{right} = e^{iL_s\frac{\Delta t_1}{2}} e^{iL_u\frac{\Delta t_1}{2}} e^{iL_z\frac{\Delta t_1}{2}} e^{iL_y\frac{\Delta t_1}{2}} e^{iL_x\frac{\Delta t_1}{2}} \quad (3.3.119)$$

for the rightmost propagator. Note that

$$\left( e^{iL_1\frac{\Delta t_1}{2}} \right)_{left}^{-1} = \left( e^{iL_1\frac{\Delta t_1}{2}} \right)_{right} \quad (3.3.120)$$

Inserting these approximations into 3.3.116 the overall integrator is found to be time reversible and second order. Time reversible integrators are in fact always even order and hence at least second order [Channel and Scovel 1990, Sanz-Serna 1992]. Therefore the overall molecular and atomic (or group) discrete time propagators are given by

$$\begin{aligned} e^{iL_{mol}\Delta t_1} &= e^{iL_s\frac{\Delta t_1}{2}} e^{iL_u\frac{\Delta t_1}{2}} e^{iL_z\frac{\Delta t_1}{2}} e^{iL_y\frac{\Delta t_1}{2}} e^{iL_x\frac{\Delta t_1}{2}} \left( e^{iG_0\Delta t_0} \right)^n \times \\ &\times e^{iL_x\frac{\Delta t_1}{2}} e^{iL_y\frac{\Delta t_1}{2}} e^{iL_z\frac{\Delta t_1}{2}} e^{iL_u\frac{\Delta t_1}{2}} e^{iL_s\frac{\Delta t_1}{2}} \quad (3.3.121) \\ e^{iL_{atom}\Delta t_1} &= e^{iL_s\frac{\Delta t_1}{2}} e^{iL_y\frac{\Delta t_1}{2}} e^{iL_x\frac{\Delta t_1}{2}} \times \end{aligned}$$

$$\begin{aligned} & \times \left( e^{iL_u \Delta t_0/2} e^{iL_z \Delta t_0/2} e^{iG_0 \Delta t_0} e^{iL_z \Delta t_0/2} e^{iL_u \Delta t_0/2} \right)^n \times \\ & \times e^{iL_x \frac{\Delta t_1}{2}} e^{iL_y \frac{\Delta t_1}{2}} e^{iL_s \frac{\Delta t_1}{2}} \end{aligned} \quad (3.3.122)$$

The propagator  $e^{iG_0 \Delta t_0}$ , defined in Eq. 3.3.107, is further split according to the usual velocity Verlet breakup of Eq. 3.3.26. Note that in case of molecular scaling the "slow" coordinates ( $\mathbf{S}; \mathbf{h}; \eta$ ) move with constant velocity during the  $n$  small time steps since there is no "fast" force acting on them in the inner integration. The explicit integration algorithm may be easily derived for the two propagators in Eqs. 3.3.122 and 3.3.122 using the rule in Eq. 3.3.27 and its generalisation:

$$\begin{aligned} e^{a\mathbf{y}\nabla_{\mathbf{y}}} f(\mathbf{y}) &= f(\mathbf{y}e^{\mathbf{a}}) \\ e^{\mathbf{a}\mathbf{y}\nabla_{\mathbf{y}}} f(\mathbf{y}) &= f(e^{\mathbf{a}}\mathbf{y}) \end{aligned} \quad (3.3.123)$$

Where  $a$  and  $\mathbf{a}$  are a scalar and a matrix, respectively. The exponential matrix  $e^{\mathbf{a}}$  on the right hand side of Eq. 3.3.123 is obtained by diagonalization of  $\mathbf{a}$ .

As stated before the dynamics generated by Eqs. 3.3.71–3.3.74 or 3.3.82–3.3.85 in the *NPT* ensemble is not Hamiltonian and hence we cannot speak of symplectic integrators [Toxvaerd 1993] for the  $t$ -flow's defined by Eqs. 3.3.122, 3.3.122. The symplectic condition Eq. 3.3.12 is violated at the level of the transformation 3.3.64–3.3.69 which is not canonical. However, the algorithms generated by Eqs. 3.3.122, 3.3.122 are time reversible and second order like the velocity Verlet. Several recent studies have shown [Procacci and Berne 1994b, Martyna et al. 1996, Marchi and Procacci 1998] that these integrators for the non microcanonical ensembles are also stable for long time trajectories, as in case of the symplectic integrators for the *NVE* ensemble.

### Group Scaling and Molecular Scaling

We have seen in Sec. 3.3.2 that the center of mass or molecular pressure is equivalent to the atomic pressure. The atomic pressure is the natural quantity that enters in the virial theorem [Allen and Tildesley 1989] irrespectively of the form of the interaction potential among the particles. So, in principle it is safer to adopt atomic scaling in the extended system constant pressure simulation. For systems in confined regions, the equivalence between atomic or true pressure and molecular pressure (see Sec. 3.3.2) holds for any definition of the molecular subsystem irrespectively of the interaction potentials.

In other words we could have defined virtual molecules made up of atoms selected on different real molecules. We may expect that, as long as the system, no matter how its unities or particles are defined, contains a sufficiently large number of particles, generates a distribution function identical to that generated by using the "correct" atomic scaling. From a computational standpoint molecular scaling is superior to atomic scaling. The fastly varying Liouvillean in Eq. 3.3.115 for the atomic scaling contains the two terms  $iL_z, iL_u$ . These terms are slowly varying when molecular scaling is adopted and are assigned to the slow part of the Liouvillean in Eq. 3.3.114. The inner part of the time propagation is therefore expected to be more expensive for the multiple time step integration with atomic scaling rather than with molecular scaling. Generally speaking, given the equivalence between the molecular and atomic pressure, molecular scaling should be the preferred choice for maximum efficiency in the multiple time step integration.

For large size molecules, such as proteins, molecular scaling might be inappropriate. The size of the molecule clearly restricts the number of particles in the MD simulation box, thereby reducing the statistics on the instantaneous calculated molecular pressure which may show nonphysically large fluctuations. Group scaling [Marchi and Procacci 1998] is particularly convenient for handling the simulation of macromolecules. A statistically significant number of groups can be selected in order to avoid all problems related to the poor statistics on molecular pressure calculation for samples containing a small number of large size particles. Notwithstanding, for solvated biomolecules and provided that enough solvent molecules are included, molecular scaling again yields reliable results. In Ref. [Marchi and Procacci 1998] Marchi and Procacci showed that the scaling method in the *NPT* ensemble does not affect neither the equilibrium structural and dynamical properties nor the kinetic of non equilibrium MD. For group-based and molecular-based scaling methods in a system of one single molecule of BPTI embedded in a box of about a 1000 water molecules, they obtained identical results for the system volume, the Voronoi volumes of the proteins and for the mean square displacement of both solvent and protein atoms under normal and high pressure.

### Switching to Other Ensembles

The *NPT* extended system is the most general among all possible extended Lagrangians. All other ensemble can be in fact obtained within the same

computational framework. We must stress [Marchi and Procacci 1998] that the computational overhead of the extended system formulation, due to the introduction and handling of the extra degrees of freedom of the barostat and thermostat variables, is rather modest and is negligible with respect to a  $NVE$  simulation for large samples ( $N_f > 2000$ ) [Procacci and Berne 1994b, Martyna et al. 1996, Marchi and Procacci 1998]. Therefore, a practical, albeit inelegant way of switching among ensembles is simply to set the inertia of the barostat and/or thermostat to a very large number. This must be of course equivalent to decouple the barostat and/or the thermostat from the true degrees of freedom. In fact, by setting  $W$  to infinity<sup>10</sup> in Eqs. 3.3.71–3.3.74 we recover the  $NVT$  canonical ensemble equations of motion. Putting instead  $Q$  to infinity the  $NPH$  equations of motion are obtained. Finally, setting both  $W$  and  $Q$  to infinity the  $NVE$  equations of motion are recovered.

Switching to the  $NPT$  isotropic stress ensemble is less obvious. One may define the kinetic term associated to barostat in the extended Lagrangian as

$$K = \frac{1}{2} \sum_{\alpha\beta} W_{\alpha\beta} s^2 \dot{h}_{\alpha\beta}^2 \quad (3.3.124)$$

such that a different inertia may in principle be assigned to each of 9 extra degrees of freedom of the barostat. Setting for example

$$W_{\alpha\beta} = W \quad \text{for } \alpha \leq \beta \quad (3.3.125)$$

$$W_{\alpha\beta} = \infty \quad \text{for } \alpha > \beta \quad (3.3.126)$$

one inhibits cell rotations [Marchi and Procacci 1998].

This trick does not work, unfortunately, to change to isotropic stress tensor. In this case, there is only one independent barostat degrees of freedom, namely the volume of the system. In order to simulate isotropic cell fluctuations a set of five constraints on the  $h$  matrix are introduced which correspond to the conditions:

$$\frac{h_{\alpha\beta}}{h_{11}} - \frac{h_{\alpha\beta}^0}{h_{11}^0} = 0$$

---

<sup>10</sup>The value of  $W$  which works as "infinity" depends on the "force" that is acting on barostat coordinate expressed by the Eq. (4.25), i.e. on how far the system is from the thermodynamic equilibrium. For a system near the thermodynamic equilibrium with  $N_f \simeq 10000$  a value of  $W = 1020$  a.m.u. is sufficient to prevent cell fluctuations.



$$\dot{h}_{\alpha\beta} - \frac{h_{\alpha\beta}^0}{h_{11}^0} \dot{h}_{11} = 0 \quad \text{for } \alpha \leq \beta \quad (3.3.127)$$

with  $\mathbf{h}_0$  being some reference  $\mathbf{h}$  matrix. These constraints are implemented naturally in the framework of the multi time step velocity Verlet using the RATTLE algorithm which evaluates iteratively the constraints force to satisfy the constraints on both coordinates  $\mathbf{h}$  and velocities  $\dot{\mathbf{h}}$  [Marchi and Procacci 1998]. In Ref. [Marchi and Procacci 1998] it is proved that the phase space sampled by the *NPT* equations with the addition of the constraints Eq. 3.3.127 correspond to that given by *NPT* distribution function.

### 3.3.3 Multiple Time Steps Algorithms For Large Size Flexible Systems with Strong Electrostatic Interactions

In the previous sections we have described how to obtain multiple time step integrators given a certain potential subdivision and have provided simple examples of potential subdivision, based on the inter/intra molecular separation. Here, we focus on the time scale separation of model potentials of complex molecular systems. Additionally, we provide a general potential subdivision applying to biological systems, as well as to many other interesting chemical systems including liquid crystals. These systems are typically characterised by high flexibility and strong Coulombic intermolecular interactions. Schematically, we can then write the potential  $V$  as due to two contributions:

$$V = V_{bnd} + V_{nbn}. \quad (3.3.128)$$

Here, the "bonded" or intramolecular part  $V_{bnd}$  is fast and is responsible for the flexibility of the system. The "non bonded" or intermolecular (or inter-group) term  $V_{nbn}$  is dominated by Coulombic interactions. The aim of the following sections is to describe a general protocol for the subdivision of such forms of the interaction potential and to show how to obtain reasonably efficient and transferable multiple time step integrators valid for any complex molecular system.

### Subdivision of the "Bonded" Potential

As we have seen in Sec. 3.3.1 the idea behind the multiple time step scheme is that of the reference system which propagates for a certain amount of time under the influence of some unperturbed reference Hamiltonian, and then undergoes an impulsive correction brought by the remainder of the potential. The exact trajectory spanned by the complete Hamiltonian is recovered by applying this impulsive correction onto the "reference" trajectory. We have also seen in the same section that, by subdividing the interaction potential, we can determine as many "nested" reference systems as we wish. The first step in defining a general protocol for the subdivision of the bonded potential for complex molecular systems consists in identifying the various time scales and their connection to the potential. The interaction bonded potential in almost all popular force fields is given as a function of the stretching, bending and torsion internal coordinates and has the general form

$$V_{bnd} = V_{stretch} + V_{bend} + V_{tors}, \quad (3.3.129)$$

where

$$\begin{aligned} V_{stretch} &= \sum_{Bonds} K_r (r - r_0)^2 \\ V_{bend} &= \sum_{Angles} K_\theta (\theta - \theta_0)^2 \\ V_{tors} &= \sum_{Dihedrals} V_\Phi [1 + \cos(n\Phi - \gamma)]. \end{aligned} \quad (3.3.130)$$

Here,  $K_r$  and  $K_\theta$  are the bonded force constants associated with bond stretching and angles bending respectively, while  $r_0$  and  $\theta_0$  are their respective equilibrium values. In the torsional potential,  $V_{tors}$ ,  $\Phi$  is the dihedral angle, while  $K_\Phi$ ,  $n$  and  $\gamma$  are constants.

The characteristic time scale of a particular internal degrees of freedom can be estimated assuming that this coordinate behaves like a harmonic oscillator, uncoupled from the rest the other internal degrees of freedom. Thus, the criterion for guiding the subdivision of the potential in Eq. 3.3.129 is given by the characteristic frequency of this uncoupled oscillator. We now give, for each type of degree of freedom, practical formula to evaluate the harmonic frequency from the force field constants given in Eq. 3.3.130.

**Stretching:** The stretching frequencies are given by the well known expression

$$\nu_s = \frac{1}{2\pi} \left( \frac{K_r}{\mu} \right)^{\frac{1}{2}} \quad (3.3.131)$$

where  $\mu$  is reduced mass.

**Bending:** We shall assume for the sake of simplicity that the uncoupled bending frequencies depends on the masses of the atom 1 and 3 (see Fig. 3.3.3), that is mass 2 is assumed to be infinity. This turns out to be in general an excellent approximation for bending involving hydrogens and a good approximation for external bendings in large molecules involving masses of comparable magnitude. The frequency is obtained by writing the Lagrangian in polar coordinates for the mechanical system depicted in Fig. 3.3.3. The Cartesian coordinates are expressed in terms of the polar coordinates as

$$x_1 = r_{12} \sin(\alpha/2) \quad y_1 = r_{12} \cos(\alpha/2) \quad (3.3.132)$$

$$x_3 = r_{32} \sin(\alpha/2) \quad y_3 = r_{32} \cos(\alpha/2) \quad (3.3.133)$$

where the distance  $r_{32}$  and  $r_{12}$  are constrained to the equilibrium values. The velocities are then

$$\dot{x}_1 = -r_{12} \cos(\alpha/2) \frac{\dot{\alpha}}{2} \quad \dot{y}_1 = r_{12} \sin(\alpha/2) \frac{\dot{\alpha}}{2} \quad (3.3.134)$$

$$\dot{x}_3 = -r_{32} \cos(\alpha/2) \frac{\dot{\alpha}}{2} \quad \dot{y}_3 = r_{32} \sin(\alpha/2) \frac{\dot{\alpha}}{2} \quad (3.3.135)$$

The Lagrangian for the uncoupled bending is then

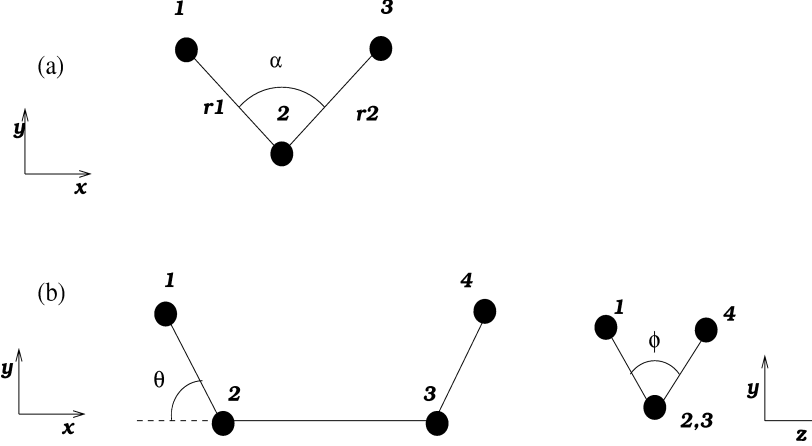
$$\mathcal{L} = \frac{1}{2} (m_1 \dot{x}_1^2 + m_1 \dot{y}_1^2 + m_3 \dot{x}_3^2 + m_3 \dot{y}_3^2) + V_{bend} \quad (3.3.136)$$

$$= \frac{1}{8} (m_1 r_{12}^2 + m_3 r_{32}^2) \dot{\alpha}^2 + \frac{1}{2} k_\theta (\alpha - \alpha_0)^2. \quad (3.3.137)$$

The equation of motion  $\frac{d}{dt} \frac{\delta \mathcal{L}}{\delta \dot{\alpha}} - \frac{\delta \mathcal{L}}{\delta \alpha} = 0$  for the  $\alpha$  coordinate is given by

$$\ddot{\alpha} + \frac{4K_\theta}{I_b} (\alpha - \alpha_0)^2 = 0. \quad (3.3.138)$$

Figure 3.4: Bending and dihedral angles



Where,  $I_b = m_1 r_{12}^2 + m_3 r_{32}^2$  is the moment of inertia about an axis passing by atom 3 and perpendicular to the bending plane. Finally, the uncoupled bending frequency is given by

$$\nu_b = \frac{1}{2\pi} \left( \frac{4K_\theta}{m_1 r_{12}^2 + m_3 r_{32}^2} \right)^{\frac{1}{2}} \quad (3.3.139)$$

**Torsion:** We limit our analysis to a purely torsional system (see Fig. 3.4) where atoms 2 and 3 are held fixed, and all bond distances and the angle  $\theta$  are constrained to their equilibrium values. The system has only one degree of freedom, the dihedral angle  $\Phi$  driven by the torsional potential  $V_\Phi$ . Again we rewrite the kinetic energy in terms of the bond distances, the dihedral angle and the constant bend angle  $\theta$ . For the kinetic energy, the only relevant coordinates are now those of atoms 1 and 4:

$$\begin{aligned} x_1 &= d_{12} \cos \theta + \frac{d_{23}}{2} & x_4 &= d_{34} \cos \theta + \frac{d_{23}}{2} \\ y_1 &= d_{12} \sin \theta \cos(\Phi/2) & y_4 &= d_{34} \sin \theta \cos(\Phi/2) \\ z_1 &= d_{12} \sin \theta \sin(\Phi/2) & z_4 &= d_{34} \sin \theta \sin(\Phi/2). \end{aligned} \quad (3.3.140)$$

The Lagrangian in terms of the dihedral angle coordinate is then

$$\mathcal{L} = \frac{1}{8} I_t \dot{\Phi}^2 - V_\Phi [1 + \cos(n\Phi - \gamma)], \quad (3.3.141)$$

where

$$I_t = \sin^2 \theta (m_1 d_{12}^2 + m_4 d_{34}^2). \quad (3.3.142)$$

Assuming small oscillations, the potential may be approximated by a second order expansion around the corresponding equilibrium dihedral angle  $\Phi_0$

$$V_{tors} = \frac{1}{2} \left( \frac{\delta^2 V_{tors}}{\delta \Phi^2} \right)_{\Phi=\Phi_0} (\Phi - \Phi_0)^2 = \frac{1}{2} V_\Phi n^2 (\Phi - \Phi_0)^2 \quad (3.3.143)$$

Substituting 3.3.143 into Eq. 3.3.141 and then writing the Lagrange equation of motion for the coordinate  $\Phi$ , one obtains again a differential equation of a harmonic oscillator, namely

$$\ddot{\Phi} + \frac{4V_\Phi n^2}{I_t} (\Phi - \Phi_0) = 0. \quad (3.3.144)$$

Thus, the uncoupled torsional frequency is given by

$$\nu_t = \frac{n}{2\pi} \left( 4 \frac{V_\Phi}{\sin^2 \theta (m_1 d_{12}^2 + m_4 d_{34}^2 \Delta)} \right)^{\frac{1}{2}}. \quad (3.3.145)$$

For many all-atom force fields, improper torsions [Wiener et al. 1986, van Gunsteren and Berendsen 1987] are modelled using a potential identical to that of the proper torsion in Eq. 3.3.130 and hence in these cases Eq. 3.3.145 applies also to the improper torsion uncoupled frequency, provided that indices 1 and 4 refer to the lighter atoms. In Fig. 3.5 we report the distribution of frequencies for the hydrated protein Bovine Pancreatin Trypsin Inhibitor (BPTI) using the AMBER [Cornell et al. 1995] force field. The distributions might be thought as a density of the uncoupled intramolecular states of the system. As we can see in the figure there is a relevant degree of overlap for the various internal degrees of freedom. For example, "slow" degrees of freedom such as torsions may be found up to 600 wavenumber, well inside the "bending" region; these are usually improper or proper torsions involving hydrogens. It is then inappropriate to assign such "fast" torsions involving hydrogens to a slow reference system. We recall that in a multiple time simulation the integration of a supposedly slow degree of freedom with a excessively large time step is enough to undermine the entire simulation. In

Figure 3.5: Density of the uncoupled (see text) states for stretching, bending, proper and improper torsion obtained with the AMBER force field on the protein bovine pancreatic trypsin inhibitor (BPTI). Frequencies were calculated according to Eqs. 3.3.131,3.3.139,3.3.145.

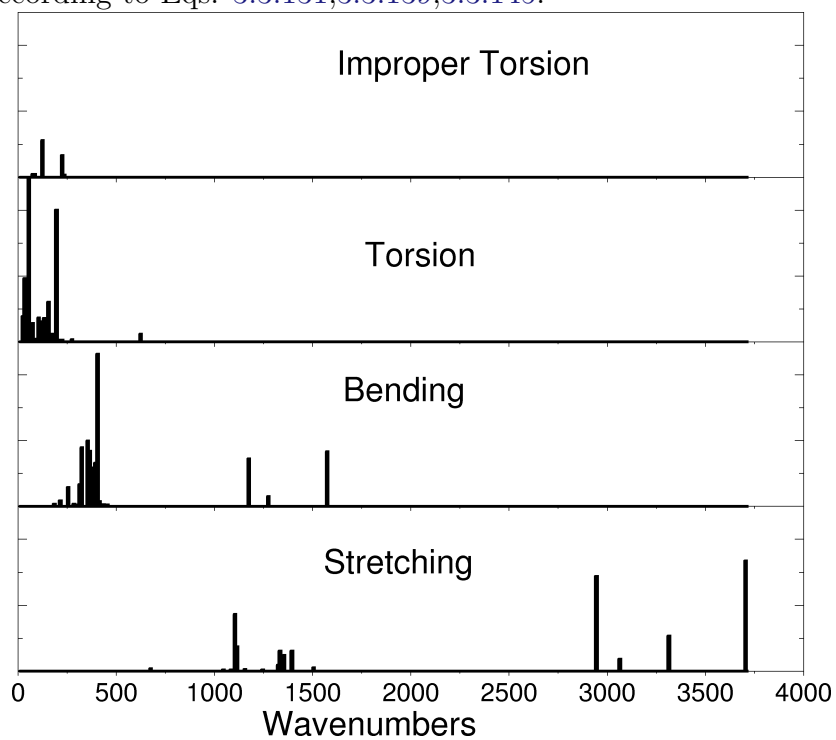


Fig. 3.3.4 we also notice that almost all the proper torsions fall below  $350\text{ cm}^{-1}$ . An efficient and simple separation of the intramolecular AMBER potential [Marchi and Procacci 1998] assigns all bendings stretching and the improper or proper torsions involving hydrogens to a "fast" reference system labelled  $n_0$  and all proper torsions to a slower reference system labeled  $n_1$ . The subdivision is then

$$\begin{aligned} V_{n_0} &= V_{stretch} + V_{bend} + V_{i-tors} + V_{p-tors}^{(h)} \\ V_{n_1} &= V_{p-tors} \end{aligned} \quad (3.3.146)$$

Where with  $V_{p-tors}^{(h)}$  we indicate proper torsions involving hydrogens. For the reference system  $V_{n_0}$ , the hydrogen stretching frequencies are the fastest motions and the  $\Delta t_{n_0}$  time step must be set to 0.2–0.3 fs. The computational burden of this part of the potential is very limited, since it involves mostly two or three body forces. For the reference system  $V_{n_1}$ , the fastest motion is around  $300\text{ cm}^{-1}$  and the time step  $\Delta t_{n_1}$  should be set to 1–1.5 fs. The computational effort for the reference system potential  $V_{n_1}$  is more important because of the numerous proper torsions of complex molecular systems which involve more expensive four body forces calculations. One may also notice that some of the bendings which were assigned to the  $n_0$  reference system fall in the torsion frequency region and could be therefore integrated with a time step much larger than  $\Delta t_{n_0} \simeq 0.2 - 0.3$ . However, in a multiple time step integration, this overlap is just inefficient, but certainly not dangerous.

### 3.3.4 The smooth particle mesh Ewald method

Before we discuss the non bonded multiple time step separation it is useful to describe in some details one of the most advanced techniques to handle long range forces. Indeed, this type of non bonded forces are the most cumbersome to handle and deserve closer scrutiny.

In the recent literature, a variety of techniques are available to handle the problem of long range interactions in computer simulations of charged particles at different level of approximation [Barker and Watts 1973, Barker 1980, Allen and Tildesley 1989]. In this section, we shall focus on the Ewald summation method for the treatment of long range interactions in periodic systems [Ewald 1921, deLeeuw et al. 1980, Hansen 1986]. The Ewald method gives the exact result for the electrostatic energy of a periodic system consisting of an infinitely replicated neutral box of charged particles. The method

is the natural choice in MD simulations of complex molecular system with PBC.

The Ewald potential [deLeeuw et al. 1980] is given by the sum of two terms, one in the direct lattice:

$$V'_{qd} = \frac{1}{2} \sum_{ij}^N q_i q_j \sum_n \frac{1}{|r_{ij} + r_n|} \operatorname{erfc}(\alpha|r_{ij} + r_n|) \quad (3.3.147)$$

and the other in the reciprocal lattice:

$$V_{qr} = \left[ \frac{1}{2\pi V} \sum_{\mathbf{m} \neq 0}^{\infty} \frac{\exp(-\pi^2|\mathbf{m}|^2/\alpha^2)}{\mathbf{m}^2} S(\mathbf{m})S(-\mathbf{m}) - \frac{\alpha}{\pi^{1/2}} \sum_i q_i^2 \right] - V_{intra}. \quad (3.3.148)$$

with

$$S(\mathbf{m}) = \sum_i^N q_i e^{(2\pi i \mathbf{m} \cdot \mathbf{r}_i)} \quad (3.3.149)$$

$$V_{intra} = \sum_{ij-excl} q_i q_j \frac{\operatorname{erf}(\alpha r_{ij})}{r_{ij}}, \quad (3.3.150)$$

where,  $r_i$  is the vector position of the atomic charge  $q_i$ ,  $r_{ij} = r_i - r_j$ ,  $r_n$  is a vector of the direct lattice,  $\operatorname{erfc}(x) = \pi^{-1/2} \int_x^\infty e^{-t^2} dt$  is the complementary error function,  $\operatorname{erf}(x) = 1 - \operatorname{erfc}(x)$ ,  $V$  the unit cell volume,  $\mathbf{m}$  a reciprocal lattice vector and  $\alpha$  is the Ewald convergence parameter. In the direct lattice part, Eq. 3.3.147, the prime indicates that intramolecular excluded contacts are omitted. In addition, in Eq. 3.3.148 the term  $V_{intra}$  subtracts, in direct space, the intra-molecular energy between bonded pairs, which is automatically included in the right hand side of that equation. Consequently, the summation on  $i$  and  $j$  in Eq. 3.3.150 goes over all the excluded intramolecular contacts. We must point out that in the Ewald potential given above, we have implicitly assumed the so-called "tin-foil" boundary conditions: the Ewald sphere is immersed in a perfectly conducting medium and hence the dipole term on the surface of the Ewald sphere is zero [deLeeuw et al. 1980]. For increasingly large systems the computational cost of standard Ewald summation, which scales with  $N^2$ , becomes too large for practical applications. Alternative algorithms which scale with a smaller power of  $N$  than



standard Ewald have been proposed in the past. Among the fastest algorithms designed for periodic systems is the particle mesh Ewald algorithm (PME) [Darden et al. 1993, Essmann et al. 1995], inspired by the particle mesh method of Hockney and Eastwood [Hockney 1989]. Here, a multidimensional piecewise interpolation approach is used to compute the reciprocal lattice energy,  $V_{qr}$ , of Eq. 3.3.148, while the direct part,  $V_{qd}$ , is computed straightforwardly. The low computational cost of the PME method allows the choice of large values of the Ewald convergence parameter  $\alpha$ , as compared to those used in conventional Ewald. Correspondingly, shorter cutoffs in the direct space Ewald sum  $V_{qd}$  may be adopted. If  $\mathbf{u}_j$  is the scaled fractional coordinate of the  $i$ -th particle, the charge weighted structure factor,  $S(\mathbf{m})$  in Eq. 3.3.150, can be rewritten as:

$$S(\mathbf{m}) = \sum_{j=1}^N q_j \exp \left[ 2\pi i \left( \frac{m_1 u_{j1}}{K_1} + \frac{m_2 u_{j2}}{K_2} + \frac{m_3 u_{j3}}{K_3} \right) \right] \quad (3.3.151)$$

Where,  $N$  is the number of particles,  $K_1, K_2, K_3$  and  $m_1, m_2, m_3$  are integers. The  $\alpha$  component of the scaled fractional coordinate for the  $i$ -th atom can be written as: <sup>11</sup>

$$u_{i\alpha} = K_\alpha \mathbf{k}_\alpha \cdot \mathbf{r}_i, \quad (3.3.152)$$

where  $k_\alpha, \alpha = 1, 2, 3$  are the reciprocal lattice basic vectors.

$S(\mathbf{m})$  in Eq. 3.3.151 can be looked at as a discrete Fourier transform (FT) of a set of charges placed irregularly within the unit cell. Techniques have been devised in the past to approximate  $S(\mathbf{m})$  with expressions involving Fourier transforms on a regular grid of points. Such approximations of the weighted structure factor are computationally advantageous because they can be evaluated by fast Fourier transforms (FFT). All these FFT-based approaches involve, in some sense, a smearing of the charges over nearby grid points to produce a regularly gridded charge distribution. The PME method accomplishes this task by interpolation. Thus, the complex exponentials  $\exp(2\pi i m_\alpha u_{i\alpha}/K_\alpha)$ , computed at the position of the  $i$ -th charge in Eq. 3.3.151, are rewritten as a sum of interpolation coefficients multiplied by their values at the nearby grid points. In the smooth version of PME

---

<sup>11</sup>The scaled fractional coordinate is related to the scaled coordinates in Eqs 3.3.51, 3.3.81 by the relation  $s_{i\alpha} = 2u_{i\alpha}/K_\alpha$ .

(SPME) [Essmann et al. 1995], which uses cardinal B-splines in place of the Lagrangian coefficients adopted by PME, the sum is further multiplied by an appropriate factor, namely:

$$\exp(2\pi i m_\alpha u_{i\alpha}/K_\alpha) = b(m_\alpha) \sum_k M_n(u_{i\alpha} - k) \exp(2\pi i m_\alpha k/K_\alpha), \quad (3.3.153)$$

where  $n$  is the order of the spline interpolation,  $M_n(u_{i\alpha} - k)$  defines the coefficients of the cardinal B-spline interpolation at the scaled coordinate  $u_{i\alpha}$ . In Eq. 3.3.153 the sum over  $k$ , representing the grid points, is only over a finite range of integers, since the functions  $M_n(u)$  are zero outside the interval  $0 \leq u \leq n$ . It must be stressed that the complex coefficients  $b(m_i)$  are independent of the charge coordinates  $u_i$  and need be computed only at the very beginning of a simulation. A detailed derivation of the  $M_n(u)$  functions and of the  $b_\alpha$  coefficients is given in Ref. [Essmann et al. 1995]. By inserting Eq. 3.3.153 into Eq. 3.3.151,  $S(\mathbf{m})$  can be rewritten as:

$$S(\mathbf{m}) = b_1(m_1)b_2(m_2)b_3(m_3)\mathcal{F}[Q](m_1, m_2, m_3) \quad (3.3.154)$$

where  $\mathcal{F}[Q](m_1, m_2, m_3)$  stands for the discrete FT at the grid point  $m_1, m_2, m_3$  of the array  $Q(k_1, k_2, k_3)$  with  $1 \leq k_i \leq K_i$ ,  $i = 1, 2, 3$ . The gridded charge array,  $Q(k_1, k_2, k_3)$ , is defined as:

$$Q(k_1, k_2, k_3) = \sum_{i=1, N} q_i M_n(u_{i1} - k_1) M_n(u_{i2} - k_2) M_n(u_{i3} - k_3) \quad (3.3.155)$$

Inserting the approximated structure factor of Eq. 3.3.154 into Eq. 3.3.148 and using the fact that  $\mathcal{F}[Q](-m_1, -m_2, -m_3) = K_1 K_2 K_3 \mathcal{F}^{-1}[Q](m_1, m_2, m_3)$ , the SPME reciprocal lattice energy can be then written as

$$\begin{aligned} V_{gr} &= \frac{1}{2} \sum_{m_1=1}^{K_1} \sum_{m_2=1}^{K_2} \sum_{m_3=1}^{K_3} B(m_1, m_2, m_3) C(m_1, m_2, m_3) \times \\ &\times \mathcal{F}[Q](m_1, m_2, m_3) \mathcal{F}[Q](-m_1, -m_2, -m_3) \end{aligned} \quad (3.3.156)$$

$$\begin{aligned} &= \frac{1}{2} \sum_{m_1=1}^{K_1} \sum_{m_2=1}^{K_2} \sum_{m_3=1}^{K_3} \mathcal{F}^{-1}[\Theta_{rec}](m_1, m_2, m_3) \mathcal{F}[Q](m_1, m_2, m_3) \times \\ &\times K_1 K_2 K_3 \mathcal{F}^{-1}[Q](m_1, m_2, m_3), \end{aligned} \quad (3.3.157)$$

with

$$B(m_1, m_2, m_3) = |b_1(m_1)|^2 |b_2(m_2)|^2 |b_3(m_3)|^2 \quad (3.3.158)$$

$$C(m_1, m_2, m_3) = (1/\pi V) \exp(-\pi^2 m^2 / \alpha^2) / m^2 \quad (3.3.159)$$

$$\Theta_{rec} = \mathcal{F}[BC]. \quad (3.3.160)$$

Using the convolution theorem for FFT the energy 3.3.157 can be rewritten as

$$V_{qr} = \frac{1}{2} \sum_{m_1=1}^{K_1} \sum_{m_2=1}^{K_2} \sum_{m_3=1}^{K_3} \mathcal{F}^{-1}[\Theta_{rec} \star Q](m_1, m_2, m_3) \mathcal{F}[Q](m_1, m_2, m_3) \quad (3.3.161)$$

We now use the identity  $\sum_{\mathbf{m}} F(A)(\mathbf{m})B(\mathbf{m}) = \sum_{\mathbf{m}} A(\mathbf{m})F(B)(\mathbf{m})$  to arrive at

$$V_{qr} = \frac{1}{2} \sum_{m_1=1}^{K_1} \sum_{m_2=1}^{K_2} \sum_{m_3=1}^{K_3} (\Theta_{rec} \star Q)(m_1, m_2, m_3) Q(m_1, m_2, m_3) \quad (3.3.162)$$

We first notice that  $\Theta_{rec}$  does not depend on the charge positions and that  $M_n(u_{i\alpha} - k)$  is differentiable for  $n > 2$  (which is always the case in practical applications). Thus the force on each charge can be obtained by taking the derivative of Eq. 3.3.162, namely

$$F_{i\alpha}^{(qr)} = -\frac{\delta V_{qr}}{\delta r} = \sum_{m_1=1}^{K_1} \sum_{m_2=1}^{K_2} \sum_{m_3=1}^{K_3} \frac{\delta Q(m_1, m_2, m_3)}{\delta r_{i\alpha}} (\Theta_{rec} \star Q)(m_1, m_2, m_3). \quad (3.3.163)$$

In practice, the calculation is carried out according to the following scheme: i) At each simulation step one computes the grid scaled fractional coordinates  $u_{i\alpha}$  and fills an array with  $Q$  according to Eq. 3.3.155. At this stage, the derivative of the  $M_n$  functions are also computed and stored in memory. ii) The array containing  $Q$  is then overwritten by  $\mathcal{F}[Q]$ , i.e.  $Q$ 's 3-D Fourier transform. iii) Subsequently, the electrostatic energy is computed via Eq. 3.3.157. At the same time, the array containing  $\mathcal{F}[Q]$  is overwritten by the product of itself with the array containing  $BC$  (computed at the very beginning of the run). iv) The resulting array is then Fourier transformed

to obtain the convolution  $\Theta_{rec} \star Q$ . v) Finally, the forces are computed via Eq. 3.3.163 using the previously stored derivatives of the Mn functions to recast  $\delta Q/\delta r_{i\alpha}$ .

The memory requirements of the SPME method are limited.  $2K_1K_2K_3$  double precision reals are needed for the grid charge array  $Q$ , while the calculation of the functions  $M_n(u_{i\alpha} - j)$  and their derivatives requires only  $6 \times n \times N$  double precision real numbers. The  $K_\alpha$  integers determines the fineness of the grid along the  $\alpha$ -th lattice vector of the unit cell. The output accuracy of the energy and forces depends on the SPME parameters. The  $\alpha$  convergence parameter, the grid spacing and the order  $n$  of the B-spline interpolation. For a typical  $\alpha \simeq 0.4 \text{ \AA}^{-1}$  relative accuracies between  $10^{-4}$  –  $10^{-5}$  for the electrostatic energy are obtained when the grid spacing is around  $1 \text{ \AA}$  along each axis, and the order  $n$  of the B-spline interpolation is 4 or 5. A rigorous error analysis and a comparison with standard Ewald summation can be found in Refs. [Essmann et al. 1995] and [Petersen 1995]. For further readings on the PME and SPME techniques we refer to the original papers [Darden et al. 1993, Petersen 1995, Lee et al. 1995, Essmann et al. 1995]. The power of the SPME algorithm, compared to the straightforward implementation of the standard Ewald method, is indeed astonishing. In Fig. 3.6 we report CPU timing obtained on a low end 43P/160MH IBM workstation for the evaluation of the reciprocal lattice energy and forces via SPME for cyanobiphenyl as a function of the number of atoms in the system. Public domain 3-D FFT routines were used. The algorithm is practically linear and for 12000 particles SPME takes only 2 CPU seconds to perform the calculation. A standard Ewald simulation for a box  $64 \times 64 \times 64 \text{ \AA}^3$  (i.e. with a grid spacing in  $k$  space of  $k = 2\pi/64 \simeq 0.01 \text{ \AA}^{-1}$ ) for the same sample and at the same level of accuracy would have taken several minutes.

### 3.3.5 Subdivision the Non Bonded Potential

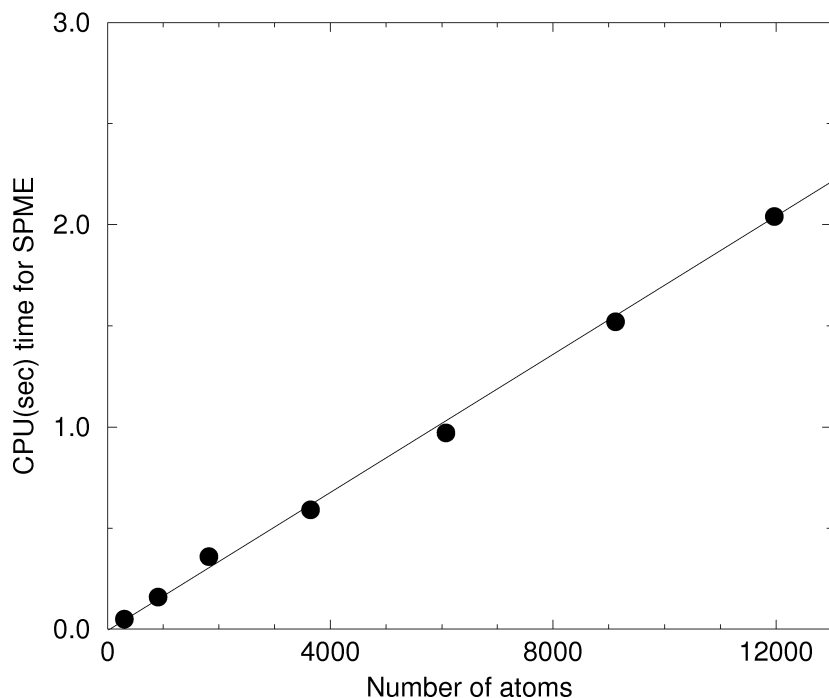
In addition to the long range electrostatic contributions,  $V_{qr}$  and  $V_{qd}$ , given in Eqs. 3.3.147, 3.3.148, more short range forces play a significant role in the total non bonded potential energy. The latter can be written as:

$$V_{nbn} = V_{vdw} + V_{qr} + V_{qd} + V_{14}. \quad (3.3.164)$$

Where,  $V_{vdw}$  is the Lennard-Jones potential, namely

Figure 3.6: CPU time versus number of particles for the SPME algorithm as measured on a 43P/160MH IBM workstation

### PME tests on 5CB



$$V_{vdw} = \sum_{i < j}^{N'} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right) \right]. \quad (3.3.165)$$

Here, the prime on the sum indicates that interactions between atoms separated by less than three consecutive bonds must be omitted. The term  $V_{14}$  is typical for force fields of complex molecular systems [Wiener et al. 1986, Brooks et al. 1983]. While non bonded forces between atoms involved in the same covalent bond or angle bending interaction are generally excluded, the potential between atoms separated by three covalent bonds is retained and readjusted in various ways. In all cases, the  $V_{14}$  term remains in general a very stiff and, hence, a fastly varying term. The computational cost of the  $V_{14}$  contribution is very small compared to other non bonded interactions. Thus, it is safer to assigns this potential term to the slowest intramolecular

reference system potential  $V_{n1}$  of Eq. 3.3.146.

The  $V_{qr}$  reciprocal lattice term, including the correction due to the excluded or partially excluded (i.e. the electrostatic part of  $V_{14}$ ) interactions cannot be split when using SPME and must be assigned altogether to only one reference system. The time scale of the potential  $V_{qr}$  depends on the convergence parameter  $\alpha$ . Indeed, this constant controls the relative weights of the reciprocal lattice energy  $V_{qr}$ , and of the direct lattice energy  $V_{qd}$ . By increasing  $\alpha$ , one increases the weight of the reciprocal lattice contribution  $V_{qr}$  to the total Coulomb energy. When using SPME the cost of the reciprocal lattice sums is cut down dramatically and, therefore, the use of large  $\alpha$ 's becomes helpful to reduce the computational burden of the direct lattice calculation. For a value of  $\alpha$  increased beyond a certain limit, there is no longer a computational gain, since the pair distances must always be evaluated in direct space until convergence of the Lennard–Jones energy (usually occurring at a 10 Å cutoff). Furthermore, the larger is  $\alpha$ , the more short–ranged and fastly varying becomes the potential  $V_{qr}$ , thus requiring short time steps to integrate correctly the equations of motion. A good compromise for  $\alpha$ , valid for cell of any shape and size, is  $\alpha = 0.4 - 0.5$ .

The direct space potential is separated [Procacci et al. 1996, Marchi and Procacci 1998] in three contributions according to the interaction distance. The overall non bonded potential breakup is therefore

$$\begin{aligned}
 V_{n1} &= V_{14} \\
 V_m &= V_{vdw}^{(1)} + V_{qd}^{(1)} \\
 V_l &= V_{vdw}^{(2)} + V_{qd}^{(2)} + V_{qr} \\
 V_h &= V_{vdw}^{(3)} + V_{qd}^{(3)},
 \end{aligned}
 \tag{3.3.166}$$

where the superscripts  $m$ ,  $l$ ,  $h$  of the direct space term  $V_{vdw}$  and  $V_{qd}$  refer to the short, medium and long range non–bonded interactions, respectively. The  $m$ –th reference system includes non–bonded direct space interactions at short range, typically between 0 to 4.3–5.3 Å.  $V_l$  contains both the medium range direct space potential, with a typical range of 4.3–5.3 to 7.3–8.5 Å, and the reciprocal space term,  $V_{qr}$ . Finally, the  $h$ –th reference system, which is the most slowly the varying contains, the remaining direct space interactions from 7.3–8.3 Å to cutoff distance. As the simulations proceeds the particles seen by a target particle may cross from one region to an other, while the

Table 3.2: Potential breakup and relative time steps for complex systems with interactions modelled by the AMBER[Cornell et al. 1995] force field and electrostatic computed using the SPME method

Component	Contributions	Spherical Shells	Time step
$V_{n0}$	$V_{stretch} + V_{bend} +$ $+V_{i-tors} + V_{p-tors}^{(h)}$	-	$\Delta t_{n0} = 0.33 fs$
$V_{n1}$	$V_{p-tors} + V_{14}$	-	$\Delta t_{n1} = 1.0 fs$
$V_m$	$V_{LJ}^{(1)} + V_{qd,\alpha=0.43}^{(1)}$	$0 < r < 4.5 \text{ \AA}$	$\Delta t_m = 2.0 fs$
$V_l$	$V_{LJ}^{(2)} + V_{qd,\alpha=0.43}^{(2)} +$ $+V_{qr,\alpha=0.43}$	$4.5 \leq r < 7.5 \text{ \AA}$	$\Delta t_l = 4.0 fs$
$V_h$	$V_{LJ}^{(3)} + V_{qd,\alpha=0.43}^{(3)}$	$7.5 \leq r < 10.0 \text{ \AA}$	$\Delta t_h = 12.0 fs$

number of two body contacts in one distance class [Grubmuller et al. 1991] or reference system potential must be continuously updated. Instabilities caused by this flow across potential shell boundaries are generally handled by multiplying the pair potential by a group-based switching function [Procacci and Marchi 1996]. Thus, at any distance  $r$  the direct space potential  $V$  can be written schematically as:

$$V = V_1 + V_2 + V_3 \quad (3.3.167)$$

with

$$V_1 = VS_1 \quad (3.3.168)$$

$$V_2 = V(S_2 - S_1) \quad (3.3.169)$$

$$V_3 = V(S_3 - S_2) \quad (3.3.170)$$

where  $S_j$  is the switching function for the three shells,  $j = m, l, h$  defined as:

$$S_j(R) = \left\{ \begin{array}{ll} 1 & R_{j-1} \leq R < R_j \\ S_{3p}^{(j)} & R_j \leq R < R_j + \lambda_j \\ 0 & R_j + \lambda_j < R \end{array} \right\} \quad (3.3.171)$$

Here,  $R$  is the inter-group distance and  $j$  is the healing interval for the  $j$ -th shell. While  $R_0$  is zero,  $R_1 = R_m$ ,  $R_2 = R_l$ , and  $R_3 = R_h$  are the

short, medium, long range shell radius, respectively. The switching  $S_{3p}^{(j)}(R)$  is 1 at  $R_j$  and goes monotonically to 0 at  $R_j + \lambda_j$ . Provided that  $S^{(j)}_{3p}$  and its derivatives are continuous at  $R_j$  and  $R_j + \lambda_j$ , the analytical form of  $S_{3p}$  in the healing interval is arbitrary [Tuckerman et al. 1991; 1992, Procacci and Marchi 1996, Procacci et al. 1996]. The full breakup for an AMBER type force field along with the integration time steps, valid for any complex molecular system with strong electrostatic interactions, is summarise in Tab. 3.2. The corresponding five time steps integration algorithm for the  $NVE$  ensemble is given

$$e^{iL\Delta t_h} = \exp \left[ -\frac{\delta V_h}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_h}{2} \right] \cdot \left\{ \exp \left[ -\frac{\delta V_l}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_l}{2} \right] \left\{ \exp \left[ -\frac{\delta V_m}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_m}{2} \right] \right. \right. \\ \left. \left\{ \exp \left[ -\frac{\delta V_{n1}}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_{n1}}{2} \right] \left\{ \exp \left[ -\frac{\delta V_{n0}}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_{n0}}{2} \right] \exp \left[ \dot{\mathbf{r}}_i \frac{\delta}{\delta r_i} \Delta t_0 \right] \right. \right. \right. \\ \left. \left. \exp \left[ -\frac{\delta V_{n0}}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_{n0}}{2} \right] \right\}^{n_{n0}} \exp \left[ -\frac{\delta V_{n1}}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_{n1}}{2} \right] \right\}^{n_{n1}} \\ \left. \exp \left[ -\frac{\delta V_m}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_m}{2} \right] \right\}^{n_m} \exp \left[ -\frac{\delta V_l}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_l}{2} \right] \right\}^{n_l} \exp \left[ -\frac{\delta V_h}{\delta r_i} \frac{\delta}{\delta p_i} \frac{\Delta t_h}{2} \right], \quad (3.3.172)$$

where  $n_l = \Delta t_h / \Delta t_l$ ,  $n_m = \Delta t_l / \Delta t_m$ ,  $n_{n1} = \Delta t_m / \Delta t_{n1}$ ,  $n_{n0} = \Delta t_{n1} / \Delta t_{n0}$ . The explicit integration algorithm can be easily derived applying the five-fold discrete time propagator 3.3.172 to the state vector  $\{\mathbf{p}, \mathbf{q}\}$  at time 0 using the rule Eq. 3.3.27. The efficiency and accuracy for energy conservation of this r-RESPA symplectic and reversible integrator have been discussed extensively in Refs. [Procacci et al. 1996; 1997a]. Extension of this subdivision to non  $NVE$  simulation is described in Ref. [Marchi and Procacci 1998].

### Electrostatic Corrections for the Multiple Time Step Simulation

In flexible molecular systems of large size, the Ewald summation presents computational problems which are crucial to constructing efficient and stable multiple time step integrators [Stuart et al. 1996, Procacci et al. 1997a]. We have seen that intra-molecular Coulomb interactions between bonded atoms or between atoms bonded to a common atom are excluded in most of the standard force fields for protein simulation. In any practical implementation of the Ewald method, the intra-molecular energy  $V_{intra}$  is automatically included in the reciprocal space summation and is subtracted in direct space (see Eqs. 3.3.150, 3.3.148. In actual simulations the reciprocal space sum is computed with a finite accuracy whereas the intra-molecular term  $V_{intra}$ , due to the excluded Coulombic interactions, is computed exactly.



This clearly prevents the cancellation of the intra-molecular forces and energies. When the stretching and bending forces are integrated explicitly, the intra-molecular term due to the excluded contacts varies rapidly with time and so does the cancellation error. Consequently, instability may be observed when integrating the reciprocal lattice forces in reference systems with large time steps. The correction due to the truncation can be evaluated by approximating the reciprocal lattice sum for the excluded contacts in Eq. 3.3.148 to an integral in the 3-dimensional  $k$  space and evaluating this integral from the cutoff  $k_{cut} \equiv 2\pi|m|_{max}$  to infinity in polar coordinates. The neglected reciprocal lattice intra-molecular energy is then [Procacci et al. 1998]

$$V_{corr} = \frac{\alpha}{p_i^{1/2}} \text{erfc}(k_{cut}/2\alpha) \sum_i q_i^2 + \sum_{ij-excl.} q_i q_j \chi(r_{ij}, k_{cut}, \alpha) \quad (3.3.173)$$

with

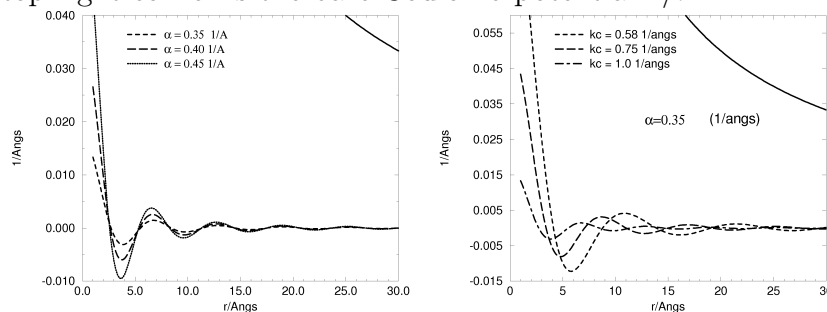
$$\chi(r, k_{cut}, \alpha) = \frac{2}{\pi} \int_{k_{cut}}^{\infty} e^{-k^2/4\alpha^2} \frac{\sin(kr)}{kr} dk. \quad (3.3.174)$$

The first constant term in 3.3.173 refers to the self energy, while the second accounts for the intra-molecular excluded interactions. This correction must be included in the same reference systems to which  $V_{qr}$  is assigned, e.g.  $V_l$  in our potential separation (see Table 2).

In principle the correction in Eq. 3.3.173 applies only to standard Ewald and not to the reciprocal lattice energy computed via SPME. We can still, however, use the correction Eq. 3.3.173, if a spherical cutoff  $k_{cut}$  is applied to SPME. This can be done easily by setting  $\exp(-\pi^2 \mathbf{m}^2 / \alpha^2) m^2 = 0$  for  $2\pi m > k_{cut} \equiv f_f \pi N_f / L$  where  $L$  is the side length of the cubic box and  $N_f$  is the number of grid points in each direction. The factor  $\alpha$  must be chosen slightly less than unity. This simple device decreases the effective cutoff in reciprocal space while maintaining the same grid spacing, thus reducing the B-spline interpolation error (the error in the B-spline interpolation of the complex exponential is, indeed, maximum precisely at the tail of the reciprocal sums [Essmann et al. 1995]). In Ref. [Procacci et al. 1998] the effect of including or not such correction in electrostatic systems using multiple time step algorithms is studied and discussed thoroughly.

The potential  $\chi(r, k_{cut}, \alpha)$  yields, in direct space, the neglected reciprocal energy due to the truncation of the reciprocal lattice sums, and must, in

Figure 3.7: The correction potential  $\chi(r, k_c, \alpha)$  as a function of the distance for different values of the parameters  $\alpha$  (left) and  $k_c$  (right). The solid line on the top right corner is the bare Coulomb potential  $1/r$



principle, be included for each atom pair distance in direct space. Thus, the corrected direct space potential is then

$$V'_{qd} = \frac{1}{2} \sum_{ij} q_i q_j \left[ \sum_n \frac{\text{erfc}\alpha |r_{ij} + r_n|}{|r_{ij} + r_n|} + \chi(|r_{ij} + r_n|, k_{cut}, \alpha) \right] \quad (3.3.175)$$

which is then split as usual in short–medium–long range according to 3.3.166. The correction is certainly more crucial for the excluded intramolecular contacts because  $V_{corr}$  is essentially a shortranged potential which is non negligible only for intramolecular short distances. For systems with hydrogen bonds, however, the correction is also important for intermolecular interactions.

In Fig. 3.7 the correction potential is compared to the Coulomb potential (solid line in the top right corner) for different value of the reciprocal space cutoff  $k_c$  and of the convergence parameter  $\alpha$ . For practical values of  $\alpha$  and  $k_c$ , the potential is short ranged and small compared to the bare  $1/r$  Coulomb interaction. In the asymptotic limit  $V_{corr}$  goes to zero as  $\sin(ar)/r^2$  where  $a$  is a constant. This oscillatory long range behaviour of the correction potential  $V_{corr}$  is somewhat nasty: In Fig. 3.8 we show the integral

$$I(r, k_c, \alpha) = \int_0^r \chi(x, k_c, \alpha) x^2 dx \quad (3.3.176)$$

as a function of the distance. If this integral converges then the  $\chi(r, k)$  is absolutely convergent in 3D. We see that the period of the oscillations in  $I(r)$  increases with  $k_c$  while  $\alpha$  affects only the amplitude. The total energy is hence

Figure 3.8: The integral  $I(r)$  of Eq. 3.3.176 as a function of the distance for different values of the parameters  $\alpha$  (left) and  $k_c$  (right)

again conditionally convergent, since the limit  $\lim_{r \rightarrow \infty} I(r)$  does not exist. However, unlike for the  $1/r$  bare potential, the energy integral remains in this case bounded. Due to this, a cutoff on the small potential  $V_{corr}$  is certainly far less dangerous than a cutoff on the bare  $1/r$  term. In order to verify this, we have calculated some properties of liquid water using the SPC model [Rahman and Stillinger 1971] from a 200 ps MD simulation in the  $NPT$  ensemble at temperature of 300 K and pressure of 0.1 MPa with i) a very accurate Ewald sum (column EWALD in Table 3.3.4), ii) with inaccurate Ewald but corrected in direct space using Eq. 3.3.175 (CORRECTED) and iii) with simple cutoff truncation of the bare Coulomb potential and no Ewald (CUTOFF). Results are reported in Table 3.3.4 We notice that almost all the computed properties of water are essentially independent, within statistical error, of the truncation method. The dielectric properties, on the contrary, appear very sensitive to the method for dealing with long range tails: Accurate and inaccurate Ewald (corrected in direct space through 3.3.175 yields, within statistical error, comparable results whereas the dielectric constant predicted by the spherical cutoff method is more than order of magnitude smaller. We should remark that method ii) (CORRECTED) is almost twice as efficient as the "exact" method i).

Table 3.3: Properties of liquid water computed from a 200 ps simulation at 300 K and 0.1 MPa on a sample of 343 molecules in PBC with accurate Ewald ( $\alpha = 0.35 \text{ \AA}^{-1}$ ,  $k_c = 2.8 \text{ \AA}^{-1}$ ) and no correction Eq. 3.3.173 (column EWALD), with inaccurate Ewald ( $\alpha = 0.35 \text{ \AA}^{-1}$ ,  $k_c = 0.9 \text{ \AA}^{-1}$ ) but including the correction Eq. 3.3.173 and with no Ewald and cutoff at 10.0  $\text{\AA}$ .  $R_{0-0}$  is the distance corresponding to the first peak in the Oxygen–Oxygen pair distribution function.

	EWALD	CORRECTED	CUTOFF
Coulomb energy (KJ/mole)	$-55.2 \pm 0.1$	$-55.1 \pm 0.1$	$-56.4 \pm 0.1$
Potential energy (KJ/mole)	$-46.2 \pm 0.1$	$-46.1 \pm 0.1$	$-47.3 \pm 0.1$
Heat Capacity (KJ/mole/K)	$74 \pm 24.5$	$94 \pm 22.0$	$87 \pm 23.2$
Volume ( $cm^3$ )	$18.2 \pm 0.1$	$18.3 \pm 0.1$	$18.1 \pm 0.1$
Volume Fluctuation ( $\text{\AA}^3$ )	$136.9 \pm 3.5$	$147.0 \pm 3.5$	$138.7 \pm 3.5$
$R_{0-0}$ ( $\text{\AA}$ )	$2.81 \pm 0.01$	$2.81 \pm 0.01$	$2.81 \pm 0.01$
Dielectric constant	$59 \pm 25.8$	$47 \pm 27.3$	$3 \pm 2$

# Chapter 4

## Azobenzene in organic solvents

### 4.1 Introduction

It is well known that **azobenzene** (AB) and its derivatives can undergo a major structural change upon irradiation with light, transforming from the longer *trans*-isomer to the shorter, bent, *cis*-isomer and vice versa. This variation, coupled with the high stability of the *cis* form and the reversibility of the isomerization, can be exploited in the design of materials with photo-switchable physical properties [Finkelmann et al. 2001a, Ikeda and O.Tsutsumi 1995, Yu et al. 2003, Camacho-Lopez et al. 2004] for photonic [Tong et al. 2005] and micro- and nano-scale device [Lansac et al. 1999, Hugel et al. 2002, Banerjee et al. 2003, Buguin et al. 2006, Muraoka et al. 2006] applications.

The azobenzene photophysics at the root of the conformational change has been very extensively studied both theoretically and experimentally for over two decades, and various essential features are now understood, at least in the gas phase. The process, that typically occurs in the picosecond time scale, can involve a ( $n, \pi^*$ ) or a ( $\pi, \pi^*$ ) **absorption**, depending on the excitation wavelength [Ciminelli et al. 2004, Satzger et al. 2004, Ishikawa et al. 2001, Chang et al. 2004]. In the most common experimental conditions (a near UV ( $\pi, \pi^*$ ) excitation), a three-state mechanism seems to take place, with promotion from the fundamental state  $S_0$  of the *trans* isomer to the second **singlet** excited state  $S_2$ , followed by a decay to the first **singlet** state  $S_1$ , and finally by a decay, either **non-radiative decay** via conical intersection ( $S_0/S_1$  CI) or **radiative decay** by weak fluorescence to the  $S_0$  state. The process

can be even more complex and some authors have recently pointed out the importance of other [singlet](#) and [triplet](#) states [Ciminelli et al. 2004, Cembran et al. 2004]. By comparison, the photophysics of the process following the  $(n,\pi^*)$  absorption from the ground state in the visible ( $\lambda = 440 - 480$  nm) is simpler as it involves only the first excited state  $S_1$  [Chang et al. 2004, Cattaneo and Persico 1999, Diau 2004].

In addition to the photophysical aspects, the intramolecular mechanism of the isomerization process involves two basic pathways: torsion (changing the dihedral angle  $Ph-N=N-Ph$ ), that requires a reduction of the order of the nitrogen-nitrogen double bond, and inversion, that implies a wide increase of the  $Ph-N=N$  bending angles with an exchange of the position of the lone pair of one of the nitrogens. The two mechanisms have often been considered in alternative, and their relative contributions to the isomerization is still controversial even in the gas phase, although some precious clarification have been provided by recent works [Ciminelli et al. 2004, Cembran et al. 2004]. Besides, the existence of a third mechanism, “concerted inversion”, has been suggested to be active in case of  $(\pi,\pi^*)$  excitation, but it is believed to produce either the *trans* isomer or dissociation [Diau 2004].

In the more general case the photoactive molecule can be considered to undergo a mixed mechanism that involves both processes and that reduces to pure [torsion](#) or [inversion](#) only in the limiting cases. Surprisingly enough, the *trans-cis* quantum yield for the  $(n,\pi^*)$  is nearly double than that of the  $(\pi,\pi^*)$  in *n*-hexane [Cembran et al. 2004], while according to the standard wisdom (Kasha rule) they should be the same. Given that in all practical applications the AB photoisomerization takes place in solution or in a polymer, it is somehow disappointing that the vast majority of the theoretical information available only refers to isomerization in the gas phase, where the conformational change is not hindered by the environment and where we can expect that the mechanism can be different.

Part of the difficulty in studying the solvent environment effects on the *trans-cis* isomerization is due to the need of building a model of the process that combines the essential photophysics with an atomistic description of the guest-host system. Several publications, mostly by Persico and coworkers, already shed light on the photoisomerization dynamics of AB and azobenzenophanes in vacuum, using mixed quantum-classical simulation schemes [Ciminelli et al. 2004, Toniolo et al. 2005, Ciminelli et al. 2005, Nonnenberg et al. 2006], but to our knowledge no simulation studies of azobenzene isomerization in solution have been published so far, with the exception

of an investigation of azomethane photochemistry in water [Cattaneo and Persico 2001].

Here we wish to contribute to this challenging task and as a first step in the study of the isomerization process in condensed phases we have chosen to follow the photophysically simpler ( $n,\pi^*$ ) transition in various low molar mass organic solvents using a suitably adapted molecular dynamics (MD) simulation which allows for transitions from the ground to the excited state and back during the time evolution [Tiberio et al. 2006b]. We believe this investigation to be particularly timely, since significant experimental studies on AB photoisomerization in solution following a  $S_1$  excitation have recently appeared [Chang et al. 2004].

To allow for the possibility of coupled **torsional-inversion** pathways, we have modeled the ( $n,\pi^*$ ) transition of AB with a specific molecular mechanics force field containing a quantum-mechanically derived potential energy surface (PES) [Ishikawa et al. 2001] for the relevant bending and **torsional** degrees of freedom, in both the electronic states involved. We have then studied by classical MD simulations an AB molecule either isolated (in vacuum) or dissolved in one of four solvents (**anisole**, ***n*-hexane**, **methyl-*n*-pentyl ether**, and **toluene**), analyzing the molecular movements occurring during the isomerization, the AB structural reorganization, the *trans-cis* isomerization quantum yield and the effect of solvent on favoring either the rotational or the **inversion** channel.

The paper is organized as follows: in the next section we introduce the models adopted for the azobenzene ground and excited states, then we describe our procedure for modeling with “virtual experiments” the transitions between the two states. We also discuss the modeling and parameterization of the various solvents and provide details of the simulation conditions. In the latest sections we describe our simulations results, discussing the isomerization mechanism, and its modifications when going from vacuum to solvents and providing a comparison with experimental data when available.

## 4.2 Models and Simulations

### 4.2.1 Azobenzene ground and excited states

We have modeled the AB molecule at fully atomistic level starting from the AMBER molecular mechanics force field (FF) [Cornell et al. 1995, Wang

et al. 2004]. This FF is expected to describe with sufficient accuracy the intermolecular interactions and the molecular shape near the equilibrium geometry, but it is known to fail in describing properly a molecule in a distorted geometry, which unfortunately is the case for the isomerization process. The *trans-cis* isomerization directly influences at least four internal degrees of freedom: the *Ph-N=N-Ph* torsional angle ( $\phi$ ), the two *Ph-N=N* bending angles ( $\theta$ ), and the *N=N* bond length. Here we have re-parameterized the standard torsional and bending contributions with appropriate functions of the torsional angle  $\phi$  and of one of the bending angles  $\theta$  (Figure 4.1) in the ground and in the first singlet excited state. We have employed as far as possible the *ab initio* CASSCF ( $\theta, \phi$ ) potential energy surfaces for the two electronic states  $S_0$  and  $S_1$  obtained by Ishikawa *et al.* [Ishikawa et al. 2001]. We have computed the energies and the forces for arbitrary ( $\theta, \phi$ ) points, sampling these surfaces with a uniform grid of  $10 \times 10$  points with  $105^\circ \leq \theta \leq 180^\circ$  and  $0^\circ \leq \phi \leq 180^\circ$ , and using a second order Lagrange interpolation. As the *ab initio* surfaces in Ref. [Ishikawa et al. 2001] were calculated only for bending angles greater than  $105^\circ$ , we have also extrapolated the portion of energy surface for  $\theta < 105^\circ$  with a set of least squares parabolas optimized in the range  $105^\circ \leq \theta \leq 110^\circ$  with  $\phi$  fixed at the grid values.

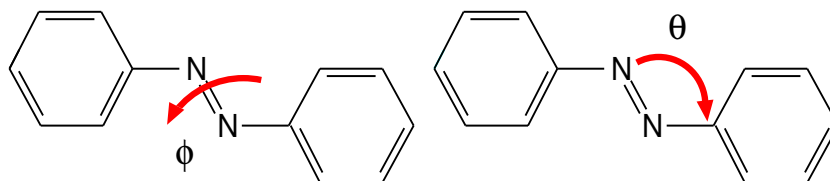


Figure 4.1: Scheme of the azobenzene (AB) molecule showing the *Ph-N=N-Ph* torsional angle  $\phi$ , and the *Ph-N=N* bending angle  $\theta$ .

In the quantum mechanical PES, the intramolecular electrostatic and dispersive contributions between the azobenzene atoms are implicitly included. To correct the FF, avoiding counting these terms twice when the standard sum over charges and atoms is performed in the simulation [Cheung et al. 2002], we have subtracted from the *ab initio* PES  $U_n^{ai}(\theta, \phi)$  for the  $n = 0, 1$  states the molecular mechanics PES calculated with the ground state atomic charges. This procedure implies of course a further approximation, particularly in the treatment of the solute-solvent electrostatic interactions in the  $S_1$  state. This is related to the difficulty of charge determination for all the pos-



sible geometry and generally the charge interaction are completely neglected in solute–solvent interaction in isomerization studies [Cattaneo and Persico 2001].

In practice, to evaluate the required contributions we have preliminarily computed a Boltzmann distribution  $P_c(\theta, \phi; T)$  for the [torsional](#) and bending degrees of freedom by performing a constant volume, 10 ns–long [MD](#) simulation of an isolated AB molecule at  $T = 2000$  K using a FF with all the energy terms involving explicitly  $\phi$  or  $\theta$  set to zero. The high temperature has been chosen to allow a thorough exploration of the PES during the trajectory. We have then obtained the correction  $U_c(\theta, \phi)$  to the total conformational energy through an [inversion](#) of the distribution as in reference [Berardi et al. 2005]

$$U_c(\theta, \phi) = -k_B T [\ln P_c(\theta, \phi; T) + \ln A], \quad (4.2.1)$$

where the scaling constant  $A$  has been chosen in order to set the lowest energy point to zero, and  $k_B$  is the Boltzmann constant. The final force field contribution which replaces the [torsional](#) term for the  $Ph-N=N-Ph$  dihedral and the bending term for the  $Ph-N=N$  angle in the  $n$ -th electronic state is therefore

$$U_n(\theta, \phi) = U_n^{ai}(\theta, \phi) - U_c(\theta, \phi). \quad (4.2.2)$$

### 4.2.2 Azobenzene excitation and decay

We have proceeded to model the transition between the electronic states in a simplified way, considering the  $S_0 \rightarrow S_1$  process to take place within the Franck–Condon regime, i.e. as a vertical transition occurring instantaneously at a given simulation time with fixed nuclei positions. During the [MD](#) simulation, we imagine the system to be exposed to radiation of suitable wavelength and we schedule an excitation event (photon absorption) at regular time intervals of 10 ps along the ground state main trajectory. When this happens, the FF parameterization for the AB molecule is switched from that of the ground state to that of the first [singlet](#) excited state and a new secondary trajectory is spawned from the main one and followed. In the secondary trajectory the AB molecule can then move and change its [conformation](#) according to the new PES and the solvent environment constraints. In each of these conformational states the AB molecule has a certain chance of decaying back to the  $S_0$  state. Assuming the existence of a conical inter-

section with the ground state in the neighborhood of the minimum of the  $S_1$  PES, the transition probability from  $S_1$  to  $S_0$  is modelled with an exponential energy gap law:

$$P_{1\rightarrow 0}(\theta, \phi) = K \exp \left\{ -\chi [U_1^{ai}(\theta, \phi) - U_0^{ai}(\theta, \phi)] \right\}, \quad (4.2.3)$$

where  $U_1^{ai}(\theta, \phi)$  and  $U_0^{ai}(\theta, \phi)$  are respectively the energies for the excited and ground states in the given  $(\theta, \phi)$  conformation (see Equation 4.2.2), while  $K$  and  $\chi$  are empirical constants, described later. In practice, during the trajectory on the  $S_1$  surface a uniformly distributed random number is sampled every femtosecond and the transition to the fundamental state is accepted (or not) using Equation 4.2.3 and the von Neumann's rejection criterion[von Neumann 1951].

This simple approach, which ignores the dynamic coupling between the electronic states, is inspired by the theoretical work of Englman, Jortner, Henry and Siebrand [Englman and Jortner 1970, Henry and Siebrand 1973]. These authors provided a physical interpretation for radiationless transition laws often observed in experiments [Henry and Siebrand 1973, Kitamura et al. 1999]. Following the classification of reference [Englman and Jortner 1970], the  $(n, \pi^*)$  AB photoisomerization falls in the so-called strong coupling limit (two electronics states with large horizontal displacements, crossing close to the minimum of the higher one), and in the low temperature regime (Eq. 4.10 in ref.[Englman and Jortner 1970]).

At the conical intersection ( $\phi \approx 88^\circ$ ,  $\theta \approx 130^\circ$ ), the *ab initio* PESs computed in ref.[Ishikawa et al. 2001] do not touch (i.e.  $U_1^{ai} - U_0^{ai} \approx 1.6$  kcal/mol ), and we have set the preexponential factor  $K$  in order to achieve a transition to the value of  $1 \text{ fs}^{-1}$  in this region. This choice eventually defines the proximity of the CI as a small portion of the  $S_1$  surface (with approximate bounds  $86^\circ < \phi < 93^\circ$ , and  $125^\circ < \theta < 131^\circ$ ) in which the probability of decaying is greater than 1%. Due to the interpolation algorithm described in section 4.2.1, this CI is partially devoid of the cusp shape which is expected from a fully quantomechanical treatment. The  $\chi$  constant can be considered as an estimate of the opening of the  $S_0/S_1$  molecular funnel: a small  $\chi$  value gives a large intersection area in the  $(\theta, \phi)$  surface and vice versa. In this regard, we have observed that increasing  $\chi$  increases both the permanence time in the  $S_1$  state, and the isomerization quantum yield achieved in our computations.

A more rigorous approach to the nonadiabatic dynamics in solution, us-

ing for instance Ehrenfest [Li et al. 2005] or surface hopping methods [Tully and Parandekar 2005], would make this study computationally too demanding, even at semiempirical level, as the total simulated time here has been of 20 ns for every solvent sample. Moreover, this strategy would introduce a trade-off, since while providing a better estimate of the transition probability, the PESs calculated “on-the-fly” would be probably less reliable than the CASSCF ones used here. We have then opted for an empirical scheme, determining a value of  $\chi$  which gives a reasonable agreement between the computed isomerization quantum yield and decay times and the corresponding experimental values in *n*-hexane [Chang et al. 2004, Bortolus and Monti 1979, Lednev et al. 1998, Lu et al. 2002]. After the parameterizing procedure we have applied the same model (i.e. force field and  $\chi$  value) to the isomerization in other solvents.

### 4.2.3 Solvents and solutions

We have studied the *trans*-*cis* isomerization of AB either isolated in vacuum or dissolved in one of four isotropic solvents: *n*-hexane, employed to parameterize the decay model, methyl-*n*-pentyl ether (MPE), toluene and anisole. The solvents have a different density, viscosity and polarity (see Table 1) and later on we shall try to connect these features to their effects on the isomerization process. For all the compounds studied we have preliminarily performed a quantum mechanical DFT B3LYP/6-31G\*\* geometry optimization, and determined the atomic point charges using the ESP scheme with the additional constraint of reproducing the total dipole moment [Besler et al. 1990]. These atomic charges have been used to complete the parameterization of the FF prior to the computation of the correction term (Equation 4.2.1) for azobenzene. We have further assumed the  $(\theta, \phi)$  PES of the AB electronic states, i.e. the intramolecular energy for *torsion* and bending, to be unaffected by solute-solvent interactions. This necessary, even if seemingly drastic, approximation is to some extent confirmed experimentally (see References [Rau 2003, Kobayashi et al. 1987]). We should stress, however, that the interaction of AB with the solvent is very important as it affects the total energy of the system and the trajectories in the  $\theta, \phi$  surfaces, influencing the *trans*-*cis* conversion.

#### 4.2.4 Simulation conditions

Each of the solutions used consisted of one *trans*-AB molecule surrounded by 99 solvent molecules, all treated at fully atomistic level, and contained in a cubic box with periodic boundary conditions (PBC). We have found this relatively small sample sizes to be sufficient to describe about two solvation shells and thus adequate for the short range solvent effects expected in these isotropic systems. All the solution samples have been simulated with a suitably modified MD code ORAC [Procacci et al. 1997b], and equilibrated employing constant *NPT* conditions using a Nosé–Hoover thermostat [Nosé 1984, Hoover 1985], and an isotropic Parrinello–Rahman barostat [Parrinello and Rahman 1980] ( $P = 1$  bar), PBCs, and 1 fs time step. We have chosen for the simulation temperature a value  $T = 320$  K above room temperature and below the boiling point of the solvents, to take advantage of the higher solute mobility and reduced solvent viscosity and indirectly favor the isomerization process. The MD equilibration was continued until the density  $\rho$  and other thermodynamics observables (e.g. energy) were observed to fluctuate around a constant average value for at least 1 ns. In Table 4.1 we report the equilibrium densities obtained for our solutions and we see that they compare rather well with experimental values for the pure solvents. The molecular dynamics simulations in vacuum were performed for AB in a cubic box with sides of 100 Å with canonical (constant *NVT*) conditions at  $T = 320$  K using a Nosé–Hoover thermostat [Nosé 1984, Hoover 1985] and PBC.

After this preliminary stage, we have continued the MD simulations under the same conditions for 10 ns, extracting a configuration every 10 ps, for a total of  $N = 1000$  different starting points for the excitation experiments. Each excitation “experiment” consists of enforcing a vertical electronic transition,  $S_0 \rightarrow S_1$ , which simulates the change of PES occurring experimentally upon irradiation with light, and following the ensuing  $S_1 \rightarrow S_0$  relaxation dynamics (see Equation 4.2.3) for a 20 ps time window, sufficient to span the standard experimental time scale of the isomerization process [Chang et al. 2004]. Figure 4.2 shows a graphical representation of a typical trajectory. As each experiment provides a different estimate of the permanence time in the excited state, determined by the statistical sampling of the hopping probability  $P_{1 \rightarrow 0}$  (see Equation 4.2.3), a large number of excitation events is necessary to compute the ensemble-averaged observables which are presented in the following.

Table 4.1: The physical properties of the compounds studied: experimental [Lide 2004–2005] densities ( $\text{g cm}^{-3}$ ), viscosities (cP), boiling temperatures (K), and dielectric constants; and calculated densities ( $\text{g cm}^{-3}$ ), and molecular dipoles (D).

	$\rho^{(a)}$	$\eta^{(a)}$	$T_b^{(a)}$	$\epsilon^{(a)}$	$\rho^{(b)}$	$\mu^{(c)}$
<i>trans</i> -AB	—	—	566	—	—	0.0
<i>cis</i> -AB	—	—	—	—	—	3.19
anisole	0.99 <sup>(d)</sup>	0.747 <sup>(e)</sup>	427	4.3 <sup>(f)</sup>	1.01	1.06
<i>n</i> -hexane	0.64 <sup>(e)</sup>	0.240 <sup>(e)</sup>	342	1.89 <sup>(f)</sup>	0.66	0.0
MPE	0.75 <sup>(d)</sup>	0.3 – 0.5 <sup>(g)</sup>	372	1.3 – 4.0 <sup>(h)</sup>	0.73	1.31
toluene	0.84 <sup>(e)</sup>	0.424 <sup>(e)</sup>	384	2.38 <sup>(f)</sup>	0.89	0.34

<sup>(a)</sup> Experimental value [Lide 2004–2005] for the pure solvent;

<sup>(b)</sup> MD simulation of a model solution of 99 solvent molecules and 1 AB solute, at  $T = 320$  K, and  $P = 1$  bar;

<sup>(c)</sup> Computed in vacuum for an isolated molecule at B3LYP/6–31G\*\* level;

<sup>(d)</sup> Measured at  $T = 298.15$  K;

<sup>(e)</sup> Measured at  $T = 323.15$  K;

<sup>(f)</sup> Measured in the temperature range  $T = 293.15 - 298.15$  K;

<sup>(g)</sup> Estimated considering the smaller and larger experimental values of methyl ether series;

<sup>(h)</sup> Estimated from chemically similar compounds.

## 4.3 Results and discussion

We start presenting our results for the quantum yield and the decay kinetics in vacuum and in *n*-hexane for various values of the empirical decay parameter  $\chi$ , computed for a subset of 500 MD experiments, and comparing them with available theoretical results for the yield in vacuum [Ciminelli et al. 2004] and with experimental data for yield [Rau 2003] and decay times in *n*-hexane [Chang et al. 2004]. Without attempting a perfect fit that would be inappropriate considering the various approximations of the treatment, a comparison with these data provides a reasonable estimate for  $\chi$ . We shall then use the chosen value of  $\chi$  to calculate and predict results for AB in the other solvents for the complete set of  $N = 1000$  MD experiments, before proceeding to analyzing in detail the isomerization mechanism.

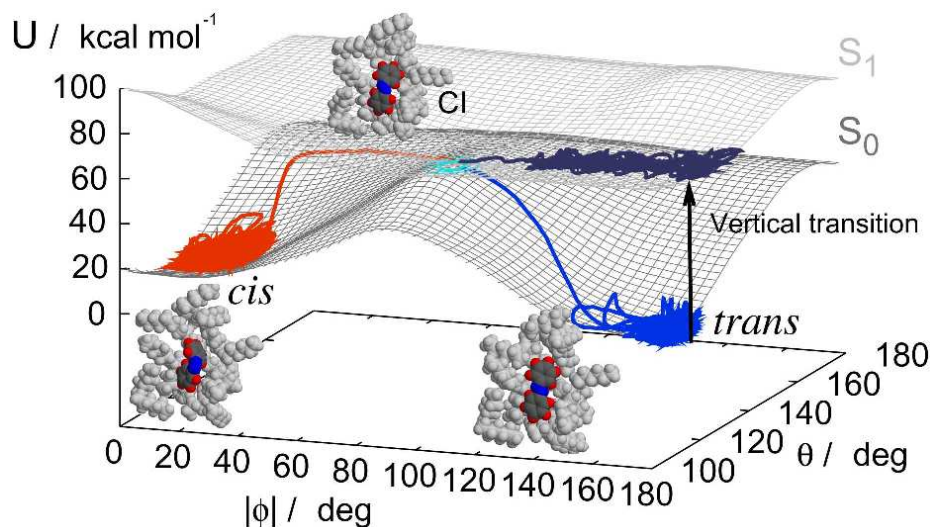


Figure 4.2: The potential energy surfaces for the  $S_0$  and  $S_1$  states of AB used in this work [Ishikawa et al. 2001], with superimposed representative MD trajectories in the excited state (black), and after one successful (final *cis* isomer, red), and one unsuccessful (final *trans* isomer, blue) isomerization process of AB in *n*-hexane, at  $T = 320$  K and  $P = 1$  bar. We also show the snapshots of an AB molecule with the closest solvent neighbors in typical MD configurations of the *trans*- and *cis*-isomers, and at the conical intersection. The cyan circle defines the region in  $S_1$  with transition probability greater than 1%

### 4.3.1 Photoisomerization quantum yield

The efficiency of the *trans*-*cis* process can be measured by the photoisomerization quantum yield,  $\Phi = N_{cis}/N$ , where  $N_{cis}$  is the number of virtual experiments yielding a *cis* isomer out of the total number of experiments (here  $N = 500$ ), which are in turn equivalent to the number of absorbed photons. In Table 4.2 we report our results for various values of the empirical parameters  $\chi$  in Equation 4.2.3, expressed in units of  $\chi_0 = 0.376$  mol/kJ. Comparing the simulation quantum yields with the values recently calculated for this transition in vacuum ( $\Phi = 0.33 - 0.46$  [Ciminelli et al. 2004]) and with the experimental values in *n*-hexane ( $\Phi = 0.20 - 0.25$  [Bortolus and Monti 1979, Rau 2003]), we see that a reasonable agreement is obtained only

for  $\chi \geq 3\chi_0$ , so lower values will be neglected from now on. We would like to stress that, in comparing with experimental work, we are only looking at this stage for a semi-quantitative agreement, also in consideration of the fact that our photoisomerization study proceeds only via the [singlet](#) process while other mechanisms [Cembran et al. 2004] could, at least in principle, provide additional contributions to the experimental value of  $\Phi$ .

Table 4.2: The *trans-cis* photoisomerization quantum yields  $\Phi$  computed from the 500 MD virtual experiments in vacuum and *n*-hexane using different values of  $\chi$ . The largest standard deviation was  $\sigma = 0.02$ . The experimental value in *n*-hexane for an excitation at  $\lambda = 439$  nm is  $\Phi = 0.20-0.25$  [Bortolus and Monti 1979, Rau 2003].

$\chi/\chi_0$	1	2	3	4	5
<i>n</i> -hexane	0.01	0.07	0.12	0.19	0.23
vacuum	0.08	0.29	0.48	0.49	0.42

### 4.3.2 Kinetic model for isomerization

We have analyzed the kinetics of the isomerization process by studying and modeling the time dependence of the average populations of the excited and ground states, similarly to what is often done when analyzing experimental data. In all excitation/relaxation experiments we have found that the decay takes place at least 50–100 fs after the  $S_0 \rightarrow S_1$  hopping, which is essentially the time necessary for an excited molecule in a certain [conformation](#)  $AB_1$ , far from the  $S_0/S_1$  conical intersection and thus unable to effectively decay by non radiative mechanism, to transform into a reactive form  $AB_1^*$  (in the neighborhood of the  $S_0/S_1$  CI) which has a larger probability of decaying to the ground state ( $AB_0$ ). This simple kinetic scheme is the standard one adopted to analyze experimental femtosecond [fluorescence](#) data [Lu et al. 2002] and takes the form of two first order consecutive irreversible reactions



where  $\tau_1$  and  $\tau_2$  are the characteristic times. Integration of the corresponding set of kinetic equation gives (for  $AB_1(0) = 1$ )

$$AB_1(t) = e^{-t/\tau_1}, \quad (4.3.5)$$

$$AB_1^*(t) = \frac{\tau_2}{\tau_1 - \tau_2} \left( e^{-t/\tau_1} - e^{-t/\tau_2} \right), \quad (4.3.6)$$

$$AB_0(t) = 1 - \frac{\tau_1 e^{-t/\tau_1} - \tau_2 e^{-t/\tau_2}}{\tau_1 - \tau_2} = 1 - AB_1(t) - AB_1^*(t). \quad (4.3.7)$$

Table 4.3: The decay times  $\tau_1$ , and  $\tau_2$  (in ps) for AB in vacuum and *n*-hexane computed from the 500 MD virtual photoisomerization experiments with different values of  $\chi$ . The experimental fluorescence decay times in *n*-hexane are  $\tau_1 = 0.24 - 0.6$  ps, and  $\tau_2 = 1.7 - 2.6$  ps [Chang et al. 2004, Lednev et al. 1998].

	$\chi/\chi_0$	vacuum	<i>n</i> -hexane
$\tau_1$	3	$0.05 \pm 0.03$	$0.50 \pm 0.02$
	4	$0.31 \pm 0.01$	$1.19 \pm 0.02$
	5	$0.91 \pm 0.01$	$2.04 \pm 0.05$
$\tau_2$	3	$0.33 \pm 0.03$	$1.59 \pm 0.03$
	4	$0.89 \pm 0.02$	$3.26 \pm 0.02$
	5	$2.27 \pm 0.02$	$9.73 \pm 0.07$

The decay times  $\tau_1$ , and  $\tau_2$  shown in Table 4.3 have been derived by fitting with Equation 4.3.7 our observed population  $AB_0$  for various values of  $\chi$ . We see that the decay times significantly depend on  $\chi$  and we can now proceed to compare our results with the experimental decay times obtained by time resolved femtosecond fluorescence data in *n*-hexane. These generally show two different components obtained from a bi-exponential fit: a fast one in the range  $\tau_1 = 0.24 - 0.6$  ps followed by a slower one in the range  $\tau_2 = 1.7 - 2.6$  ps [Chang et al. 2004, Lednev et al. 1998] with the latter one exhibiting the largest increase when increasing the probing wavelength. Given the single minimum shape of the  $S_1$  surface, and the fact that the probability of radiative decay from  $S_1$  to  $S_0$  is very low for AB [Fujino et al. 2001, Lu et al. 2005], it is possible to compare our radiationless decay times with the radiative ones registered at high wavelengths, i.e. fluorescence times giving information on the experimental dynamics of the molecules when moving towards the minimum of the excited state.



It is comforting to see that the simulation results for  $\chi = 3\chi_0$  are in good, semi-quantitative, agreement with the experimental values. Since the main focus of this study was on clarifying the dynamical mechanism of AB isomerization, and not that of reproducing exactly the decay probability, and considering also the quite reasonable agreement with the experimental quantum yield and decay times in *n*-hexane obtained for  $\chi = 3\chi_0$ , only this value of the parameter has been used to perform the MD simulations with the other solvents, obtaining the results in Table 4.4, and in the rest of the analysis.

Table 4.4: The average decay times  $\tau_1$ , and  $\tau_2$  (in ps), and the AB *trans-cis* photoisomerization quantum yields  $\Phi$  computed from the 1000 MD virtual photoisomerization experiments in the various solvents at  $T = 320$  K,  $P = 1$  bar, and assuming  $\chi = 3\chi_0$ .

	vacuum	anisole	<i>n</i> -hexane	MPE	toluene
$\tau_1$	$0.08 \pm 0.02$	$0.48 \pm 0.07$	$0.43 \pm 0.02$	$0.52 \pm 0.02$	$0.41 \pm 0.01$
$\tau_2$	$0.35 \pm 0.03$	$5.80 \pm 0.07$	$1.73 \pm 0.03$	$0.98 \pm 0.01$	$0.74 \pm 0.01$
$\Phi$	$0.51 \pm 0.01$	$0.06 \pm 0.01$	$0.11 \pm 0.01$	$0.14 \pm 0.01$	$0.13 \pm 0.01$

The physical interpretation of the bi-exponential decay obtained is not straightforward, particularly since the kinetic equation for  $AB_0(t)$  is symmetric under  $\tau_1, \tau_2$  exchange. For a proper attribution of the individual relaxation times, we have thus computed, in a separate series of simulations in which the decay to the ground state was forbidden, the average torsional angle evolution in the excited state for the first 2 ps after the excitation. The rapid variation of this geometry indicator just after the photon absorption suggests that the first process ( $AB_1 \rightarrow AB_1^*$ ) is the fastest, and that it can be identified with the conformational changes necessary to reach the region near the minimum of the  $S_1$  PES. We have found this process to last between 0.3 – 0.5 ps in all the solvents investigated (see  $\tau_1$  in Table 4.4). The solute-solvent interactions become more important when the molecule moves around the minimum near  $S_0/S_1$  CI, waiting for reaching a  $(\theta, \phi)$  point favorable for the decay. We identify this process as the slow one, whose characteristic times  $\tau_2$  are shown in Table 4.4.

As for the solvent effects on the yield, we see that for the other solvents we obtain yields similar to *n*-hexane, except for anisole, where we observe the lowest simulated yield. From experimental work it is known that, for the

( $n, \pi^*$ ) band, the yield  $\Phi$  is fairly constant for low to weakly polar solvents (from 0.25 in *n*-hexane to 0.26 in ethyl bromide) [Bortolus and Monti 1979] and this is also what we observe if we compare results in hexane and methyl-*n*-pentyl ether or toluene. In the case of anisole, however, other factors, in particular viscosity, which is significantly larger than in the other solvents, seem instead to dominate with respect to polarity.

### 4.3.3 Geometry modifications during the isomerization process

We now wish to examine the details of the *trans*-*cis* transformation in solution as obtained from our computer simulation results. We have monitored the isomerization mechanism with various geometrical indicators related to the structural reorganization of AB that we have evaluated from the trajectories of the excitation/relaxation experiments. These include the absolute value of the *Ph*-*N=N*-*Ph* torsional angle ( $|\phi|$ ), the average of the selected *Ph*-*N=N* bending angle ( $\theta$ ), the distance  $r_{4,4'}$  between the outmost (4, 4') carbons of the phenyl rings, and the average module of the four *C-C-N=N* torsional angles  $|\gamma|$  which monitors the concerted rotation of the phenyl groups. The dynamic evolution of these indicators from the initial *trans* values is reported in Figure 4.3, where for clarity we do not plot the curves in MPE and toluene because they are similar to those in hexane. The averages have been computed considering all the experiments and separating the values for the trajectories leading to successful (*cis* final isomer) and unsuccessful (*trans* final isomer) photoisomerization outcomes, as indicated by the lateral bars on the right hand side of the plots of Figure 4.3.

We notice first that the AB isomerization in vacuum differs from the one in solution in that the trajectories are already well separated at 0.5 ps into those leading to *cis* and *trans*. For all other cases we can observe, looking at Figure 4.3-[a], 4.3-[c] and 4.3-[d], that the geometrical indicators after 0.5 ps in the excited state are still more similar to those of the starting *trans* isomer than that of the *cis*. The exception is the bending angle (Figure 4.3-[b]), that within the first 0.5 ps in the excited state has attained values closer to the *cis* conformer ones before relaxing to its final value.

We can see that the *trans*-*cis* isomerization requires a large rotation of a phenyl group from  $|\gamma| = 4 - 10^\circ$  to  $|\gamma| = 30 - 40^\circ$  and thus if this rotation is hindered, e.g. by steric interactions with the neighboring solvent

molecules, the isomerization is correspondingly slowed down (Figure 4.3–[d]). The distance indicator  $r_{4,4'}$  is strictly related to the **torsional** angle  $\phi$  (see Figure 4.3–[a] and 4.3–[c]) and shows directly the shortening of the AB when going from the *trans* to *cis* conformer, that turns out to be about 2.25 Å in solution and 2.5 Å in vacuum. The *cis* isomer in vacuum is more “compact” than in solution and its phenyl groups (Figure 4.3–[d]) are rotated of five additional degrees in opposite directions away from the double bond plane, to minimize the intramolecular steric repulsion between the phenyl rings. Conversely, in vacuum the *trans* isomer exhibits a more planar average structure with respect to that in solution: a possible explanation is that the flatter **conformation** of the *cis* isomer increases the contact surface and accordingly the number of non-bonded interactions with the solvent. On the other hand the same solvent interaction with the phenyls stabilizes higher  $|\gamma|$  values than in a vacuum for the *trans* isomer.

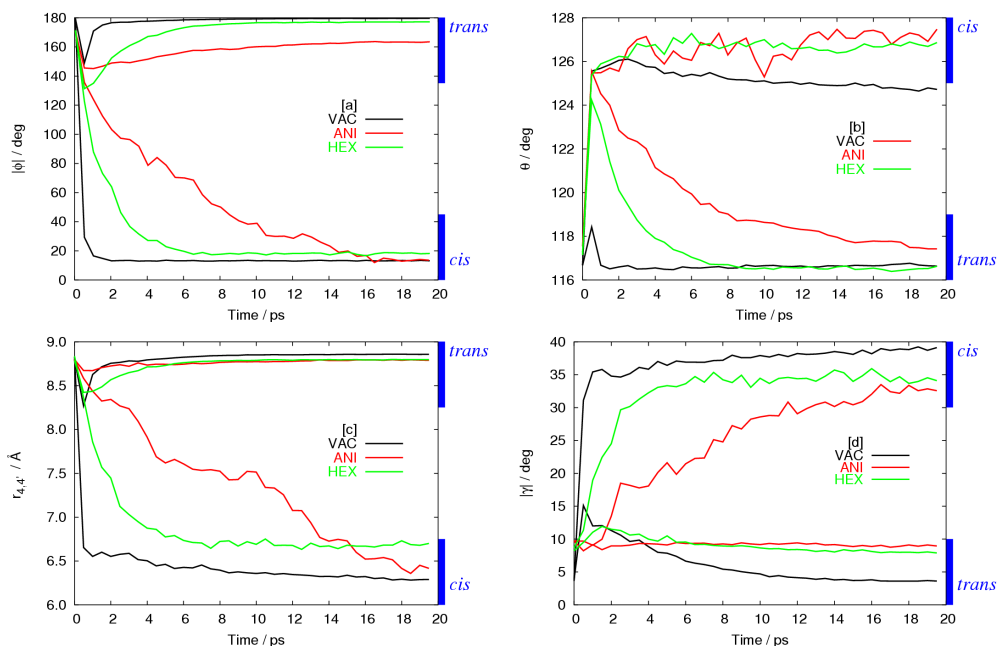


Figure 4.3: The evolution of averaged conformational indicators from the 1000 MD virtual photoisomerization experiments: [a] *Ph-N=N-Ph* torsional angle  $|\phi|$ ; [b] *N=N-C* bending angles  $\theta$ ; [c] phenyl-phenyl distance  $r_{4,4'}$ ; and [d] *C-C-N=N* torsional angles  $|\gamma|$ .

### 4.3.4 Permanence time in the excited state

We have also measured the average permanence time in the excited state as a function of the isomerization product, shown in Table 4.5. These times should be roughly proportional to the sum  $\tau_1 + \tau_2$ , and as  $\tau_1$  is similar for all solvation conditions, the main correlation is with  $\tau_2$ . Besides, we see that the permanence times change from one solvent to the other, but that the isomerization products are fairly independent on these times suggesting that the key step for a successful isomerization resides in the molecular **conformation** and the atomic velocities when the decay takes place (since these are conserved when the electronic state changes from  $S_1$  to  $S_0$ ). This observation and a direct inspection of the actual trajectories indicates the dominance of only one principal pathway in  $S_1$  state.

Table 4.5: The average permanence times  $\langle t_{S_1} \rangle$  (in ps) of the AB molecule in the  $S_1$  state for successful (*cis*) and unsuccessful (*trans*) MD virtual photoisomerization experiments ( $N = 1000$ ) in vacuum and in four different solvents.

	vacuum	anisole	<i>n</i> -hexane	MPE	toluene
<i>cis</i>	$0.44 \pm 0.04$	$6.2 \pm 0.5$	$1.9 \pm 0.1$	$1.4 \pm 0.1$	$1.3 \pm 0.1$
<i>trans</i>	$0.45 \pm 0.01$	$5.5 \pm 0.1$	$2.3 \pm 0.1$	$1.6 \pm 0.1$	$1.2 \pm 0.1$

### 4.3.5 Isomerization mechanism

We now proceed to assess the likelihood of the **torsion** or inversion *trans-cis* isomerization channels by following the time evolution of the molecular geometry from the MD simulations. To do this, we have first defined three classes of isomerization pathways in the  $(\theta, \phi)$  surfaces according to the upper value,  $\theta_{max}$ , of the bending angle values explored by the AB molecule during its  $\theta, \phi$  trajectory in both electronic states (see Figure 4.4): *pure torsion* (if  $\theta_{max} < 140^\circ$ ); *mixed* (if  $140^\circ \leq \theta_{max} < 160^\circ$ ); and *pure inversion* (if  $\theta_{max} \geq 160^\circ$ ). Then, we have evaluated the relative importance of each of these channels computing their occurrence for successful and unsuccessful events by the ratios  $\Theta_m = N_m/N_{cis}$  and  $\Theta_m = N_m/N_{trans}$ , reported in Table 4.6, where  $m$  refers to the isomerization mechanism (*tor, mix, inv*). It is relevant to remark that analyzing the preliminary results of the MD simulations used

to parameterize equation 4.2.3, we have also verified that the isomerization mechanism discussed in this section is scantily influenced by the specific values of  $\chi$ , supporting the validity of the following analysis.

In the top of table 4.6 we show the mechanism ratio for unsuccessful isomerizations (*trans-trans* process) and we observe a predominancy of the torsional movement (from 96% to 99%) in all the simulation conditions, and a small occurrence of the mixed one, while inversion was observed only in a single experiment in toluene. More interestingly, looking at the *trans-cis* ratios (table 4.6, bottom) we notice first that the  $\Theta_m$  ratios in vacuum indicate that the two favorite pathways for isomerization are the torsional and mixed ones, with a very rare occurrence of inversion, in agreement with the semiempirical surface hopping simulations by Ciminelli *et al.* [Ciminelli *et al.* 2004]. We also find that, even if the ratios exhibit different values in the various solvents, the most probable pathways in condensed phase always correspond to a mixed mechanism. Pure inversion and pure torsional pathways seem to be hindered by steric solvent interactions and consequently less likely to occur. In vacuum, where these limitations do not exist, the contribution of pure torsion is 2 – 3 times larger than that in *n*-hexane, MPE and toluene solutions, and the effect is even more marked in the more dense and viscous anisole, where torsion is almost negligible. Hence, we deduce that the inversion movement has a larger energy barrier and that this pathway can be followed only when the solvent significantly increases the height of the torsional barrier: in solution, a mixed mechanisms becomes the most efficient way to reduce the solvent steric hindrance during the decay process. The effect of more viscous solvents, like anisole, goes in the direction of increasing the pure inversion contribution, as observed experimentally in the comparison of femtosecond fluorescence results in *n*-hexane and ethylene glycol [Chang *et al.* 2004].

The analysis of MD simulations trajectories (see Fig. 4.4 for some examples) also gives insights of the state dependence of the inversion and rotation movements. When the molecule is excited, it generally moves torsionally towards the  $S_1$  surface energy minimum and the conical intersection region, shown as a circular region in Fig. 4.4. Only after the decay to the  $S_0$  state, the bending angle  $\theta$  can reach values close to  $180^\circ$ ; i.e. the inversion movement takes place primarily in the  $S_0$  state, at least under these excitation conditions.

This finding supports the “cold isomerization” model described by Diau and coworkers [Lu *et al.* 2002], that depicts a rotational pathway in the  $S_1$  state for an  $(n, \pi^*)$  excitation. The inversion mechanism in the excited state

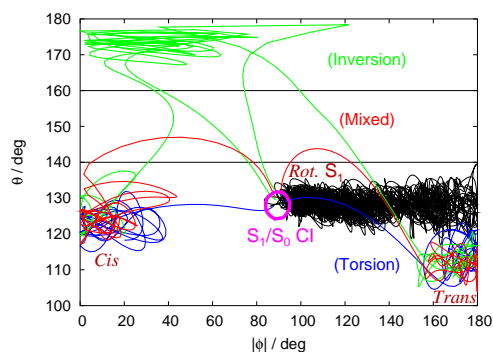


Figure 4.4: Examples of three isomerization pathways of AB in *n*-hexane corresponding to the **torsion** (blue), **mixed** (red), and **inversion** (green) mechanisms.

becomes possible only if the molecule starts from a higher vibrational state (“hot isomerization”) having enough energy to overcome the **inversion** energy barrier; this second mechanism is expected to be important only after a decay from a higher excited state, as it can happen in  $(\pi, \pi^*)$  isomerization.

### 4.3.6 Geometrical species identification

The next step will be to define, by average geometrical value, the conformational species introduced in the kinetic model (equation 4.3.4). The procedure to obtain this specific parameters calculate, for a fixed time  $t$ , the average of geometrical contribution to each species present to the total averaged parameter. The weight of each contribution is obtained directly from kinetic equations 4.3.7 substituting for each solvent determination the decay time obtained in section 4.3.2. We perform a conjuncted fitting for each

Table 4.6: The isomerization pathway indicators  $\Theta_m = N_m/N_{trans}$  (top) and  $\Theta_m = N_m/N_{cis}$  (bottom) computed from the 1000 MD virtual photoisomerization experiments of AB in vacuum and in four different solvents.

$m$	vacuum	anisole	<i>n</i> -hexane	MPE	toluene
<i>trans-trans</i> process					
<i>tor</i>	$0.98 \pm 0.01$	$0.99 \pm 0.01$	$0.96 \pm 0.01$	$0.97 \pm 0.01$	$0.96 \pm 0.01$
<i>mix</i>	$0.02 \pm 0.02$	$0.01 \pm 0.02$	$0.04 \pm 0.02$	$0.03 \pm 0.02$	$0.038 \pm 0.02$
<i>inv</i>	$0.00 \pm 0.02$	$0.00 \pm 0.02$	$0.00 \pm 0.02$	$0.00 \pm 0.02$	$0.002 \pm 0.02$
<i>trans-cis</i> process					
<i>tor</i>	$0.58 \pm 0.02$	$0.05 \pm 0.10$	$0.25 \pm 0.05$	$0.36 \pm 0.05$	$0.42 \pm 0.05$
<i>mix</i>	$0.41 \pm 0.03$	$0.84 \pm 0.10$	$0.65 \pm 0.05$	$0.59 \pm 0.05$	$0.53 \pm 0.05$
<i>inv</i>	$0.01 \pm 0.01$	$0.11 \pm 0.10$	$0.10 \pm 0.05$	$0.05 \pm 0.05$	$0.05 \pm 0.05$

geometrical parameter ( $|\phi|$ ,  $\theta$ ,  $r_{4,4'}$  and  $|\gamma|$ ) of both *trans* and *cis* series of data as follow:

$$F_{trans}(t) = P_{AB_1} * AB_1(t) + P_{AB_1^*} * AB_1^*(t) + P_{AB_0}^{trans} * AB_0(t) \quad (4.3.8)$$

$$F_{cis}(t) = P_{AB_1} * AB_1(t) + P_{AB_1^*} * AB_1^*(t) + P_{AB_0}^{cis} * AB_0(t) \quad (4.3.9)$$

where  $P_{AB_1}$  is the specified geometrical parameter ( $|\phi|$ ,  $\theta$ ,  $r_{4,4'}$  or  $|\gamma|$ ) in  $S_1$  state for  $AB_1$  specie identical for *trans* and *cis* geometrical trajectory,  $P_{AB_1^*}$  is the specified geometrical parameter in  $S_1$  for  $AB_1^*$  specie identical for *trans* and *cis*,  $P_{AB_0}^{trans}$  is the parameter in the  $S_0$  state for *trans*  $AB_0$  spacie and  $P_{AB_0}^{cis}$  the parameter for  $S_0$  state for *cis*  $AB_0$  specie. The merit function for fitting data is defined as the sum of equations 4.3.8 and 4.3.9 and was minimized the variance by 4 parameter simplex algorithm. In table 4.7 are shown the fitting results and an example of graphical fitting result in figure 4.5. The fit parameter obtained from this procedure reproduce correctly the geometrical parameter trend for both isomers, all solvents and for  $|\phi|$ ,  $\theta$ ,  $r_{4,4'}$  diagrams. This procedure seems to falls for  $|\gamma|$  expecially in the *trans* case. This indicate a different kinetic behavior of phenyl torsions slower respect the kinetic decay. The quantum yields are strictly related to the  $|\phi^{AB_1^*}|$  and  $r_{4,4'}^{AB_1^*}$  values: lower values of azoic torsional angle and para-para distance in  $AB_1^*$  conformation reduce the final quantum yield. We can also quantify the different conformations obtained for isomers in ground state: in vacuum the mean

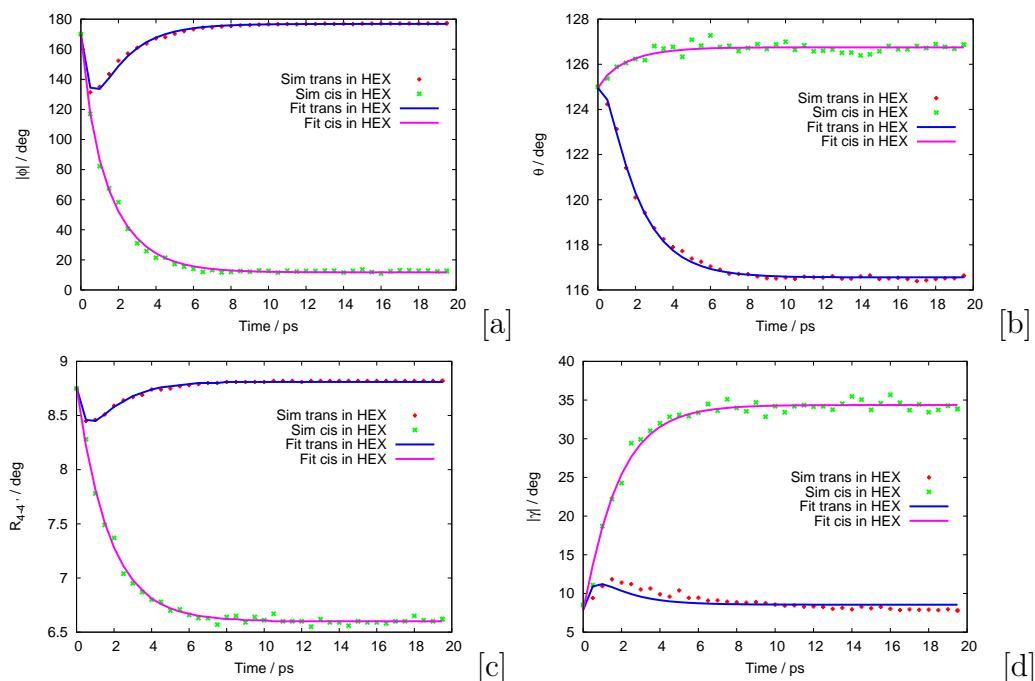


Figure 4.5: Geometrical means parameter fitting for hexane using functional forms in equation 4.3.8 and 4.3.9.

conformation of *trans* isomer is more planar respect the conformation in solvent showing a  $|\phi^{AB_0^{cis}}|$  greater than 7% in anisole and only 1% in the other solvents. Otherwise the *cis* conformer and less contracted in vacuum having a  $|\gamma^{AB_0^{trans}}|$  parameter about 18% greater than anisole and 5% respect the other solvents. We can also confirm the similarity of  $AB_1$  specie respect the  $AB_0$  *trans* one indicating the presence of the excited Franck–Condon adduct in the kinetic process. Bending angle values are less sensitive parameter for species description and are not strictly related to quantum yield. Perhaps in anisole are observed greater values.

## 4.4 Conclusions

In this work we have analyzed the mechanism for the photoisomerization of azobenzene in various organic solvents when excited in the visible ( $(n, \pi^*)$  *trans*–*cis* transition). In order to do this, we have proposed a method for the



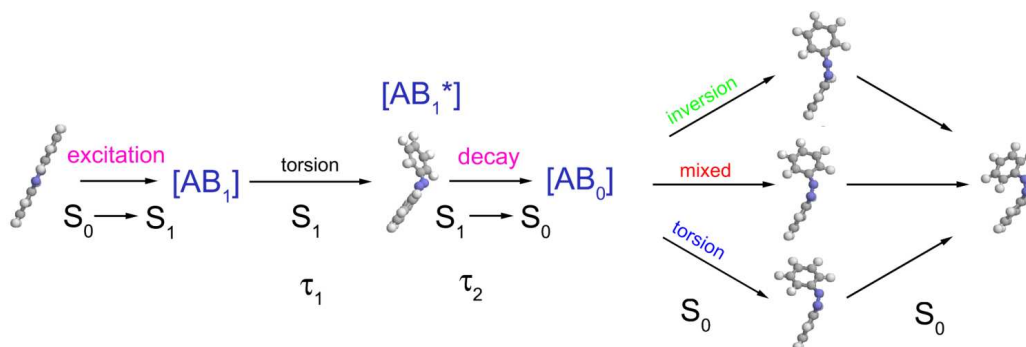


Figure 4.6: Summary of the *trans-cis* AB isomerization process in solution as emerged from the MD virtual experiments of this study: excitation, torsional movement in the  $S_1$  excited state (lifetime  $\tau_1$ ), decay at the CI to the ground state (lifetime  $\tau_2$ ), and possible relaxation pathways in the  $S_0$  state.

molecular dynamics simulation of photoresponsive molecules in solution that has the feature, essential for this problem, of allowing for their electronic excitation and statistical decay while they move in the crowded solvent environment.

Using this methodology we have been able to follow the conformational changes of AB in the presence of a solvent and to evaluate conversion efficiency and decay dynamics. The isomerization quantum yields obtained from the simulations indicate that the presence of the solvent strongly reduces the isomerization yield with respect to the gas phase. Amongst the solvent features, viscosity and density seem to play a significant role and further affect the isomerization quantum yield. The kinetic analysis of the decay times shows a bi-exponential behavior, in agreement with ultrafast spectroscopy results, and we have identified in the fast process the geometry reorganization in the  $S_1$  state occurring just after the excitation, and in the slow process the lifetime of the molecule near the conical intersection with the  $S_0$  state (see Figure 4.6). The average decay times are fairly independent on the photo product obtained after decay, thus indicating similar trajectories in the excited state.

The simulations also show that, while in vacuum the isomerization follows prevalently a torsional mechanism, as already found by other authors [Ciminelli et al. 2004], the dominant mechanism in solution is a mixed torsional-inversion one. Indeed the pure inversion seems to occur only after the decay into the  $S_0$

state, while in the excited state it appears unlikely, as a high energy barrier must be overcome. A higher solvent viscosity increases the pure inversion contribution, albeit the principal mechanism remains the mixed one.

Although a number of approximations have been introduced to make the study computationally feasible, our approach is quite general and can be applied for various levels of quantum mechanics description of the electronic states (e.g. using [PCM](#) solvent-specific potential energy surfaces) and transition probabilities, and to more systematic studies of the effect of specific solvent physical and chemical properties on isomerization.

Table 4.7: Geometrical descriptor for  $AB_1$ ,  $AB_1^*$ ,  $AB_0$  *trans* and  $AB_0$  *cis* kinetic species.

VAC				
Kinetic times: $\tau_1 = 0.08$ ps; $\tau_2 = 0.35$ ps			Quantum yield: 0.51	
Species	$AB_1$	$AB_1^*$	$AB_0$ <i>trans</i>	$AB_0$ <i>cis</i>
$ \phi $ (deg)	178.11	71.70	178.11	12.40
$\theta$ (deg)	116.59	125.10	116.59	125.78
$r'_{4-4}$ (Å)	8.83	7.15	8.83	6.39
$ \gamma $ (deg)	5.80	28.32	5.83	37.51
ANI				
Kinetic times: $\tau_1 = 0.48$ ps; $\tau_2 = 5.80$ ps			Quantum yield: 0.06	
Species	$AB_1$	$AB_1^*$	$AB_0$ <i>trans</i>	$AB_0$ <i>cis</i>
$ \phi $ (deg)	163.16	140.18	165.38	10.08
$\theta$ (deg)	121.21	125.32	120.57	127.29
$r'_{4-4}$ (Å)	8.78	8.67	8.80	6.75
$ \gamma $ (deg)	9.36	7.96	9.41	33.14
HEX				
Kinetic times: $\tau_1 = 0.43$ ps; $\tau_2 = 1.73$ ps			Quantum yield: 0.11	
Species	$AB_1$	$AB_1^*$	$AB_0$ <i>trans</i>	$AB_0$ <i>cis</i>
$ \phi $ (deg)	175.99	106.56	176.89	11.74
$\theta$ (deg)	116.96	127.20	116.50	126.69
$r'_{4-4}$ (Å)	8.82	8.22	8.81	6.60
$ \gamma $ (deg)	7.94	12.96	8.56	34.36
MPE				
Kinetic times: $\tau_1 = 0.52$ ps; $\tau_2 = 0.98$ ps			Quantum yield: 0.14	
Species	$AB_1$	$AB_1^*$	$AB_0$ <i>trans</i>	$AB_0$ <i>cis</i>
$ \phi $ (deg)	176.64	91.07	177.15	12.52
$\theta$ (deg)	117.00	129.15	116.47	126.47
$r'_{4-4}$ (Å)	8.83	8.02	8.82	6.58
$ \gamma $ (deg)	7.96	14.84	8.94	34.83
TOL				
Kinetic times: $\tau_1 = 0.41$ ps; $\tau_2 = 0.74$ ps			Quantum yield: 0.13	
Species	$AB_1$	$AB_1^*$	$AB_0$ <i>trans</i>	$AB_0$ <i>cis</i>
$ \phi $ (deg)	176.68	105.78	176.74	12.81
$\theta$ (deg)	116.76	129.29	116.47	126.50
$r'_{4-4}$ (Å)	8.82	8.20	8.81	6.58
$ \gamma $ (deg)	8.28	11.12	8.90	34.73



# Chapter 5

## 4,n-alkyl,4'-cyano-biphenyls simulations

### 5.1 Introduction

In the last years the liquid crystal materials began to be very important for scientific and technological applications and the atomistic simulation can be a useful method for the understanding their microscopic properties. The advantages of full atomistic (or coarse grained) simulation models are the complete (or partial) knowledge of molecular details that can suggest to a synthetic chemist the modification necessary for obtaining of the searched chemical-physical properties. The limitation up to now was related to computational power available. The exponential increase of the computer performance begins to be sufficient to allow studies like transition temperature determination and time correlation functions of the molecular reorientation, still impossible to realise only few years ago.

Following the promising results of the study of the odd–even effect in the transition temperature of an aminocinnamate series [Berardi et al. 2004], we focus here on obtaining and characterising the phase transitions of the series of 4,n-alkyl,4'-cyano-biphenyls for  $n = 4, 5, 6, 7, 8$ , starting from a “hot” and isotropic configuration and inducing order simply by a temperature reduction. For this purpose, we have derived a modification of the AMBER/OPLS united atoms (UA) force field to simulate a sample of  $N=250$  molecules at  $NPT$  conditions ( $N=250$  molecules,  $P=1$  atm) scanning temperature around the experimental nematic-isotropic transition temperature ( $T_{NI}$ ).

## 5.2 Simulation models

The difficulty in the study of mesogenic properties of this class of compound is the strong necessity to simulate the greatest number of molecules as possible. The properties of a phase are determined by the enormous number of interaction between the molecules. The main contributions are directly related to the coulombic and dispersive interaction generally considered between single couples of atoms. The estimation of these interactions are often chose considering the general chemical properties of similar atoms studied in other context. For example for an aliphatic chain there are well known intramolecular parameters obtained from simulations of boiling point and vapourisation enthalpy and specific head of alkanes. These set of parameters are general and in some, case are considered universal. This approach could be dangerous in the particular case where the target properties involve very small energy change, like the nematic to isotropic phase transition of a liquid crystal. In this situation slight energy evaluation errors could produce large errors in the searched observable. The approach used in this work will be to complete the standard force field present in literature and improve the interaction that cannot be considered as universal. All the simulations are performed using the united atoms models for increasing the number of molecules and to try to reduce all the problems related to the sample size. On the other hand a less detailed model reduces the number of information about the atomic detail, but permits us to simulate at a wider number of temperature and subsequently calculate and predict the temperature dependence of our properties.

### 5.2.1 Force field

We have performed a preliminary quantomechanics analysis over all the molecules studied. The aims are to obtain all the FF terms not included or not adequate in the standard AMBER. The FF model used in this work include the following potential energy expressions:

$$U_{\text{total}} = U_{\text{bonds}} + U_{\text{angle}} + U_{\text{dihed}} + U_{\text{LJ}} + U_{\text{charge}} \quad (5.2.1)$$

$$U_{\text{bonds}} = \sum_{\text{bonds}} K_r^{t_i t_j} \left( r_{ij} - r_{eq}^{t_i t_j} \right)^2 \quad (5.2.2)$$

$$U_{\text{angles}} = \sum_{\text{angle}} K_{\theta}^{t_i t_j t_k} \left( \theta_{ijk} - \theta_{eq}^{t_i t_j t_k} \right)^2 \quad (5.2.3)$$

$$U_{\text{dihed}}^{\text{std}} = \sum_{\text{dihed}} V_{\phi}^{t_i t_j t_k t_l} [1 + \cos(n^{t_i t_j t_k t_l} \phi_{ijkl} - \gamma^{t_i t_j t_k t_l})] \quad (5.2.4)$$

$$U_{\text{dihed}}^{\text{ext}} = \sum_{\text{dihed}} V_{\phi}^{t_i t_j t_k t_l} \left[ K_0 + \sum_{n=1}^6 K_n^{\text{cos}} \cos(n\phi_{ijkl}) + \sum_{n=1}^3 K_n^{\text{sin}} \sin(n\phi_{ijkl}) \right] \quad (5.2.5)$$

$$U_{\text{LJ}} = 4 \sum_{i < j} f_{LJ}^{1,4} \epsilon_{t_i t_j} \left[ \left( \frac{\sigma_{t_i t_j}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{t_i t_j}}{r_{ij}} \right)^6 \right] \quad (5.2.6)$$

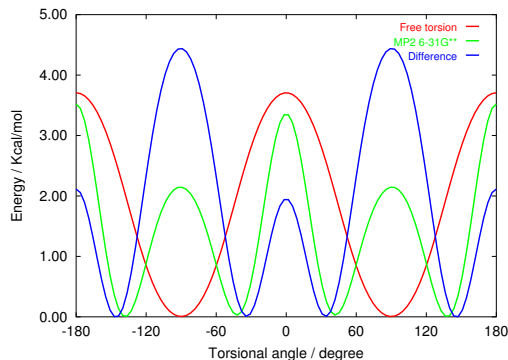
$$U_{\text{charge}} = \sum_{i < j} f_q^{1,4} \frac{q_i q_j}{r_{ij}} \quad (5.2.7)$$

The first calculation were performed to obtain punctual atomic charges on each UA center for each species. We use a quantum mechanical DFT B3LYP/6–31G\*\* geometry optimisation, and determined the atomic point charges or **united atoms** charges of atomic groups, for  $CH_n$ , group using the ESP scheme with the additional constraint of reproducing the total dipole moment [Besler et al. 1990].

The standard energy term for the torsional degree of freedom ( $U_{\text{dihed}}^{\text{std}}$  Eq. 5.2.4), are extended adding a more general expression ( $U_{\text{dihed}}^{\text{ext}}$  Eq. 5.2.5) able to better reproduce the torsional angle between the two phenyl groups. For this purpose we fitted the torsional potential energy surface (tPES) parameters ( $K_0$ ,  $K_n^{\text{cos}}$  and  $K_n^{\text{sin}}$ ) in Eq. 5.2.5 computing the quantomechanical torsional (MP2/6 – 31G\*\*) energy for the single molecule of 5CB (Fig. 5.1). This terms for the other molecules of serie is considered invariant.

For a correct estimation of the new potential term, we have to know the difference between the quantomechanical potential (total energy) and the total molecular mechanics energy in function of  $\phi$ , calculated with a null torsional term. To do this, we compute the Boltzmann  $P(\phi, T)$  distribution from a simulation in vacuum at the temperature of 2000 K set the standart torsional term for phenyls to zero. After this operation we derive the correction terms using  $U(\phi)_c = -k_B T (\ln P(\phi, T) + \ln A)$  where  $U(\phi)_c$  is the correction to be subtracted to the QM torsion curve to obtain the new tPES term,  $k_B$  is the Boltzmann constant and A is a constant term to set the minimum of tPES to zero. Similar treatement was done for alkyl–chain torsional

Figure 5.1: Phenyl–phenyl torsional potential.



degrees of freedom.

## 5.2.2 Tuning the molecular force field

Preliminary runs for a sample of 250 molecules of 5CB described with the standard UA AMBER/OPLS FF parameter set for the aromatic region and NERD parameter set for aliphatic part revealed that it was not possible to obtain a **nematic isotropic** (NI) transition for simulations at temperature close to the experimental value. For this reason we undertook a large scanning in temperature above (until 500 K) and below (until 200K) the experimental transition value of 308.7 K ( $T_{NI}^{5CB}$ ) and we obtained an increase of order in a condensed phase for temperatures about 120 K higher than the experimental value (Fig. 5.2), similarly to what observed by other authors for 8CB [McDonald and Hanna 2004; 2006].

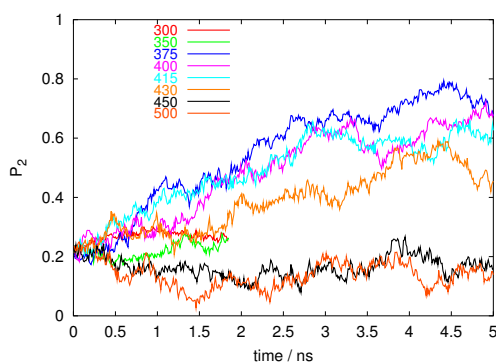
The overestimation of the transition temperature means that intermolecular interaction energy is overestimated and becomes necessary, for reproducing correctly the experimental NI transition, reduce the amount of this energy. In the force fields like AMBER the intermolecular interactions is described by the Lennard–Jones and the electrostatic terms (equations 5.2.6 and 5.2.7). To determine the relative influence of the two terms on the transition temperature we ran some trial simulations of 5CB molecules, setting to zero the atomic charges. Even if the removal of the electrostatic part produced remarkable differences in the atom–atom radial distribution at close range, the  $T_{NI}$  temperature estimated in these conditions resulted only about



5 K lower than the one obtained with the full set of charges.

After this test, we decided to leave unaltered the electrostatic term and to adjust only the Lennard–Jones parameters  $\sigma$  and  $\epsilon$  of aromatic and aliphatic carbons, which were modified to reproduce in a first step the experimental transition temperature of 5CB, and subsequently the  $T_{NI}$  of all the members of the series ( $n = 4 - 8$ ).

Figure 5.2: Order parameter evolution using standard LJ parameters



In the choice of the best LJ parameter set it is possible to find many different combinations of  $\epsilon$  and  $\sigma$  able to reproduce the correct temperature trend of the density, also considering that we have four atom types to play with. In this parameter space, presumably only a restricted number of sets are able to reproduce the experimental  $T_{NI}$ . The effect of FF on properties depending from a high number of molecules, like density or phase transition temperatures, are in practice impossible to predict quantitatively only with the knowledge of FF, but a series of simulations (with equilibration and production phase) is required to calculate and compare simulations results with the experimental data. As the determination of the  $T_{NI}$  from a number of independent simulations equilibrating at different temperatures is difficult to automatize, the optimization has been done heuristically by systematic reduction of  $\epsilon$  terms using the same  $\Delta\epsilon$  ( $\Delta\epsilon = -0.1 \rightarrow \Delta T_{NI} \simeq 30K$ ) both for aliphatic and aromatic carbons to reduce the simulated transition temperature, followed by a increasing of the same  $\Delta\sigma$  values to keep the density trend similar (for  $\Delta\sigma = 0.05 \rightarrow \Delta\rho \simeq 0.01$ ) to the experimental one. After many iterations and a satisfactory result for 5CB, we have simulated all the members of series  $n$ CB (from  $n = 4-8$ ), this time adjusting the

$\epsilon_{aliphatic}/\epsilon_{aromatic}$  and  $\sigma_{aliphatic}/\sigma_{aromatic}$  ratios, taking into account that the ratio of aliphatic/aromatic carbons is different for each  $nCB$ .

### 5.3 Phase transition determination

The increasing of order in the samples is very slow (see next section for the sagacities used for reducing it), thus implying that we needed more than 15-20 ns for obtaining **nematic** phases; even longer times were needed for 4CB. Also experimentally the 4CB shows difficult to give the order transition, showing only a monotropic NI transition on cooling. In addition, when the simulation temperatures are too low ( $< 60 - 80$  K) with respect to the simulation  $T_{NI}$ , the system remains locked in a meta-stable state and this glassy sample is unable to self assemble (if the starting configuration is **isotropic**) or to melt (if the starting configuration is **nematic**). The simple check of simulation properties like density or energy is thus not sufficient to determine if the system is under cooled (not equilibrated) or not (equilibrated). The non zero values of order parameter for temperature above the NI transition derive principally from the relatively small size of the samples. The theoretical lower value for our systems can be calculated using the simple relation  $\langle P_2 \rangle_{min} = 1/\sqrt{N}$  where  $P_2$  is the second order Legendre polynomial of the function  $\cos_2\theta$  with  $\theta$  the angle between the molecular orientation with respect the director  $\vec{D}$  of the sample,  $N$  is the number of molecules and in our case  $\langle P_2 \rangle_{min} \simeq 0.06$ . This minimum value for the order parameter is reached for temperatures 15–20 K higher than the transition temperature. An explanation could be found in the strong tendence of the molecular dipoles to form dimers also in the **isotropic** phase far from transition. This feature can represent the experimental pre–transitional effect. In practice is possible to observe the presence of small cluster of ordered molecules. During the simulation these aggregates (composed of 10–30 molecules) can induce a small local order for short range of time increasing the mean value of the order parameter. In a big sample these small aggregates (much smaller than well known **nematic** domains) are oriented randomly and the resulting phase order is null. In our case the presence of a few of this structure influence the order of all the phase. effect is also due to the small dimension of the samples.

Another trouble, in the determination of the transition temperature, is the wide oscillation of order parameter at temperature close to the transition. These oscillation are strictly related to the range of thermostat applied. On

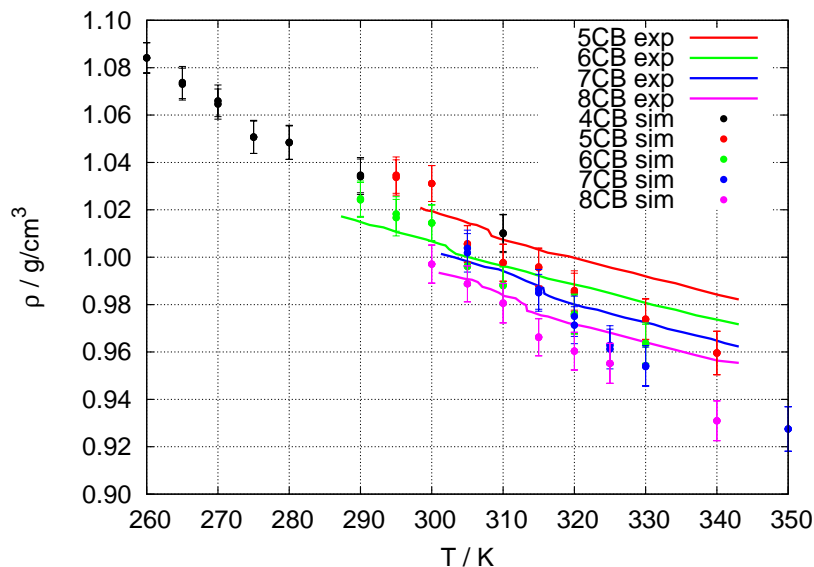
the other hand, this simulation parameter could be enough big to reduce the risk of locking our systems in a meta-stable states. After these considerations we decided to use a thermostat range of 10 K during equilibration phase and 5 K during the production one.

## 5.4 Simulation details

The samples used in the MD simulation were composed of 250 *nCB* **united atoms** molecules in a cubic box, a system size sufficient to describe about four solvation shells. Every sample has been equilibrated by simulation in *NPT* condition ( $P=1$  atm,  $T=260-350$  K) using velocity scaling thermostat, **isotropic** Parrinello-Rahman barostat[Parrinello and Rahman 1980] and periodic boundary conditions (PBC). The Particles Mesh Ewald parameters used are  $\alpha = 0.4 \text{ \AA}^{-1}$ , grid parameter 30  $\text{\AA}$  for each box side and 3<sup>th</sup> order spline, the scaling factor for 1-4 electrostatic intramolecular interaction is set to 0.8333 and 0.5 for the 1-4 intramolecular Lennard Jones. The time step used is 10 *fs* with multiple time step of 1 *fs* for intramolecular time step and 1 *fs* for non bonded interactions with 10  $\text{\AA}$  shall. The neighbor list was updated every 50 fs with a cut-off of 1.5  $\text{\AA}$ , and the Smoothed Particle Mesh Ewald parameters used are  $\alpha = 0.4 \text{ \AA}^{-1}$ , with cutoff radius of 30  $\text{\AA}$  and using a 3 - *th* order interpolation splines.

For each *nCB* we started from a hot and disordered configuration at temperature of 360 K and we cooled instantaneously at the desired temperature (from 280 to 350 K). We did a series of eight temperatures for each homologue around the experimental transition temperature with a temperature difference of 10 K between each simulation. During the equilibration phase, we have changed some simulation conditions for reducing the equilibration time. The thermostat scaling range was set to 10 K to permit a temperature oscillation able to jump meta-stable states. We have also modified the coupling frequency of barostat from 200 to 500  $\text{cm}^{-1}$  to increase the frequency of volume adjustments. We have found that the modification of this two parameters reduces about 10 times the equilibration time from an **isotropic** to **nematic** phase. During the production phase we have run a series of seven temperatures around the simulated transition temperature obtained with a temperature separation of 5 K. The thermostat range are now set to 5 K and the coupling of barostat are set to 200  $\text{cm}^{-1}$ . The time evolution of order parameters obtained from the simulation are showed in Fig. 5.4, the relative

Figure 5.3: Temperature dependence of density value of simulated and experimental  $n$ CB series. Experimental data for 4CB not available.



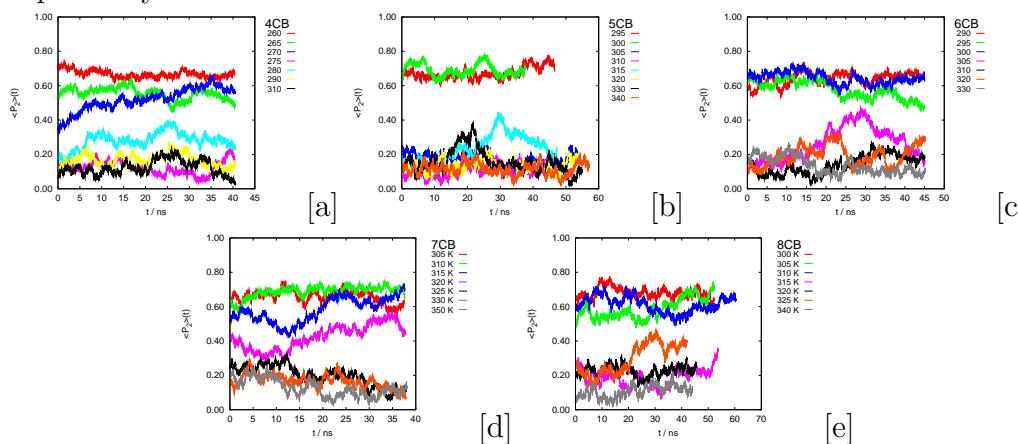
phase diagrams in Fig. 5.5.

## 5.5 Results

### 5.5.1 Phase transitions

The first results obtained from simulation are the determination of the  $T_{NI}$  with an error of  $\pm 5 K$  respect the experimental values and density comparable (around 2–5%) with real value for all the molecule of the series using the same force field. This result is very important considering that all the samples are obtained from disordered configuration simply reducing the simulation temperature without external field and considering both dispersive and electrostatic interactions. From the phase diagram is clear to observe the small odd–even effect in the series smaller than 10  $K$  and for this reason difficult to reproduce by simulations. The stability of the nematic phase are complete for temperature below 10–15  $K$  the phase transition (see Fig. 5.4). The same stability consideration, for the order in the sample, can be done

Figure 5.4: Order parameter evolution during production run. The time evolution of order parameter of sample just below and just up the transition temperature shows sensible oscillation during time. This requires very long simulation (more than 40 ns for each temperature) to obtain good averaged properties. From [a] to [e] are shown the evolution from 4 to 8CB respectively.



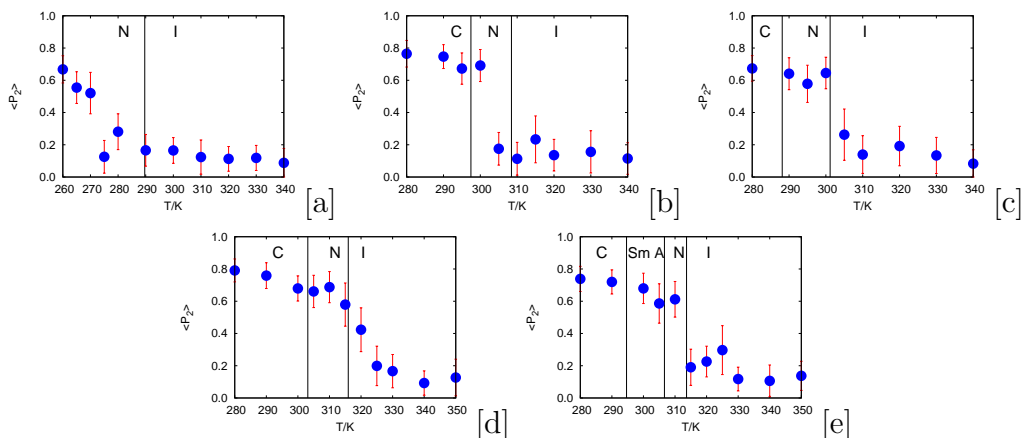
for the transition. The simulations around the transition shows wide oscillation of order parameter. This problem can be reduced by narrowing of thermostating range of simulation. Experimentally the temperature resolution normally achievable is around  $\pm 0.1 K$  much lower until respect to our value of  $\pm 5 K$ . Despite this, we preferred to do not reduce this range to be sure that our samples was not locked in a metastable state unable to reach the thermodynamic equilibrium.

## 5.5.2 Radial distributions

For a detailed description of the order properties and structure of the phases simulated, we have used as customary radial distribution functions (rdf)  $g_0(r)$ ,  $g_1(r)$  and  $g_2(r)$  (see App. D).

In the case of atomistic simulation, is possible to define in many different way the reference point for the calculation of each radial distribution. This choice is strictly related to the meaning of the distribution function and can be tuned for the observation of specific interactions between the molecules.

Figure 5.5: Simulated phase diagrams for  $n$ CB series. The averaged order parameter for 4CB are not clearly, but the monotropic behavior of experimental ([BDH 1986a]) transition indicates the small tendency to obtain stable **nematic** phase for this compound. The transition temperatures are well defined from 5 to 8CB (Fig. [b]–[e]) and very close to the experimental values (vertical lines in the diagrams [BDH 1986b, Roche 1988, BDH 1986c;d]).

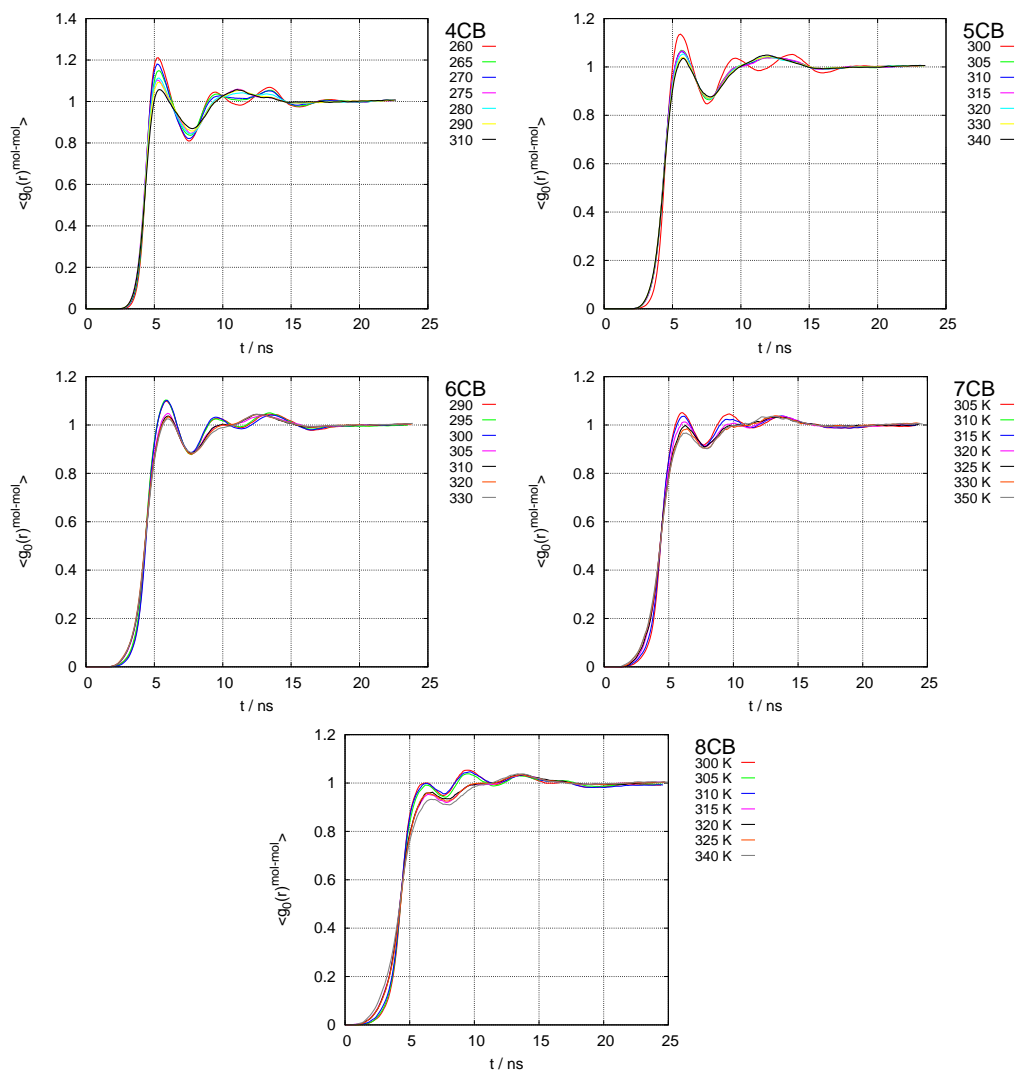


In the following we show the radial distribution obtained using as reference center the center of mass, the nitrogen atom and the cyano group for all the simulated  $n$ CB series members and temperatures.

### Center of mass distributions

From rdf computed with respect to the centers of mass of molecules (Fig. 5.6) we observe the presence of three/four sharp maxima (5–6, 9–10, 14–15 and 18–19 Å increasing with  $n$ ) slightly for all the  $n$ CB series, indicating the high order in the **nematic** phases respect what is observed in the **isotropic** phases where only two wide solvation spheres the first at 5–6 Å and the second from 12 to 14 Å. In the series is present a sensible constant decrease of first peak intensity for **nematic** and **isotropic** phases depending on the increase of molecular flexibility with the increase of alkyl chain length. This means that the side–side interaction is reduced by the deformation of flexible parts of molecules that tend to separate the molecules. The second peak in the **nematic** phase is less sensitive than the first one, indicating that the center of mass distance for antiparallel configurations that is less interested by the

alkyl conformation.

Figure 5.6:  $g_0(r)$  mass center–mass center distributions.



**$N_1-N_1$  and  $N_1 - C_2$  bond  $N_1 - C_2$  bond distributions**

The rdf of nitrogen atoms (Fig. 5.7) are very structured. This is due to the strong charge interactions between the highest charged atom region (on  $N_1$ ,  $C_2$  and  $C_3$ ) where is located the main contribution to the molecular dipole moment. This property, in a first approximation, is slightly affected by the length of alkyl chains, but only by the structure of the phase. In the **nematic** phase the first maximum at 3.25 Å is very pronounced and sharp, but completely disappears in the **isotropic** phase. This derives from the antiparallel configuration at low temperature. The increase of thermal energy above the transition breaks or induces wide oscillations of the dipolar couples formed in the **nematic**, producing a large reduction of  $N_1-N_1$  rdf. Other two maxima can be well identified in the **nematic** at 8.5 and at 12.5 Å, and are present, but less pronounced, also in the **isotropic** phase. The attribution of these peaks is difficult because they seem to be the overlap of different contributions, but they could be related to the head-tail dipolar conjugation.

To obtain more details of the morphology of the phases, we focus our attention on the 0 - *th* and 1 - *st* order rdf functions of  $N_1 - C_2$  bonds. These two properties combined can clarify the positions and the orientation of the molecular dipoles. As observed from the  $N_1-N_1$  rdf, the dependence of  $g_0(r)^{N_1-C_2}$  and  $g_1(r)^{N_1-C_2}$  on the different molecule of the series is small for the same reasons explained before. In all the  $g_0(r)^{N_1-C_2}$  (Fig. 5.8) are present a first maximum at 3.25 Å that is without any doubt due to the antiparallel configuration of the first neighbors strongly evident in the  $g_1(r)^{N_1-C_2}$  diagrams (Fig. 5.9) where is present a minimum (negative value of  $g_1(r)$  function corresponds to an antiparallel alignment and viceversa) at the same distance. In the 0 - *th* order functions in **nematic** phases is present a small relative maximum around 7 Å, related to the parallel configuration of head-tail configurations. This kind of configuration is not very frequent as revealed from the intensity of the  $g_0(r)$  maximum but becomes more important with higher terms of series. Are also visible in **nematic** phase other shells at 9, 13 and 17.5 Å due to the succession of parallel-antiparallel molecules.

Finally we want to spend some words about the second rank rdf. First of all the a long range order is observed inside the **nematic** phase and the tendency to the  $g_2(r)^{N_1-C_2}$  reach to a constant value. This feature is characteristic to a really ordered **nematic** phase. Theoretically [Zannoni 2001] the plateau value must be equal to the  $\langle P_2 \rangle^2$  of the sample and is satisfied in our

simulations. The oscillations due to the alternation of antiparallel/parallel molecules are also visible in these plots.

Figure 5.7:  $g_0(r)$   $N_1-N_1$  distributions.

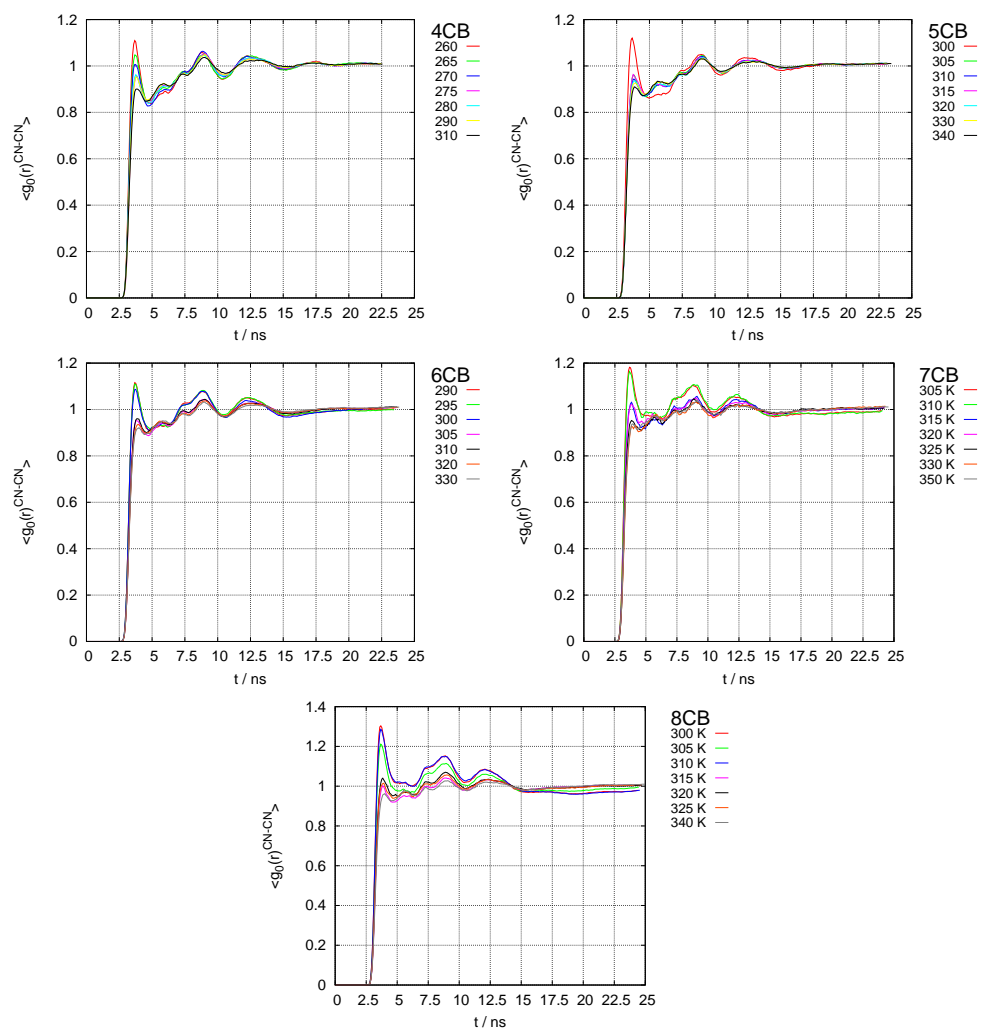


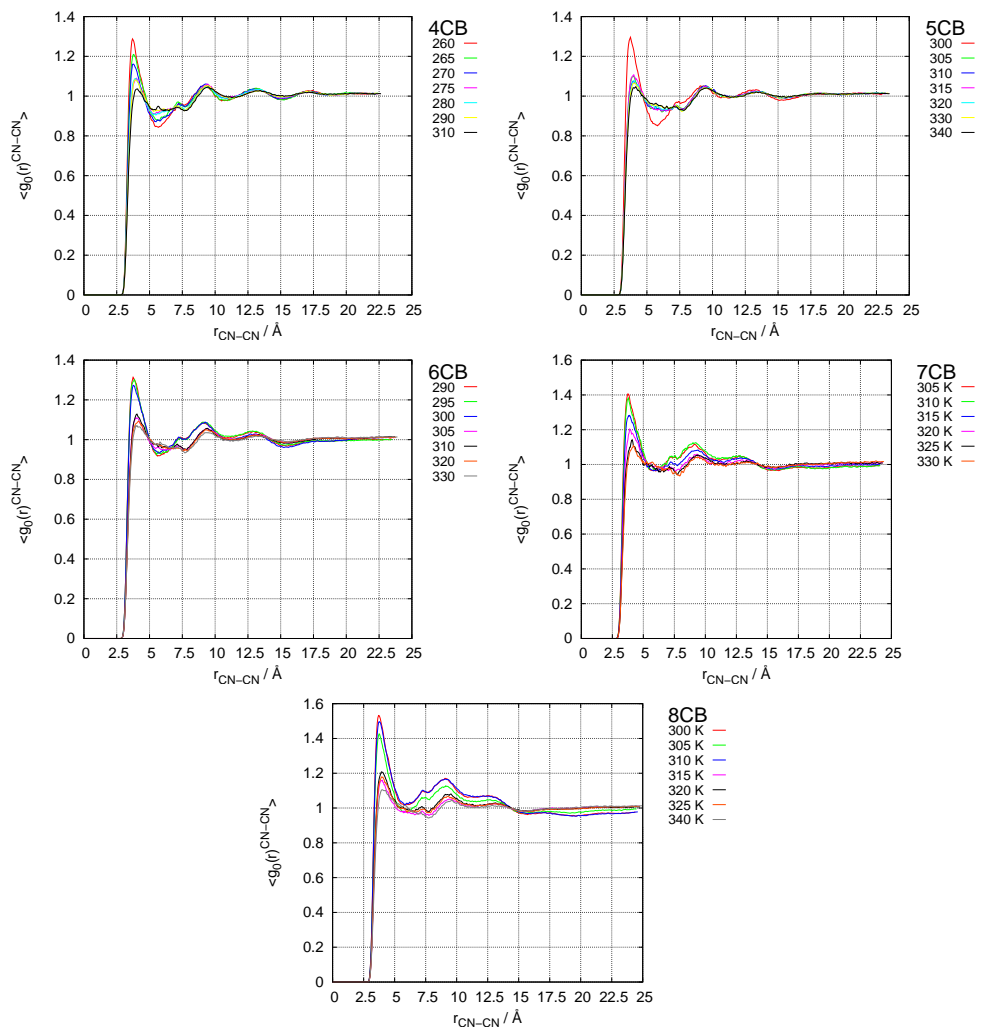
Figure 5.8:  $g_0(r)$   $N_1 - C_2$  bond  $N_1 - C_2$  bond distributions.

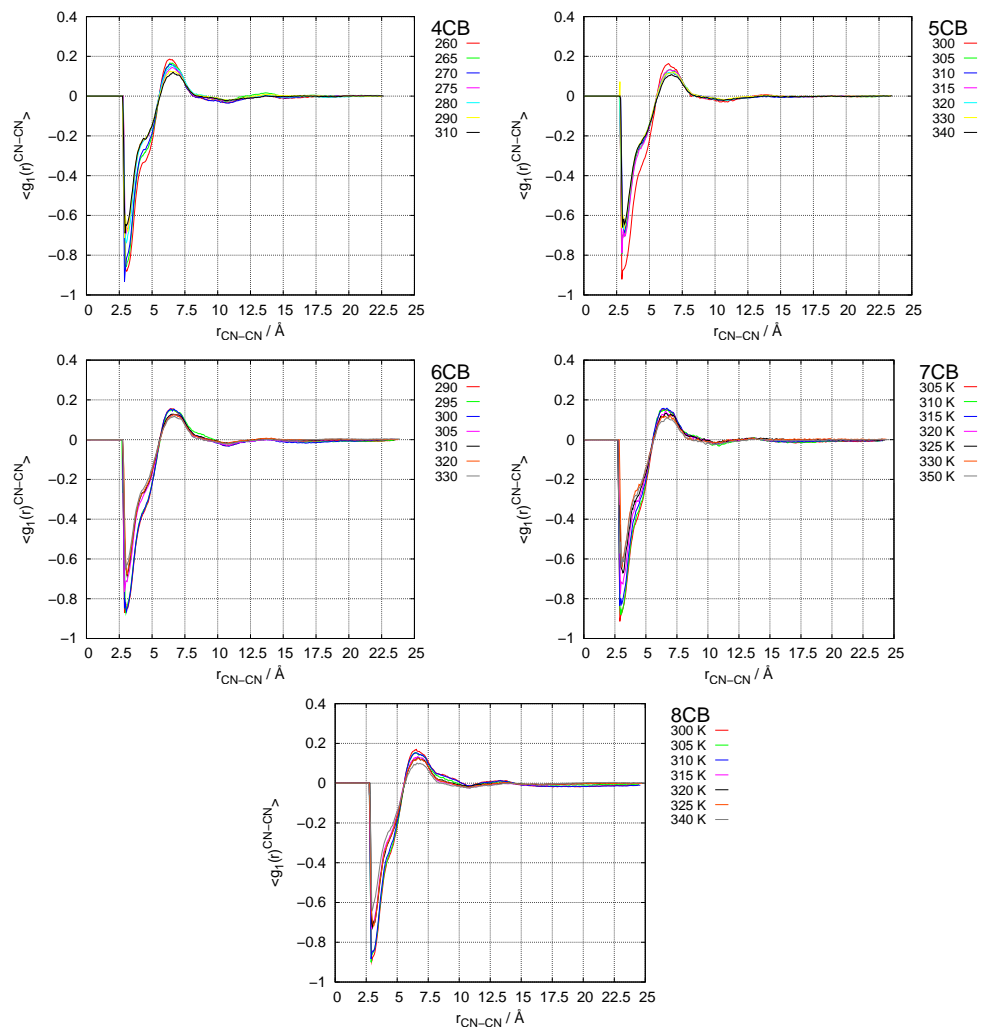
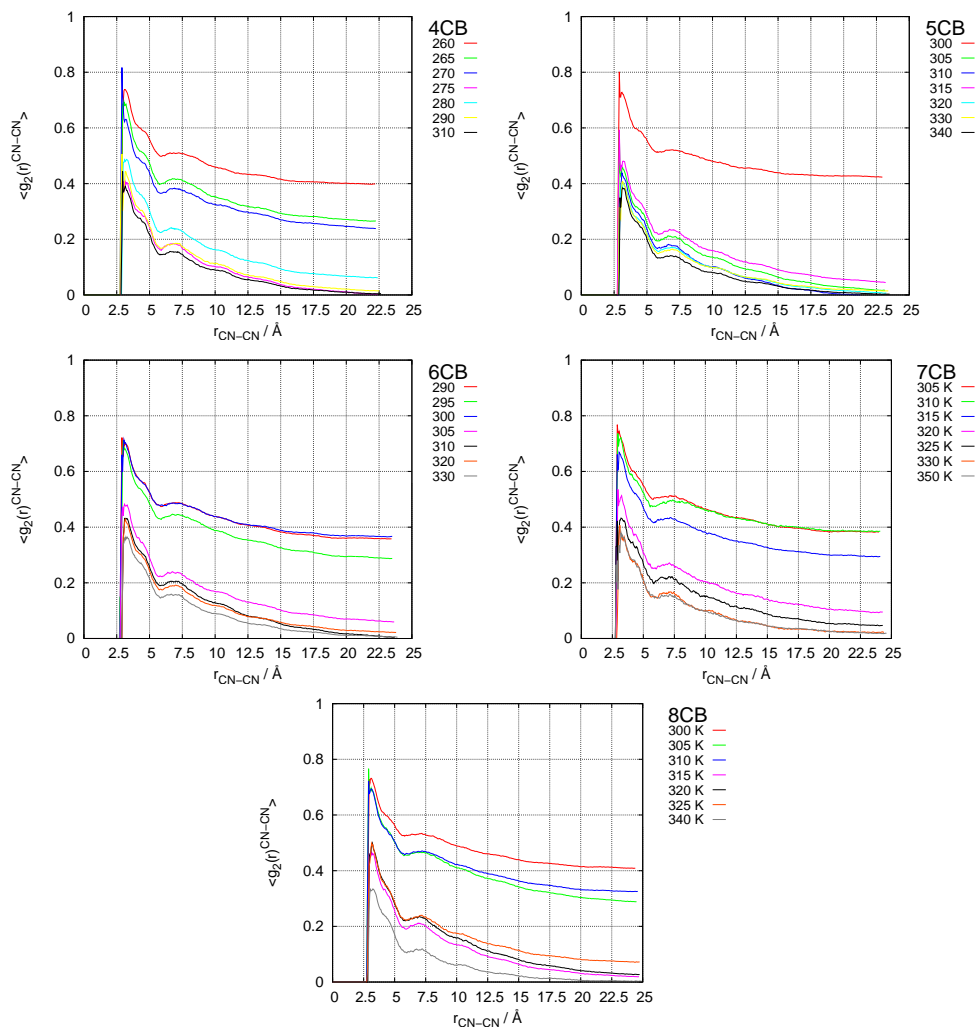
Figure 5.9:  $g_1(r)$   $N_1 - C_2$  bond  $N_1 - C_2$  bond distributions.

Figure 5.10:  $g_2(r)$   $N_1 - C_2$  bond  $N_1 - C_2$  bond distributions.

## 5.5.3 Length analysis

Figure 5.11: Averaged long and short axes molecular lengths.

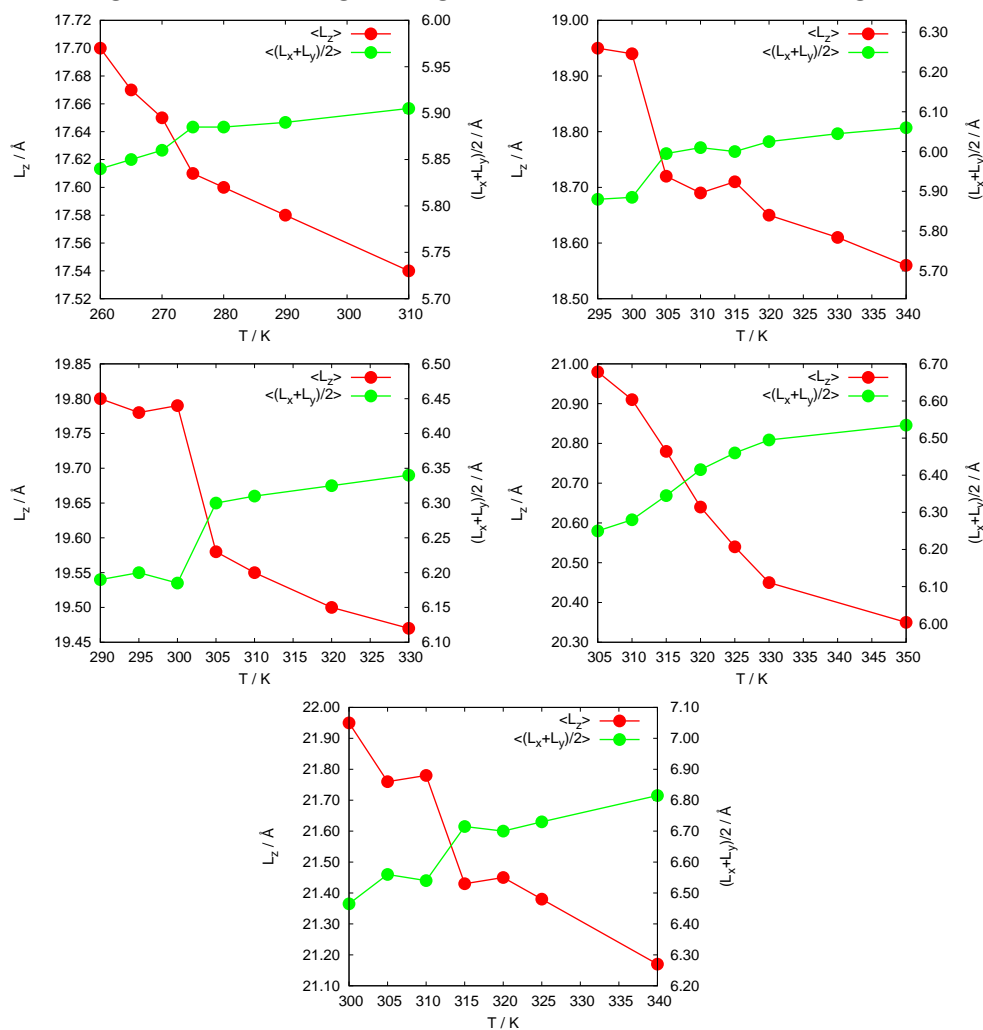


Figure 5.12: zLong axis length distributions.

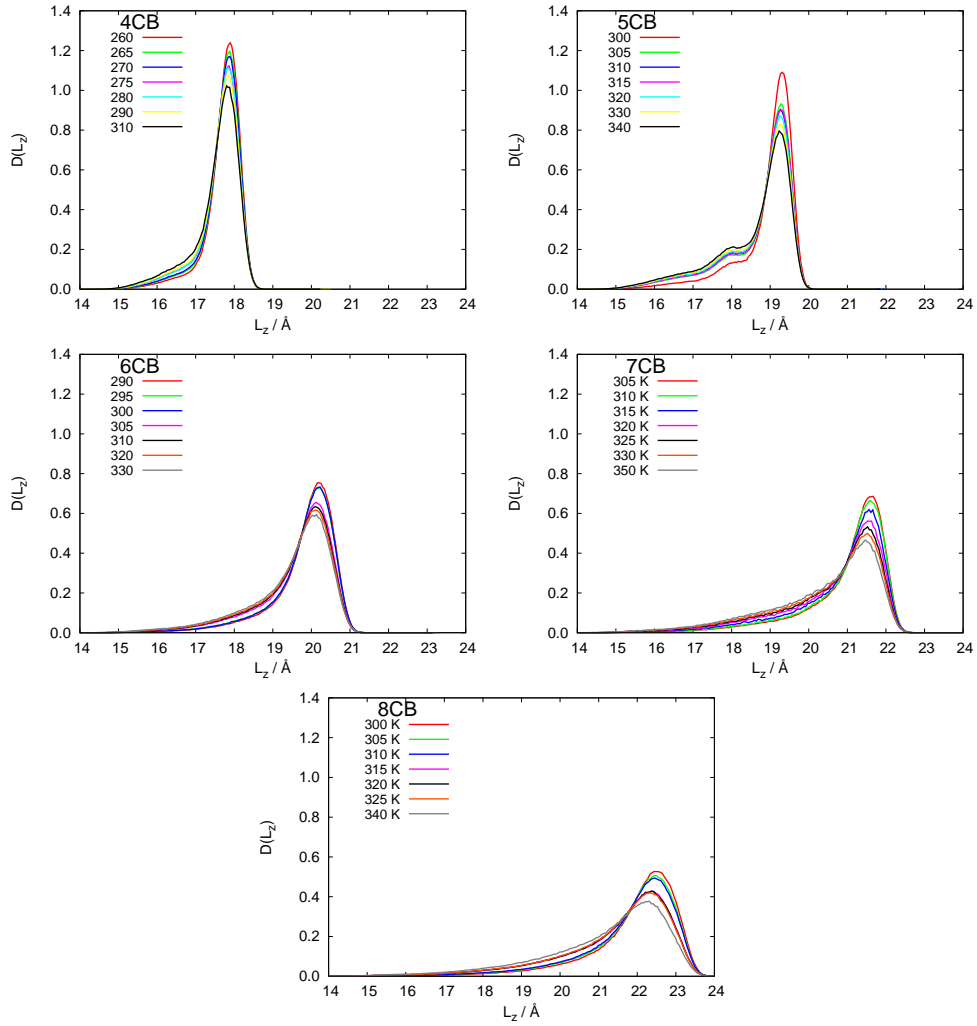


Figure 5.13: y-axis length distributions.

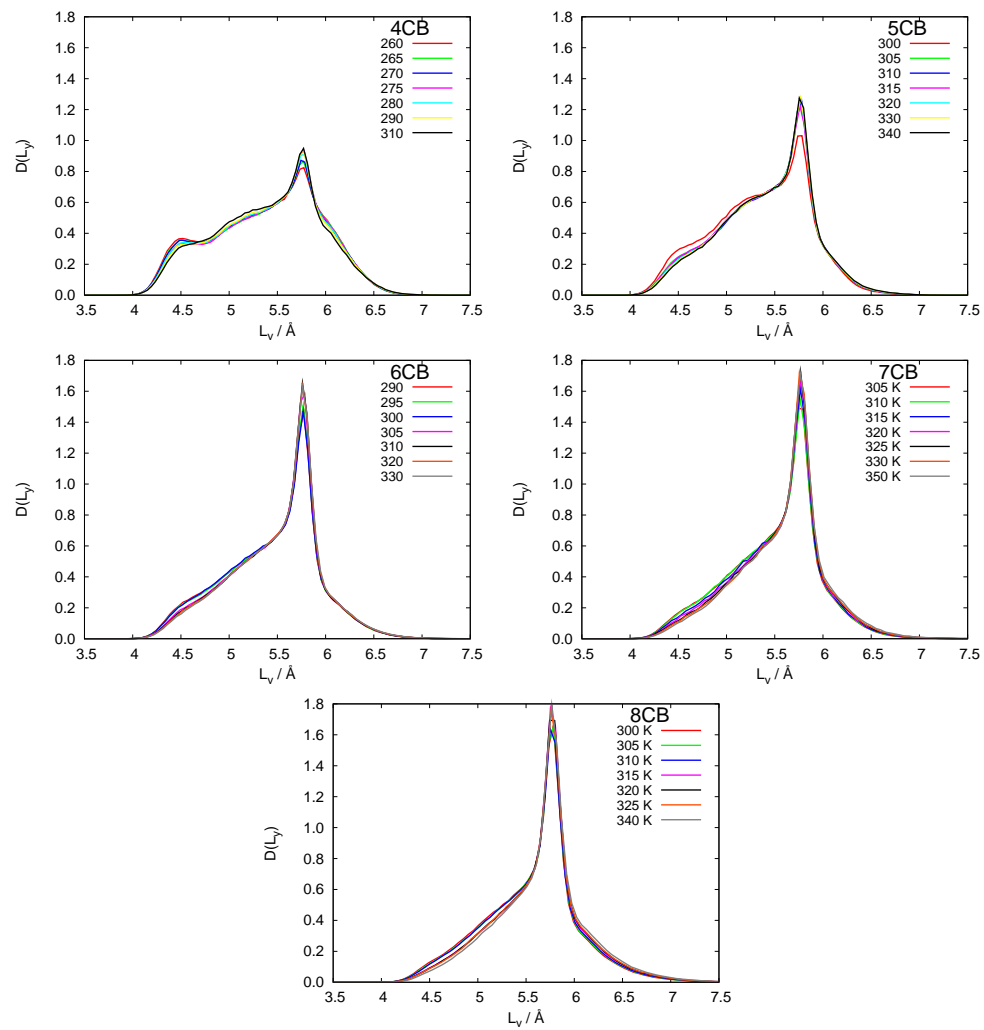
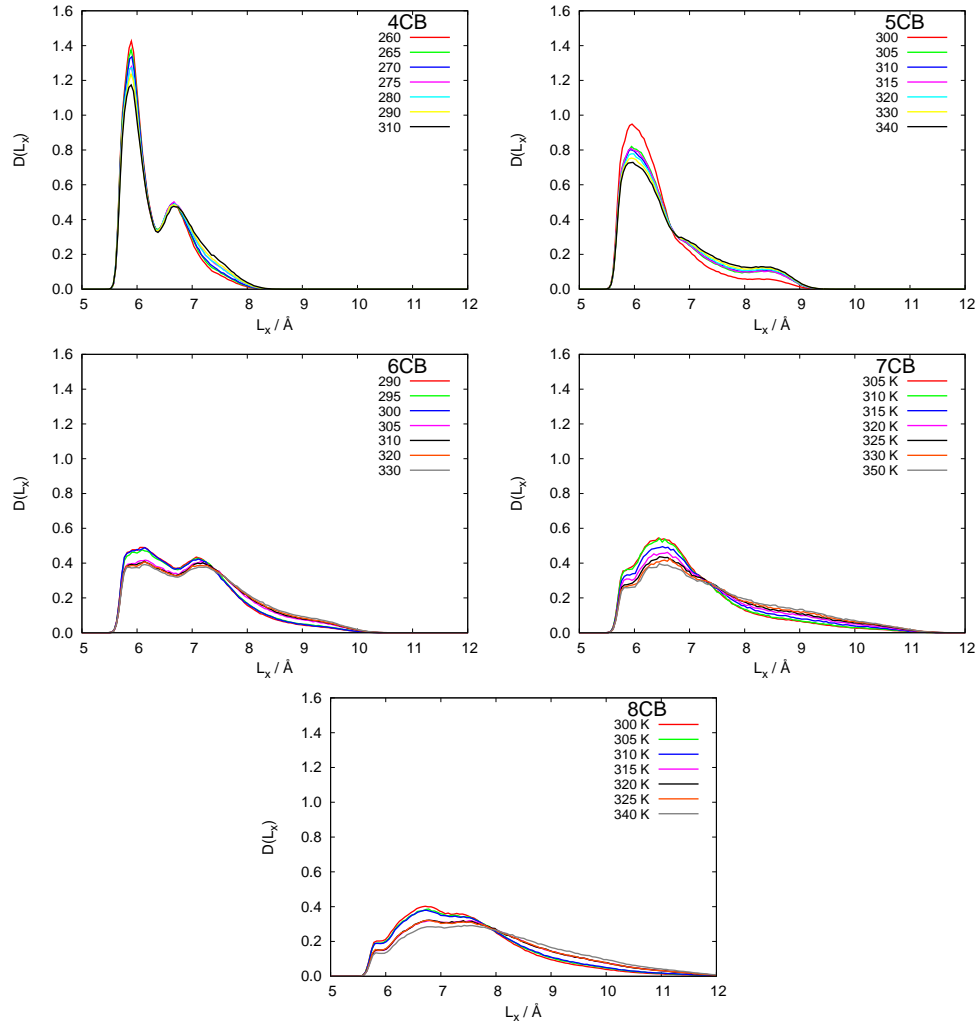




Figure 5.14: x-axis length distributions.



### 5.5.4 Smectic phase

As surprising result from simulation, is the presence of the smectic phase for 8CB. For temperatures below the transition is observed the presence of a layering of molecules. In the samples of 250 molecules, the layer formation is difficult to be studied because the thickness of layer is comparable to the box side. For a more detailed investigation we prepared a bigger sample of 8CB by replicating the samples at 300 (**smectic**), 310 (**nematic**) and 315 (**isotropic**) K for a total of 2000 8CB molecules and we equilibrated for 1 ns. The configuration at T=300 K is reported in Fig. 5.16 on the left where, on the right, all the  $C_6$  carbons are selected showing anisotropic distribution of molecular center of mass. The distance between the layers have been calculated computing the  $g_0(z)$  function. This distance is approximately around 29–30 Å similar to experimental results[Urban et al. 2005] of about 31 Å. From  $g_0(z)$  is also possible observe the degree of stratification for different atoms (or atom groups) used for the calculation. We observe the best separation between the terminal groups of alkyl chain  $g_0(z)^{C20}$ . This behavior cannot be explained explainable considering the strong tendence of 8CB molecules to arrange in antiparallel way, but we must consider also that the parallel alignment from dipoles placed in consecutive way that contributes to reduce the spatial distribution of dipoles.

This result was not considered for the parametrization of FF, it is only a consequence of molecular interactions.

We observe a small tilt angle between the phase **director** and the normal of the planes. Each layer is formed of anti-parallel 8CB molecules that put out of layers their alkyl chains so each layer is on average apolar and the electrostatic interaction between the dipoles is maximized.

Figure 5.15:  $g_0(z)$  for big 8CB samples at 300, 310 and 315 K, computed with respect to different atoms.

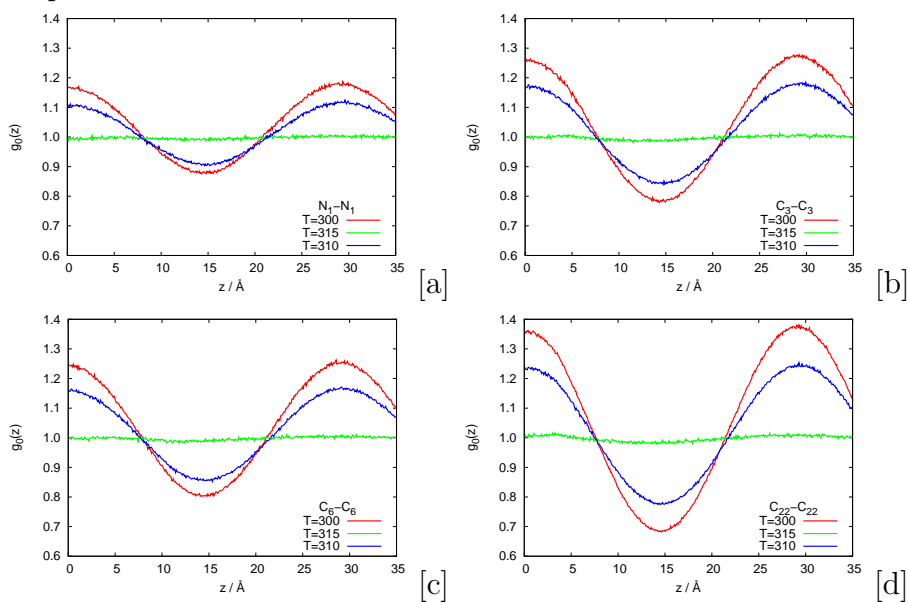
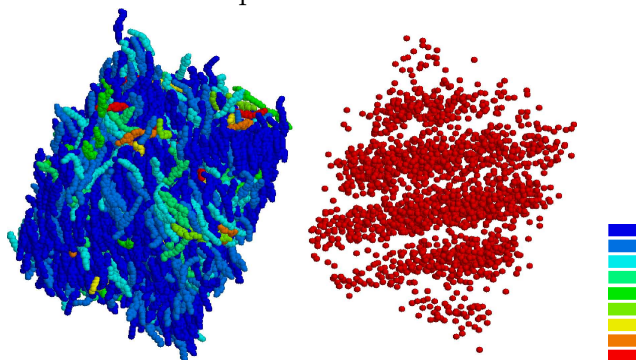


Figure 5.16: Smectic phase obtained for 8CB at 300 K.



## 5.6 Conclusions

The most relevant result obtained from our work is the obtaining of a UA force field able to reproduce the temperature dependence of phase change of a series of nematogenic molecules. The validity of our FF reside inside the possibility to generate the order inside the simulation samples simply by reduction of simulation temperature that, surprising enough after many years from the birt of atomistic simulation applied to the liquid crystal field, was obtained for the first time in our group. The importance of this work is not related to the simple reproduction of the transition temperature of this nematogenic serie of molecule. This results will be very usefull to the next studies on liquid crystals or molecule dissolved inside its. The comparison with experimental works give us the confirms that our simulation are able to correctly describe the properties of a real system and can give us more informations unreachible from experiment. The next step will be the study of more complex liquid crystals systems based to investigate the inner process inside this interesting matrix.

# Chapter 6

## Azobenzene photo–isomerization in 5CB

### 6.1 Introduction

The order of the nematic phase sensibly interacts with the isomerization process of azobenzene, changing the isomerization quantum yield and the permanence time in excited state. This can be observed in the polarised experiments where molecular movements are more or less restricted by the alignment of solvent molecules.

### 6.2 Model

We introduce a new criterion with respect to the isomerization model showed in Chap. 4 inserting the polarised excitation probability of the molecule. We assume the excitation probability proportional to the angle between light polarisation axis  $\vec{D}$  and transition moment from ground and excited state ( $\vec{\mu}_{0\rightarrow 1}$ ) on azo–groups [Pedersen et al. 1998] (figure 6.1), using the relation in equation 6.2.1:

$$P_{0\rightarrow 1}(\theta, \phi) \propto (\vec{\mu}_{0\rightarrow 1} \cdot \vec{D})^2 \quad (6.2.1)$$

We performed experiment with unpolarized light and light perpendicularly and parallelly polarized with respect to the phase director  $\vec{D}$  defined by the orientation of the nematic solvent 5CB.

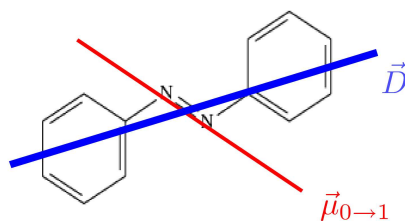


Figure 6.1: Transition moment of azo-benzene molecule used in the excitation model

## 6.3 Results

### 6.3.1 Order parameter $P_2$

We are able to follow the inertial order parameter  $P_2$  during the isomerization process both for azobenzene and 5CB. From this analysis the average  $P_2$  of the solvent is not subjected to important variation due to the isomerization process. On the other hand the  $P_2$  of azobenzene changes a lot, indicating a fast rotation of molecules. The rotation of molecules can be in favour of an increase or decrease the order parameter, but generally the molecules retain a parallel alignment respect the director of the sample.

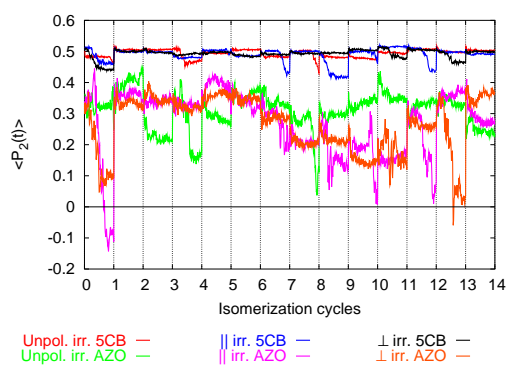


Figure 6.2: Average order parameter  $P_2$  during illumination process.

The order parameter variation seems to be independent to the polarisation conditions of light.

### 6.3.2 Photo-isomerization quantum yield and permanence time in $S_1$

From quantum yield determination we observe a greater value for ordered azobenzene molecules (irradiation parallel to director sample). The experiment with unpolarized light seem to be approximately the average between the ones with parallel and perpendicular light. Observing the number of excited molecules for each experiment, there is a small amount of molecules (about 12%) oriented at  $45^\circ$  to director, thus unable to be excited from a parallel or perpendicular irradiation. These molecules can show different behaviour during sample excitation with respect the aligned or perpendiculars oriented ones. Our aim is to try and separate each contribution by combining the three isomerization series of experiments.

Different permanence times in  $S_1$  for successful (*trans-cis*) and unsuccessful (*trans-trans*) processes in polarised experiments are observed with perpendicular excitation. In parallel experiments, like observed in isotropic solvent, the mean permanence times are similar, suggesting similar pathways in the excited state, and leave the successful decision of isomerization to the ground state shape and to the geometrical structure of molecules at non-radiative decay. Follow these guidelines the difference of permanence time in perpendicular experiments indicate different pathway in excited state for successful or not isomerization process. From another point of view we can believe that the perpendicular excitation strongly depends on the starting configuration at the excitation process. The shortness of decay time could be related to an increase of energy interaction with the nematic solvent. From mean field theory the perpendicular angular orientation is related to an increase of energy interaction between solute and anisotropic solvent. This further energy contribution gives to azobenzene molecules, perpendicularly oriented, a possibility to overcome the energy barrier and sequentially reach fastly the conical intersection and experience the non-radiative decay.

To observe the geometrical differences between the pathways in the different irradiation conditions we calculate the distribution of geometrical species present during isomerization process. The methodology used is to compute the distribution of the angle between molecular inertial axes and a reference axes fixed on molecules. We have chosen the  $N=N$  bond for azobenzene and  $C\equiv N$  bond for 5CB one. These geometrical parameters does not give details of the molecular geometry but give us a general conformational shape of molecules during isomerization process where many degree of freedom are

Table 6.1: The isomerization pathway indicator  $\Theta_m$  computed from the 1000 MD virtual photoisomerization experiments of AB in vacuum and in four different solvents.

Illumination	% exc AZO	$\Phi_{t-c}$	$\langle t_{S_1}^{cis} \rangle$	$\langle t_{S_1}^{trans} \rangle$
Non-Polarised light	100	0.1569	4.42 ps	4.37 ps
Polarised light $\text{---}$	46	0.1765	4.40 ps	4.19 ps
Polarised light $\perp$	42	0.1294	1.62 ps	3.92 ps

involved. Is observed the same position of maxima in each polarisation case indicating the same kinds of intermediate in photo-reaction. This indicate that the geometrical pathway is identical in each experiment and the order of the phase simply induces a faster decay process.

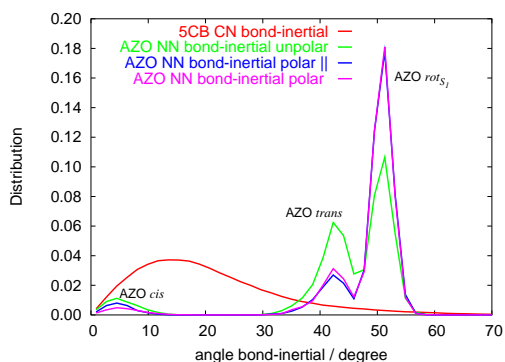


Figure 6.3: Conformer distributions during photo-isomerization

### 6.3.3 Geometrical analysis

Following the time evolution of geometrical parameters we observe a faster molecular rearrangements in perpendicular polarisation experiments. In unpolarized experiments all geometrical modifications seem to be slower with respect to the other ones probably due to the contribution of  $45^\circ$  oriented molecules, moving much slower than the others. The parameters that mainly show this difference are  $Ph-N=N-Ph$  torsion angle,  $N=N-C-C$  torsion angles and  $r_{4,4'}$  distance. For bending angles we observe a strange trend in unpolarized experiment where the relaxation curve for  $trans-trans$  process is slower when compared with respect the other one indicating the difficult, of tilted



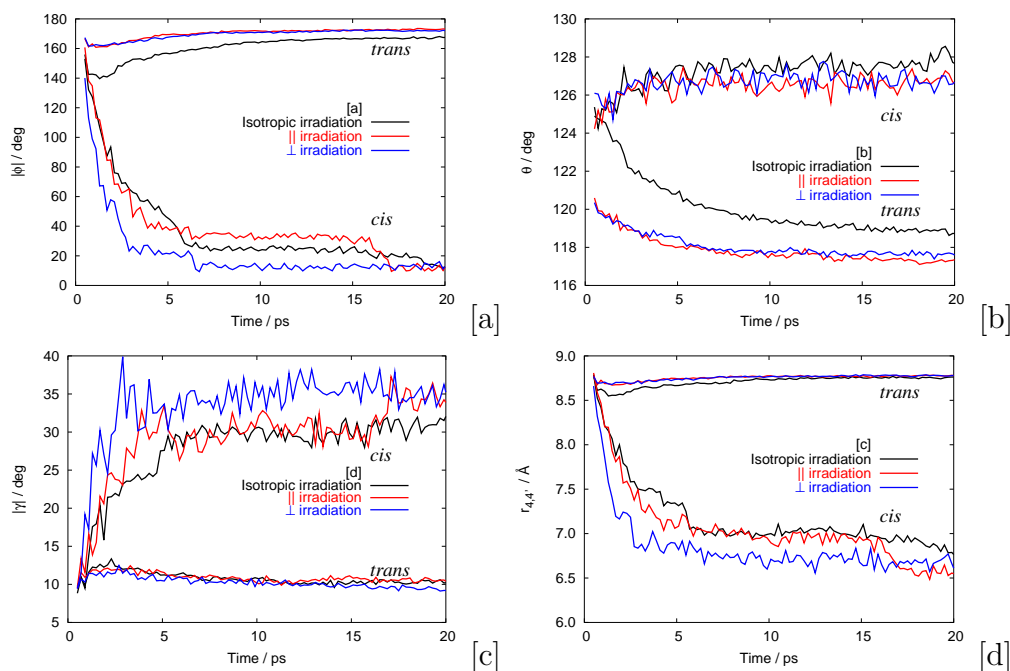


Figure 6.4: Average geometrical parameters. [a] *Ph-N=N-Ph* torsion angle. [b] *N=N-Ph* bending angles. [c] *N=N-C-C* torsion angles. [d]  $r_{4,4'}$  distance.

molecules to come back to the starting configuration. The phenyl rotation required to obtain isomerization is faster for polarised experiments.

An explanation of the faster behaviour of perpendicular molecules can reside in the appearance of a non-zero molecular dipole moment of azobenzene for bent conformations. For a molecule perpendicular to the director at the beginning of the bending the forming dipole moment of azo-molecules, interact in the antiparallel way with the strong 5CB dipole moment. This interaction added to the reduction of the elastic deformation of nematic phase reduce the geometrical relaxation time in the excited state. These interaction probably stabilise the excited azobenzene form reducing the mobility on  $S_1$  PES near the conical intersection resulting in a lower quantum yield of photo-isomerization process.

### 6.3.4 Isomerization mechanism

The mechanism ratio calculated from simulation indicates a prevalence of a mixed mechanism in each polarisation case studied. It is also observed a greater amount of pure torsional contribution in non-polarised and perpendicular light simulations while in the parallel polarised light this contribution is 2–3 times lower. Also in this case the unpolarized results seem to be the average of parallel and perpendicular light irradiation. Anywhere we have measured similar inversion contribution. Probably the aligned molecules have less sterical constraint with respect to the one perpendicular.

Table 6.2: The isomerization pathway indicator  $\Theta_m$  computed from the 1000 MD virtual photoisomerization experiments of AB in vacuum and in four different solvents.

solvent	Torsional	Tors.+Inv.	Inversion
Non-Polarised light	25%	63%	12%
Parallel light	10%	80%	10%
Perpendicular light	33%	56%	11%

## 6.4 Conclusions

We had performed a new strategy of simulation of photo-isomerization processes in liquid crystals joining the metodological approach and the results obtained in the Chap.s 4 and 5. For the moment, we have only informations about the femtosecond dynamic just after the light absorbtion, but we can extend the study to longer timescale until the equilibrium phenomena (see experimental details in the next chapter) able to strongly modify the chemical and physical properties of the ordered matrix. The knowledge of the response of the photo-modulation process driven by an ordered can be usefull to the advance and

# Chapter 7

## Azobenzene in liquid crystal: ESR study

### 7.1 Introduction

Photoresponsive liquid crystal systems, formed by a nematic liquid crystal (LC) doped with azobenzene derivatives, are of great interest in view of their application as optical-switching and image-storing materials [Ichimura 2000] and as model systems for more complex photoresponsive polymers [Finkelmann et al. 2001b]. The azo- derivative *trans-cis* reversible photoisomerization can produce two types of photomodulation of the LC mesophase [Tsutsumi et al. 1997, Ikeda 2003]: changes in the LC director orientation (“order-order” change) and shift of the phase transitions (“order-disorder” change). Qualitatively, the working principle of the photochemical phase transition is based on the change of molecular shape of the photochromic guest molecule: the *trans* form of the azobenzene derivative has a rod-like shape similar to that of the host LC, while the *cis* isomer destabilizes the LC phase because of its bent molecular shape. However, a detailed investigation of the relation between the azo-derivative molecular features and the changes in *trans*-host and respectively *cis*-host solute-solvent system is, to our knowledge, still lacking.

## 7.2 Experimental

The liquid crystal 4-cyano-4'-n-pentylbiphenyl (5CB) was used as host matrix (Fig. 7.1). This compound exhibits, on heating, the following phase sequence: K-N 295.7 K, N-I 308.5 K. It was purchased from Merck KGaA (Darmstadt, Germany) and was used without any further purification.

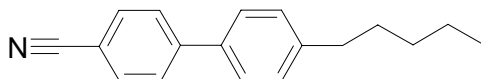
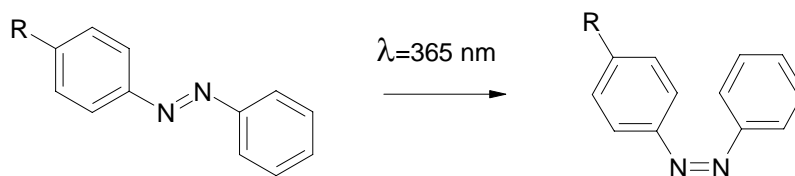


Figure 7.1: Chemical structure of 5CB LC.

The structure of the photo-active units used in this study was 4-R-phenylazobenzene (see Fig. 7.2), where R = H, F, Br, CH<sub>3</sub>, CF<sub>3</sub>, OCH<sub>3</sub>, *On*-Bu, *Ot*-Bu. None of the azo-derivatives considered showed a liquid crystal phase. Azobenzene was purchased from Aldrich and used without further purification. The azo-derivatives were synthesized according to the Mills reaction [Boyer 1969], in which nitrosobenzene combines with primary arylamines in glacial acetic acid to give unsymmetrical azo-compounds. The synthesis and characterization of 4-*Ot*-Bu-phenylazobenzene is original and is reported in the App. ??.

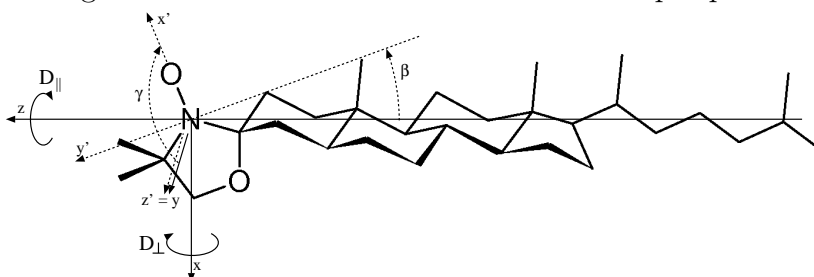
Figure 7.2: *trans-cis* photo-isomerization scheme of 4-R-phenylazobenzene (R = H, F, Br, CH<sub>3</sub>, CF<sub>3</sub>, OCH<sub>3</sub>, *On*-Bu, *Ot*-Bu).



The nitroxide free radical we used in this study was the 3 $\beta$ -DOXYL-5 $\alpha$ -cholestane (CSL), purchased from Aldrich. It was chosen because it has been proven [Meirovitch and Freed 1984, Arcioni et al. 2004] to be a reliable probe to monitor the order and the dynamics of the LC system, due to its size, rigidity and rod-like shape analogous to that of the 5CB. The CSL structure is shown in Fig. 7.3 together with the chosen ordering ( $x, y, z$ , solid line) and magnetic ( $x', y', z'$ , dashed line) molecular frames and the principal

components  $D_{\perp}$  (reorientation of the molecular long axis) and  $D_{\parallel}$  (rotation around the long axis) of the rotational diffusion tensor. The details about the CSL spectroscopic parameters have been discussed elsewhere [Meirovitch and Freed 1984, Arcioni et al. 2004].

Figure 7.3: Chemical structure of the CSL spin probe.



Samples were prepared by adding one of the eight different azo-derivatives indicated above to a solution of the CSL spin probe in the 5CB liquid crystal at a concentration of  $6 \times 10^{-4} g_{CSL}/g_{5CB}$ . Two values of the azo-derivative mole fraction were studied:  $x = 0.01$  and  $x = 0.07$  (hereafter, referred respectively as  $x_1$  and  $x_2$ ). Given the solubility of the spin probe and of the azo-derivatives in 5CB at the concentrations studied, mechanical mixing at a temperature slightly above  $T_{NI}$  was sufficient to prepare uniform samples without adding any solvent.

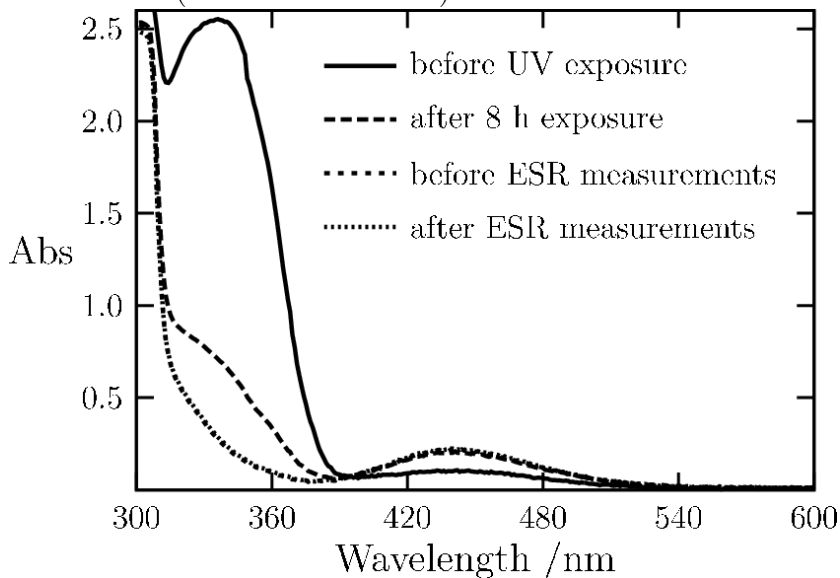
Fresh samples not requiring photo-isomerization (*trans* isomers) were inserted into glass capillaries of 1.8 mm internal diameter and stored in the dark at room temperature until ESR measurements. To obtain the *cis* isomers, samples were isomerized as described later in this section. Immediately after the isomerization, they were inserted into the capillaries and ESR spectra were recorded.

We acquired ESR spectra with a Bruker ESP300E spectrometer equipped with an ER 041XG microwave X-band (9.5 GHz) Gunn Diode bridge and a rectangular ER 4102 cavity. The samples were thermostated with a nitrogen flux through a variable temperature unit Bruker B-VT 2000. The temperature was monitored with a calibrated type T thermocouple (Comark Ltd.) located in contact with the capillary containing the sample. The temperature stability during the measurements was better than  $\pm 0.05$  K. We registered ESR spectra on heating taking the sample from the nematic to the isotropic phase, in a range of about 20 K around the  $T_{NI}$ .

The *trans-cis* photo-isomerization reactions were carried on by exposing the samples to UV-light ( $\lambda=365$  nm) in a quartz cuvette with a UVGL-58 MINERALIGHT Lamp (UVP). During the exposure, we recorded UV-visible absorption spectra with a Cary 1E (Varian) spectrophotometer to verify the complete *trans-cis* photo-isomerization. UV-visible spectra were taken before any irradiation and after different times of exposure, dissolving 5CB azo-doped samples in cyclohexane to achieve a concentration of the azo-derivative of  $2.5 \times 10^{-4}$  M. The *trans* isomer spectra exhibited features characteristic of *trans*-azobenzene type molecules, i.e. a strong absorption band at about 340 nm, due to  $\pi \rightarrow \pi^*$  transition, and a much weaker band centered at 450 nm, due to  $n \rightarrow \pi^*$  transition. Irradiation at 365 nm caused a decrease of the  $\pi \rightarrow \pi^*$  band intensity and a build-up of the  $n \rightarrow \pi^*$  band, which was slightly shifted toward shorter wavelengths. This indicated that the isomerization was taking place. We considered the process successful when we could not observe any further change by comparing two consecutive UV-visible spectra (see Fig. 7.4). This typically required between 10 and 24 hours, depending on the azo-compound concentration. After the ESR measurements of the samples doped with the *cis* azo-derivatives, a further UV-visible absorption spectrum was taken and compared to the last one taken immediately before the ESR measurements. In all the cases, these spectra were essentially identical, indicating that the *cis*-isomers did not significantly reverse to the *trans* conformation during the ESR measurements.

It is well known that the presence of non-mesomorphic solutes in a nematic liquid crystal can lead, in a temperature range close to the  $T_{NI}$ , to the appearance of a biphasic region, consisting of both nematic and isotropic phases which coexist in equilibrium [Humphries and Luckhurst 1976, Martire et al. 1976, Oweimreen 2001]. Therefore, we finally performed polarizing microscope observations under crossed polarizers to verify the existence of a two-phase region in our samples and to find out the limits of the coexisting region, if any. Every LC mixture was held in a flat quartz cell with  $25 \mu\text{m}$  spacing. The inner surface of the cell was coated with polyvinyl formal and then rubbed to achieve homogeneous alignment of the LC molecules parallel to the rubbing direction. The sample was then irradiated with UV-light ( $\lambda=365$  nm) to cause the isomerization, when needed, and examined with a ORTHOPLAN Leitz polarizing microscope equipped with a thermostating stage. For each system, the temperature range explored was the same considered in the ESR measurements. These investigations allowed a better understanding of the sample behavior around the  $T_{NI}$  and the appropriate

Figure 7.4: Absorption spectra recorded after different times of exposure of the sample containing 5CB and 4-*O*-*t*-Bu-phenylazobenzene dissolved in cyclohexane (see text for details).



choice of the ESR spectra fitting model (see Results section).

## 7.3 Results

ESR spectra were analyzed using a software package which combines an implementation of a modified Gauss-Newton-Marquardt non-linear least-squares fitting [Bevington 1969], developed in our group [Arcioni et al. 1993], with the slow tumbling theory Fortran codes for the simulation of the ESR line shapes, developed by Freed and collaborators [Freed 1972; 1976, Schneider and Freed 1989].

The choice of the fitting model was done accordingly to the results obtained by microscope investigations, which revealed two distinguished behaviors in the nematic phase across the  $T_{NI}$ . Basically, pure 5CB and samples doped with *trans* isomers completely turned, at the transition, into the isotropic phase, whilst mixtures containing *cis* isomers exhibited, close to the  $T_{NI}$ , a biphasic region in which both nematic and isotropic phases coexisted.

The ESR spectra associated to samples that did not show any sign of

phase separation near the  $T_{NI}$  were fitted using a monodomain model.

Otherwise, if the samples showed a biphasic region close to the  $T_{NI}$ , their ESR nematic spectra were fitted with a two-site model, in which we assumed that the experimental spectra were made up of two contributions, nematic and isotropic, which affected the ESR lineshape with different weight. Every spectrum was thus related to two parameter sets linked by a weighting factor, expressing the fraction of nematic ( $f_N$ ) or isotropic ( $f_I$ ) contribution in each spectrum, with  $f_N + f_I = 1$ .

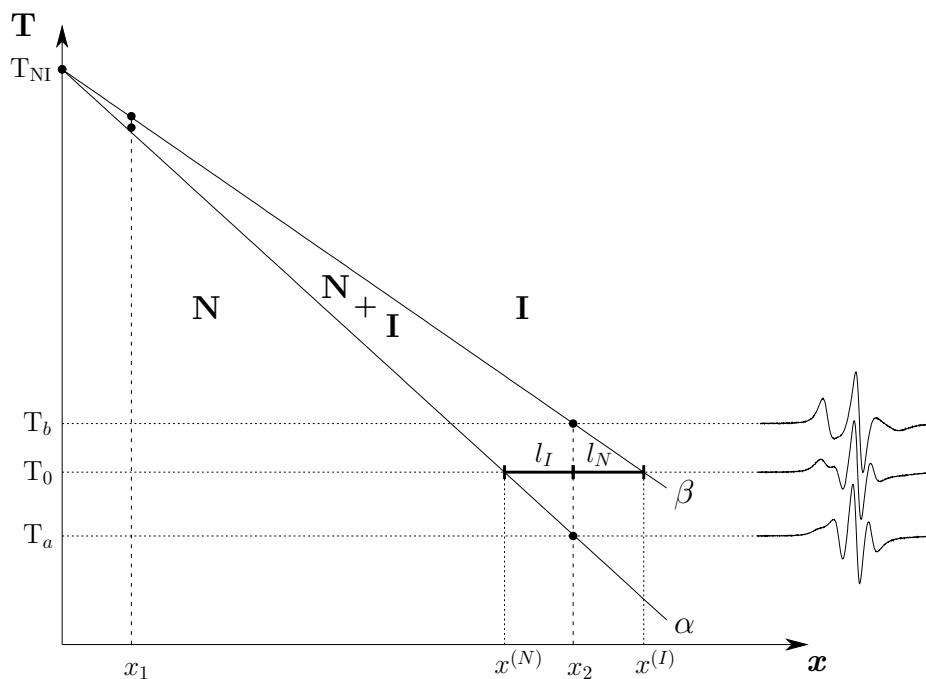
The relative amount of each phase could be estimated by using the lever rule in a temperature-composition phase diagram. We illustrate the method employed in Fig. 7.5, where we report as example the phase diagram of 5CB doped with 4-OCH<sub>3</sub>-phenylazobenzene in its *cis* form. By inspecting the ESR-lineshapes we have detected the biphasic region limits, marked in Fig. 7.5 as dots on the isopleth lines at  $x_1$  and at  $x_2$ , and we have then built the  $\alpha$  and  $\beta$  lines, which represent respectively the composition of the nematic and the isotropic coexisting phases. Considering for simplicity the heating process of the nematic binary system at the mole fraction  $x_2$ , at  $T_a$  the isotropic phase begins to appear while at  $T_b$  the nematic phase completely disappears.

At every temperature  $T_0$  included into the coexisting range, we traced the corresponding tie line (of total length  $l_N + l_I$ ) to apply the lever rule:  $l_I f_N = l_N f_I$ . We thus calculated the values of  $f_N$  ( $= l_N / (l_N + l_I)$ ) and  $f_I$  ( $= l_I / (l_N + l_I)$ ) which represent the relative amount of nematic and isotropic phase standing in equilibrium within the biphasic region. Also, the intersection between the  $\alpha$  and  $\beta$  lines with the tie line allowed to find out the composition of respectively the nematic,  $x^{(N)}$ , and isotropic,  $x^{(I)}$ , coexisting phases, but the role of the solute concentration will be discussed later.

To establish a notation common to the different samples, hereafter we refer to  $T_{NI}$  as to the highest temperature at which the nematic phase is still present in the sample and to  $\Delta T_{NI}$ , the  $T_{NI}$  shift, as to the difference between the  $T_{NI}$  of the pure 5CB and that of the system considered. As reported in Table 7.1, all the azo-doped samples studied showed a  $T_{NI}$  lower than that observed in the pure 5CB and a  $\Delta T_{NI}$  increasing with the concentration of the azo-derivative and larger, at the same concentration, for the *cis* isomers. We found, in particular,  $\Delta T_{NI}$  values for the *trans* isomers in the range 0.2-1.6 K at  $x_1$  mole fraction, increasing at  $x_2$  to values in the range 0.7-8.0 K and, for the *cis* isomers, values in the ranges 1.7-3.7 K and 7.9-19.5 K at  $x_1$  and  $x_2$  mole fraction, respectively.



Figure 7.5: Temperature-composition phase diagram, relative to the 5CB doped with *cis* 4-OCH<sub>3</sub>-phenylazobenzene samples, showing the features considered in applying the lever rule to obtain the  $f_N$  and  $f_I$  values, see text for details. We also report the experimental ESR spectra collected respectively at the offset ( $T_a$ ), in the middle ( $T_0$ ) and at the end ( $T_b$ ) of the two-phase region.



The presence of the azo-derivatives had also the effect of lowering the order degree of the samples in the N phase. Also in this case, the effect was much larger for the *cis* isomers and increased with the concentration of the azo-derivatives.

In Fig. 7.6(a) we show our results for the order parameter  $\langle P_2 \rangle$  of the spin probe against reduced temperature  $T^*$  ( $= T/T_{NI}$ ) in samples containing the *trans* isomers at  $x_1$  mole fraction. As we can see, the values are essentially identical to those relative to the pure 5CB, and this can be also observed at the higher  $x_2$  mole fraction, with only a very small decrease of the  $\langle P_2 \rangle$  values in the samples containing the -*Ot*-Bu and -F azo-derivatives (see Fig. 7.6(b)). For the *cis* isomers, the deviations are relatively larger. As shown in Fig. 7.7(a), even at  $x_1$  mole fraction all the samples show an order

Table 7.1: Nematic-isotropic phase transition temperatures ( $T_{NI}$ ) and  $T_{NI}$  shifts relative to the pure 5CB ( $\Delta T_{NI}$ ) for LC samples doped with various *trans* and *cis* *p*-azobenzene derivatives with a different R substituent at the two concentrations studied ( $x_1$  and  $x_2$ ).

R	$T_{NI}$ /K				$\Delta T_{NI}$ /K			
	<i>trans</i>		<i>cis</i>		<i>trans</i>		<i>cis</i>	
	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$	$x_1$	$x_2$
<i>Ot</i> -Bu	307.3	301.3	304.8	289.0	1.2	7.2	3.7	19.5
OCH <sub>3</sub>	308.3	307.5	306.0	289.5	0.2	1.0	2.5	19.0
<i>On</i> -Bu	308.1	307.8	304.9	291.9	0.4	0.7	3.6	16.6
CH <sub>3</sub>	308.1	306.4	305.6	294.5	0.4	2.1	2.9	14.0
F	307.4	302.7	305.8	295.6	1.1	5.8	2.7	12.9
H	306.9	300.5	305.7	295.7	1.6	8.0	2.8	12.8
CF <sub>3</sub>	307.5	302.0	306.8	299.5	1.0	6.5	1.7	9.0
Br	308.0	305.2	305.9	300.6	0.5	3.3	2.6	7.9
5CB	308.5							

degree lower than that of the pure 5CB and this decrease becomes even more evident at  $x_2$  mole fraction for all the azo-derivatives (see Fig. 7.7(b)) and is particularly large in samples containing the -OCH<sub>3</sub>, - *On*-Bu and -*Ot*-Bu derivatives.

The  $\langle P_2 \rangle$  temperature dependence, in this latter case (Fig. 7.7(b)), seems to have an anomalous trend. In the temperature range far from the corresponding  $T_{NI}$ , the order parameter of the various samples decreases with increasing temperature, as usually happens in thermotropic liquid crystals, but, approaching the  $T_{NI}$ , the order degree tends to increase a bit and then to decrease again, eventually showing the discontinuity characteristic of the nematic-isotropic phase transition, once the  $T_{NI}$  is reached.

The reason for this behavior relies on the formation of the biphasic region and on the consequent partition of the azo-derivatives within the two phases. To clarify this point, it is useful to refer again to the phase diagram in Fig. 7.5, considering the isopleth at  $x_2$  and a heating process raising the system temperature from  $T_a$  to  $T_b$ . Within the biphasic region ( $T_a < T < T_b$ ), the solute is distributed between the nematic and the isotropic phase in different proportions. At the generic temperature  $T_0$ , for instance, the azo-

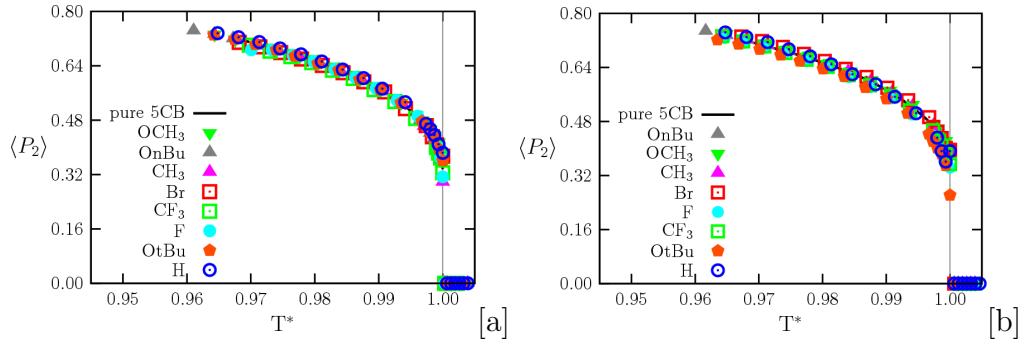


Figure 7.6: Order parameter  $\langle P_2 \rangle$  of the spin probe against reduced temperature  $T^*$  for the samples doped with the *trans* azo-derivatives at (a)  $x_1$  and (b)  $x_2$  mole fraction compared to the pure 5CB.

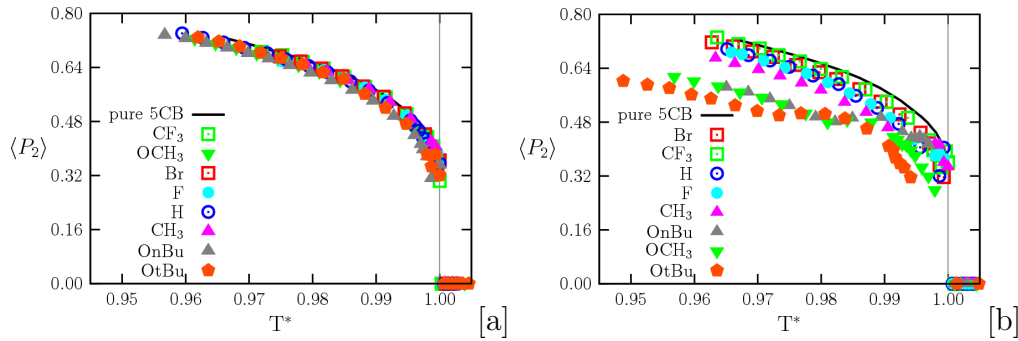


Figure 7.7: Order parameter  $\langle P_2 \rangle$  of the spin probe against reduced temperature  $T^*$  for the samples doped with the *cis* azo-derivatives at (a)  $x_1$  and (b)  $x_2$  mole fraction compared to the pure 5CB.

compound mole fraction dissolved into the nematic phase,  $x^{(N)}$ , is smaller than the initial concentration  $x_2$  at temperature  $T_a$ . Thus, in the coexisting region, the order of the nematic phase present in the system is the result of the competition of two temperature related opposite (contributions - processes - effects). On the one hand, in fact, increasing the temperature causes a higher disorder of the nematic phase and a lower order parameter but, on the other hand, it is the driving force for the diminishing of the *cis* isomer concentration in the nematic fraction, resulting in a smaller perturbation of the system and thus in an increase of the order degree. The  $\langle P_2 \rangle$  behavior shown in Fig. 7.7(b) is the result of the balance of these two terms.

Surprisingly, the dynamics seems not dependent on the nature and con-

centration of the different azo-derivatives, whilst it is strictly connected to the phase of the system (N or I). The temperature dependence of the tumbling diffusion coefficient  $D_{\perp}$  of the spin probe in the nematic and in the isotropic phase is well represented by two different Arrhenius-type behavior.

## 7.4 Discussion

It would be interesting to relate the changes of  $\langle P_2 \rangle$  and  $T_{NI}$ , from the values observed in the pure 5CB, to the molecular-level interactions occurring between the different azo-derivatives and the LC host.

A first, general observation is that the *cis* isomers (V-shaped) always produced a  $T_{NI}$  shift larger than the *trans* (rod-like), revealing that the main role on the nematic phase stability is played by the solute molecular shape. This is in agreement with experimental studies and simulations dealing with the orientational ordering in nematic liquid crystals: the short-range interactions, which depend on the size and on the molecular shape, dominate the orientational behavior of molecules in nematic solvents [Burnell and de Lange 1998, Syvitski and Burnell 2000].

Another indication of the role played by the short-range interactions might come from a more detailed comparison of the differences of the  $\langle P_2 \rangle$  dependence on  $T^*$  and of the  $T_{NI}$  shifts, due to the type of the azo-derivative, for each isomer conformation and at the two mole fractions,  $x_1$ ,  $x_2$  studied.

At  $x_1$ ,  $x_2$ , *trans* and at  $x_1$  *cis*, the destabilization results essentially in a shift of the whole phase diagram to a lower temperature. In fact, the  $\langle P_2 \rangle$  dependence on  $T^*$  is essentially unaffected for the *trans* at  $x_1$ ,  $x_2$  (Fig. 7.6) and it begins to show only a small decrease for the *cis* at  $x_1$  (Fig. 7.7(a)).

At  $x_2$ , *cis*, instead, the shift of the phase diagram to a lower temperature is no longer able to compensate the destabilization introduced by the azo-derivatives. In this case, the order of the phase is clearly lower than that of the pure 5CB and the differences among the  $\langle P_2 \rangle$  vs.  $T^*$  plots, are significantly enhanced (Fig. 7.7(b)).

Interestingly, in this case we also observe that the different azo-derivatives affect in the same way the  $\langle P_2 \rangle$  dependence on the  $T^*$  and the  $T_{NI}$ , i.e., the azo-derivative which causes the largest shift of  $T_{NI}$  is also depressing the order of the phase the most.

By listing the derivatives, starting from the one which causes the largest  $T_{NI}$  shift, we observe the following sequence: *Ot*-Bu, OCH<sub>3</sub>, *On*-Bu, CH<sub>3</sub>, F,

H, CF<sub>3</sub>, Br (this is also the sequence that has been used in Table 7.1). If we consider the  $\langle P_2 \rangle$  deviations, again starting from the largest one, the sequence of the derivatives is the same, apart the inversion of CF<sub>3</sub> with Br. Therefore, it would seem that, at least above a certain concentration of the *cis* isomer, a destabilizing contribution to the host-guest interactions due essentially to the size of the substituent, becomes dominant (*Ot*-Bu, OCH<sub>3</sub> and *On*-Bu) and it affects the order of the phase as well as the shift of the  $T_{NI}$  in *almost* the same way. Clearly, these short-range interactions alone are not sufficient to account for the sequence of stability observed, otherwise the unsubstituted azobenzene, would have occupied the last position in the previous sequence. At a lower concentration, longer-range interactions might become relatively more important and determine a different sequence of stability. In fact, the sequence at  $x_1$  of the *cis* isomers is the following: *Ot*-Bu, *On*-Bu, CH<sub>3</sub>, H, F, Br, OCH<sub>3</sub>, CF<sub>3</sub> which looks quite different from the one observed at  $x_2$ . The sequence of stability of the *trans* isomers at  $x_2$  is also different: H, *Ot*-Bu, CF<sub>3</sub>, F, Br, CH<sub>3</sub>, OCH<sub>3</sub>, *On*-Bu. Here it seems that the destabilizing contribution due to the larger substituents is even less important. Moreover, this sequence remains quite similar at  $x_1$ : H, *Ot*-Bu, F, CF<sub>3</sub>, Br, CH<sub>3</sub>, *On*-Bu, OCH<sub>3</sub>.

Another important contribution, responsible of the orientational ordering, involves long-range, solute-solvent electrostatic interactions.

To investigate the role of these interactions, we compared the dipole moment and the quadrupole tensor of both the solvent and the solute molecules. For all the compounds studied, we performed a preliminary quantum mechanical DFT B3LYP/6-31G\*\* geometry optimization and frequency calculation. Then, by using this optimized geometry, we have determined the atomic point charges, the dipoles and the quadrupoles using a more extended basis set (*AUG-cc-PVTZ*) and the Merz-Singh-Kollman scheme, with the additional constraint of reproducing the total molecular dipole moment [Cornell et al. 1995]. The results of these calculations are reported in Table 7.2 and Table 7.3.

Unfortunately, it was not possible to find a simple correlation between the degree of destabilization caused by the various azo-derivatives and their dipole moments or quadrupole tensors compared to those of the 5CB liquid crystal. On the one hand it seems that it is not sufficient to simplify the electrostatic host-guest interaction using low rank multipole terms. On the other hand, the use of terms of higher rank, to better describe the charge distributions, would go beyond our scope of finding *simple* correlations.

Table 7.2: Calculated B3LYP/AUG-cc-PVTZ total dipole moment in Debye and projection over the long (x), middle (y) and short (z) inertial molecular axis.

R	Dipole moments and components /D							
	<i>trans</i>				<i>cis</i>			
	$\mu$	$\mu_x$	$\mu_y$	$\mu_z$	$\mu$	$\mu_x$	$\mu_y$	$\mu_z$
O <i>t</i> -Bu	1.66	-0.96	-0.11	1.35	3.26	-2.23	-2.37	0.15
OCH <sub>3</sub>	1.88	1.42	-1.23	0.00	3.89	-2.55	-2.85	0.70
O <i>n</i> -Bu	2.50	2.26	-1.05	0.00	4.38	3.60	2.43	0.55
CH <sub>3</sub>	0.80	-0.80	-0.06	0.00	3.63	-1.10	-3.45	0.20
F	1.52	-1.50	0.20	0.00	2.34	0.69	-2.22	0.22
H	0.00	0.00	0.00	0.00	3.16	-0.00	-3.16	0.00
CF <sub>3</sub>	3.63	3.63	0.18	-0.07	2.65	-1.96	-1.77	-0.24
Br	1.89	-1.88	0.14	0.00	2.39	0.54	-2.30	0.31
5CB	6.32	6.26	0.60	0.63				

The importance of the role played by long-range interactions on the molecular orientation mechanism in nematics has been discussed in several works by Burnell *et al.* (see, e.g. [Syvitski and Burnell 1997] and refs. therein) and it has been related to the average electric field, arising from the charge distribution of the solvent molecules surrounding the solute, and the corresponding field gradient, experienced by the solute.

For each molecular species studied, the electrostatic potential (EP) distributions were calculated at the molecular surface obtained considering the standard Lennard-Jones radius of each atom, using the following relation:

$$EP(\vec{r}) = \sum_i \frac{q_i}{r_i} \quad (7.4.1)$$

where  $\vec{r}$  is the position on the molecular surface,  $q_i$  is the point charge located on the  $i$ -th atom and  $r_i$  is the distance between the molecular center of charge and  $\vec{r}$ .

Our approach was to try a qualitative comparison between the EP map of 5CB and the map of each azo-derivative by checking for the absence or the presence of similar features which could be then related to a lower or a higher degree of destabilization caused by that derivative.

In Fig. 7.8(a) we show the EP map at the molecular surface of the 5CB

Table 7.3: Calculated B3LYP/*AUG-cc-PVTZ* traceless quadrupole moments in Debye for the inertial molecular axis.

R	Quadrupole tensor /D					
	<i>trans</i>			<i>cis</i>		
	$Q_{xx}$	$Q_{yy}$	$Q_{zz}$	$Q_{xx}$	$Q_{yy}$	$Q_{zz}$
<i>O</i> <i>t</i> -Bu	13.03	0.16	-13.19	6.37	-7.41	1.03
OCH <sub>3</sub>	13.60	-0.12	-13.47	12.18	-11.30	-0.87
<i>O</i> <i>n</i> -Bu	14.24	-1.58	-12.65	5.29	-6.41	1.11
CH <sub>3</sub>	12.90	-0.29	-12.60	6.63	-8.16	1.53
F	1.47	6.14	-7.61	1.23	-6.82	5.58
H	10.86	1.09	-11.95	6.05	-7.84	1.79
CF <sub>3</sub>	0.45	6.82	-7.28	-4.63	-2.66	7.30
Br	5.26	4.32	-9.59	-1.68	-3.69	5.38
5CB	-35.84	22.35	13.49			

liquid crystal, compared to the maps of the *trans* unsubstituted azobenzene (b) and of the *trans* 4-*O**n*-Bu-phenylazobenzene (c), which represent, respectively, the first and the last (at  $x_2$ , or the penultimate at  $x_1$ ) derivatives in the sequence of stability presented above.

The EP distribution of 5CB appears to be strongly unbalanced. We observe two different sites: a strong negative site located on the CN group, and a more positive one on the alkyl region. Therefore, it seems reasonable to assume that the 5CB molecules will have a stronger electrostatic interaction among themselves, via their dipolar site and with other similar sites present on solute molecules.

Among the *trans* derivatives, the unsubstituted azobenzene (Fig. 7.8(b)) destabilizes the N phase the most. We notice that it shows an EP map perfectly balanced between the two rings, therefore its electrostatic interaction with 5CB molecules is expected to be relatively low since it could be regarded as a sort of “insulator” of the electrostatic interaction of the 5CB molecules among themselves, thus destabilizing the N phase.

At the opposite end of the sequence of stability we find the *trans* 4-*O**n*-Bu-phenylazobenzene (Fig. 7.8(c)) which exhibits, instead, an unbalanced EP map, with a well evident positive charge in the alkyl region.

The above considerations work quite well for the whole sequence of stability of the *trans* derivatives at both concentrations studied. So, even if it is

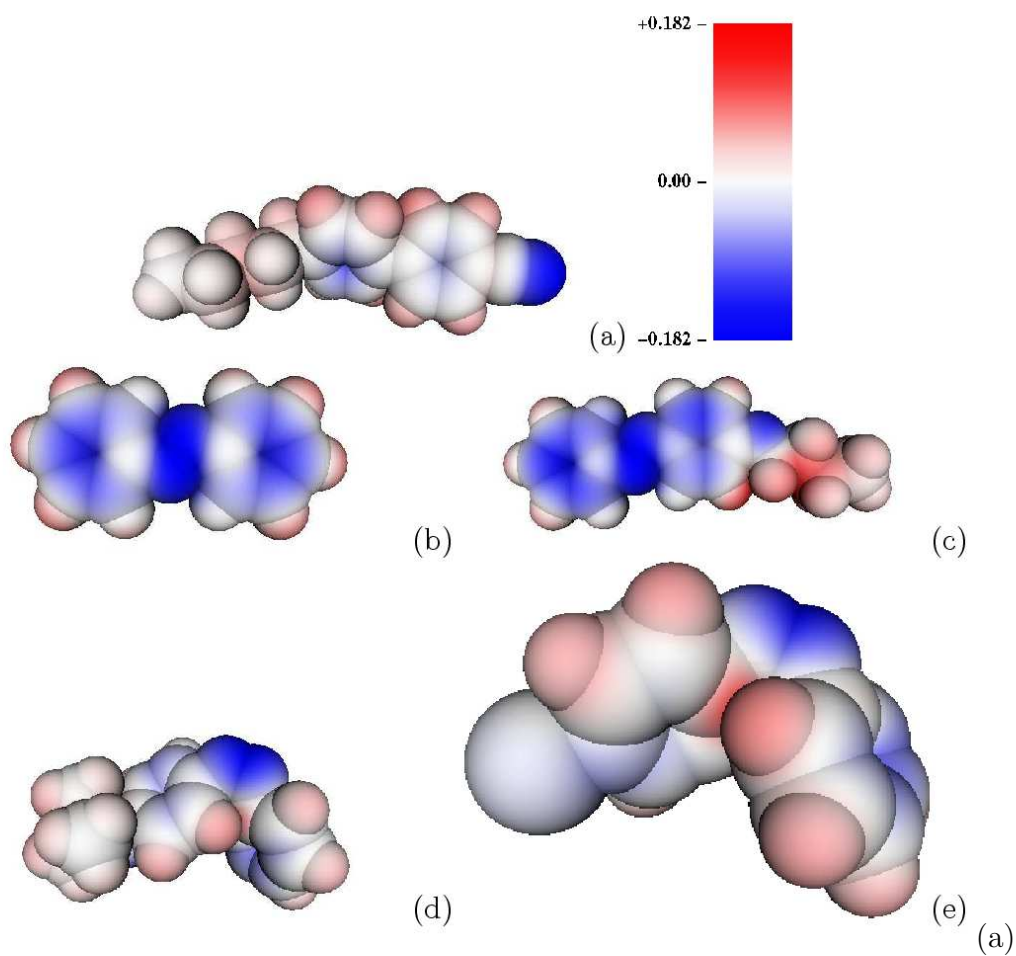


Figure 7.8: Electrostatic potential at the molecular surface for the 5CB (a), *trans* azobenzene (b), *trans* 4-*On*-Bu-phenylazobenzene (c), *cis* 4-*Obu*-phenylazobenzene (d) and *cis* 4-Br-phenylazobenzene (e). Colour coded as from the palette (to be inserted). Geometry and charges obtained at B3LYP/6-31G\* level.



difficult to evaluate the relative contributions of other source of destabilization, the electrostatic one, seems to be an important driving mechanism of the sequence of stability observed.

Far from being an exhaustive rationalization of the molecular-level interactions responsible of the different stabilities observed in the 5CB N phase, we value this set of qualitative observations as a starting point for a more in-depth theoretical study which, we believe, would require a much more robust approach like, e.g., that possible via Molecular Dynamics simulation of the guest-host interactions. A similar approach goes, of course, well beyond the scope of this article and it would represent a complete new study in itself which we plan to carry on in a following work.

## 7.5 Conclusions

Even at the lower mole fraction studied, the presence of the azo-derivative caused, in all cases, a depression of the  $T_{NI}$ . The *cis* isomer slightly reduced the order of the 5CB nematic phase and, as expected, induced a larger  $T_{NI}$  shift with respect to the *trans* isomer. At the higher mole fraction, we observed an analogous behavior with enhanced effects. The presence of the phase separation before the  $T_{NI}$  induced a singular effect on the solvent order, showing an increase of the  $\langle P_2 \rangle$  during the heating process.

Any correlation between the solute disordering effect and the molecular dipole and quadrupole was observed. Some qualitative consideration about the EP simmetry could be done. The dynamics appeared to not have a dependence on the nature of the different derivatives.



## Chapter 8

# Orientation of rigid solutes in 5CB: MD study

The aim of this work is to study by Molecular Dynamics simulations the orientational order of small rigid solutes dissolved in a liquid crystalline phase. Experimentally this kind of informations are available by NMR spectroscopy and mainly from dipolar coupling data. Up to now this study is one of the first atomistic investigations about the alignment processes of solutes in a LC matrix. In the literature of the last decades a large number of theoretical models for the rationalization of the experimental measurements has been published, but none of them is able to explain completely the chemical-physic process involved. In this study we have tried to investigate the nature of this phenomenon and we have compared our simulation results with the available experimental data and theoretical models with the ambitious aim to explain the main host-guest orientational effects into a liquid cristalline phase.

We believe to reach these targets taking advantage of the possibility of changing inside the Force Field the parameters that allow us to estimate the contribution of alignment processes separately for dispersive and charge interactions. In this point of view after the validation of simulations with experimental observables, we have performed a series of simulations removing the partial atomic charges on solute molecules to estimate only the short-range dispersive (Lennard-Jones) force effects inside the LC solvent.

## 8.1 Dipolar couplings

The first step of our work is to validate simulation results with experimental ones comparing the calculated dipolar coupling with the NMR investigation. It is in fact possible to correlate the entity of the order of solutes dissolved in a nematic phase by the knowledge of their dipolar couplings. From simulation we can access to this quantities using the relation:

$$D_{ij} = \frac{\mu_0}{16\pi^2} \gamma_i \gamma_j \hbar T_{ij} \quad (8.1.1)$$

where  $D_{ij}$  is measured in Hz and  $\mu_0 = 4\pi \cdot 10^{-7} T^2 J^{-1} m^3$  is the magnetic permeability in vacuum;  $\gamma = g\mu_N/\hbar$  is the nuclear gyromagnetic ratio expressed in radiants. In this expression are present the nuclear factor  $g$ , constant for each nucleos, and the nuclear magneton  $\mu_N = e\hbar/2m_p = 5.051 \cdot 10^{-27} JT^{-1}$ .  $\mathbf{T}$  is the cartesian tensor of dipolar coupling.

$$\begin{aligned} \mathbf{T}^{(2)}(r) = \nabla \nabla \frac{1}{r} &= \begin{pmatrix} \frac{\partial^2 r^{-1}}{\partial x^2} & \frac{\partial^2 r^{-1}}{\partial x \partial y} & \frac{\partial^2 r^{-1}}{\partial x \partial z} \\ \frac{\partial^2 r^{-1}}{\partial y \partial x} & \frac{\partial^2 r^{-1}}{\partial y^2} & \frac{\partial^2 r^{-1}}{\partial y \partial z} \\ \frac{\partial^2 r^{-1}}{\partial z \partial x} & \frac{\partial^2 r^{-1}}{\partial z \partial y} & \frac{\partial^2 r^{-1}}{\partial z^2} \end{pmatrix} \\ &= -\frac{1}{r^5} \begin{pmatrix} 3x^2 - r^2 & 3xy & 3xz \\ 3yx & 3y^2 - r^2 & 3yz \\ 3zx & 3zy & 3z^2 - r^2 \end{pmatrix} \end{aligned} \quad (8.1.2)$$

The molecules of interest, in our work, have cylindrical simmetry  $D_{\infty h}$ , so that the off-diagonals components are zero. In addition for cylindrical simmetry we have the relation:

$$\text{Tr} [\mathbf{T}(r)] = 0 \quad (8.1.3)$$

and

$$T_{xx} = T_{yy} = -\frac{1}{2}T_{zz} \quad (8.1.4)$$

we can only consider the absolute value of the bigger component  $T_{zz}$  to obtain the dipolar coupling for the other ones.

This relations has validity for an effective  $D_{\infty h}$  symmetry. In the atomistic simulation the internal degrees of freedom can breake the molecular

symmetry so that in our calculations we do not obtain any strictly zero off-diagonal elements. This problem can be by-passed analysing a very large number of molecular conformations so that the statistical fluctuations can be considered negligible and it follows that the minimal acceptable length of simulation can be established by the values of these matrix elements.

So we have from the Eq. 8.1.1 the more useful form:

$$D_{ij}^{zz} = \frac{\mu_0}{16\pi^2} \gamma_i \gamma_j \hbar \frac{\langle 3 \cos^2(\theta) - 1 \rangle}{r^3} = \frac{\mu_0}{8\pi^2} \gamma_i \gamma_j \hbar \frac{\langle P_2 \rangle}{r^3} \quad (8.1.5)$$

that give us the possibility to extract, considering a rigid molecule model, the order parameter of an internuclear vector as:

$$\langle P_2 \rangle = \frac{8\pi^2}{\mu_0 \gamma_i \gamma_j \hbar} r^3 D_{ij}^{zz} \quad (8.1.6)$$

## 8.2 Mean Field models

The theoretical models used in literature for the interpretation and the prevision of solute orientation behavior in a liquid crystal solvent, are based on the Mean Field theory [Matteo et al. 2000]. This theoretical approach is built on the fundamental assumption that it is possible to substitute the specific interaction between the molecules (solute-solvent) by an effective mean field, generated from the solvent, that interacts with a single molecule of solute placed in a solvent cavity. This approach, given an  $N$  particle system, reduces the number of couple interaction terms from  $\frac{1}{2}N(N-1)$  (or more considering three body potential terms) to only a single effective potential term. This simplification reduces drastically the simulation time with respect to the explicit simulation models like MD.

After the formulation of the model, it is possible to obtain a ( $U_{MF}$ ) potential depending on a series of parameters extracted from experiments and from more sophisticated theoretical calculations. These parameters become the constraints representing our knowledge about the macroscopic observables of interest. The variables of the system are chosen in order to calculate a constrained maximum of the entropy distribution, [Jaynes 1957] obtaining the solving distribution function (in this case the orientational ones) and the  $MF$  potential.

### 8.2.1 Maier Saupe model

The Maier–Saupe model [Straley 1974] is the father of all the  $MF$  models applied to solute molecules dissolved in liquid crystal. It introduced the order parameter referred to the solute as follows:

$$S = \langle P_2(\beta) \rangle, \quad (8.2.7)$$

where  $\beta$  is the angle between the nematic phase director and the principal molecular axis,  $P_2(\beta) = \langle 3(\cos^2\beta - 1)/2 \rangle$  is the averaged ensemble of second order Legendre polynomial.

Supposing the pair potential for interacting particles to be:

$$U_{ij} = \epsilon P_2(\beta_{ij}) = \frac{1}{2}\epsilon(3 \cos^2 \beta_{ij} - 1) \quad (8.2.8)$$

where  $\beta_{ij}$  is the angle between the principal axes of molecules  $i$  and  $j$ , the angular distribution function for solute molecule is derived from the  $MF$  potential explicitly as:

$$U_{MF} = \epsilon S P_2(\cos \beta_{ij}) \quad (8.2.9)$$

### 8.2.2 Surface tensor model

In this model, conceived from [A.Ferrarini and P.L.Nordio 1998], the solute–solvent interaction is estimated from the angular dependence of the normal vector of the single surface element placed on solute molecular surface and the mean field generated from the solvent that changes as  $P_2(\cos\theta_{\hat{n}})$ .

It is assumed that the normal vector  $\hat{s}$ , related to a infinitesimal molecular surface element  $dS$ , tends to align perpendicularly to the nematic director  $\hat{n}$ . This behavior is related to the energy contribution

$$dU_{st} = \epsilon P_2(\hat{s} \cdot \hat{n}) dS \quad (8.2.10)$$

where  $P_2$  is the second order Legendre polynomial and  $\epsilon$  is the intensity of the interaction. Integrating over all the molecular surface  $S$  the global potential [Matteo et al. 2000] is obtained.

$$U_{st}^{(\Omega)} = \epsilon \int_S P_2(\hat{s} \cdot \hat{n}) dS \quad (8.2.11)$$

For uniaxial molecules the orientational dependence is generally parametrized by the angle  $\Omega$  of principal inertial molecular axis with respect the nematic director.

From Eq. 8.2.11 it is possible to notice that both the surface tensor and the potential  $U_{st}$  are proportional to the total molecular surface so that, for small molecules, the short range energy contribution slightly contributes to the total orientational potential.

### 8.3 Samples preparation

The MD solution sample simulated are composed from 238 molecules of 5-cyanobiphenyl (5CB solvent), 2 molecules of 1,3,5-trichlorobenzene (TCB standard) and 10 molecules of solute. We have studied the following solutes:

benzene (BEN), 1,3,5-tribromobenzene (TBB), 2,4-hesadiene (ESA), acetylene (ENE), propyne (INO), acetonitrile (NIT) and biphenyl (BIF).

All the samples are obtained from a box of 250 5CB molecules with the substitution method showed in App. F.

Experimentally the complexity of NMR spectra requires a different series of measures for each solute. This is due to the solute<sub>A</sub>-solute<sub>B</sub> interactions that interfere with dipolar coupling determination. Since the orientational order of solute in liquid crystal is widely influenced from specific interaction with solvent and from the perturbation effect on the solutes, as well as from concentration, temperature and solvent purity, it becomes necessary the presence of a molecule as an orientational reference to be able to compare the small effects present in different experiments. The solutes are small and rigid molecules simple to be studied experimentally with respect the bigger and flexible liquid cristal molecules. In the NMR spectra is possible to discern, above the band of 5CB molecules, the sharp lines due to the solute.

Our simulation samples have been built to be chemically similar to the experimental results present in literature[Weaver et al. 1987]. For this reason we have inserted also the standard molecule (TCB), but we do not really need iti for our determination of dipolar couplings, for doing which we only need to know the atomic positions, available directly from simulations.

## 8.4 Simulation details

The standard and the solutes have been described with fully atomistic detail, the atomic charges have been computed [Besler and and P. A. Kollman 1990] at minimum energy geometry using the B3LYP density functional technique and a  $6-311^{++}G^{**}$  basis set [various authors 2000][et al. 2004]. The solvent molecules have been described in the `united atoms` detail (for a total sample dimensions of 250 molecules) obtained from our previous work [Tiberio et al. 2006a] discussed in Chap. 5. After the insertion of each solute we have performed an equilibration run of 2 ns at chosen thermodynamic conditions ( $NPT$  conditions 290 and 295 K using Velocity Scaling thermostat, at  $P=1$  atm by isotropic Parrinello-Rahman barostat).

In order to investigate the contributions of different intermolecular forces we have performed two series of simulations. First of all we considered the complete force field contribution and compared the results with the experimental data. After we removed the electrostatic contributions of solute molecules leaving only the Lennard–Jones interactions. This, in the simulation, can be easily done setting to zero the atomic charges. The differences between this results can be used to confirm or not many hypothesis proposed in the mean field theory in which the dipole–dipole contributions are neglected. Experimentally a solute that cannot establish long range electrostatic interactions is defined as a “magic” one [Burnell and Lange 1998]. So, in our simulations, we transform our common solutes in *magic solutes*.

## 8.5 Results and discussion

The presence of the solute molecules inside the nematic phase produces, in both cases, a sensible alteration in the solvent order. In our simulations the stronger effects due to the solute are observed for TBB and ESA. In these samples we notice at 295 K that the time dependence of the phase order sensibly oscillates from nematic close to isotropic values (from  $\langle P_2 \rangle = 0.5 - 0.8$  to  $\langle P_2 \rangle = 0.1 - 0.2$ ) indicating the proximity to the  $T_{NI}$ . This means that these solutes decrease the nematic–isotropic transition temperature of 5CB from 308.5 to 295 K. The presence of the solute, like is generally observed for freezing point depression, tends to reduce the thermal stability of more ordered phases so that the transition temperatures of solutions are lowered with respect to the pure solvent ones. Differently from cryoscopic effects



related to isotropic solvents, in which the concentration of solute gives the main contribution, in the nematic–isotropic transition we also have to consider the molecular shape. In our series of simulations, having the same molar concentration, the solute molecular shape (including solute-solvent charge interactions) is the “parameter” that determines the different behavior of each solution.

### 8.5.1 Order parameters

We start to discuss about the order parameter, that is the alignment of solutes, defining the molecular axes chosen for each solute. The common attribution of molecular principal axis, for approximately uniaxial molecules, is the principal axis of inertia, i.e. the one with lowest inertia moment for a rod-like molecule. In this study we can do it for prolate solutes like ESA, INO, ENE, NIT and BIF. For the other ones, BEN, TBB and TCB, this attribution is not correct because these molecules cannot be defined as “rod-shaped”, but as “disk-shaped”, or oblate ones. For oblate molecules the best choice of principal molecular axis falls to the shorter one (or better the greater inertial moment). So in the following we implicitly refer to the molecular axis as described before: for prolate molecule the longer molecular axis (prolate  $I_z < I_x = I_y$ ); and for oblate solutes the shorter one (oblate  $I_z > I_x = I_y$ ).

#### Simulations with charges

From the comparison with experimental data (at 294 K except for BEN at 296 K) available (Tab. 8.1) we notice that there is good agreement with our simulation values (at 295 K) (Tab. 8.2).

The biggest deviations have been found in the TBB and ESA simulations where the closeness to the predicted simulation phase transition gives us a large deviation from experiments even though the signs of the obtained order parameters are correct. It seems clear that the difference order parameter of solvent depends to the presence of a different interaction with solutes. Oblate molecules (BEN, TBB and TCB) tend to arrange themselves perpendicularly with respect to the sample director (or in other words with phenyl group parallel to 5CB phenyls orientation). This can be easily understood considering their molecular shape. It is more difficult to explain the relative low tendency of ENE (the more “rod-like” shaped molecule). This can be ascribed to the different molecular section with respect the 5CB. The lower

Table 8.1: † Experimental order parameters for solutes in 5CB at 294 K.  
 \* Experimental order parameters for benzene in 5CB at 296 K. Data from [Weaver et al. 1987]. Experimental data for BIF and NIT not available.

Campione	$S(exp)$	$S(calc)$	$S(calc)$
BEN*	-0.14	-0.14	-0.12
ENE†	0.17	0.14	0.17
ESA†	0.35	0.37	0.44
INO†	0.17	0.15	0.16
TBB†	-0.22	-0.21	-0.22
TCB†	-0.21	-0.19	-0.21

widht of the ESA molecule can induce an higher orientational freedom that tends to disorder the sample. Similar topics can be considered for INO, but contemplating the flexible portion of 5CB. The importance of charges interactions is clear analysing the order parameter of BIF. The molecular shape of this solute is identical to the rigid fragment of 5CB molecule, but the order parameter is noticeably lower (about 0.2–0.3). In this case the solute has not any dipole moment, so that the necessity to be strictly antiparallel with the solvent molecule vanishes. This reduces the order inside the phase.

Table 8.2: Simulated inertial order parameters of solvent, standard and solutes at 290 and 295 K.

Solution	5CB		Standard		Solute	
	290 K	295 K	290 K	295 K	290 K	295 K
BEN	0.60	0.46	-0.29	-0.19	-0.23	-0.18
BIF	0.63	0.55	-0.28	-0.25	0.33	0.29
ENE	0.51	0.50	-0.23	-0.22	0.16	0.16
ESA	0.59	0.32	-0.26	-0.15	0.29	0.14
INO	0.58	0.50	-0.26	-0.22	0.22	0.17
NIT	0.59	0.49	-0.26	-0.24	0.18	0.14
TBB	0.56	0.37	-0.25	-0.16	-0.25	-0.15

### Simulations without charges

In this section we can directly evaluate the effects of solute–solvent charge interaction influence on the alignment of solutes inside the nematic phase.

Excluding the electrostatic interactions (setting parameter to zero the partial charges on the solute molecules) we observe a general order reduction at 290 K (from 0.03 to 0.1 Fig. 8.1[a]). Probably the charge interactions tend to reduce the rotational and translational degrees of freedom of solutes causing a little increase of order in the LC matrix.

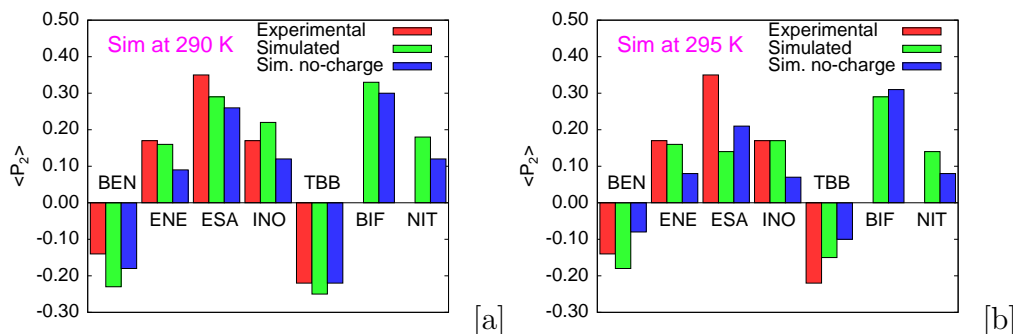
Table 8.3: Simulated inertial order parameters of solvent, standard and solutes at 290 and 295 K for no-charged solutes.

Solution	5CB		Standard		Solute		$\Delta \langle (P_2)_{solute} \rangle$	
	290 K	295 K	290 K	295 K	290 K	295 K	290 K	295 K
BEN	0.57	0.24	-0.29	-0.13	-0.18	-0.08	-0.05	-0.10
BIF	0.52	0.55	-0.23	-0.24	0.30	0.31	0.03	-0.02
ENE	0.52	0.50	-0.24	-0.24	0.09	0.08	0.07	0.08
ESA	0.56	0.49	-0.26	-0.21	0.26	0.21	0.03	-0.07
INO	0.59	0.35	-0.26	-0.18	0.12	0.07	0.10	0.10
NIT	0.62	0.45	-0.29	-0.26	0.12	0.08	0.06	0.06
TBB	0.55	0.24	-0.25	-0.15	-0.22	-0.10	-0.03	-0.05

The situation is a bit different at 295 K where a small increase of order with respect to the “normal” solutes (see Fig. 8.1[b]) is observed for ESA (BIF simulation are not considered because the differences are smaller than simulation order parameter error ( $\Delta \langle (P_2)_{solute}^{290K} \rangle - \Delta \langle (P_2)_{solute}^{295K} \rangle \rangle < 0.05$ ). As observed for solutes with a complete force field description, also in this case at T=295 K we notice a wide order parameter oscillations in time for BEN, INO and TBB indicating the closeness to simulation phase transition for these solutions, while in the previous series of simulations these oscillations were observed for TBB and ESA. By the combination of these results it seems clear that the inversion in the sign of the difference between order parameters  $\Delta \langle (P_2)_{solute} \rangle = \langle (P_2^{charged})_{solute} \rangle - \langle (P_2^{no-charged})_{solute} \rangle$  at 295 K with respect to T=290 K, showed from ESA, is simply due to the closeness to the phase transition in charged simulations, which is not observed in the no-charged case. This shows us a small direct influence on  $T_{NI}$  of charge interactions, which can be estimated as an oscillation of about  $\pm 5$  K around the standard one. TBB destabilizes the nematic phase similarly both in charged and no-charged solutions; it is probably due to its shape (the bromine groups have dimension comparable to a  $CH_3$  group) while the local polarization of  $C-Br$  bonds seems to have a marginal rule. The opposite situation, with respect to

ESA simulations, can be observed in the solution with INO. In this case the charge suppression induces a strong reduction of the nematic phase stability and the  $\langle P_2 \rangle(t)$  oscillations in the without charges simulation indicate that the phase transition is approaching.

Figure 8.1: Graphical comparison among experimental data, simulations and simulations with no-charged solutes at 290 [a] and 295 K [b]



Only in the BIF solution case we can assume that the order is not influenced by the presence of charges on the molecule. In fact the differences between the order parameters, both for solute and solvent, are in practice not important. So BIF is the only case, inside our solute group pool, in which the mean field assumption about the negligible effect of charge interactions can be correctly applied even if for all the solutes the predominant ordering effect is due to the shape and to LJ interaction more than electrostatic effects. Concerning the order parameters of the standard, we can observe a similarity of behavior in all the solutions except for the case of closeness to the N-I phase transition. The averaged order of TCB seems to be slightly greater of TBB. This can be ascribed to the similarity of dimension between TCB molecule and the phenyl groups of 5CB. Moreover, the comparison between the order parameter of standard and solutes must be done considering the different number of molecules present (2 of standard and 10 of solute) so that the errors over the order parameters are different.

## 8.5.2 Dipolar coupling

Using the results (or better the orientations) obtained in the previous section we want to compare the dipolar couplings obtained from simulations with the experimental ones, through two different relations in Eq. 8.1.1 and through:

$$D_{i,j} = -2 \frac{\mu_0}{16\pi^2} \gamma_i \gamma_j \hbar \frac{P^2}{r^3}. \quad (8.5.12)$$

The dipolar coupling for the solvent must be considered as an upper limit of our values because is obtained for pure 5CB [Stevensson et al. 2001]. The results obtained from Eq. 8.1.1 at 290 K are in better agreement with experimental data than the estimation done using Eq. 8.5.12 excepting INO for which we predict a wrong dipolar coupling with opposite sign. Anyway in other cases the larger differences are observed for ENE and TBB probably related to the bigger tendency to destabilize the simulated nematic phase just observed in order parameter analysis. In the simulations without charges we can add the no-charged BIF simulation in the list of greater destabilization solute while for other solutes the intensity of dipolar couplings can be considered identical to the respective simulations with charges.

Table 8.4: Simulated dipolar coupling in Hz of 5CB related to C2 C3 atoms, obtained at 290 K and 295 K, to be compared with the experimental value of -1299 Hz obtained for pure 5CB [Stevensson et al. 2001].

Solution	Charged solutes				No-charged solutes			
	$D_{ij}$		$D_{ij}^\dagger$		$D_{ij}$		$D_{ij}^\dagger$	
	290 K	295 K	290 K	295 K)	290 K	295 K	290 K	295 K)
BEN	-1217	-948	-1208	-940	-1165	-489	-1156	-485
BIF	-1290	-1125	-1280	-1116	-1060	-1132	-1052	-1123
ENE	-1050	-1025	-1042	-1017	-1062	-1026	-1054	-1018
ESA	-1202	-664	-1193	-659	-1145	-1008	-1136	-1000
INO	-1192	-1023	-1184	-1015	-1203	-710	-1193	-705
NIT	-1201	-1001	-1191	-993	-1275	-921	-1265	-914
TBB	-1157	-747	-1149	-742	-1126	-479	-1117	-476

†calculated from equation 8.5.12

## 8.6 Comparison with mean field models

Now we can analyze the coherence among our data and two mean field models: Maier-Saupe and Surface tensor. This comparison is important for

Table 8.5: Simulated dipolar coupling in Hz of H4 H8 atoms for the standard. Experimental data to compare with: -213.23 Hz in 1132; -156.60 Hz in magic mixture; -148.12 in EBBA [der Est et al. 1987].

Solution	Charged		No-charged	
	290 K	295 K	290 K	295 K
BEN	-232	-166	-221	-152
BIF	-186	-224	-193	-178
ENE	-153	-120	-196	-129
ESA	-188	-100	-187	-188
INO	-236	-169	-121	-107
NIT	-188	-171	-245	-177
TBB	-182	-155	-174	-105

confirming the validity and the application range of these approaches that completely exclude the solvent detail expressed only through a field averaged over all the present couple interactions. This comparisons will be done by considering the mean field in our simulations as generated from 5CB expressed in terms of  $\langle P_2^{5CB} \rangle$  and the tumbling energy necessary to rotate the solute molecule perpendicularly with respect to its principal axis.

### 8.6.1 Maier–Saupe

In the Maier–Saupe model (see App. E.0.1) the  $\epsilon$  parameter should be, in principle, temperature independent and related only from the solute–solvent couple. In our simulation we have calculated the tumbling energy  $U_\beta$  by inversion of angular distribution function  $U_\beta = -RT \log P(\beta)$ , and than we compare with the  $U_{MF}$  energy Eq. E.0.3 with  $S = \langle P_2^{sim} \rangle$ . From data analysis (see Tab. 8.7 and Fig. 8.2) we can define three different groups of chemical behavior:

1) prolate molecules with conformational rigidity (INO and NIT) with dipole moment. We obtain a good agreement and a low standard deviation confirming the availability of MS model to reproduce the simulation data. The  $\epsilon$  values are quite similar. This means that in this case the molecular details are not fundamental to the determination of solute order inside the LC matrix.

2) prolate molecules with conformational rigidity (BIF, ESA and ENE) without dipole moment. In this case we obtain a good agreement with MS

Table 8.6: Mean solute dipolar couplings (in Hz) from [der Est et al. 1987] and from simulations

Solution	Atoms				Charged		No-charged	
		1132	magic§	EBBA	$D_{ij}^{290K}$	$D_{ij}^{295K}$	$D_{ij}^{290K}$	$D_{ij}^{295K}$
BEN	$H_6H_{10}$	-191.73	-137.15	-88.46	-174	-132	-181	-66
BIF	$H_9H_{11}$	-	-	-	-2791	-2406	-148	-56
ENE	$H_3H_4$	-644.6	-298.56	145.8	-547	-538	-840	-562
ESA	$H_{10}H_{11}$	-120.51	-101.06	-74.36	-99	-49	-287	-271
INO	$H_4H_5$	2853.10	1656.75	595.23	-1273	-1033	-112	-100
NIT	$H_4H_6$	-	-	-	-1055	-921	-829	-652
TBB	$H_6H_2$	-213.56	-159.2	-119.90	-209	-99	-2557	-2602

§ 55%peso 1132, 45% EBBA

model (low fit  $\sigma_\epsilon$ , but the  $\epsilon_{ENE}$  parameter is not similar to the others. In this case the molecular a similar chemical properties cannot be done and we must consider the structural details for each solute separately.

3) oblate molecules (BEN and TBB) without dipole moment. The oblate shape of molecules probably is the main responsible to the worst fitting parameter errors obtained in this case. The solute-solvent interactions are very different and the local charge distributions determine different  $\epsilon$  parameter. The Maier-Saupe model cannot represent satisfactory the orientational behavior of this group of molecules.

Table 8.7:  $U_{MF}$  parameters fitted from simulation distribution functions with errors

Soluto	$\epsilon$	$\sigma_\epsilon$
BEN	$-7.5 \pm 0.21$	0.13
BIF	$-2.8 \pm 0.03$	0.03
ESA	$-2.8 \pm 0.03$	0.03
ENE	$-5.2 \pm 0.10$	0.03
INO	$-4.7 \pm 0.06$	0.02
NIT	$-4.2 \pm 0.06$	0.03
TBB	$-4.5 \pm 0.10$	0.07

## 8.6.2 Surface tensor

We have tried to investigate the capabilities of Surface Tensor model to reproduce our simulation data. First we observe a good agreement of this MF model to predict correctly the sign of the order parameter simply consider the molecular surface obtained from Eq. 8.6.13. For oblate molecules is predicted a perpendicular alignment while for prolate a parallel one.

Next we proceed to calculate the molecular surface and integral in Eq. 8.6.13 numerically by Ferrarini's methods [A.Ferrarini and P.L.Nordio 1998, A.Ferrarini et al. 1992]. To simplify the functional form of we approximate the integral in Eq. 8.6.13 with a polynomial:

$$\int_S P_2(\mathbf{n} \cdot \mathbf{s}) dS = a_0 + a_1 \cos^2 \beta + a_2 \cos^4 \beta + a_3 \cos^6 \beta \quad (8.6.13)$$

where  $\beta$  is the angle between the molecular principal axis and the sample director. In Tab. 8.8 are shown the results for the molecules in study. The trend of this Equation are showed in Fig. 8.3[a] and in Fig. 8.3[b] the function are divided by molecular surface to better compare the different solutes contribution. We can observe that the surface tensor model can not only discriminate oblate to prolate molecule but also establish from more or less tendency to molecules to be prolate or oblate. This means (see Fig. 8.3[b]) that we have a orientational depending instrument to compare the different interaction due to the shape having very different dimension (for example BIF with respect to ENE). From this preliminary results the most interacting prolate solute with 5CB is BIF either in terms of total surface exposed to solvent either in terms of molecular anisotropy. Interesting is the behavior of ENE. It shows a total energy interaction (see Fig. 8.3[a]) similar to INO, but a surface reduced one similar to ESA(see Fig. 8.3[b]). This means that the interaction per surface unit of ENE with solvent is lower than the INO one, but the larger surface dimension of the first one finally increases the total energy interaction of ENE at level of INO.

Finally, by fitting on simulation tumbling energy obtained from solute orientational distribution functions, we fit the  $\xi$  parameter, in Eq. ?? that should be independent to the solute observed, depending only on solvent properties and temperature. In Tab. 8.9 we report the obtained parameter. We can observe at T=290 K similar parameter for BEN, ENE, INO and NIT, but BIF, ESA and TBB seems to have a different series of parameter. At T=295 K the BIF, ENE, INO, NIT e TBB show similar value. Suddenly ee



Table 8.8: Molecular surface ( $\text{\AA}^2$ ) calculated using the method of Ferrarini *et al* [A.Ferrarini and P.L.Nordio 1998, A.Ferrarini et al. 1992], and fitting of integral in Eq. ?? by relation Eq. 8.6.13

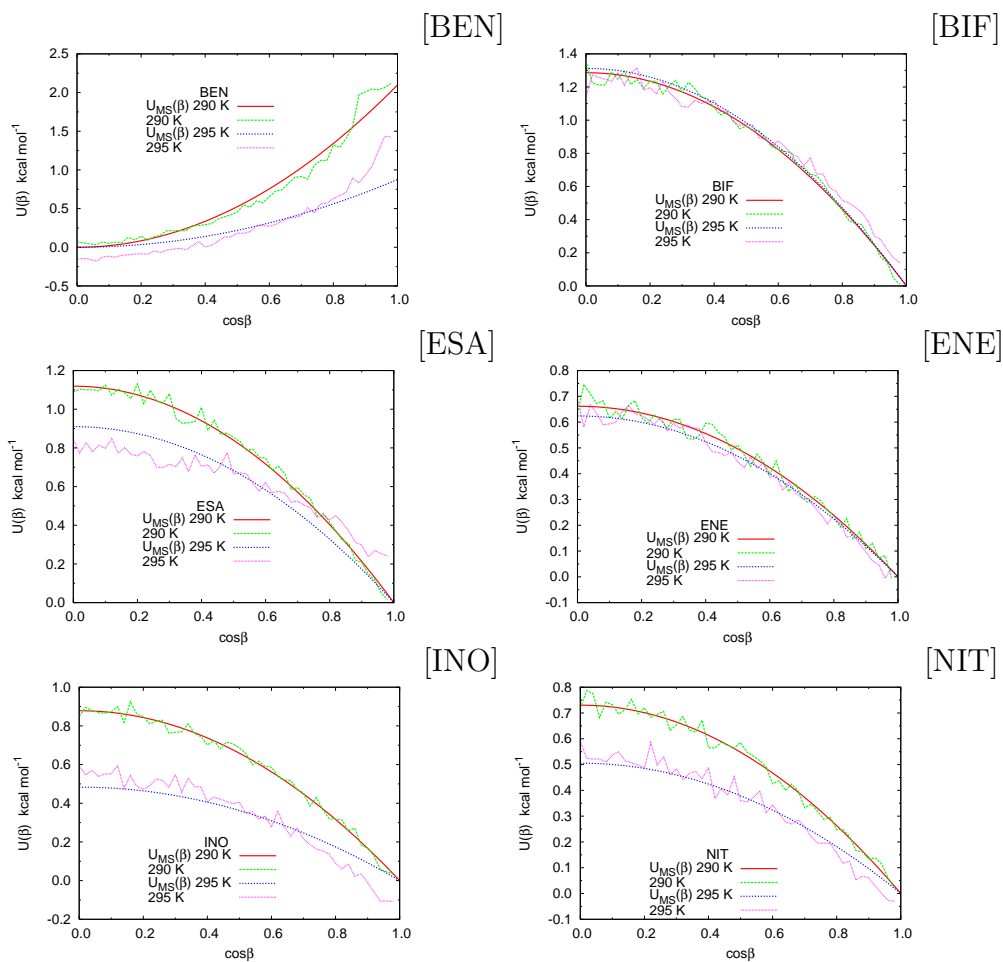
solute	surface	$a_0$	$a_1$	$a_2$	$a_3$
BEN	$112.7 \pm 1.1$	-24.0994	72.30	-0.01365	+0.01153
BIF	$186.5 \pm 1.4$	-43.9238	131.76	+0.04852	-0.04139
ENE	$54.8 \pm 1.0$	-6.9231	20.77	+0.00559	-0.00500
ESA	$140.6 \pm 0.8$	-19.7446	59.25	-0.09196	+0.07814
INO	$76.6 \pm 0.5$	-8.6370	25.91	+0.01695	-0.01436
NIT	$70.6 \pm 0.7$	-6.7664	20.29	+0.03420	-0.02920
TBB	$194.5 \pm 0.7$	-34.3307	103.00	-0.02468	+0.02116
TCB	$166.9 \pm 0.5$	-31.5886	94.75	+0.07346	-0.06263

obtain correct trend only for three solute ENE, INO and NIT. After these considerations we can assert that the model can give us only a correct qualitative description of solute solvent interaction but fails in the quantitative predictions for molecules strongly different in shape.

Table 8.9:  $\xi$  constant value and standard deviation in ( $\text{kcal } \text{\AA}^{-2} \text{ mol}^{-1}$ ) of Eq. ?? obtained from fitting of fixed function defined in Eq. 8.6.13 over the tumbling energy simulation data obtained from Boltzman inversion of rotational distribution function.

solute	$\xi$		$\sigma_\xi$	
	290 K	295 K	290 K	295 K
BEN	+0.0300	+0.0196	0.11	0.08
BIF	-0.0098	-0.0086	0.04	0.04
ENE	-0.0321	-0.0321	0.03	0.03
ESA	-0.0190	-0.0101	0.03	0.03
INO	-0.0341	-0.0272	0.02	0.03
NIT	-0.0363	-0.0296	0.03	0.03
TBB	+0.0155	+0.0097	0.03	0.04

Figure 8.2: Orientational tumbling energy for Maier–Saupe model functions fitted over simulation data.



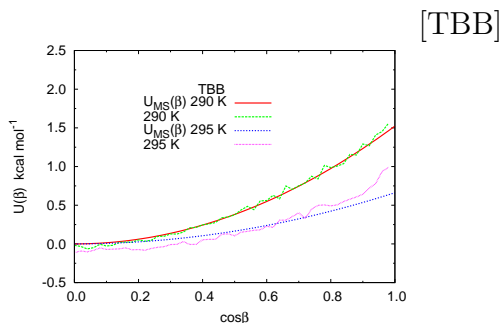
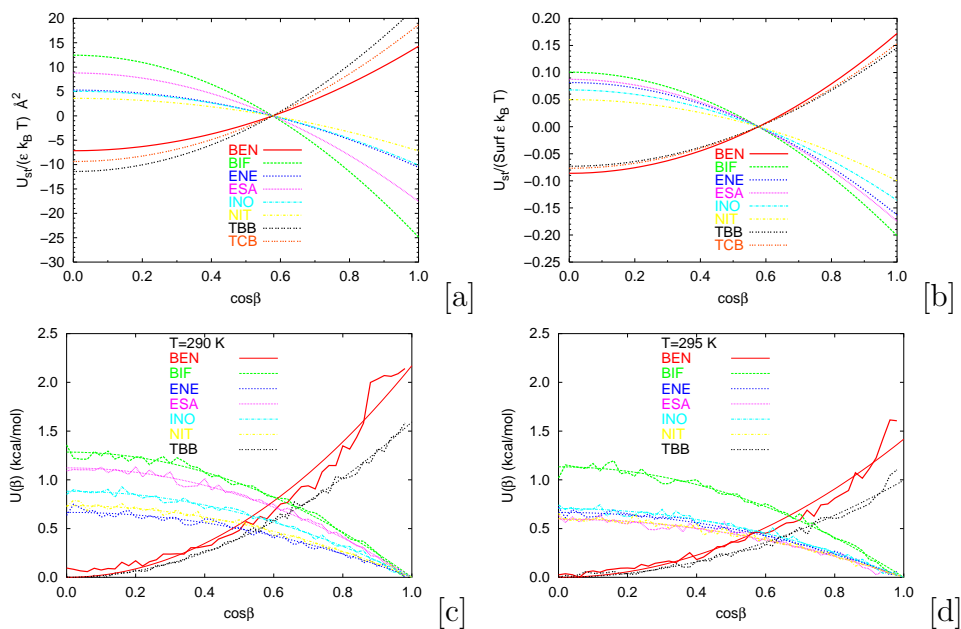


Figure 8.3: [a] Integral function from Eq. ?? versus molecular orientation. [b] Integral function reduced for molecular surface from Eq. ?? versus molecular orientation. [c]  $U_{ST}$  tumbling orientational energy versus molecular orientation fitted on simulation data at  $T=290$  K. [d]  $U_{ST}$  tumbling orientational energy versus molecular orientation fitted on simulation data at  $T=290$  K.





# Chapter 9

## Avian prion hexarepeat

### 9.1 Introduction

The prion protein ( $PrP^C$ ) is a copper binding glycoprotein that can misfold into a sheet rich pathogenic isoform ( $PrP^{Sc}$ ) leading to prion diseases, particularly in mammals, as opposed to avians. Mammal and avian proteins show different histidine and proline containing N-terminal tandem repeats:  $PHGGGWGQ$  and  $PHNPGY$ . The NMR conformational analysis of avian  $PrP^C$  N-terminal region has been hampered until now by its great flexibility. Here the effect of prolines and pH on conformational changes was determined to clarify the role of histidine and tyrosine residues on the secondary structures of the single hexarepeat  $PHNPGY$  (1 – *HexaPY*) with both N and C-termini blocked by acetylation and amidation respectively, by means of NMR and of an MD study at various pH. An NMR study of the hexarepeat in trifluoroethanol was also carried out to investigate the dependence of the conformation on the solvent polarity. We found a better defined bent structure in trifluoroethanol than in water and a strong conformation dependence on pH: local conformations such as type I beta turn and elements of  $PPII$  are present at acid and neutral pH, while at basic conditions unordered dihedral regions dominate.

Introduction The prion protein ( $PrP^C$ ) is a cell surface glycosylphosphatidylinositol (GPI) anchored glycoprotein in [Prusiner 1997, Prusiner et al. 1998]. Its biological function has not yet been identified, but it has been established that a conformational change gives rise to the sheet rich and pathogenic isoform  $PrP^{Sc}$  ([Prusiner 1997],[Prusiner et al. 1998]). Avian species also express prion protein, although no evidence of neurode-

generative disorders have been reported among them [Procacci and Marchi 1996] and chicken prion protein (*ChPrP<sup>C</sup>*) has initially been described to induce acetylcholinereceptor activity ([Harris et al. 1991; 1993]). This protein shares about 33% of primary sequence identity with the mammalian one and some essential features are conserved ([Wopfner et al. 1999]). In particular both proteins possess: i) multiple *N* glycosylated sites; ii) an aminoterminal signal sequence that is removed in the mature protein; iii) a carboxyterminal signal that is eliminated when the mature protein is linked to GPI, and iv) an *N*-terminal domain featured by tandem amino acid repeats (*PHNPGY* in avians, *PHGGWGQ* in mammals) followed by a highly conserved hydrophobic core ([Wopfner et al. 1999, Harris et al. 1991]). Interestingly, it has been reported that the *N*-terminal domain is essential for efficient endocytosis via clathrincoated pits in cultured neuroblastoma cells, embryonic neurons and glia for chicken as well as mammalian prion protein ([Shyng et al. 1994; 1995]). Differently from mammalian *PrP<sup>C</sup>*, the digestion of *ChPrP<sup>C</sup>* with trypsin or proteinase K produces peptide fragments stable to further proteolysis. One of these fragments comprises the 49129 sequence, consisting of a large part of the *N*-terminal domain ([Marcotte and Eisenberg 1999]). This resistance to proteolysis may suggest a compact domain of the proline/glycine rich *N*-terminus, although the hexarepeats amino acid sequence seems to show no tendency towards a particular structured conformation such as  $\alpha$  or  $\beta$ , but rather to have a randomcoil conformation ([Marcotte and Eisenberg 1999]). More recently, NMR results showed that the *ChPrP<sup>C</sup>* globular domain contains three helices and two short antiparallel sheets, with a spatial arrangement similar to that observed in mammalian *PrP<sup>C</sup>* ([Lysek et al. 2004, Calzolari et al. 2005]). The *N*-terminal domain results to be flexible and unstructured and the 5073 residues, encompassing the tandem hexarepeat sequence, could not be individually assigned yet by NMR ([Lysek et al. 2004, Calzolari et al. 2005]). It is also known that the *cis-trans* isomerization in *XxxPro* peptide groups is one of the rate determining steps in protein folding. In native and folded proteins the *XxxPro* exists essentially in either *cis* or *trans* form because the interactions with neighbouring groups can favour one of them, but an equilibrium exists in unfolded proteins. In small peptides containing proline residues, the two forms are both present because are almost isoenergetic, with an observed increase of *cis* $\rightarrow$ *trans* isomerization going from polar to nonpolar solvents ([Radzicka et al. 1992, Jhon and Kang 1999]). In solution, polyproline peptides have been shown to exist predominantly in the polyproline II (*PPII*) conformation, a lefthanded 31 helix

in which all the residues are in the *trans* form. This conformation plays an important role in several biological processes such as signal transduction, cell motility and immune response ([Kay et al. 2000, Chandra et al. 2004]). Different short peptides not containing proline are supposed to partially adopt this conformation, but it seems quite sure that proteins domain having an intrachain PXXP motif fold into a *PPII* helix. The adoption of a *PPII* conformation is a prerequisite for the recognition of a peptide or protein fragment by “Src homology domain 3” (SH3). One of these motifs has been identified in mammalian prion proteins as well as avian and it corresponds to the region 101104 in mouse PrP<sup>C</sup> and 107111 in chicken *PrP<sup>C</sup>* ([Lysek and Wthrich 2004]). However, as indicated by NMR data, the N-terminal proline rich region of *ChPrP<sup>C</sup>* is unstructured and this can suggest, besides the wide and conformational space of the N-terminal backbone, the presence of an equilibrium between *cis* and *trans* form in each *XxxPro* peptide groups. On the other hand the stabilization of one isomer due to side chains of other residues groups, particularly histidine and tyrosine, could also be possible. In this context, Molecular Dynamics (MD) studies are useful to provide support for, and to clarify, the different conformational states populated in a faster (nanoseconds) time scale than that of NMR, which typically explores time scales a few orders of magnitude longer. Therefore, both the the combined NMR and MD studies represent a powerful tool to investigate the conformational equilibria of peptides and proteins in solution, once the accuracy of MD simulations in predicting experimental results has been validated. Here we report an NMR and a pH dependent MD study concerning the conformation in water of the single hexarepeat *PHNPGY* (1 – *HexaPY*) with both *N* and *C*-termini blocked by acetylation and amidation respectively. An NMR study of the hexarepeat in trifluoroethanol was also carried out to investigate the dependence of the conformation on the solvent polarity. This peptide may represent a starting model that can be used for successive studies concerning two and four hexarepeats. The different behavior of mammals *PrP<sup>C</sup>* and chicken *PrP<sup>C</sup>* protein in the interaction with protease ([Marcotte and Eisenberg 1999]) could be suggested on the basis of the results concerning the molecular features of their *N*-terminal domain.

## 9.2 Materials and methods NMR Measurements.

The peptide under study was synthesized as previously reported ([Mendola et al. 2005]). All the experiments were carried out at 500 MHz on a VARIAN NOVA UNITY PLUS located at the Department of Chemical Sciences, Catania (Italy), and on a VARIAN UNITY 500 spectrometer, located at the Department of Environmental Sciences, Caserta (Italy). Spectra were processed using the VARIAN VnmrJ and XEASY ([Bartels et al. 1995]) software. Sample solutions (5mM) were prepared in  $\text{TFEd}_3/\text{H}_2\text{O}$  (80/20 v/v),  $\text{H}_2\text{O}/\text{D}_2\text{O}$  90/10 (v/v) or neat  $\text{D}_2\text{O}$ . NMR spectra for the threedimensional structure determination were collected at 300 K and referenced to external TMS ( $\delta = 0$  ppm); the pH of the aqueous solution was adjusted to 4.2. The dependence of the amide chemical shifts on the temperature was observed in the range 300311 K. Furthermore, for all the solvent systems the possible occurrence of aggregation was verified by analyzing the cross peak patterns in ROESY spectra, recorded at a peptide concentration of 0.8 mM. Deuterated  $\text{D}_2\text{O}$  (99.9% relative isotopic abundance) and TFE $d_3$  (99%) were purchased from Cambridge Isotope Laboratories. Mono (1D) and two dimensional (2D) spectra were accumulated with a spectral width of 6000 Hz in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  and 4800 Hz in  $\text{TFE}/\text{H}_2\text{O}$  (80/20 v/v). 2D experiments DQFCOSY ([Rance et al. 1983]), TOCSY ([Braunschweiler and Ernst 1983]), ROESY ([Griesinger and Ernst 1987]) and NOESY ([Kumar et al. 1980]) were recorded in the phase sensitive mode using the States Haberkorn method. Water suppression was achieved by DPFGE sequence ([Hwang and Shaka 1995]). TOCSY, NOESY and ROESY spectra were acquired with mixing time of 70, 250 and 150 ms, respectively. Typically, 64 transients of 4K data points were collected for each of the 256 increments; the data were zero filled to 1K in 1. Squared shifted sinebell functions were applied in both dimensions prior to Fourier transformation and baseline correction.

### 9.2.1 Structure Calculations

Due to unfavourable correlation time ( $\tau_c$ ) NOESY experiments were scarcely informative and not suitable for structure calculations. Therefore experimental distance restraints for structure calculations were derived from the cross-peak intensities in ROESY spectra, recorded in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  and  $\text{TFE}/\text{H}_2\text{O}$



(80/20 v/v). The ROESY cross peaks were manually integrated using the XEASY software ([Bartels et al. 1995]) and converted to upper distance constraints according to an inverse sixth power peak volumetodistance relationship for the backbone and to an inverse fourth power function for side chains, by using the CALIBA module of the CYANA program ([Herman et al. 2002]). Distance constraints together with the obtained scalar coupling constants were then used by the GRIDSEARCH module, implemented in CYANA, to generate a set of allowable dihedral angles. Structure calculations, which used the torsion angle dynamics protocol of CYANA, were then started from 100 randomized conformers. The 20 conformers with the lowest CYANA target function were further refined in vacuo by means of unrestrained energy minimization, using the GROMOS 96 ([Gunsteren et al. 1996]) force field with the program SPDB viewer ([Guex and Peitsch 1997]). Several cycles of steepest descent were repeated until the energy difference between two successive steps was less than 103 kJ/mol. The structure analysis has been performed with the program MOLMOL ([Koradi et al. 1996]). Moreover, the 20 conformers with the lowest CYANA target function were minimized in a water shell and in a box with TFE/H<sub>2</sub>O 80/20. The energy minimization was carried out with a conjugate gradient method having a tolerance in energy of 105 kJ mol<sup>-1</sup>, using ORAC 4.0 program ([Procacci et al. 1997b]) and the Amber95 force field ([Cornell et al. 1995]).

### 9.2.2 Molecular Dynamics

All the simulation were run in water using ORAC 4.0 program ([Procacci et al. 1997b]) and the Amber94 force field (FF) ([Cornell et al. 1995]), that have been already used for proteins similar to our one ([Barducci et al. 2005]); a cubic box containing one 1 – *HexaPY* chain and 1184 water molecules was used applying periodic boundary conditions ([Schlick 2002]) in the isothermal-isobaric ensemble (NPT, P=1 atm, T=300 K); temperature control was achieved using a Nosé-Hoover thermostat [Nosé 1984, Hoover 1985] and SPC model ([Berendsen et al. 1981]) was used for water. The ESP charges of deprotonated tyrosine, not available in the Amber94 FF, were calculated at the HF/631G\* level for the acetyl tyrosinamide after geometry optimization, following the Amber charge fitting philosophy. An rRESPA multiple timestep algorithm with a potential subdivision specifically tuned for proteins (36) was used for integrating the equations of motion, using a time step equal to 10 fs. The total simulation time was about one hundred nanoseconds

for each of the twelve simulations of the four isomers in the three different protonation states. Each sample was equilibrated for at least 30 ns to avoid the dependence on the starting configuration and each trajectory was analyzed after the equilibration, checking that both volume and total energy fluctuated around an average constant value, without systematic drifts. The trajectory analysis is performed on rather long, 73 ns production runs, with configurations stored every 5 ps. To assess if the force field was suitable to predict *PPII* conformation, we also carried out a 30 ns simulation of 9mer polyL proline in the zwitterion form in the isothermal isobaric ensemble (NPT, P=1 atm, T=300 K) using a cubic box with 2123 explicit water molecules. The *PPII* structure ( $\alpha = 75^\circ$ ,  $\beta = 145^\circ$ ) ([Adzhubei and Stemnberg 1993]) was retained for the whole simulation time, with the exception of the C-terminal proline, whose average angles resulted in equilibrium between 30 and 150 values.

## 9.3 Results and Discussion

### 9.3.1 NMR Analysis and Solution Structure Determination

The solution conformations of the 1 HexaPY hexapeptide were studied by proton NMR spectroscopy in  $H_2O/D_2O$  90/10 and in TFE/ $H_2O$  (80/20 v/v) to support the study in water, at pH = 4.2. Identification of the complete spin systems in both solvents was accomplished by homonuclear Jcorrelated 2D techniques such as TOCSY and DQFCOSY. These techniques readily allowed for the identification of the spin systems but were not sufficient to determine their position in the sequence. To unambiguously assign all the amino acids and their position in the peptide chain, sequential NOE connectivities between backbone protons were utilized. Complete chemical shift assignments in  $H_2O/D_2O$  and TFE/ $H_2O$  (80/20 v/v) are reported in Table 9.1 and 9.2 respectively.

$^3J_{NH-CH}$  and  $^3J_{CH-CH}$  coupling constants and temperature coefficients are reported in Tab. 9.3. A peculiar characteristic of the 1 - *HexaPY* peptide is the presence of two proline residues in the sequence; the first one is located at the beginning of the chain, the other proline occupies a central

Table 9.1: Proton chemical shifts (ppm) for the major conformer [a] and the minor conformer [b] of 1 – *HexaPY* at 300 K in  $H_2O/D_2O$  (90/10, v/v). The values for the Acetyl and CONH2 groups are 2.10 ppm and 7.55/7.05 ppm [a], 2.10 ppm and 7.55/7.05 ppm [b].

	AA	NH	$\alpha CH$	$\beta CH$	$\gamma CH$	Others
[a]	<i>Pro</i> <sup>1</sup>	/	4.31	2.20/1.91	1.79	$\delta CH_2$ 3.61
	<i>His</i> <sup>2</sup>	8.58	4.71	3.24/3.13	/	<i>H</i> (2) 8.59; <i>H</i> (4) 7.27
	<i>Asn</i> <sup>3</sup>	8.47	4.95	2.82/2.66	/	$\delta NH_2$ 7.60/6.93
	<i>Pro</i> <sup>4</sup>	/	4.41	2.25/2.02	1.95	$\delta CH_2$ 3.85/3.73
	<i>Gly</i> <sup>5</sup>	8.41	3.91/3.84	/	/	/
	<i>Tyr</i> <sup>6</sup>	7.91	4.52	3.05/2.95	/	<i>H</i> (2,6) 7.13 ; <i>H</i> (3,5) 6.84
[b]	AA	NH	$\alpha CH$	$\beta CH$	$\gamma CH$	Others
	<i>Pro</i> <sup>1</sup>	/	4.47	2.34/1.92	1.72	$\delta CH_2$ 3.45
	<i>His</i> <sup>2</sup>	8.76	4.70	3.20/3.14	/	<i>H</i> (2) 8.61; <i>H</i> (4) 7.29
	<i>Asn</i> <sup>3</sup>	8.63	4.77	2.82/2.65	/	$\delta NH_2$ 7.60/6.95
	<i>Pro</i> <sup>4</sup>	/	4.20	2.15/1.85	1.65	$\delta CH_2$ 3.55
	<i>Gly</i> <sup>5</sup>	8.49	3.90	/	/	/
	<i>Tyr</i> <sup>6</sup>	8.00	4.47	3.08/2.95	/	<i>H</i> (2,6) 7.13 ; <i>H</i> (3,5) 6.80

position in the  $AcPro^1His^2Asn^3Pro^4Gly^5Tyr^6NH_2$  sequence. As a consequence,  $Pro^4$  is the major responsible of the possible occurrence of more conformers in solution. The hexapeptide shows a well resolved 1D 1H spectrum in TFE/ $H_2O$ , characterized by a good dispersion of the proton resonances and, surprisingly, a single predominant set of sharp peaks for the amide protons. All these features are indicative of the existence of a main conformer, with the minor one populated less than 5%. These percentages change in  $H_2O/D_2O$  where two observable sets of resonances, presumably due to the *cis-trans* isomerization of the  $XxxPro$  bond, are clearly present in the proton spectrum in  $H_2O/D_2O$ ; the relative populations of the two conformerspeaks resulted equal to 83% and 17%. The hexapeptide assumes a *trans* conformation for the AcPro and AsnPro peptide bonds in TFE/ $H_2O$ , as indicated by the strong NOEs between H2,3 of  $Pro^1$  and Ac, and between H3 of  $Pro^4$  and H of  $Asn^3$ ; in  $H_2O/D_2O$ , both the  $XxxPro$  bonds of the major conformer also assume a *trans* conformation as confirmed by the presence of two NOE contacts between the H of  $Pro^1$  and the Acgroup, and between

Table 9.2: Proton chemical shifts (ppm) for 1 – *HexaPY* at 300 K in TFE/ $H_2O$  (80/20, v/v). The values for the Acetyl and CONH2 groups are 2.07 ppm and 7.14/6.54 ppm, respectively.

AA	NH	$\alpha CH$	$\beta CH$	$\gamma CH$	Others
<i>Pro</i> <sup>1</sup>	/	4.30	2.18/1.88	1.96	$\delta CH_2$ 3.60/3.52
<i>His</i> <sup>2</sup>	7.93	4.68	3.29/3.13	/	<i>H</i> (2) 8.43; <i>H</i> (4) 7.21
<i>Asn</i> <sup>3</sup>	8.16	4.92	2.81/2.68	/	$\delta NH_2$ 7.30/6.48
<i>Pro</i> <sup>4</sup>	/	4.41	2.25/1.95	2.00	$\delta CH_2$ 3.73
<i>Gly</i> <sup>5</sup>	8.19	3.82/3.79	/	/	/
<i>Tyr</i> <sup>6</sup>	7.57	4.54	3.08/2.95	/	<i>H</i> (2,6) 7.09 ; <i>H</i> (3,5) 6.81

the H of *Pro*<sup>4</sup> and H of *Asn*<sup>3</sup>. The presence of a less populated conformer is presumably due to a *cis* arrangement of the *Asn*<sup>3</sup>*Pro*<sup>4</sup> peptide bond since an *Ac* – *Pro*<sup>1</sup> NOE contact is not observed in the ROESY spectrum and, moreover, the H resonance of *Asn*<sup>3</sup> is obscured by the water signal. All the amide protons present a linear and negative dependence of the amide chemical shifts on the temperature. In particular, the backbone amide protons of residues *His*<sup>2</sup> and *Tyr*<sup>6</sup> in TFE/ $H_2O$  and *Tyr*<sup>6</sup> in  $H_2O/D_2O$  show the lowest temperature gradients in both solvent systems, where the scalar coupling constants,  ${}^3J_{NH-CH}$ , for residues *His*<sup>2</sup>, *Asn*<sup>3</sup> and *Tyr*<sup>6</sup> show values in the range 68 Hz (Tab. 9.3). On the basis of  ${}^3J_{CH-CH}$  coupling constants and *CH* – *CH*, *NH* – *CH* NOEs, the favored conformations of the side chains were identified ([Wagner et al. 1987]): a *trans* arrangement resulted for all the amino acids except for *Asn*<sup>3</sup> which appears to prefer a *gauche* conformation in TFE/ $H_2O$ . These data were confirmed by an analysis of rotamer populations ([Jardetzky and Roberts 1981]) reported in Tab.s 9.1[a] and ??[b]. After a careful analysis of the ROESY spectra, 50 protonproton NOE cross peaks were assigned and integrated for 1HexaPY in TFE/ $H_2O$  (80/20 v/v), and 37 in  $H_2O/D_2O$ , for the major conformer. A list of the relevant distances derived from the integration of the ROESY peaks is reported in Tab.s 9.2 and 9.2; it should be noted that in both solvents a correlation between the *NH* of *Gly*<sup>5</sup> and the *NH* of *Tyr*<sup>6</sup>, corresponding to a calculated distance of about 3 Å, is clearly observed. Furthermore, only in TFE/ $H_2O$  a mediumlong range NOE value could be assigned as a weak correlation between the *Pro*<sup>4</sup> *CH* and the *Tyr*<sup>6</sup> amide proton. Four 3JNHH and ten 3JHH coupling constants were extracted from the 1H monodimensional spectra of

Table 9.3: Temperature coefficients (ppb/K) and  $^3J$  coupling constants (Hz) for 1-*HexaPY* in TFE/ $H_2O$  (80/20, v/v) and in  $H_2O/D_2O$  (90/10, v/v). In  $H_2O/D_2O$  data are reported only for the major conformer.

AA	TFE/ $H_2O$ (80/20)				$H_2O/D_2O$ (90/10)			
	$\Delta\delta/\Delta t$	$^3J_{NH_\alpha CH}$	$^3J_{\alpha NH_\beta CH}$	$^3J_{\alpha NH_\beta' CH}$	$\Delta\delta/\Delta t$	$^3J_{NH_\alpha CH}$	$^3J_{\alpha NH_\beta CH}$	$^3J_{\alpha NH_\beta' CH}$
<i>Pro</i>	-	-	5.3	8.5	-	-	5.1	8.5
<i>His</i>	-4.9	7.8	5.5	7.5	-5.6	7.8	6.1	8.6
<i>Asn</i>	-6.7	6.8	7.4	6.3	-6.5	7.0	6.3	8.1
<i>AsnNH<sub>2</sub>γ</i>	-7.0/6.2	-	-	-	-4.6/-4.6	-	-	-
<i>Pro</i>	-	-	5.1	8.5	-	-	5.5	8.5
<i>Gly</i>	-5.7	5.8/5.8	-	-	-6.3	6.0	-	-
<i>Tyr</i>	-4.5	7.5	6.3	8.0	-4.5	7.3	7.2	8.4
<i>CONH<sub>2</sub></i>	-/-6.2	-	-	-	-4.6/-3.7	-	-	-

1 – *HexaPY* either in TFE/ $H_2O$  (80/20 v/v) and  $H_2O/D_2O$  (Tab. 9.3). Stereospecific assignments for *His*<sup>2</sup> ( $CH_2$  protons) and *Asn*<sup>3</sup> ( $CH_2$  and  $CH_2$  protons) of 1 – *HexaPY* in TFE/ $H_2O$  (80/20 v/v) were derived from the input data, using the GRIDSEARCH module of CYANA software. The final input files for the structure calculation contained 19 meaningful distance constraints (15 intraresidue and 4 shortrange) and 20 angle constraints for the 1 HexaPY in  $H_2O/D_2O$ , and concerning the study in TFE/ $H_2O$  (80/20 v/v), 27 meaningful distance constraints (15 intraresidue, 10 short and 2 mediumrange) and 22 angle constraints. These constraints were then used to generate a total of 100 structures and among them the 20 structures with the lowest target function values were selected and energy minimized. The backbone superpositions of the best 20 CYANA conformers of 1 – *HexaPY* in  $H_2O/D_2O$  (target function 0.77) and TFE/ $H_2O$  (80/20 v/v) (target function 1.05) minimized in vacuo and in a solvent shell are reported in Fig. 9.1 and 9.2 respectively, a stick drawing of the structures are reported in Fig. 9.1 of the Supplemental Materials. The NMR structure of 1 – *HexaPY* in  $H_2O/D_2O$  is reasonably defined and elements of *PPII* conformation are present, which include the *Asn*<sup>3</sup> and *Pro*<sup>4</sup> residues; the presence of *Gly*<sup>5</sup> provokes the break of this secondary structure and allows *Tyr*<sup>6</sup> to experience a high conformational freedom.

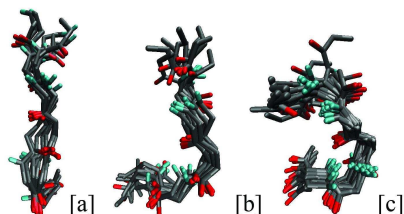


Figure 9.1: Superposition of the 20 energy conformers of 1 – *HexaPY* in  $H_2O$  (a) and TFE/ $H_2O$  (80/20) (b) minimized in vacuo and aligned according with the minimal root mean square deviation of the backbone atoms residues. Only the polypeptide backbone is shown.

In TFE/ $H_2O$  (80/20 v/v), the NMR structure of 1 – *HexaPY* is better defined, with an increase of the tendency of polyproline *II* secondary structure which involves residues *Pro*<sup>1</sup>, *His*<sup>2</sup>, *Asn*<sup>3</sup> and *Pro*<sup>4</sup>.

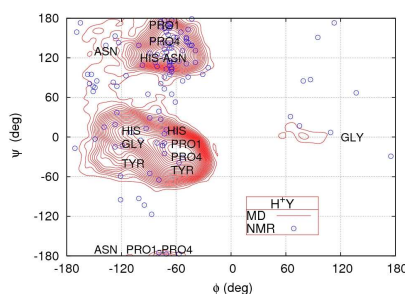


Figure 9.2: Superposition of the 20 energy conformers of 1 – *HexaPY* in H<sub>2</sub>O (a) and in TFE/H<sub>2</sub>O (80/20) (b) minimized in solvent and aligned according with the minimal root mean square deviation of the backbone atoms residues. Only the polypeptide backbone is shown.

### 9.3.2 Molecular Dynamics simulations

The NMR study above discussed gives a quite clear picture on the averaged conformation adopted by the 1 – *HexaPY* peptide chain. However, such a study has been limited to acidic pH values and it is not of much help for a complete interpretation of the CD spectra, previously reported by some of us ([Mendola et al. 2005]) and concerning acid, neutral and basic pH values. Consequently, in order to further investigate the pH dependence of 1 – *HexaPY* conformational states in aqueous solution, a fully atomistic Molecular Dynamics (MD) study was carried out in water. We noticed first that, due to its partial double bond character, the torsional barrier around the peptide bond *XxxPro* is not thermally overcome ([Cheng and Bovey 1977, Delaney and Madison 1982]) in the nanosecond time scale, typical of MD simulations. Consequently, the different isomers, originated from the AsnPro and AcPro torsion angles (N and C-termini were capped with acetyl and amide groups, respectively), were simulated separately. For each isomer, three different protonation states were studied to mimic the different pH conditions. Taking into consideration the pKa of the single amino acids, by assuming acid pH in the simulation, the histidine is protonated and tyrosine is in its neutral form (labelled  $H^+Y$ ), at neutral pH both histidine, in the form, and tyrosine are in their neutral states (labelled  $HY$ ) and finally at basic pH histidine is in the neutral form and tyrosine is deprotonated (labelled  $HY^-$ ). One chloride ion and one sodium ion were added at acid and basic pH conditions respectively, to ensure charge neutrality in the simulation box.

Table 9.4: Average dihedral angles obtained from the 20 NMR structures minimized in vacuo for 1 – *HexaPY* in  $H_2O/D_2O$  (90/10, v/v) [a] and in TFE/ $H_2O$  (80/20, v/v) [b].

	AA	$\Phi$	$\Psi$		AA	$\Phi$	$\Psi$
	<i>Pro</i> <sup>1</sup>	-75 ± 1	80 ± 43		<i>Pro</i> <sup>1</sup>	-75 ± 1	-177 ± 1
	<i>His</i> <sup>2</sup>	-92 ± 78	98 ± 17		<i>His</i> <sup>2</sup>	-54 ± 1	173 ± 1
[a]	<i>Asn</i> <sup>3</sup>	-73 ± 7	142 ± 4	[b]	<i>Asn</i> <sup>3</sup>	-52 ± 1	125 ± 1
	<i>Pro</i> <sup>4</sup>	-73 ± 5	124 ± 30		<i>Pro</i> <sup>4</sup>	-75 ± 1	126 ± 30
	<i>Gly</i> <sup>5</sup>	7 ± 88	0 ± 78		<i>Gly</i> <sup>5</sup>	79 ± 87	-1.5 ± 15
	<i>Tyr</i> <sup>6</sup>	-93 ± 26	94 ± 50		<i>Tyr</i> <sup>6</sup>	-100 ± 28	65 ± 67

### 9.3.3 Ramachandran analysis

The Ramachandran maps of all the isomeric forms, calculated from the simulations, present strong variations as a function of pH, the most striking feature being the presence of distinct regions of local minima, (see Tab.s 9.4 and 9.6). The trans conformation of proline is known to stabilize two kinds of typical dihedral regions, namely type *I* turn ( $i+1=60, i+1=30, i+2=90, i+2=0$ ) ([Hutchinson and Thornton 1994]) and polyproline II ( $=75, =145$ ) ([Adzhubei and Sternberg 1993]). On the contrary, the cis conformation of proline, stabilizes a special type of turn, called VI ([Che and Marshall 2006, Keller et al. 1998]). Inside this family of turns, there are three subgroups, namely VIa1 ( $i+1=60, i+1=120, i+2=90, i+2=0$ ), VIa2 ( $i+1=120, i+1=120, i+2=60, i+2=0$ ), VIb ( $i+1=135, i+1=135, i+2=75, i+2=160$ ) ([C.Chou 2000]). To correctly assign the different conformations, the coexistence of more amino acid residues with the typical and dihedral angles of a secondary structure is required and this cannot be inferred simply from the Ramachandran map. In this context, to estimate the percentage of the different conformers during the simulation, we calculated the percentage of type *I* turn and *PPII* considering both 2, 3, 4 consecutive residues with the typical type *I* turn and *PPII* dihedrals; for the *cis* and the *trans-cis* isomers we considered also the type VIb turn. For all the and dihedral angles we considered the standard range of  $\pm 30$ . During the MD simulations, no more than two consecutive residues with typical type *I* turn and dihedrals were found. This happens because the asparagine prefers to adopt *PPII* dihedrals, thus acting as a point breaker, preventing the formation of four residues turns as in more rigid backbones. Therefore, considering the short



sequence of the hexarepeat, we calculated also the percentage related to two consecutive amino acids which possess at the same time type *I* turn dihedrals. From this analysis, type *I* turn results the principal conformer in the ProGly sequence of the *trans-trans* isomer when the tyrosine is not deprotonated,  $\text{pH} < 10$ . This kind of structure is also present in the ProHis and GlyTyr sequences, although it is not stable during the simulation.

Furthermore, the percentage of such a conformation decreases as the pH increases (Fig. 9.3), giving raise to unordered conformations, which we identified by the change of angles minima of glycine respectively from 97.1, in the  $H^+Y$  and in the  $HY$  form, to 137.15 in the  $HY^-$  form, (see Tab.s 9.4 and 9.6), thus destabilizing the type *I* turn structure. This result explains

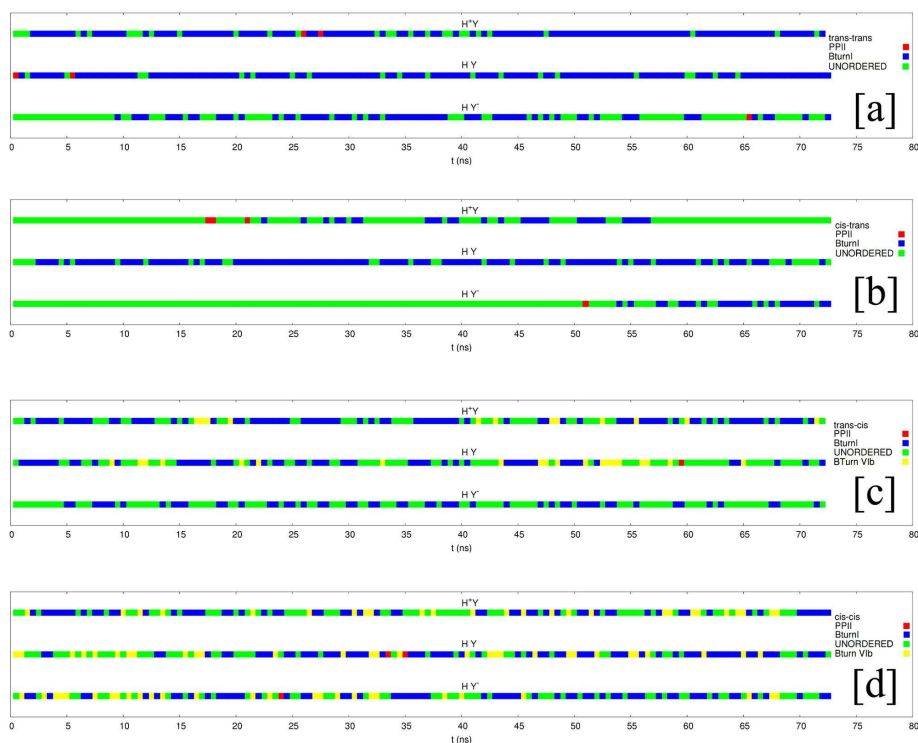


Figure 9.3: Time evolution of 1 – *HexaPY* conformers for each of the four isomers (*trans-trans* [a], *cis-trans* [b], *trans-cis* [c], *ciscis* [d]) in the three different protonation states obtained from MD simulations. The calculation is based on at least two consecutive residues in the same conformation concerning type *I* turn and three consecutive residues for *PPII*.

Table 9.5: Average dihedral angles obtained from the 20 NMR structures minimized in solvent for 1 – *HexaPY* in water [a] and in TFE/H<sub>2</sub>O 80/20 [b]

	AA	$\Phi$	$\Psi$		AA	$\Phi$	$\Psi$
	<i>Pro</i> <sup>1</sup>	-71 ± 10	103 ± 60		<i>Pro</i> <sup>1</sup>	-69 ± 7	155 ± 42
	<i>His</i> <sup>2</sup>	-90 ± 45	89 ± 28		<i>His</i> <sup>2</sup>	-87 ± 28	148 ± 33
[a]	<i>Asn</i> <sup>3</sup>	-65 ± 21	142 ± 9	[b]	<i>Asn</i> <sup>3</sup>	-39 ± 51	117 ± 18
	<i>Pro</i> <sup>4</sup>	-68 ± 7	138 ± 45		<i>Pro</i> <sup>4</sup>	-72 ± 8	135 ± 40
	<i>Gly</i> <sup>5</sup>	-173 ± 70	1 ± 56		<i>Gly</i> <sup>5</sup>	62 ± 52	20 ± 30
	<i>Tyr</i> <sup>6</sup>	-111 ± 40	67 ± 64		<i>Tyr</i> <sup>6</sup>	-124 ± 33	110 ± 66

the CD spectra blue shift for the minimum around 205 nm at pH 10 and 298 K previously reported ([Mendola et al. 2005]). A type VIIb turn is present in the AsnPro residues, but only in the *trans-cis* and in the *ciscis* isomers; such a turn is destabilized by an increasing of pH, disappearing at basic pH in the *trans-cis* isomer (Fig. 9.3). Polyproline II is not very frequent during the (un)folding, thus being a local conformation, as found for polyalanine ([Makowska et al. 2006]), and concerning 1 – *HexaPY*, interconverts to turn very frequently. The region in which *PPII* conformation seems to be more likely for 1 HexaPY is the ProHisAsn, especially of the *cis-trans* isomer, and less frequently we found it also in the ProHisAsnPro sequence (only for the *trans-trans* isomer in the H+Y form), when the imidazole ring of the histidine is protonated, recently proposed to have an effect on the conformational transition of prion protein ([Langella et al. 2004]). As confirmed by NMR investigations in water (Tab.s 9.4 and 9.5 and Fig. 9.4), the polyproline region is not related to the full length peptide.

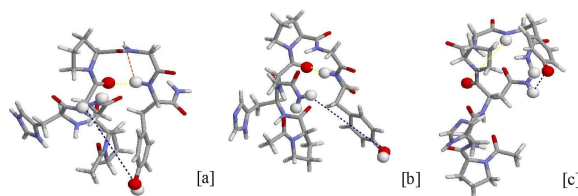


Figure 9.4: Superposition of the 20 NMR structures minimized in water (a) and the MD (b, c) distributions of the 1 – *HexaPY* *trans-trans* isomer.

On the contrary, *PPII* dihedrals are found in the AsnPro region or,

as underlined from MD simulations, only for single amino acids, such as asparagine, the two proline residues and in lower amount the histidine. As it can be seen from Fig. 9.4, our NMR and MD Ramachandran maps show a nice agreement thus suggesting a good capability of MD simulations to predict 1 – *HexaPY* conformational behavior.

### 9.3.4 Intrapeptide interactions

The presence of hydrogen bonds between carbonyl oxygens and amide hydrogens belonging to different amino acids was estimated following the DSSP criteria ([Kabsch and Sander 1983]). To investigate their persistence we also calculated the variation of a given OH distance as a function of time. Furthermore, we monitored the angle between the  $C = O$  and  $NH$  vectors ( $\theta$ ), in order to understand the orientation of the two atoms involved in the hydrogen bond and to calculate the angle/OH distance/energy surface so to decide if a hydrogen bond was present. The hydrogen bond energy used in this work is purely electrostatic and it is expressed as follows ([Kabsch and Sander 1983]):

$$E_{HB} = 332 \frac{q_1 q_2}{r_{CN} r_{CH} r_{OH} r_{NO}} \quad (9.3.1)$$

where  $q_1$  and  $q_2$  are respectively  $0.42e$  and  $0.20e$ , being  $e$  the unit electron charge, 332 is the conversion factor from  $e^2/$  to kcal/mol and  $r_{CN}$ ,  $r_{CH}$ ,  $r_{OH}$  and  $r_{NO}$  are the interresidue distances expressed in Å. A hydrogen bond is considered to be present when the DSSP conditions are reached, namely  $E_{HB}$  equal or less than 0.5 kcal/mol, equal or less than 60 and  $r_{OH}$  equal or less than 3 Å. We only found  $i - i + 3$  hydrogen bonds involving the asparagine residue, and exclusively when the *XXXPro* torsion angle is in *trans* conformation (Tab. 9.6). The *cis* proline, instead, destabilizes a type *I* turn in favour of type *VI* turn conformation, which, in fact, does not form intra-chain hydrogen bonds. Another striking feature is the interaction between the phenolate oxygen of tyrosine and the side chain amide hydrogen of asparagine, found by looking at the variation in time of the distance reported in Fig. 9.5.

This weak bond seems possible only when tyrosine is deprotonated and mainly for the *trans-trans* isomer. This interaction determines a tilt of the glycine residue, breaking the turn structure (Fig. 9.6) and giving rise to an unordered structure; this results indicates the reason of the conforma-

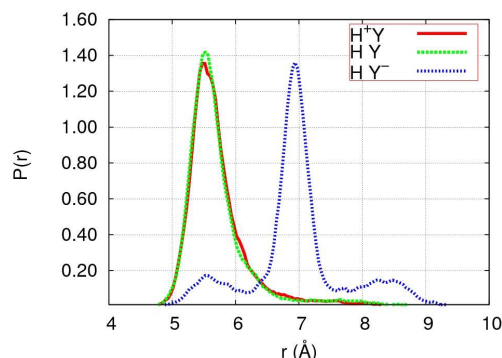


Figure 9.5: Time evolution of the distance between the phenolate oxygen of tyrosine and the side chain amide hydrogens of asparagine of 1 – *HexaPY trans-trans* in the three different protonation states.

tional change of glycine, described in the Ramachandran analysis, and the blue shift observed in the CD spectra recorded at basic pH ([Mendola et al. 2005]). A useful indicator of the conformational behaviour of a peptide and of the presence of intrachain interactions is offered by the  $C_iC_{i+3}$  distances, that in this particular case allow the observation of some intriguing features related both to the *cis/trans* proline conformation and to the position of asparagine residues. In the first two regions, (*PHNP* and *HNPG*), the  $C_iC_{i+3}$  distances, due to the break point of asparagine, are quite large, while in the *NPGY* *C*-terminal sequence the distances are found to be smaller (Fig. 9.7), underlining the existence of a turn conformation, whose distance, as a general criterion, has to be less than 7 Å ([Richardson 1981, Rose et al. 1985]). The presence of the turn conformation in the *NPGY* region is confirmed by the hydrogen bond screening, which shows an  $i - i + 3$  hydrogen bond involving asparagine and tyrosine residues. Another turn region was detected, despite in lower extent, in the *AcPHN* *N*-terminal sequence (see Tab. 9.6) where asparagine is not an intrachain residue, as in the *NPGY* sequence. Although the asparagine and dihedrals are typical of polyproline *II* conformation, the spatial features of turn conformation are preserved, giving rise to two bent regions. The probability of turn conformation in this region is very high when tyrosine is not deprotonated and the fourth proline is in *trans* conformation, mainly in the *trans-trans* isomer and in lower amount, in the *cis-trans* isomer at neutral and acid pH conditions. In the *trans-trans* isomer, at basic conditions, the interaction between the phenolate and the

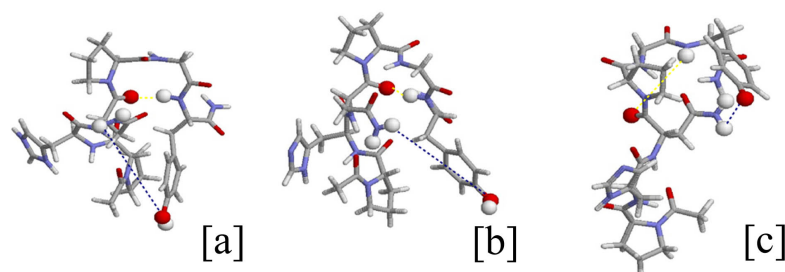


Figure 9.6: 1 – *HexaPY* *trans–trans* typical conformations in the  $H^+Y$  [a],  $HY$  [b] and  $HY^-$  [c] form obtained from MD simulations. It is worth to note the short distance between the phenolate oxygen of tyrosine and the amide side chain hydrogen of asparagine residue (blue line) only in the  $HY^-$  form, where tyrosine is deprotonated. In the  $H^+Y$  and  $HY$  forms, instead, a turn structure is stabilized by an intrachain  $ASN_{C=O}TYR_{NH}$   $H$ -bond (yellow line).

side chain hydrogens of asparagine, causing the tilt of glycine, breaks the turn structure, thus increasing the distance in the  $NPGY$  region (Fig. 9.7), and consequently the hydrogen bond percentage decreases abruptly. The presence of two such bent regions were found also in the NMR structures at acid pH minimized in water, as can be noticed from Fig. 9.8, thus confirming the Molecular Dynamics simulation results and the relative rigidity of the hexarepeat backbone.

## 9.4 Conclusions

The solution structure determination of flexible proteins, including the  $N$ -terminal region of prion protein, is generally hard because of the wide regions of conformational space they are allowed to explore. In this context, the combined use of Molecular Dynamics simulations supported by NMR investigations represents a powerful tool for the understanding of the conformational states of peptides and proteins in solution. Here, we have examined the Avian Prion Hexarepeat (1 – *HexaPY*) and we have found that besides the good agreement between NMR and MD Ramachandran maps, the results of our Molecular Dynamics conformational studies are consistent with NMR structure determination. These indicated, for the *trans–trans* iso-

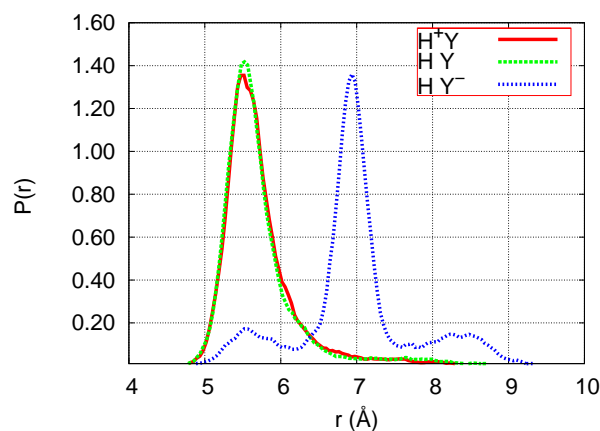


Figure 9.7:  $C_i C_{i+3}$  distance distribution concerning the *NPGY* region of 1 – *HexaPY trans* trans in the three different protonation states. It can be noticed the short distance in the  $H^+Y$  and  $HY$  forms and the larger distance distribution in the  $HY^-$  form.

mer in  $H_2O$ , an averaged conformation with element of *PPII* in the  $Asn^3$  and  $Pro^4$  residues, while a more defined propensity toward a bent conformation was observed in the presence of TFE. Interestingly, the presence of a given secondary structure inside a small region of the peptide is not related to a more extended sequence. The *AsnPro* region or also individual residues, such asparagine, spend more time with polyproline dihedrals, but the local conformation does not propagate along the backbone, as it was recently reported for the polyalanine peptide ([Zagrovic et al. 2005]). Notably, MD simulations were successful in predicting the hydrogen bonds distributions, hard to understand simply from NMR investigations. In particular, we found an  $i - i + 3$  turn structure inside the AcPHN region and particularly in the *NPGY* sequence at physiological pH, although the asparagine does not possess type *I* turn angles, but angles typical of polyproline *II* structure. Following general practice, minimization was normally carried out in vacuo. However, consistently with suggestions from a recent work ([Calzolari et al. 2005]) we also obtained the structure minimized in a solvent shell. Following this approach, we have found that the conformational features of the NMR structures are consistent with Molecular Dynamics results allowing further energetically favoured bent conformations in water. Moreover, MD calculations were performed for all four isomers in water at acid, neutral and basic

Table 9.6: pH dependent MD hydrogen bond percentages for 1 – *HexaPY* isomers found between  $i$  and  $i+3$  residues.

$H$ -bond $i - i + 3$	em trans-trans	em cis-trans	em trans-cis	em cis-cis
$H^+Y$				
$ACE_{C=O}-ASN_{NH}$	76%	0%	57%	0%
$ASN_{C=O}-TYR_{NH}$	68%	21%	0%	0%
$HY$				
$ACE_{C=O}-ASN_{NH}$	84%	0%	38%	0%
$ASN_{C=O}-TYR_{NH}$	70%	64%	0%	0%
$HY^-$				
$ACE_{C=O}-ASN_{NH}$	49%	0%	22%	0%
$ASN_{C=O}-TYR_{NH}$	8%	14%	0%	0%

pH conditions, extending the study to conditions not amenable using only NMR investigations. An interesting dependence of the peptide shape on the tyrosine residue deprotonation was also detected, mainly in the *trans-trans* isomer. This is the driving force for the unordered structure, which causes a shortening of the distance between the phenolate oxygen of tyrosine and the amide side chain hydrogens of asparagine thus tilting the glycine residue. The regions of type *I* turn (destabilized at basic pH) are in equilibrium with unordered dihedrals conformations, which instead increase at basic pH and, concerning the *PHNP* sequence, a very little amount of polyproline *II* conformer, probably because the *PXXP* region is not an intrachain motif, thus destabilizing a full polyproline *II* structure. The presence of two such turns inside the hexarepeat fragment at physiological pH gives a better defined structure with respect to the *PHGGWGQ* mammal repeats. It was reported that in the octarepeat fragment there is the presence of unordered conformations ([Gustiananda et al. 2004]) and turn like structure in the *GWGQ* *C*-terminal region at pH 6.2 ([Zahn. 2003]), like the glycine *C*-terminal region of 1 – *HexaPY*. The larger number of glycine residues in the octarepeat region probably increases the unordered conformations in the mammal homologue because of the introduction of more degrees of freedom. On the other hand, the presence of one proline next to the asparagine in the hexarepeat sequence contributes to the turn formation. This can give a more defined structure to avian hexarepeats, which could be linked to the protease resistant domain of the *N*-terminal avian prion protein region. Moreover, secondary structural algorithms predicted that hexarepeats could form a se-

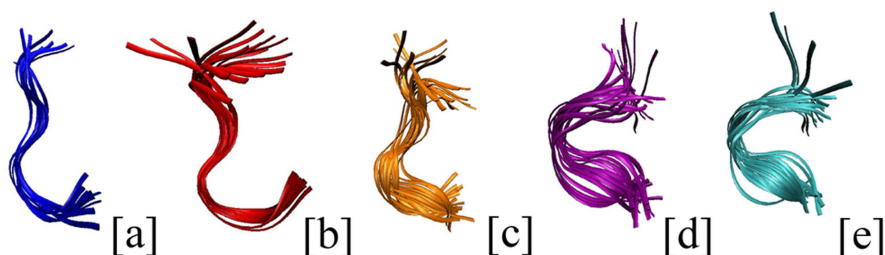


Figure 9.8: From left to right: 1 – *HexaPY* principal conformers obtained from NMR structures minimized in a water shell at acid pH [a], 1 – *HexaPY* *trans–trans* [b], *cis–trans* [c], *trans–cis* [d] and *ciscis* [e] isomers obtained from MD simulations in the H+Y form sampled every 2.5 ns. 1 – *HexaPY* *trans–trans* MD isomer shows the turn conformer, underlined also in the NMR structures [a].

ries of turns and this could play a crucial role in the endocytosis processes ([Bazan et al. 1987]). Our results confirm this hypothesis giving a strong evidence of this turn conformation adopted by the *PHNPGY* avian hexarepeat sequence, especially in the *NPGY* region, which is also the same conformation adopted by the *NPXY* internalization signal of the LDL receptor ([Bansal and Gierasch 1991]). Furthermore, the tyrosinecontaining motifs, essential for coated pitmediated endocytosis, have been found to adopt an  $i - i + 3$  turn conformation especially when present in the last position of the hexarepeat ([Paccaud et al. 1993]). These results reinforce the idea of a possible function of endocytosis carried out by the *N*-terminal repeat region of the prion protein, which seems to be less flexible and thus probably having a more resistant domain than that of the mammal analogue. Finally we can observe that a combined experimental NMR and theoretical MD approach has provided us with a deeper understanding of the chemical and physical processes leading to the actual conformational distribution in solution that would be difficult, or perhaps impossible, to reach using only a laboratory or simulation approach.



# List of Tables

3.1	Energy conservation ratio R for various integrators (see text). The last three entries refer to a velocity Verlet with bond constraints. $\langle V_i \rangle$ and $\langle V_m \rangle$ are the average value of the intramolecular and intermolecular energies (in KJ/mole), respectively. CPU is given in seconds per picoseconds of simulation and $\Delta t$ in fs. Single time step velocity Verlet with $\Delta t = 4.5$ fs is unstable. . . . .	40
3.2	Potential breakup and relative time steps for complex systems with interactions modelled by the AMBER[Cornell et al. 1995] force field and electrostatic computed using the SPME method	75
3.3	Properties of liquid water computed from a 200 ps simulation at 300 K and 0.1 MPa on a sample of 343 molecules in PBC with accurate Ewald ( $\alpha = 0.35 \text{ \AA}^{-1}$ , $k_c = 2.8 \text{ \AA}^{-1}$ ) and no correction Eq. 3.3.173 (column EWALD), with inaccurate Ewald ( $\alpha = 0.35 \text{ \AA}^{-1}$ , $k_c = 0.9 \text{ \AA}^{-1}$ ) but including the correction Eq. 3.3.173 and with no Ewald and cutoff at 10.0 $\text{\AA}$ . $R_{0-0}$ is the distance corresponding to the first peak in the Oxygen–Oxygen pair distribution function. . . . .	80
4.1	The physical properties of the compounds studied: experimental [Lide 2004–2005] densities ( $\text{g cm}^{-3}$ ), viscosities (cP), boiling temperatures (K), and dielectric constants; and calculated densities ( $\text{g cm}^{-3}$ ), and molecular dipoles (D). . . . .	89

4.2	The <i>trans-cis</i> photoisomerization quantum yields $\Phi$ computed from the 500 MD virtual experiments in vacuum and <i>n</i> -hexane using different values of $\chi$ . The largest standard deviation was $\sigma = 0.02$ . The experimental value in <i>n</i> -hexane for an excitation at $\lambda = 439$ nm is $\Phi = 0.20 - 0.25$ [Bortolus and Monti 1979, Rau 2003]. . . . .	91
4.3	The decay times $\tau_1$ , and $\tau_2$ (in ps) for AB in vacuum and <i>n</i> -hexane computed from the 500 MD virtual photoisomerization experiments with different values of $\chi$ . The experimental fluorescence decay times in <i>n</i> -hexane are $\tau_1 = 0.24 - 0.6$ ps, and $\tau_2 = 1.7 - 2.6$ ps [Chang et al. 2004, Lednev et al. 1998]. . . . .	92
4.4	The average decay times $\tau_1$ , and $\tau_2$ (in ps), and the AB <i>trans-cis</i> photoisomerization quantum yields $\Phi$ computed from the 1000 MD virtual photoisomerization experiments in the various solvents at $T = 320$ K, $P = 1$ bar, and assuming $\chi = 3\chi_0$ . . . . .	93
4.5	The average permanence times $\langle t_{S_1} \rangle$ (in ps) of the AB molecule in the $S_1$ state for successful ( <i>cis</i> ) and unsuccessful ( <i>trans</i> ) MD virtual photoisomerization experiments ( $N = 1000$ ) in vacuum and in four different solvents. . . . .	96
4.6	The isomerization pathway indicators $\Theta_m = N_m/N_{trans}$ (top) and $\Theta_m = N_m/N_{cis}$ (bottom) computed from the 1000 MD virtual photoisomerization experiments of AB in vacuum and in four different solvents. . . . .	99
4.7	Geometrical descriptor for $AB_1$ , $AB_1^*$ , $AB_0$ <i>trans</i> and $AB_0$ <i>cis</i> kinetic species. . . . .	103
6.1	The isomerization pathway indicator $\Theta_m$ computed from the 1000 MD virtual photoisomerization experiments of AB in vacuum and in four different solvents. . . . .	132
6.2	The isomerization pathway indicator $\Theta_m$ computed from the 1000 MD virtual photoisomerization experiments of AB in vacuum and in four different solvents. . . . .	134
7.1	Nematic-isotropic phase transition temperatures ( $T_{NI}$ ) and $T_{NI}$ shifts relative to the pure 5CB ( $\Delta T_{NI}$ ) for LC samples doped with various <i>trans</i> and <i>cis</i> <i>p</i> -azobenzene derivatives with a different R substituent at the two concentrations studied ( $x_1$ and $x_2$ ). . . . .	142

7.2	Calculated B3LYP/ <i>AUG-cc-PVTZ</i> total dipole moment in Debye and projection over the long (x), middle (y) and short (z) inertial molecular axis. . . . .	146
7.3	Calculated B3LYP/ <i>AUG-cc-PVTZ</i> traceless quadrupole moments in Debye for the inertial molecular axis. . . . .	147
8.1	† Experimental order parameters for solutes in 5CB at 294 K. * Experimental order parameters for benzene in 5CB at 296 K. Data from [Weaver et al. 1987]. Experimental data for BIF and NIT not available. . . . .	158
8.2	Simulated inertial order parameters of solvent, standard and solutes at 290 and 295 K. . . . .	158
8.3	Simulated inertial order parameters of solvent, standard and solutes at 290 and 295 K for no-charged solutes. . . . .	159
8.4	Simulated dipolar coupling in Hz of 5CB related to C2 C3 atoms, obtained at 290 K and 295 K, to be compared with the experimental value of -1299 Hz obtained for pure 5CB [Stevansson et al. 2001]. . . . .	161
8.5	Simulated dipolar coupling in Hz of H4 H8 atoms for the standard. Experimental data to compare with: -213.23 Hz in 1132; -156.60 Hz in magic mixture; -148.12 in EBBA [der Est et al. 1987]. . . . .	162
8.6	Mean solute dipolar couplings (in Hz) from [der Est et al. 1987] and from simulations . . . . .	163
8.7	$U_{MF}$ parameters fitted from simulation distribution functions with errors . . . . .	163
8.8	Molecular surface ( $\text{\AA}^2$ ) calculated using the method of Ferrarini <i>et al</i> [A.Ferrarini and P.L.Nordio 1998, A.Ferrarini et al. 1992], and fitting of integral in Eq. ?? by relation Eq. 8.6.13 . . . . .	165
8.9	$\xi$ constant value and standard deviation in ( $\text{kcal \AA}^{-2} \text{ mol}^{-1}$ ) of Eq. ?? obtained from fitting of fixed function defined in Eq. 8.6.13 over the tumbling energy simulation data obtained from Boltzman inversion of rotational distribution function. . . . .	165

9.1	Proton chemical shifts (ppm) for the major conformer [a] and the minor conformer [b] of 1- <i>HexaPY</i> at 300 K in $H_2O/D_2O$ (90/10, v/v). The values for the Acetyl and CONH2 groups are 2.10 ppm and 7.55/7.05 ppm [a], 2.10 ppm and 7.55/7.05 ppm [b]. . . . .	175
9.2	Proton chemical shifts (ppm) for 1- <i>HexaPY</i> at 300 K in TFE/ $H_2O$ (80/20, v/v). The values for the Acetyl and CONH2 groups are 2.07 ppm and 7.14/6.54 ppm, respectively. . . . .	176
9.3	Temperature coefficients (ppb/K) and $^3J$ coupling constants (Hz) for 1- <i>HexaPY</i> in TFE/ $H_2O$ (80/20, v/v) and in $H_2O/D_2O$ (90/10, v/v). In $H_2O/D_2O$ data are reported only for the major conformer. . . . .	177
9.4	Average dihedral angles obtained from the 20 NMR structures minimized in vacuo for 1- <i>HexaPY</i> in $H_2O/D_2O$ (90/10, v/v) [a] and in TFE/ $H_2O$ (80/20, v/v) [b]. . . . .	180
9.5	Average dihedral angles obtained from the 20 NMR structures minimized in solvent for 1- <i>HexaPY</i> in water [a] and in TFE/ $H_2O$ 80/20 [b] . . . . .	182
9.6	pH dependent MD hydrogen bond percentages for 1- <i>HexaPY</i> isomers found between i and i+3 residues. . . . .	187
A.1	Atom charges for azobenzene simulation. . . . .	226
A.2	Bonds and bendings in azobenzene simulation . . . . .	227
B.1	Charge parameters used expressed in $e$ units . . . . .	230
B.2	Bond parameters used . . . . .	230
B.3	Bending parameters used . . . . .	231
B.4	Torsional parameters used . . . . .	231
B.5	Lennard-Jones parameters used . . . . .	231
C.1	Formula di struttura e proprietà elettrostatiche del 5CB, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6-31G** . . .	233
C.2	Formula di struttura e proprietà elettrostatiche dell'1,3,5-triclorobenzene, ottenute con il metodo B3LYP. 6-31G** . . . . .	234
C.3	Formula di struttura e proprietà elettrostatiche del 1,3,5-tribromobenzene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6-31G** .	234
C.4	Formula di struttura e proprietà elettrostatiche del 2,4-esadiene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6-31G** .	235

- C.5 Formula di struttura e proprietà elettrostatiche dell'acetonitrile, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G<sup>\*\*</sup>. . . . . 235
- C.6 Formula di struttura e proprietà elettrostatiche dell' acetilene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G<sup>\*\*</sup>. 235
- C.7 Formula di struttura e proprietà elettrostatiche del benzene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G<sup>\*\*</sup>. 236
- C.8 Formula di struttura e proprietà elettrostatiche del bifenile, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G<sup>\*\*</sup>. 237
- C.9 Formula di struttura e proprietà elettrostatiche del propino, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G<sup>\*\*</sup>. 238



# List of Figures

2.1	Application of azobenzene-like molecules. [a] Optical alignment. [b] Molecular pumps. [c] Artificial muscles. [d] Storage disk. [e] Deformable sheet [Yu et al. 2003]. . . . .	8
2.2	Sheme of a twisted Nematic display . . . . .	10
3.1	Example of computer simulation applications. Data from <a href="http://www.science.doe.gov/scidac/">http://www.science.doe.gov/scidac/</a> . [a] Vortices in a superfluid. [b] Protein dynamics. [c] Turbulent methane flame. [d] Clay-mineral geochemistry. [e] Two spheres mixing in a stream. [f] HEP particle beam halo. [g] Transport barrier dynamics. [h] Combustion turbulence modelling. [i] Fusion magnetic field. [l] Perturbation in clear-sky and cloud albedo. [m] Au-Au collision. [n] Crystal structure for $C_{36}$ solid. [o] Lattice quantum chromodynamics. [p] Binary alloy solidification. [q] Perturbed plasma density. [r] DOE Parallel Climate Model. [s] Sea surface temperature. [t] Molecular simulation of complex fluids. [u] Structural biology. [v] Nuclear theory. [z] Waveguide optics. . . . .	15
3.2	Time record of the torsional potential energy at about 300 K for a cluster of eight molecules of $C_{24}H_{50}$ obtained using three integrators: solid line integrator $E$ ; circles integrator $R3$ ; squares integrator $S1$ ; diamonds integrator $S$ (see text) . . . . .	41
3.3	Power spectra of the velocity autocorrelation function (left) and of the torsional internal coordinates (right) at 300 K for a cluster of 8 $C_{24}H_{50}$ molecules calculated with integrators $E$ , $R3$ , $S1$ and $S$ (see text) starting from the same phase space point . . . . .	43
3.4	Bending and dihedral angles . . . . .	64

3.5	Density of the uncoupled (see text) states for stretching, bending, proper and improper torsion obtained with the AMBER force field on the protein bovine pancreatic trypsin inhibitor (BPTI). Frequencies were calculated according to Eqs. 3.3.131,3.3.139,3.3.145.	66
3.6	CPU time versus number of particles for the SPME algorithm as measured on a 43P/160MH IBM workstation . . . . .	73
3.7	The correction potential $\chi(r, k_c, \alpha)$ as a function of the distance for different values of the parameters $\alpha$ (left) and $k_c$ (right). The solid line on the top right corner is the bare Coulomb potential $1/r$ . . . . .	78
3.8	The integral $I(r)$ of Eq. 3.3.176 as a function of the distance for different values of the parameters $\alpha$ (left) and $k_c$ (right) . .	79
4.1	Scheme of the azobenzene (AB) molecule showing the <i>Ph-N=N-Ph</i> torsional angle $\phi$ , and the <i>Ph-N=N</i> bending angle $\theta$ . . . . .	84
4.2	The potential energy surfaces for the $S_0$ and $S_1$ states of AB used in this work [Ishikawa et al. 2001], with superimposed representative MD trajectories in the excited state (black), and after one successful (final <i>cis</i> isomer, red), and one unsuccessful (final <i>trans</i> isomer, blue) isomerization process of AB in <i>n</i> -hexane, at $T = 320$ K and $P = 1$ bar. We also show the snapshots of an AB molecule with the closest solvent neighbors in typical MD configurations of the <i>trans</i> - and <i>cis</i> -isomers, and at the conical intersection. The cyan circle defines the region in $S_1$ with transition probability greater than 1% . . . . .	90
4.3	The evolution of averaged conformational indicators from the 1000 MD virtual photoisomerization experiments: [a] <i>Ph-N=N-Ph</i> torsional angle $ \phi $ ; [b] <i>N=N-C</i> bending angles $\theta$ ; [c] phenyl-phenyl distance $r_{4,4'}$ ; and [d] <i>C-C-N=N</i> torsional angles $ \gamma $ . . .	95
4.4	Examples of three isomerization pathways of AB in <i>n</i> -hexane corresponding to the torsion (blue), mixed (red), and inversion (green) mechanisms. . . . .	98
4.5	Geometrical means parameter fitting for hexane using functional forms in equation 4.3.8 and 4.3.9. . . . .	100



4.6	Summary of the <i>trans-cis</i> AB isomerization process in solution as emerged from the MD virtual experiments of this study: excitation, torsional movement in the $S_1$ excited state (lifetime $\tau_1$ ), decay at the CI to the ground state (lifetime $\tau_2$ ), and possible relaxation pathways in the $S_0$ state. . . . .	101
5.1	Phenyl-phenyl torsional potential. . . . .	108
5.2	Order parameter evolution using standard LJ parameters . . .	109
5.3	Temperature dependence of density value of simulated and experimental $nCB$ series. Experimental data for 4CB not available.	112
5.4	Order parameter evolution during production run. The time evolution of order parameter of sample just below and just up the transition temperature shows sensible oscillation during time. This requires very long simulation (more than 40 ns for each temperature) to obtain good averaged properties. From [a] to [e] are shown the evolution from 4 to 8CB respectively. .	113
5.5	Simulated phase diagrams for $nCB$ series. The averaged order parameter for 4CB are not clearly, but the monotropic behavior of experimental ([BDH 1986a]) transition indicates the small tendency to obtain stable nematic phase for this compound. The transition temperatures are well defined from 5 to 8CB (Fig. [b]-[e]) and very close to the experimental values (vertical lines in the diagrams [BDH 1986b, Roche 1988, BDH 1986c;d]). . . . .	114
5.6	$g_0(r)$ mass center-mass center distributions. . . . .	116
5.7	$g_0(r)$ $N_1-N_1$ distributions. . . . .	118
5.8	$g_0(r)$ $N_1 - C_2$ bond $N_1 - C_2$ bond distributions. . . . .	119
5.9	$g_1(r)$ $N_1 - C_2$ bond $N_1 - C_2$ bond distributions. . . . .	120
5.10	$g_2(r)$ $N_1 - C_2$ bond $N_1 - C_2$ bond distributions. . . . .	121
5.11	Averaged long and short axes molecular lengths. . . . .	122
5.12	zLong axis length distributions. . . . .	123
5.13	y-axis length distributions. . . . .	124
5.14	x-axis length distributions. . . . .	125
5.15	$g_0(z)$ for big 8CB samples at 300, 310 and 315 K, computed with respect to different atoms. . . . .	127
5.16	Smectic phase obtained for 8CB at 300 K. . . . .	127

6.1	Transition moment of azo-benzene molecule used in the excitation model . . . . .	130
6.2	Average order parameter $P_2$ during illumination process. . . . .	130
6.3	Conformer distributions during photo-isomerization . . . . .	132
6.4	Average geometrical parameters. [a] $Ph-N=N-Ph$ torsion angle. [b] $N=N-Ph$ bending angles. [c] $N=N-C-C$ torsion angles. [d] $r_{4,4'}$ distance. . . . .	133
7.1	Chemical structure of 5CB LC. . . . .	136
7.2	<i>trans-cis</i> photo-isomerization scheme of 4-R-phenylazobenzene (R = H, F, Br, CH <sub>3</sub> , CF <sub>3</sub> , OCH <sub>3</sub> , <i>On</i> -Bu, <i>Ot</i> -Bu). . . . .	136
7.3	Chemical structure of the CSL spin probe. . . . .	137
7.4	Absorption spectra recorded after different times of exposure of the sample containing 5CB and 4- <i>Ot</i> -Bu-phenylazobenzene dissolved in cyclohexane (see text for details). . . . .	139
7.5	Temperature-composition phase diagram, relative to the 5CB doped with <i>cis</i> 4-OCH <sub>3</sub> -phenylazobenzene samples, showing the features considered in applying the lever rule to obtain the $f_N$ and $f_I$ values, see text for details. We also report the experimental ESR spectra collected respectively at the offset ( $T_a$ ), in the middle ( $T_0$ ) and at the end ( $T_b$ ) of the two-phase region. . . . .	141
7.6	Order parameter $\langle P_2 \rangle$ of the spin probe against reduced temperature $T^*$ for the samples doped with the <i>trans</i> azo-derivatives at (a) $x_1$ and (b) $x_2$ mole fraction compared to the pure 5CB. . . . .	143
7.7	Order parameter $\langle P_2 \rangle$ of the spin probe against reduced temperature $T^*$ for the samples doped with the <i>cis</i> azo-derivatives at (a) $x_1$ and (b) $x_2$ mole fraction compared to the pure 5CB. . . . .	143
7.8	Electrostatic potential at the molecular surface for the 5CB (a), <i>trans</i> azobenzene (b), <i>trans</i> 4- <i>On</i> -Bu-phenylazobenzene (c), <i>cis</i> 4- <i>Ot</i> -Bu-phenylazobenzene (d) and <i>cis</i> 4-Br-phenylazobenzene (e). Colour coded as from the palette (to be inserted). Geometry and charges obtained at B3LYP/6-31G* level. . . . .	148
8.1	Graphical comparison among experimental data, simulations and simulations with no-charged solutes at 290 [a] and 295 K [b] . . . . .	160

8.2	Orientational tumbling energy for Maier–Saupe model functions fitted over simulation data. . . . .	166
8.3	[a] Integral function from Eq. ?? versus molecular orientation.[b] Integral function reduced for molecular surface from Eq. ?? versus molecular orientation.[c] $U_{ST}$ tumbling orientational energy versus molecular orientation fitted on simulation data at T=290 K.[d] $U_{ST}$ tumbling orientational energy versus molecular orientation fitted on simulation data at T=290 K. . . . .	167
9.1	Superposition of the 20 energy conformers of 1 – <i>HexaPY</i> in $H_2O$ (a) and TFE/ $H_2O$ (80/20) (b) minimized in vacuo and aligned according with the minimal root mean square deviation of the backbone atoms residues. Only the polypeptide backbone is shown. . . . .	178
9.2	Superposition of the 20 energy conformers of 1 – <i>HexaPY</i> in $H_2O$ (a) and in TFE/ $H_2O$ (80/20) (b) minimized in solvent and aligned according with the minimal root mean square deviation of the backbone atoms residues. Only the polypeptide backbone is shown. . . . .	179
9.3	Time evolution of 1 – <i>HexaPY</i> conformers for each of the four isomers ( <i>trans–trans</i> [a], <i>cis–trans</i> [b], <i>trans–cis</i> [c], <i>cis–cis</i> [d]) in the three different protonation states obtained from MD simulations. The calculation is based on at least two consecutive residues in the same conformation concerning type I turn and three consecutive residues for <i>PPII</i> . . . . .	181
9.4	Superposition of the 20 NMR structures minimized in water ( $\rho$ , $\sigma$ ) values and the MD ( $\rho$ , $\sigma$ ) distributions of the 1 – <i>HexaPY</i> <i>trans–trans</i> isomer. . . . .	182
9.5	Time evolution of the distance between the phenolate oxygen of tyrosine and the side chain amide hydrogens of asparagine of 1 – <i>HexaPY</i> <i>trans–trans</i> in the three different protonation states. . . . .	184

9.6	1- <i>HexaPY trans-trans</i> typical conformations in the $H^+Y$ [a], $HY$ [b] and $HY^-$ [c] form obtained from MD simulations. It is worth to note the short distance between the phenolate oxygen of tyrosine and the amide side chain hydrogen of asparagine residue (blue line) only in the $HY^-$ form, where tyrosine is deprotonated. In the $H^+Y$ and $HY$ forms, instead, a turn structure is stabilized by an intrachain $ASN_{C=O}TYR_{NH}$ H-bond (yellow line). . . . .	185
9.7	$C_iC_{i+3}$ distance distribution concerning the <i>NPGY</i> region of 1 - <i>HexaPY trans</i> in the three different protonation states. It can be noticed the short distance in the $H^+Y$ and $HY$ forms and the larger distance distribution in the $HY^-$ form.	186
9.8	From left to right: 1 - <i>HexaPY</i> principal conformers obtained from NMR structures minimized in a water shell at acid pH [a], 1 - <i>HexaPY trans-trans</i> [b], <i>cis-trans</i> [c], <i>trans-cis</i> [d] and <i>ciscis</i> [e] isomers obtained from MD simulations in the $H^+Y$ form sampled every 2.5 ns. 1 - <i>HexaPY trans-trans</i> MD isomer shows the turn conformer, underlined also in the NMR structures [a]. . . . .	188
A.1	azobenzene, anizole, hexane, MPE and toluene chemical structures . . . . .	225
B.1	Atom schemes of <i>nCB</i> series. . . . .	229
E.1	Vector and angle definitions in the surface tensor model . . . .	245

# Glossary

$\pi$	bonding molecular orbital having one nodal plane, <a href="#">81</a> , <a href="#">82</a>
$\pi^*$	anti-bonding molecular orbital having two nodal plane, <a href="#">81–83</a> , <a href="#">100</a>
$n$	non-bonding molecular orbital having, <a href="#">81–83</a> , <a href="#">100</a>
<b>absorption</b>	process by which the energy of a photon is taken up by another entity by an atom whose valence electrons make a transition between two electronic energy levels, <a href="#">81</a>
<b>CASSCF</b>	complete active space configuration self consistent field: linear combination of configuration state functions includes all that arise from a particular number of electrons in a particular number of orbitals, <a href="#">86</a>
<b>CD</b>	circular dichroism technique, <a href="#">176</a> , <a href="#">179</a> , <a href="#">181</a>
<b>conformation</b>	structure, <a href="#">85</a> , <a href="#">86</a> , <a href="#">91</a> , <a href="#">94</a> , <a href="#">95</a>
<b>director</b>	mean alignment direction of molecules in a nematic or smectic phase obtained from diagonalization of order matrix, <a href="#">124</a>
<b>fluorescence</b>	optical phenomenon in which the molecular absorption of a photon triggers the emission of another photon with a longer wavelength, <a href="#">91</a> , <a href="#">92</a> , <a href="#">97</a> , <a href="#">188</a>

<b>inversion</b>	in-plane rotation from $\theta$ to $-\theta$ of bond around an atomic center, 82, 83, 85, 97, 101, 194
<b>isomer</b>	molecules with the same chemical formula and often with the same kinds of bonds between atoms, but in which the atoms are arranged differently, 81
<b>isotropic</b>	phase of matter characterized to unordered (orientational and positional) of molecules, 103, 106, 108, 109, 112, 115, 124
<b>MD</b>	molecular dynamics: classical dynamics simulation technique, 85, 87, 88, 91–93, 95–97, 99, 187, 188, 194
<b>nematic</b>	phase of matter characterized to monodimensional orientational order of molecules along a phase director, 103, 106, 108–110, 112, 115, 124, 195
<b>non-radiative decay</b>	is the process by which an molecule in an excited state drops to a lower-energy state, through vibrational state without photon emission, 81
<b>PCM</b>	polarizable continuum model: continuous model for solvent description in quantumchemical calculations, 101
<b>radiative decay</b>	is the process by which an atom, molecule or nucleus in an excited state drops to a lower-energy state, resulting in the creation of a photon, 81
<b>singlet</b>	two or more particles prepared in a correlated state, such that the total angular momentum of the state is zero, 81, 83, 85, 89

- smectic** phase of matter characterized to mono-dimensional orientational and mono-dimensional positional of molecules, [124](#)
- torsion** formed by three consecutive bonds in a molecule and defined by the angle created by the projection of the two outer bonds on a plane that is perpendicular to the central bond, [82](#), [87](#), [96](#), [97](#), [194](#)
- torsional triplet** see torsion, [83–85](#), [93–95](#), [97](#), [99](#), [101](#), [194](#)  
two or more particles prepared in a correlated state, such that the total angular momentum of the state is 1, [81](#)
- united atoms** atomistic models in which the hydrogens atoms are not explicitly showed but considered melt with the bonded heavier atoms, [103](#), [105](#), [109](#), [153](#)





# Bibliography

- A. A. Adzhubei and M. J. E. Sternberg. Lefthanded polyproline ii helices commonly occur in globular proteins. *J. Mol. Biol.*, 229:472, 1993. 174, 180
- A.Ferrarini, G.J.Moro, P.L.Nordio, and G.R.Luckhurst. *Molec. Phys.*, 77: 1–15, 1992. 164, 165, 191
- A.Ferrarini and P.L.Nordio. *J.Chem.Soc. Perkin Trans*, 2:456, 1998. 154, 164, 165, 191
- M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Walton Street, Oxford OX2 6DP, 1989. 25, 27, 44, 58, 67
- J. K. Kjems an G. Dolling. *Phys. Rev. B*, 11:16397, 1975. 41
- A. D. MacKerell Jr. and D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick T. and Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, and M. Karplus. *J. Phys. Chem. B*, 102:3586, 1998. 24
- H. C. Andersen. *J. Chem. Phys.*, 72:2384, 1980. 44
- H.C Andersen. *J. Comput. Phys.*, 52:24, 1983. 38
- A. Arcioni, C. Bacchiocchi, I. Vecchi, G. Venditti, and C. Zannoni. *Chem. Phys. Lett.*, 396:433–441, 2004. 136, 137

- A. Arcioni, R. Tarroni, and C. Zannoni. *J. Chem. Soc. Faraday Trans.*, 89: 2815, 1993. [139](#)
- V. I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlach, Berlin, 1989. [29](#), [50](#)
- I. A. Banerjee, L. Yu, and H. Matsui. Application of host-guest chemistry in nanotube-based device fabrication: Photochemically controlled immobilization of azobenzene nanotubes on patterned  $\alpha$ -cd. *J. Am. Chem. Soc.*, 125:9542, 2003. [81](#)
- A. Bansal and L. M Gierasch. The npxy internalization signal of the ldl receptor adopts a reverseturn conformation. *Cell.*, 67:1195, 1991. [188](#)
- A. Barducci, R. Chelli, P. Procacci, and V. Schettino. Misfolding pathways of the prion protein probed by molecular dynamics simulations. *Biophys. J.*, 88:1334, 2005. [173](#)
- J. A. Barker. The problem of long-range forces in the computer simulation of condensed matter. volume 9, page 45. NRCC Workshop Proceedings, 1980. [27](#), [67](#)
- J. A. Barker and R. O. Watts. *Molec. Phys.*, 26:789, 1973. [27](#), [67](#)
- S. Baroni, S. Gironcoli, and A. Dal Corso. *Rev. Mod. Phys.*, 73:515, 2001. [19](#)
- C. Bartels, T. Xia, M. Billeter, and K. Wthrich. The program xeasy for computersupported nmr spectral analysis of biological macromolecules. *J. Biomol. NMR.*, 5:110, 1995. [172](#), [173](#)
- J. F. Bazan, R. J. Fletterick, M. P. McKinley, and S. B. Prusiner. Predicted secondary structure and membrane topology of the scrapie prion protein. *Protein Eng.*, 1:125, 1987. [188](#)
- BDH, 1986a. product information. [114](#), [197](#)
- BDH, 1986b. product information. [114](#), [197](#)
- BDH, 1986c. product information. [114](#), [197](#)
- BDH, 1986d. product information. [114](#), [197](#)

- T.L. Beck. *Rev. Mod. Phys.*, 72:1041, 2000. [18](#)
- R. Berardi, G. Cainelli, P. Galletti, D. Giacomini, A. Gualandi, L. Muccioli, and C. Zannoni. Can the  $\pi$ -facial selectivity of solvation be predicted by atomistic simulation? *J. Am. Chem. Soc.*, 127:10699, 2005. [85](#)
- R. Berardi, L. Muccioli, and C. Zannoni. Can nematic transitions be predicted by atomistic simulations? a computational study of the odd-even effect. *ChemPhysChem*, 5:104, 2004. [105](#)
- H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. In B. Pullman, editor, *Intermolecular Forces*, Dordrecht, 1981. Reider. [173](#)
- B. H. Besler and K. M. Merz Jr. and P. A. Kollman. *J. Comput. Chem.*, 11: 431, 1990. [156](#)
- B. H. Besler, K. M. Merz Jr., and P. A. Kollman. *J. Comp. Chem*, 11:431, 1990. [87](#), [107](#)
- P.R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, 1969. [139](#)
- S. Dudiy P. Hyldgaard S. Ovesson C. Ruberto E. Schröder B.I. Lundqvist, A. Bogicevic and G. Wahnström. Bridging between micro- and macroscales of materials by mesoscopic models. *Comput. Mater. Sci.*, 24:1, 2002. [23](#)
- J. J. Biesiadecki and R. D. Skeel. *J. Comput. Phys.*, 109:318, 1993. [28](#), [31](#)
- Pietro Bortolus and Sandra Monti. Cis-trans photoisomerization of azobenzene. solvent and triplet donor effects. *J. Phys. Chem*, 83:648, 1979. [87](#), [90](#), [91](#), [94](#), [190](#)
- A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. *J. Comput. Phys.*, 17:10, 1975. [21](#)
- D. Bowler and M. Gillan. *Comput. Phys. Commun.*, 112:103, 1998. [20](#)
- J.H. Boyer. *The Chemistry of the Nitro and Nitroso Groups*. John Wiley and Sons, 1969. [136](#)
- L. Braunschweiler and R. R. Ernst. Coherence transfer by isotropic mixing: Application to proton correlation spectroscopy. *J. Magn. Reson.*, 53:521, 1983. [172](#)

- D.W. Brenner. *Phys. Status Solidi*, 217:23, 2000. [20](#)
- B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. Slater, S. Swaminathan, and M. Karplus. *J. Comput. Chem.*, 4:187, 1983. [24](#), [73](#)
- A. Buguin, M. H. L. Silberzan, B. Ladoux, and P. Keller. Application of host-guest chemistry in nanotube-based device fabrication: Photochemically controlled immobilization of azobenzene nanotubes on patterned  $\alpha$ -cd. *J. Am. Chem. Soc.*, 128:1088, 2006. [81](#)
- E. E. Burnell and C. A. de Lange. 98:2359–2387, 1998. [144](#)
- E.E. Burnell and C.A. De Lange. Prediction from molecular shape of solute orientational order in liquid crystals. *Chem. Rev.*, 98:2359–2387, 1998. [156](#), [243](#), [244](#), [245](#), [246](#)
- L. Calzolari, D. A. Lysek, D. R. Perez, P. Guntert, and K. Wthrich. Prion protein nmr structures of chickens, turtles, and frogs. *Proc. Natl. Acad. Sci. USA.*, 102:651, 2005. [170](#), [186](#)
- M. Camacho-Lopez, H. Finkelmann, P.Palfy-Muhoray, and M. Shelley. Fast liquid crystal elastomer swims into the dark. *Nat. Mater.*, 3:307, 2004. [81](#)
- R. Car and M. Parrinello. *Phys. Rev. Lett.*, 55:2471, 1985. [19](#)
- G. Cardini and V. Schettino. *Chem. Phys.*, 146:147, 1990. [42](#)
- P. Cattaneo and M. Persico. An *ab initio* study of the photochemistry of azobenzene. *Phys. Chem. Chem. Phys.*, 1:4739, 1999. [82](#)
- P. Cattaneo and M. Persico. Semiclassical simulations of azomethane photochemistry in the gas phase and in solution. *J. Am. Chem. Soc.*, 123:7638, 2001. [83](#), [85](#)
- K. C.Chou. Prediction of tight turns and their types in proteins. *Anal. Biochem.*, 286:116, 2000. [180](#)
- Alessandro Cembran, Fernando Bernardi, Marco Garavelli, Laura Gagliardi, and Giorgio Orlandi. On the mechanism of the cis-trans isomerization in the lowest electronic state of azobenzene:  $s_0$ ,  $s_1$ ,  $t_1$ . *J. Am. Chem. Soc.*, 126:3234, 2004. [82](#), [91](#)

- B. R. Chandra, R. Gowthaman, R. R. Akhouri, D. Gupta, and A. Sharma. Distribution of prolinerich (pxxxp) motifs in distinct proteomes: functional and therapeutic implications for malaria and tuberculosis. *Protein Eng. Des. Sel.*, 17:175, 2004. 171
- C-W. Chang, Y-C. Lu, T-T. Wang, and E. W-G. Diau. Photoisomerization dynamics of azobenzene in solution with  $s_1$  excitation: A femtosecond fluorescence anisotropy study. *J. Am. Chem. Soc.*, 126:10109, 2004. 81, 82, 83, 87, 88, 89, 92, 97, 190
- P. J. Channel and C. Scovel. *Nonlinearity*, 3:231, 1990. 28, 31, 57
- Y. Che and G. R. Marshall. Impact of cisproline analogs on peptide conformation. *Biopolymers*, 81:392406, 2006. 180
- J.R. Chelikowsky, Y. Saad, S. Ogut, I. Vasiliev, and A. Stathopoulos. *Phys. Status Solidi*, 217:173, 2000. 18
- H. Cheng and F. A. Bovey. Cistrans equilibrium and kinetic studies of acetyl proline and glycyllproline. *Biopolymers*, 16:1465, 1977. 179
- D. L. Cheung, S. J. Clark, and M. R. Wilson. *Phys. Rev. E*, 65:051709, 2002. 84
- B. Chopard and M. Droz. *Cellular Automata Modeling of Physical Systems*. Cambridge University Press, 1998. 21
- G. Ciccotti and J. P. Ryckaert. *Comp. Phys. Report*, 4:345, 1986a. 25, 38
- G. Ciccotti and J.-P. Ryckaert. "Computer Phys. Rep.", 4:345, 1986b. 52
- C. Ciminelli, G. Granucci, and M. Persico. The photoisomerization mechanism of azobenzene: A semiclassical simulation of nonadiabatic dynamics. *Chem.-Eur. J.*, 10:2327, 2004. 81, 82, 89, 90, 97, 101
- C. Ciminelli, G. Granucci, and M. Persico. Are azobenzenophanes rotation-restricted? *J. Chem. Phys.*, 123:174317, 2005. 82
- W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.*, 117:5179, 1995. 24, 42, 65, 75, 83, 145, 173, 189

- T.A. Darden, D.M. York, and L.G. Pedersen. *J. Chem. Phys.*, 98:10089, 1993. 27, 69, 72
- H. de Raedt and B. De Raedt. *Phys. Rev. A*, 28:3575, 1983. 33
- N. G. Delaney and V. Madison. Unusual cis-trans isomerism in *n*-acetyl, *n*-methylamide derivatives of syn and anti-5-methylproline. *Int. J. Pept. Protein Res*, 19:543, 1982. 179
- S.W. deLeeuw, J. W. Perram, and E. R. Smith. "Proc. R. Soc. London A", 373:27, 1980. 27, 67, 68
- A.J. Van der Est, M.Y. Kok, and E.E. Burnell. Size and shape effects on the orientation of rigid molecules in nematic liquid crystals. *Molec. Phys.*, 60:397-413, 1987. 162, 163, 191
- G. Dhatt and G. Touzot. *The Finite Element Method Displayed*. Wiley, 1984. 22
- E. W-G. Diau. A new trans-to-cis photoisomerization mechanism of azobenzene on the  $s_1(n, \pi^*)$  surface. *J. Phys. Chem. A*, 108:950, 2004. 82
- R.M. Dreizler and E.K.U. Gross. *Density Functional Theory*. Springer, Berlin, 1998. 17
- Robert Englman and Joshua Jortner. The energy gap law for radiationless transitions in large molecules. *Molec. Phys.*, 18(2):145, 1970. 86
- U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. *J. Chem. Phys.*, 101:8577, 1995. 27, 69, 70, 72, 77
- M. J. Frisch et al. *IBM64-RS6000-G03RevB.05*. Gaussian, Inc., Pittsburgh PA, 2004. 156
- P. Ewald. *Ann. Phys.*, 64:253, 1921. 27, 67
- M. Ferrario. In M.P.Allen and D.J.Tildesley, editors, *Computer Simulation in Chemical Physics*, page 153. Kluwer Academic Publishers, 1993. 44
- M. Ferrario and J.-P. Ryckaert. *Molec. Phys.*, 78:7368, 1985. 44
- H. Finkelmann, E. Nishikawa, G. G. Pereira, and M. Warner. A new optomechanical effect in solids. *Phys. Rev. Lett.*, 87:15501, 2001a. 81

- H. Finkelmann, E. Nishikawa, G.G. Pereira, and M. Warner. *Phys. Rev. Lett.*, 87:015501(4), 2001b. 135
- Th. Frauenheim, G. Seifert, M. Elsner, Z. Hajnal, G. Jungnickel, D. Porezag, S. Suhai, and R. Scholz. *Phys. Status Solidi*, 217:41, 2000. 20
- J.H. Freed. *Electron Spin Relaxation in Liquids*. L.T. Muus and P.W. Atkins, 1972. 139
- J.H. Freed. *Spin labelling, Theory and Applications*. L.J. Berliner, 1976. 139
- D. Frenkel and B. Smit. *Understanding Molecular Simulations: From Algorithms to Applications*. Academic Press, San Diego, 1996. 19, 44, 47, 49
- T. Fujino, S. Y. Arzhantsev, and T. Tahara. Femtosecond time-resolved fluorescence study of photoisomerization of trans-azobenzene. *J. Phys. Chem. A*, 105:8123, 2001. 92
- P.L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello. *Science*, 291:2121, 2001. 20
- H. Goldstein. *Classical Mechanics*. Addison-Wesley, Reading MA, 1980. 29, 30, 47
- L. Greengard and V. Rokhlin. *J. Comput. Phys.*, 73:325, 1987. 28
- S. K. Grey, D. W. Noid, and B. G. Sumpter. *J. Chem. Phys.*, 101:4062, 1994. 28
- C. Griesinger and R. R. Ernst. Frequency offset effects and their elimination in nmr rotating frame crossrelaxation spectroscopy. *J. Magn. Reson.*, 75: 261, 1987. 172
- H. Grubmuller, H. Heller, A. Winemuth, and K. Schulten. *Mol. Simul.*, 6: 121, 1991. 26, 75
- N. Guex and M.C. Peitsch. Swissmodel and the swisspdbviewer: an environment for comparative protein modeling. *Electrophoresis*, 18:2714, 1997. 173

- W. F. Van Gunsteren, S. R. Billeter, A. A. Eising, P. H Hnenberger, P. Krger, A. E. Mark, W. R. P. Scott, and I.G. Tironi. *Biomolecular simulation: the GROMOS96 manual and user guide (BIOMOs)*., 1996. 173
- M. Gustiananda, J. R. Liggins, P. L. Cummins, and J. E. Gready. Conformation of prion protein repeat peptides probed by fret measurements and molecular dynamics simulations. *Biophys J.*, 86:2467, 2004. 187
- J.-P. Hansen. Molecular-dynamics simulation of coulomb systems in two and three dimensions. In *Molecular Dynamics Simulation of Statistical-Mechanics Systems*, Proceedings of the International School of Physics "Enrico Fermi". North Holland Physics, 1986. 67
- D. A. Harris, D. L. Falls, F. A. Johnson, and G. D. Fischbach. A prionlike protein from chicken brain copurifies with an acetylcholine receptorinducing activity. *Proc. Natl. Acad. Sci. USA.*, 88:7664, 1991. 170
- D. A. Harris, P. Lele, and W. D. Snider. Localization of the mrna for a chicken prion protein by in situ hybridization. *Proc. Natl. Acad. Sci. USA.*, 90:4309, 1993. 170
- M. Heiskanen, T. Torsti, M.J. Puska, and R.M. Nieminen. *Phys. Rev. E*, 63:245106, 2001. 18
- B. R. Henry and W. Siebrand. *Organic molecular photophysics*, chapter 4. John Wiley & Sons, London, 1973. 86
- T. Herman, P. Guntert, and K. Wthrich. Protein nmr structure determination with automated noe assignment using the new software candid and the torsion angle dynamics algorithm dyana. *J. Mol Biol.*, 319:209, 2002. 173
- G. Herzberg. *Spectra of Diatomic Molecules*. Van Nostrand, New York, 1950. 41
- R. W. Hockney. *Computer Simulation Using Particles*. McGraw-Hill, New York, 1989. 27, 69
- W.G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985. 88, 173



- T. Hugel, N. B. Holland, A. Cattani, L. Moroder, M. Seitz, and H. E. Gaub. *Science*, 296:1103, 2002. 81
- D. D. Humphreys, R. A. Friesner, and B. J. Berne. *J. Phys. Chem.*, 98:6885, 1994. 26
- R.L. Humphries and G.R. Luckhurst. 352:41–56, 1976. 138
- G. Hutchinson and J. M. Thornton. A revised set of potentials for betaturn formation in proteins. *Protein Science.*, 3:2207, 1994. 180
- T. L. Hwang and A. J. Shaka. Water suppression that works excitation sculpting using arbitrary waveforms and pulsedfield gradients. *J. Magn. Reson. A.*, 112:275, 1995. 172
- K. Ichimura. 100:1847–1873, 2000. 135
- T. Ikeda. 13:2037–2057, 2003. 135
- Tomiki Ikeda and O. Tsutsumi. Optical switching and image storage by means of azobenzene liquid-crystal films. *Science*, 268:1873, 1995. 81
- Takeshi Ishikawa, Takeshi Noro, and Takayuki Shoda. Theoretical study of the photoisomerization of azobenzene. *J. Chem. Phys.*, 115:7503, 2001. 81, 83, 84, 86, 90, 196
- O. Jardetzky and G. C. K. Roberts. *NMR in Molecular Biology*. Academic Press, New York, 1981. 176
- E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620, 1957. 153
- J. S. Jhon and Y. K. Kang. Imide *cis-trans* isomerization of *n*-acetyl *n*'methylprolineamide and solvent effects. *Phys. Chem. A.*, 103:5436, 1999. 170
- H. Jonsson, G. Mills, and K.W. Jacobsen. *Classical and Quantum Dynamics in Condensed Phase Simulations*. World Scientific, 1998. 20
- W. Kabsch and C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogenbonded and geometrical features. *Biopolymers*, 22:2577, 1983. 183

- M. Kaukonen, P.K. Sitch, G. Jungnickel, R.M. Nieminen, S. Pöykkö, D. Porezag, and Th. Frauenheim. *Phys. Rev. B*, 57:9965, 1998. 20
- B. K. Kay, P. Williamson, and M. Sudol. The importance of being proline: the interaction of prolinerich motifs in signaling proteins with their cognate domains. *FASEB J.*, 14:231, 2000. 171
- M. Keller, C. Sager, P. Dumy, M. Schutkowski, G. S. Fisher, and M. Mutter. Enhancing the proline effect: Pseudoprolines for tailoring cis/trans isomerization. *J. Am. Chem. Soc.*, 120:2714, 1998. 180
- Noburu Kitamura, Norio Sakata, Heang-Boo Kim, and Satoshi Habuchi. Energy gap dependence of the nonradiative decay rate constant of 1-anilino-8-naphthalene sulfonato in reverse micelles. *Analytical Sciences*, 15:413, 1999. 86
- S. Kobayashi, H. Yokoyama, and H. Kamei. Substituent and solvent effects on electronic absorption spectra and thermal isomerization of push-pull-substituted *cis* azobenzenes. *Chem. Phys. Lett.*, 138:333, 1987. 87
- R. Koradi, M. Billeter, and K. Wthrich. Molmol: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, 14:5155, 1996. 173
- A. Kumar, R. R. Ernst, and K. Wthrich. A twodimensional nuclear overhauser enhancement (2d noe) experiment for the elucidation of complete protonproton cross relaxation networks in biological macromolecules. *Biophys. Res. Commun.*, 95:16, 1980. 172
- D.P. Landau and K. Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge, 2000. 20
- E. Langella, R. Improtà, and V. Barone. Checking the ph-induced conformational transition of prion protein by molecular dynamics simulations: Effect of protonation of histidine residues. *Biophys. J.*, 87:3623, 2004. 182
- Y. Lansac, M. A. Glaser, N. A. Clark, and O. D. Lavrentovich. *Nature*, 398:54, 1999. 81
- I. K. Lednev, T.-Q. Ye, P. Matousek, M. Towrie, P. Foggi, F. V. R. Neuwahl, S. Umopathy, R. E. Hester, and J. N. Moore. Femtosecond time-resolved uv-visible absorption spectroscopy of trans-azobenzene on excitation wavelength. *Chem. Phys. Lett.*, 290:68, 1998. 87, 92, 190

- H. Lee, T. A. Darden, and L. G. Pedersen. *J. Chem. Phys.*, 102:3830, 1995. 27, 72
- X.S. Li, J. C. Tully, H. B. Schlegel, and M. J. Frisch. Ab initio ehrenfest dynamics. *J. Chem. Phys.*, 123:084106, 2005. 87
- D. R. Lide. *Handbook of Chemistry and Physics*. CRC Press, 2004–2005. 89, 189
- Y.C. Lu, E. W.-G. Diau, and H. Rau. Femtosecond fluorescence dynamics of rotation-restricted azobenzophenones: new evidence on the mechanism of trans  $\rightarrow$  cis photoisomerization of azobenzene. *J. Phys. Chem. A*, 109:2090, 2005. 92
- Ying-Chih Lu, Chih-Wei Chang, and Eric Wei-Guang Diau. Femtosecond fluorescence dynamics of *trans*-azobenzene in hexane on excitation to the  $s_1(n, \pi^*)$  state. *J. Chin. Chem. Soc.*, 49:693, 2002. 87, 91, 97
- D. A. Lysek, L. Calzolari, and K. Wuthrich. Nmr assignment of the chicken prion protein fragments chprp(128242) and chprp(25242). *J. Biomol. NMR*, 30:97, 2004. 170
- D. A. Lysek and K. Wuthrich. Prion protein interaction with the c-terminal sh3 domain of grb2 studied using nmr and optical spectroscopy. *Biochemistry*, 43:10393, 2004. 171
- A. D. MacKerrel, J. Wirkeiwicz-Kuczera, and M. Karplus. *J. Am. Chem. Soc.*, 117:11946, 1995. 18, 24
- J. Makowska, S. Rodziewicz-Motowid, K. Bagiska, J. A. Vila, A. Liwo, L. Chmurzynski, and H. A. Scheraga. Polyproline ii conformation is one of many local conformational states and is not an overall conformation of unfolded peptides and proteins. *Proc. Natl. Acad. Sci. USA.*, 103:1744, 2006. 182
- M. Marchi and P. Procacci. *ORAC Manual and Guide*. CECAM, Available at <ftp.cecam.fr:/pub/orac/doc/manual.ps> CECAM-ENS Lyon, 1997. 23
- M. Marchi and P. Procacci. *J. Chem. Phys.*, 109:5194, 1998. 26, 45, 46, 49, 50, 56, 58, 59, 60, 61, 67, 74, 76

- E. M. Marcotte and D. Eisenberg. Chicken prion tandem repeats form a stable, proteaseresistant domain. *Biochemistry*, 38:667, 1999. 170, 171
- D.E. Martire, G.A. Oweimreen, G.I. Agern, S.G. Ryan, and H.T. Peterson. *J. Chem. Phys.*, 64:1456–1463, 1976. 138
- G. J. Martyna, D. J. Tobias, and M. L. Klein. *J. Chem. Phys.*, 101:4177, 1994. 44, 46, 49, 50
- G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein. *Mol. Phys.*, 87:1117, 1996. 26, 44, 45, 58, 60
- A. Di Matteo, A. Ferrarini, and G.J. Moro. Effects of electrostatic interactions on orientational order of solutes in liquid crystals. *J. Phys. Chem. B*, 104:7764–7773, 2000. 153, 154, 241, 245
- A. J. McDonald and S. Hanna. Atomistic computer simulation of terraced wetting of model 8cb molecules at crystal surfaces. *Mol. Cryst. Liq. Cryst.*, 413:135/2271, 2004. 108
- A. J. McDonald and S. Hanna. Atomistic simulation of a model liquid crystal. *J. Chem. Phys.*, 124:164906, 2006. 108
- F. D. Medina and W. B. Daniels. *J. Chem. Phys.*, 64:150, 1976. 41
- E. Meirovitch and J.H. Freed. *J. Phys. Chem*, 88:4995, 1984. 136, 137
- S. Melchionna, G. Ciccotti, and B. L. Holian. *Mol. Phys.*, 78:533, 1993. 50
- D. La Mendola, R. P. Bonomo G. Impellizzeri, G. Maccarrone, G. Pappalardo, E. Rizzarelli, A. Pietropaolo, and V. Zito. Copper(ii) complexes with chicken prion repeats: influence of proline and tyrosine residues on the coordination features. *J. Biol. Inorg. Chem.*, 10:463, 2005. 172, 179, 182, 184
- T. Muraoka, K. Kinbara, and T. Aida. *Nature*, 440:512, 2006. 81
- R.M. Nieminen. *Curr. Opin. Solid State Mater. Sci.*, 4:493, 1999. 17
- C. Nonnenberg, H. Gaub, and I. Frank. First-principles simulation of the photoreaction of a capped azobenzene: The rotational pathway is feasible. *ChemPhysChem*, 7:1455, 2006. 82

- S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molec. Phys.*, 52:255, 1984. 44, 46, 47, 88, 173
- S. Nosé. In M. Meyer and V. Pontikis, editors, *Computer Simulation in Materials Science*, page 21. Kluwer Academic Publishers, 1991a. 44
- S. Nosé. *Prog. Theor. Phys. Supp.*, 103:1, 1991b. 44
- S. Nosé and M. L. Klein. *Mol. Phys.*, 50:1055, 1983. 40, 47, 56
- G. Onida, L. Reining, and A. Rubio. Electronic excitations: density-functional versus many-body greens-function approaches. *Rev. Mod. Phys.*, 74:601, 2002. 18
- P. Ordejon. *Comput. Mater. Sci.*, 12:157, 1998. 20
- G.A. Oweimreen. 105:8410–8416, 2001. 138
- J. P. Paccaud, W. Reith, B. Johansson, K. E. Magnusson, B. Mach, and J. L. Carpenter. Clathrin-coated pit-mediated receptor internalization. *J. Biol. Chem.*, 268:23191, 1993. 188
- E. Paci and M. Marchi. *J. Phys. Chem*, 104:3003, 1996. 56
- M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Letters*, 45:1196, 1980. 44, 88, 111
- J. J. Pavelites, P. A. Gao, and A. D. MacKerrel. *Biophysical J.*, 18:221, 1997. 24
- Thomas Garm Pedersen, P. S. Ramanujam, and Per Michael Johansen. Micro-actuators: When artificial muscles made of nematic liquid crystal elastome. *J. Opt. Soc. Am. B*, 15(11):2722, 1998. 129
- H.G. Petersen. *J. Chem. Phys.*, 103:3668, 1995. 72
- H.G. Petersen, D. Soelvanson, and J. W. Perram. *J. Chem. Phys*, 101:8870, 1994. 28
- P. Procacci and B. J. Berne. *J. Chem. Phys.*, 101:2421, 1994a. 26, 31, 41
- P. Procacci and B. J. Berne. *Mol. Phys.*, 83:255, 1994b. 26, 46, 48, 58, 60

- P. Procacci, T. Darden, and M. Marchi. *J. Phys. Chem.*, 100:10464, 1996. 26, 28, 41, 74, 76
- P. Procacci, T. Darden, E. Paci, and M. Marchi. *J. Comp. Chemistry*, 18:1848, 1997a. 76
- P. Procacci and M. Marchi. Taming the ewald sum in molecular dynamics simulations of solvated proteins via multiple time step algorithms. *J. Chem. Phys.*, 104:3003, 1996. 26, 41, 75, 76, 170
- P. Procacci, M. Marchi, and G. J. Martyna. *J. Chem. Phys.*, 108:8799, 1998. 77
- P. Procacci, E. Paci, T. Darden, and M. Marchi. Orac: A molecular dynamics program to simulate complex molecular systems with realistic electrostatic interactions. *J. Comput. Chem.*, 18:1848–1862, 1997b. 88, 173
- S. B. Prusiner. Prion diseases and the bse crisis. *Science*, 278:245, 1997. 169
- S. B. Prusiner, M. R. Scott, S. J. DeArmond, and F. E. Cohen. Prion protein biology. *Cell.*, 93:337, 1998. 169
- A. Radzicka, S. A. Acheson, and R. Wolfenden. *Cis/trans* isomerization at proline: Desolvation and its consequences for protein folding. 1992. 170
- A. Rahman and F. H. Stillinger. *J. Chem. Phys.*, 55:3336, 1971. 79
- M. Rance, O. W. Srensen, G. Bodenhausen, G. Wagner, R. R. Ernst, and K. Wthrich. Improved spectral resolution in cosy 1h nmr spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.*, 117:479, 1983. 172
- D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, Cambridge (UK), 1995. 19, 44
- H. Rau. *Photochromism. Molecules and system*, chapter 4. Elsevier, Amsterdam, 2003. 87, 89, 90, 91, 190
- J. S. Richardson. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.*, 34:167339, 1981. 184
- Hoffman-La Roche, 1988. product information. 114, 197

- G. D. Rose, L. M. Gierasch, and J. A. Smith. Turns in peptides and proteins. *Adv Protein Chem.*, 37:1109, 1985. 184
- R.E. Rudd and J.Q. Broughton. *Phys. Status Solidi*, 217:251, 2000. 23
- J.-P. Ryckaert and G. Ciccotti. *J. Chem. Phys.*, 78:7368, 1983. 46
- J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. *J. Comp. Phys.*, 23:327, 1977. 25, 38
- M. Saito. *J. Chem. Phys.*, 101:4055, 1994. 27
- J. M. Sanz-Serna. *Acta Numerica*, 1:243, 1992. 28, 30, 57
- H. Satzger, C. Root, and M. Braun. Excited-state dynamics of trans- and cis-azobenzene after uv excitation in the  $\pi\pi^*$  band. *J. Phys. Chem. A*, 108:6265, 2004. 81
- M. Schlick. *Molecular Modeling and Simulation An Interdisciplinary Guide*. SpringerVerlag, 2002. 173
- D.J. Schneider and J.H. Freed. *Biological Magnetic Resonance, Spin Labelling*. L.J. Berliner and J. Reuben, 1989. 139
- J. Shimada, H. Kaneko, and T. Takada. *J. Comput. Chem.*, 15:28, 1994. 28
- S. L. Shyng, J. E. Heuser, and D. A. Harris. A glycolipidanchored prion protein is endocytosed via clathrincoated pits. *J. Cell Biol.*, 125:1239, 1994. 170
- S. L. Shyng, K. L. Moulder, A. Lasko, and D. A. Harris. The nterminal domain of a glycolipidanchored prion protein is essential for its endocytosis via clathrincoated pits. *J. Biol. Chem.*, 270:14793, 1995. 170
- M.R. Sorensen and A.F. Voter. *J. Chem. Phys.*, 112:9599, 2000. 19
- B. Stevansson, A.V.Komolkin, D. Sandström, and A. Maliniak. Structure and molecular ordering extracted from residual dipolar couplings: A molecular dynamics simulation study. *J. Chem. Phys.*, 114:2332–2339, 2001. 161, 191
- J. P. Straley. Ordered phases of a liquid of biaxial particles. *Phys. Rev. A*, 10:1881, 1974. 154, 241

- W. B. Street, D.J. Tildesley, and G. Saville. *Mol. Phys.*, 35:639, 1978. 26
- S. J. Stuart, R. Zhou, and B. J. Berne. *J. Chem. Phys.*, 105:1426, 1996. 76
- Raymond T. Syvitski and E. Elliott Burnell. *Chem. Phys. Lett.*, 281:199–206, 1997. 146
- Raymond T. Syvitski and E. Elliott Burnell. *J. Chem. Phys.*, 113:3452–3465, 2000. 144
- O. Teleman and B. Joensonn. *J. Comput. Chem.*, 7:58, 1986. 26
- G. Tiberio, L. Muccioli, R. Berardi, and C. Zannoni. *to be submitted*, 2006a. 156
- G. Tiberio, L. Muccioli, R. Berardi, and C. Zannoni. How does the *trans*–*cis* photoisomerization of azobenzene take place in organic solvents? *submitted*, 2006b. 83
- X. Tong, G. Wang, A. Yavrian, T. Galstian, and Y. Zhao. *Adv. Mater.*, 17:370, 2005. 81
- A. Toniolo, C. Ciminelli, M. Persico, and T. J. Martinez. Simulation of the photodynamics of azobenzene on its first excited state: Comparison of full multiple spawning and surface hopping treatments. *J. Phys. Chem*, 123:234308, 2005. 82
- S. Toxvaerd. *Phys. Rev. B.*, 47:343, 1993. 58
- S. J. Toxvaerd. *J. Chem. Phys.*, 87:6140, 1987. 34
- H. F. Trotter. *Proc. Am. Math Soc.*, 10:545, 1959. 32
- O. Tsutsumi, T. Shiono, T. Ikeda, and G. Galli. Photochemical phase transition behavior of nematic liquid crystals with azobenzene moieties as both mesogens and photosensitive chromophores. 101:1332, 1997. 135
- M. E. Tuckerman and B. J. Berne. *J. Chem. Phys.*, 95:8362, 1991. 26
- M. E. Tuckerman, B. J. Berne, and G. J. Martyna. *J. Chem. Phys.*, 99:2278, 1993. 26, 34



- M. E. Tuckerman, B. J. Berne, and A. Rossi. *J. Chem. Phys.*, 94:1465, 1990. 26
- M. E. Tuckerman, B.J. Berne, and G.J. Martyna. *J. Chem. Phys.*, 97:1990, 1992. 26, 31, 34, 38, 76
- M. E. Tuckerman, G. J. Martyna, and B. J. Berne. *J. Chem. Phys.*, 94:6811, 1991. 26, 76
- M. E. Tuckerman, C. J. Mundy, and M. L. Klein. *Phys. Rev. Letters*, 78: 2042, 1997. 50
- M. E. Tuckerman and M. Parrinello. *J. Chem. Phys.*, 101:1302, 1994. 39
- J. C. Tully and P. V. Parandekar. Mixed quantum-classical equilibrium. *J. Chem. Phys.*, 122:094102, 2005. 87
- S. Urban, J. Przedmojsky, and J. Czub. X-ray studies of the layer thickness in smectic phases. *Liq. Cryst.*, 32:619, 2005. 126
- W.F. van Gunsteren and H. J. C. Berendsen. *Groningen Molecular Simulation (GROMOS) Library Manual*. Biomos, Groningen, 1987. 24, 65
- various authors. *Cerius 2 program manual*. Molecular Simulations Inc., 2000. 156
- J. von Neumann. Various techniques used in connection with random digits. *Appl. Math. Ser.*, 12:36, 1951. 86
- G. Wagner, W. Braun, F. T. Havel, T. Schaumann, N. Go, and K. Wthrich. Protein structures in solution by nuclear magnetic resonance and distance geometry. *J. Mol. Biol.*, 196:611, 1987. 176
- T. Waite. *Phys. Rev.*, 107:463, 1957. 23
- J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *J. Comput. Chem.*, 25:1157, 2004. 83
- A. Weaver, A.J. Van der Est, J.C.T. Rendell, G.L. Hoatson, G.S. Bates, and E.E. Burnell. Orientation mechanisms for small solutes in the nematic liquid crystal 5cb- $\alpha,\beta$ -d<sub>4</sub>. *Liq. Cryst.*, 2:633–642, 1987. 155, 158, 191

- S.J. Wiener, P.A. Kollmann and D.T. Nguyen, and D.A. Case. *J. Comput. Chem.*, 7:230, 1986. [24](#), [65](#), [73](#)
- Williamson, Hood, and Grossman. *Phys. Rev. Lett.*, 87:246406, 2001. [18](#)
- F. Wopfner, G. Weidenhofer, R. Schneider, A. von Brunn, S. Gilch, T. F. Schwarz, T. Werner, and H. M. Schatzl. Analysis of 27 mammalian and 9 avian prps reveals high conservation of flexible regions of the prion protein. *J. Mol. Biol.*, 289:1163, 1999. [170](#)
- H. Yoshida. *Phys. Letters A*, 150:262, 1990. [33](#)
- Y. Yu, M. Nakano, and T. Ikeda. Directed bending of a polymer film by light. *Nature*, 425:145, 2003. [8](#), [81](#), [195](#)
- B. Zagrovic, J. Lipfert, E. J. Sorin, I. S. Millett, W. F. van Gunsteren, S. Doniach, and V. S. Pande. Unusual compactness of a polyproline type ii structure. *Proc. Natl. Acad. Sci. USA*, 102:11698, 2005. [186](#)
- R. Zahn. The octapeptide repeats in mammalian prion protein constitute a pH dependent folding and aggregation site. *J. Mol. Biol.*, 334:477, 2003. [187](#)
- C. Zannoni. Molecular design and computer simulations of novel mesophases. *J. Mater. Chem.*, 11:2637, 2001. [117](#)
- R. Zhou and B. J. Berne. *J. Chem. Phys.*, 103:9444, 1996. [28](#)

# Appendices



# Appendix A

## Azobenzene force field

### A.1 Structures and atom labels

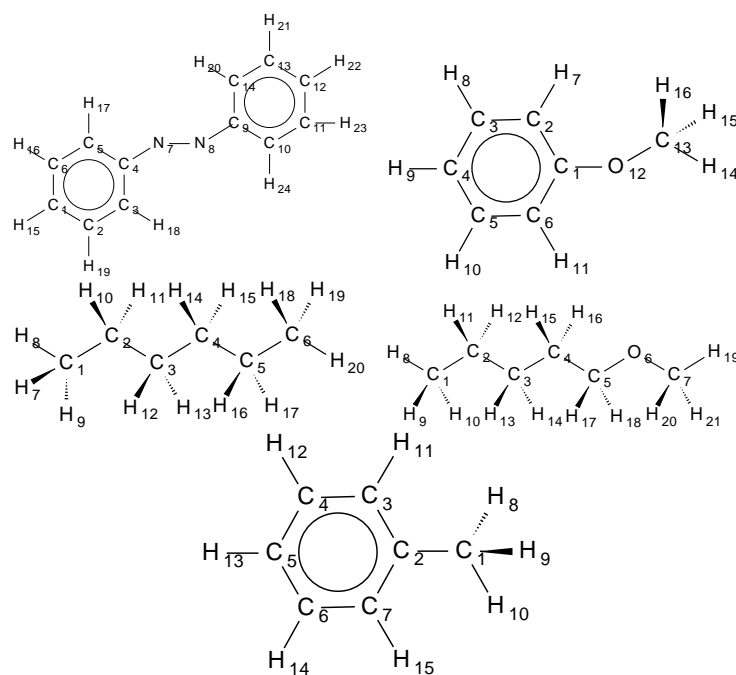


Figure A.1: azobenzene, anisole, hexane, MPE and toluene chemical structures



## A.3 Bonded force field

### A.3.1 Bond and bendings

Table A.2: Bonds and bendings in azobenzene simulation

Bonds				Bendings					Bendings				
at1	at2	Kbond	t0	at1	at2	at3	Kb	t0	at1	at2	at3	Kb	t0
CA	CY	95.9	1.540	C2	C2	C2	63.0	112.4	HC	C2	OS	50.0	109.5
CD	CD	469.0	1.400	C2	C2	C3	62.1	114.0	HC	C3	OS	50.0	109.5
CA	CD	469.0	1.400	C3	C2	C3	62.1	114.0	H1	C3	OS	50.0	107.0
C2	C2	95.9	1.540	CA	CA	CD	85.0	120.0	H1	C3	H1	35.0	109.5
C2	C3	95.9	1.540	CD	CA	CD	85.0	120.0	CA	CD	HA	50.0	120.0
CA	C2	317.0	1.510	CA	CD	CD	85.0	120.0	CD	CD	HA	50.0	120.0
CT	OS	320.0	1.410	CD	CA	C2	70.0	120.0	CA	C2	HC	50.0	109.5
CA	OS	450.0	1.360	CA	C2	C2	63.0	112.4	C2	C2	HC	50.0	109.5
CA	ND	372.0	1.220	CT	OS	CA	60.0	117.0	HC	C2	HC	35.0	109.5
CA	ND	372.0	1.220	CD	CA	ND	104.7	120.0	C2	C3	HC	50.0	109.5
ND	ND	418.1	1.220	CD	CA	ND	104.7	120.0	HC	C2	C3	50.0	109.5
C3	OS	320.0	1.420	CA	ND	ND	118.8	117.0	HC	C3	HC	35.0	109.5
C2	OS	320.0	1.420	C2	OS	C3	60.0	109.5	CD	CA	C3	70.0	120.0
CA	CA	469.0	1.400	CA	OS	C3	60.0	117.0	HC	C3	CA	50.0	109.5
CY	NY	600.0	1.206	CD	CA	OS	80.0	125.0	CD	CD	CD	63.0	120.0
H1	C3	340.0	1.090	CY	CA	CD	85.0	120.0					
CD	HA	367.0	1.080	CA	CY	NY	79.5	180.0					
HC	C2	340.0	1.090	CD	CA	C2	70.0	120.0					
HC	C3	340.0	1.090	CA	C2	C2	63.0	112.4					
C3	CA	317.0	1.510	C2	C2	OS	50.0	109.5					

### A.3.2 Proper torsion

$$E_{tor} = V_0(a_0 + a_1\cos(x) + a_2\cos(2x) + a_3\cos(3x) + a_4\cos(4x) + a_5\cos(5x))$$

$$V_0 = [kcalmol^{-1}]$$

at1	at2	at3	at4	V0	a0	a1	a2	a3	a4	a5
x	C2	C2	x	4.15	0.4671	0.1616	0.0329	0.33836	0.0	0.0
x	OS	CA	CD	3.50	0.5743	5.883e-5	-8.081e-4	2.364e-5	-1.267e-4	1.974e-5

$$E_{tor} = V_0(1 - \cos(nx + g)) \quad V_0 = [kcalmol^{-1}] \quad g = [degree]$$

at1	at2	at3	at4	V0	n	g
x	C2	CA	x	0.500	6	0.0
x	OS	CA	x	2.000	2	180.0
x	CA	ND	x	2.650	2	180.0
x	CA	ND	x	2.650	2	180.0
x	CA	CD	x	2.650	2	180.0
x	CD	CD	x	5.300	2	180.0
x	CA	CA	x	1.325	2	180.0
NY	CY	CA	CD	0.000	1	180.0
H1	C3	OS	CA	0.800	-1	0.0
C2	C2	C3	HC	0.160	3	0.0
C2	C2	OS	C3	0.383	-3	0.0
HC	C2	OS	C3	0.383	-3	0.0
C2	OS	C3	HC	0.383	-3	0.0
HC	C2	C3	HC	0.150	3	0.0
HC	C3	CA	CD	0.380	-3	180.0

## A.4 Nonbonded mixrule

$$E_{ij} = e_{ij} * ((slj/x)^{12} - (slj/x)^6)$$

$$rlj/2 = 2^{(1/6)} * slj/2 = .56123102slj$$

$$elj[Kcal/mol] = 2/1000elj[K]$$

at	rlj	slj	atomic mass
C2	1.9080	0.1094	12.01
C3	1.9080	0.1094	12.01
CA	1.85	0.120	12.01
CD	1.85	0.120	12.010
OS	1.6837	0.15	15.99
ND	1.8240	0.17	14.01
CY	1.9080	0.086	12.000
NY	1.8240	0.170	14.010
H1	1.3870	0.016	1.008
HA	1.4590	0.015	1.008
HC	1.4870	0.016	1.008



# Appendix B

## *n*CB force field

### B.1 Structures and atom labels

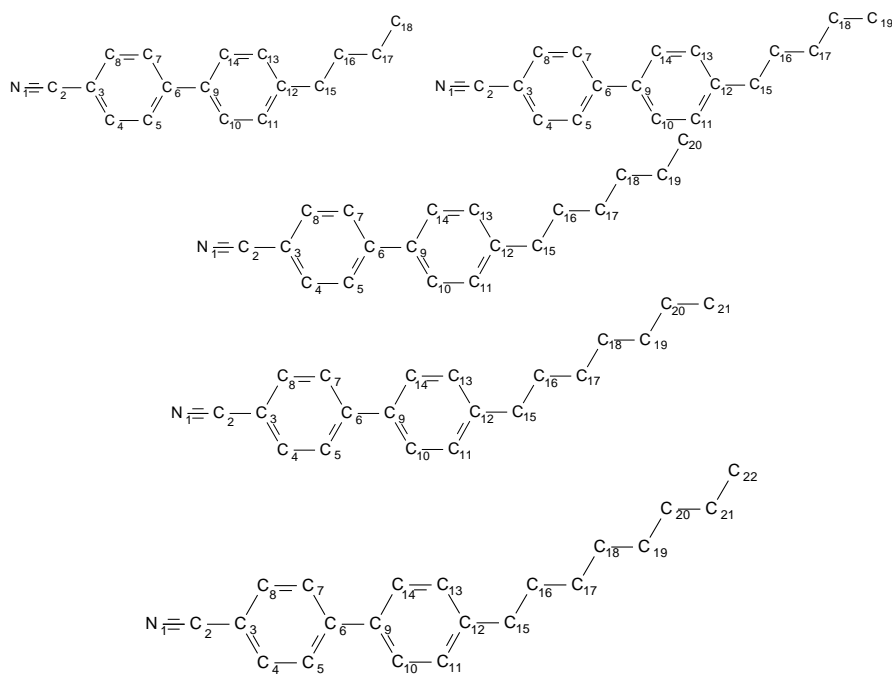


Figure B.1: Atom schemes of *n*CB series.

Table B.1: Charge parameters used expressed in  $e$  units

atom	type	4CB	5CB	6CB	7CB	8CB
N1	NY	-0.557	-0.559	-0.552	-0.556	-0.564
C2	CY	0.762	0.767	0.766	0.778	0.790
C3	CA	-0.795	-0.804	-0.850	-0.874	-0.864
C4/C8	CD	0.233	0.242	0.268	0.278	0.269
C5/C7	CD	0.181	0.156	0.136	0.134	0.137
C6	CA	-0.259	-0.211	-0.186	-0.186	-0.186
C9	CA	-0.325	-0.330	-0.306	-0.300	-0.325
C10/C14	CD	0.159	0.150	0.148	0.142	0.154
C11/C13	CD	0.184	0.196	0.182	0.187	0.186
C12	CA	-0.637	-0.640	-0.650	-0.655	-0.657
C15	C2	0.290	0.286	0.319	0.319	0.320
C16	C2	0.040	0.030	0.041	0.045	0.049
C17	C2	-0.053	-0.056	-0.101	-0.100	-0.106
C18	C3/C2	0.017	0.030	0.039	0.025	0.029
C19	C3/C2		-0.004	0.030	0.031	0.030
C20	C3/C2			-0.020	-0.001	-0.010
C21	C3/C2				-0.010	0.002
C22	C3					-0.003

Table B.2: Bond parameters used

bond	$V_0$ / kcal / mol	$r_0$
CA CY	95.9	1.42
CD CD	469.0	1.41
CA CD	469.0	1.41
CA CA	469.0	1.48
C2 C2	95.9	1.54
C2 C3	95.9	1.54
CY NY	600.0	1.16
CA C2	317.0	1.51

Table B.3: Bending parameters used

bending	$V_0$ / kcal / mol	$\theta_0$
C2 C2 C2	63.0	112.4
C2 C2 C3	62.1	114.0
C3 C2 C3	62.1	114.0
CA CA CD	85.0	120.0
CD CA CD	85.0	120.0
CA CD CD	85.0	120.0
CY CA CD	85.0	120.0
CA CY NY	79.5	180.0
CD CA C2	70.0	120.0
CA C2 C2	63.0	112.4

Table B.4: Torsional parameters used

torsion	$V_0$ / kcal / mol	$n$	$\phi_0$
* CA CD *	2.65	2	180.0
* CD CD *	5.3	2	180.0
NY CY CA CD	0.0	1	180.0
* C2 CA *	0.5	6	0.0

torsion	$V_0$ / kcal / mol	$a_0$	$b_1$	$b_2$	$b_3$	$b_4$
x C2 C2 x	4.15	0.4671	0.1616	0.0329	0.33836	0.0
x CA CA x	1.0	1.160	-0.01	-0.692	-0.001	1.064

Table B.5: Lennard–Jones parameters used

atom	description	$r_m$ / Å	$\epsilon_{ij}$ / kcal / mol	mass / u.m.a.
C2	$sp^3$ carbon with 2 hydrogens	2.035	0.0705	14.016
C3	$sp^3$ carbon with 3 hydrogens	2.050	0.105	15.024
CA	$sp^2$ carbon without hydrogens	1.95	0.0708	12.000
CD	$sp^2$ carbon with 1 hydrogen	1.95	0.0708	13.020
CY	$sp^1$ carbon	1.908	0.086	12.000
NY	$sp^1$ nitrogen	1.824	0.17	14.010



# Appendix C

## Solutes force fields

Table C.1: Formula di struttura e proprietà elettrostatiche del 5CB, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*

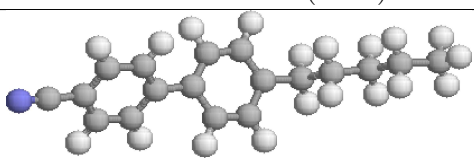
5-Cianobifenile (5CB)			UA Charges (e)			
			N1	-0.56	C14	0.15
			C2	0.77	C11	0.20
			C3	-0.80	C13	0.20
			C4	0.24	C12	-0.64
			C8	0.24	C15	0.29
			C5	0.16	C16	0.03
			C7	0.16	C17	-0.06
			C6	-0.21	C18	0.03
			C9	-0.33	C19	-0.00
			C10	0.15		
Dipolo totale			5.88 Debye			
Componenti dipolo (Debye)						
X=5.83	Y=0.58	Z=0.57				
Componenti quadrupolo (Debye Å)						
XX= -33.96	YY= 21.39	ZZ= 12.57				
XY= -4.63	XZ= -3.76	YZ= -4.10				

Table C.2: Formula di struttura e proprietà elettrostatiche dell'1,3,5-triclorobenzene, ottenute con il metodo B3LYP. 6 – 31G\*\*.

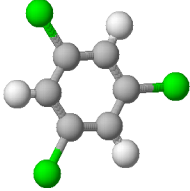
1,3,5-Triclorobenzene (TCB)			
			
Dipolo totale		0.00 Debye	
Componenti dipolo (Debye)			
X=0.00	Y= -0.00	Z=0.00	
Componenti quadrupolo (Debye Å)			
XX=0.61	YY=0.61	ZZ= -1.23	
XY=0.00	XZ=0.00	YZ=0.00	
Cariche (e)			
C1	-0.13	C7	0.07
Cl2	-0.06	H8	0.10
C3	0.07	C9	-0.12
H4	0.10	Cl10	-0.07
C5	-0.10	C11	0.10
Cl6	-0.07	H12	0.09

Table C.3: Formula di struttura e proprietà elettrostatiche del 1,3,5-tribromobenzene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*.

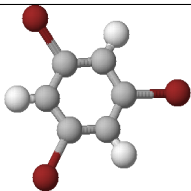
1,3,5-Tribromobenzene (TBB)			
			
Dipolo totale		0.0001 Debye	
Componenti dipolo (Debye)			
X=0.00	Y=0.00	Z=0.00	
Componenti quadrupolo (Debye Å)			
XX= 1.25	YY= 1.25	ZZ= -2.50	
XY= 0.00	XZ= 0.00	YZ= 0.00	
Cariche (e)			
C1	0.29	C7	-0.49
H2	0.07	BR8	0.00
C3	-0.39	C9	0.38
BR4	-0.01	H10	0.05
C5	0.43	C11	-0.34
H6	0.03	BR12	-0.02

Table C.4: Formula di struttura e proprietà elettrostatiche del 2,4-esadiene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*.

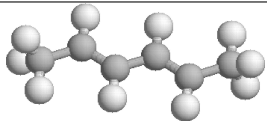
2,4-Esadiene (ESA)						
			Cariche (e)			
			C1	-0.28	C4	-0.30
			H7	0.09	C5	0.03
			H8	0.09	H12	0.19
			H9	0.09	H13	0.09
Dipolo totale			0.00	Debye		
Componenti dipolo (Debye)			C2	0.02	C6	-0.26
X=0.00	Y=0.00	Z=0.00	C3	-0.27	H14	0.08
Componenti quadrupolo (Debye Å)			H10	0.09	H15	0.08
XX= 2.79	YY= 1.21	ZZ= -4.00	H11	0.18	H16	0.08
XY=0.50	XZ=0.00	YZ=0.00				

Table C.5: Formula di struttura e proprietà elettrostatiche dell'acetonitrile, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*.


Acetonitrile (NIT)					
			Cariche (e)		
Dipolo totale			4.0530	Debye	
Componenti dipolo (Debye)			C1	-0.4688	
X= 4.0530	Y=0.0000	Z=0.0000	C2	0.4561	
Componenti quadrupolo (Debye Å)			N3	-0.5036	
XX= -2.2194	YY= 1.1098	ZZ= 1.1095	H4	0.1721	
XY=0.0001	XZ=0.0001	YZ=0.0000	H5	0.1721	
			H6	0.1721	

Table C.6: Formula di struttura e proprietà elettrostatiche dell'acetilene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*.

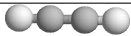
Acetilene (ENE)					
			Cariche (e)		
Dipolo totale			0.0	Debye	
Componenti dipolo (Debye)			C1	-0.2827	
X=0.0000	Y=0.0000	Z=0.0000	C2	-0.2827	
Componenti quadrupolo (Debye Å)			H3	0.2827	
XX= -2.2625	YY= -2.2625	ZZ= 0.0000	H4	0.2827	
XY=0.0000	XZ=0.0000	YZ=0.0000			

Table C.7: Formula di struttura e proprietà elettrostatiche del benzene, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G<sup>\*\*</sup>.

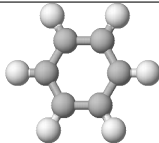
Benzene (BEN)		
		
Dipolo totale		0.0002 Debye
Componenti dipolo (Debye)		
X= -0.0002	Y=0.0000	Z=0.0002
Componenti quadrupolo (Debye Å)		
XX= 2.7263	YY= 2.7250	ZZ= -5.4513
XY= -0.0004	XZ= -0.0002	YZ= -0.0005
Cariche (e)		
C1 -0.1060	C7 -0.1060	
H2 0.1060	H8 0.1060	
C3 -0.1060	C9 -0.1060	
H4 0.1060	H10 0.1060	
C5 -0.1060	C11 -0.1060	
H6 0.1060	H12 0.1060	



Table C.8: Formula di struttura e proprietà elettrostatiche del bifenile, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*.

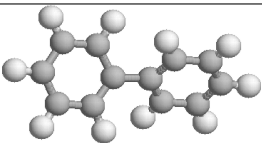
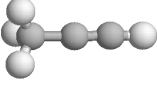
Bifenile (BIF)			
			
Dipolo totale		0.0000 Debye	
Componenti dipolo (Debye)			
X=0.0000	Y=0.0000	Z=0.0000	
Componenti quadrupolo (Debye Å)			
XX= 4.7887	YY= 3.3133	ZZ= -8.1020	
XY= -0.0001	XZ= -0.0001	YZ=0.0000	
Cariche (e)			
C1	0.0886	C12	0.0710
C2	-0.1598	C21	-0.1436
H3	0.1215	H22	0.1184
C4	-0.1277	C13	-0.1436
H5	0.1260	H14	0.1184
C6	-0.1289	C15	-0.1357
H7	0.1219	H16	0.1269
C8	-0.1277	C17	-0.1292
H9	0.1260	H18	0.1246
C10	-0.1598	C19	-0.1357
H11	0.1215	H20	0.1269

Table C.9: Formula di struttura e proprietà elettrostatiche del propino, ottenute con il metodo B3LYP. Insieme di funzioni di base: 6 – 31G\*\*.

Propino (INO)			Cariche (e)	
			C1	-0.3287
Dipolo totale			H4	0.1283
0.8543 Debye			H5	0.1283
Componenti dipolo (Debye)			H6	0.1284
X=0.8543	Y=0.0002	Z= -0.0001	C2	0.2001
Componenti quadrupolo (Debye Å)			C3	-0.5974
XX= 3.8167	YY= -1.9083	ZZ= -1.9085	H7	0.3410
XY= -0.0003	XZ= -0.0001	YZ= 0.0000		

# Appendix D

## Radial distribution functions

We report the standard radial distribution functions:

$$g_0(r) = \frac{1}{4\pi r^2 N} \langle \delta(r - r_{ij}) \rangle_{ij} \quad (\text{D.0.1})$$

$$g_1(r) = \langle \delta(r - r_{ij}) [\vec{u}_i \cdot \vec{u}_j] \rangle_{ij} \quad (\text{D.0.2})$$

$$g_2(r) = \left\langle \delta(r - r_{ij}) \left[ \frac{3(\vec{u}_i \cdot \vec{u}_j)^2 - 1}{2} \right] \right\rangle_{ij} \quad (\text{D.0.3})$$

where  $r$  is the distance from the starting center chosen,  $r_{ij}$  is the distance between the reference center  $i$  and the other center  $j$ ,  $N$  is total the number of center and  $u$  is an arbitrary direction in the molecule.



# Appendix E

## Mean Field models

The theoretical models present in literature for the orientational prevision and interpretation of solutes in liquid crystal, are generally based on Mean Field Theory [Matteo et al. 2000]. This theory substitutes the explicit solute–solvent interaction (particle–particle) with interaction the atoms of solute inside the solvent cavity with an effective mean field produced by surrounding solvent. This approach drastically reduces the number of interactions to be considered for system calculations. In the single particle models the computational cost is proportional to  $\frac{1}{2}N_{TOT}(N_{TOT} - 1)$ , where  $N_{TOT}$  is the total number of particles (solute+solvent), while in MF models the computational cost is simply proportional to  $N_{solute}$ . In the MF models is considered a potential agent on the solute molecule resulting to all the intermolecular interaction (electrostatic, dispersive, ecc. . .) without considering the specific position and orientation of solvent. After the formulation of model, it is possible to obtain the ( $U_{MF}$ ) potential depending from a series of contributions extracted from macroscopic experimental observables. This contribution becomes the constraints in the determination of orientational distribution function by entropy maximization of examined system.

### E.0.1 Maier Saupe Model

The Maier–Saupe [Straley 1974] are considered the father of MF models applied for the study of solutes dissolved in mesogens. This model introduces, as an effective potential agent on solute molecule the order parameter defined as:

$$S = \langle P_2(\beta) \rangle, \quad (\text{E.0.1})$$

where  $\beta$  is the angle between the nematic director and the principal molecular axis,  $P_2(\beta) = \langle 3(\cos^2\beta - 1)/2 \rangle$  is the averaged second order Legendre polynomial .

Supposing a couple potential between the interacting molecules  $i$  and  $j$  defined as:

$$U_{\beta_{ij}} = \epsilon P_2(\beta_{ij}) = \frac{1}{2}\epsilon(3 \cos^2 \beta_{ij} - 1) \quad (\text{E.0.2})$$

where  $\beta_{ij}$  is the angle between the molecule the principal molecular axis of the molecule  $i$  and of the molecule  $j$ , the deriving MF potential can be expressed starting from the angular distribution function as follow:

$$U_{MF} = \epsilon S P_2(\cos \beta_{ij}) \quad (\text{E.0.3})$$

In this model the parameter  $\epsilon$  are strictly related to specific solute–solvent interaction without correlation with the temperature, included in the  $S$  parameter end depending to solvent.

## E.1 MF models including the solute shape

The first models including the short range anisotropic interactions were based on the inertia tensor that implicitly take into account the shape and the molecular dimensions of the solutes. The basis idea beside this methods is that cylindrical molecules with a diameter sensibly shorter than their maximum length tend to align with this axis parallel to the phase director. In these models the solutes are modelled simply by parallelepipeds. This approximation consider the excluded volume effects and are able to describe only the interactions between identical molecules. When the system of interest are composed from different kind of species is necessary a more detailed model using for examples sphere centred on a atom groups to specify the molecular interaction. The shape anisotropy of solute interacting with the nematic field generate the orientational dependence of potential energy. Follow we show some potential energy determination in some different mean field models.

### E.1.1 C model

In the C model, proposed from Burnell and Lange [1998], the liquid crystal is represented as an elastic tube that deform itself to receive the solvent molecule. The longer axis or  $Z$  axis coincides with the nematic director. After the deformation this model take into account the elastic force induced to the solvent by the Hooke law.

$$F = -k dC(\Omega) \quad (\text{E.1.4})$$

where  $C(\Omega)$  is the circumference of solute oriented with an  $\Omega$  angle respect the nematic director and  $k$  is the deformation constant of the liquid crystalline phase. The energy related to this elastic deformation is:

$$U(\Omega) = - \int_0^{C(\Omega)} k C dC = \left[ \frac{1}{2} k C^2 \right]_0^{C(\Omega)} = \frac{1}{2} k C^2(\Omega). \quad (\text{E.1.5})$$

### E.1.2 CZ model

From result obtained to C model, is observed that the elastic constant  $k$  for long solutes tends to assume lower value than those obtained from the short ones [Burnell and Lange 1998]. To provide to this problem, is possible to add a potential term depending to the solute length as:

$$U(\Omega) = \frac{1}{2} k \left[ 1 - \varepsilon \frac{Z_p}{C(\Omega)} \right] C^2(\Omega) \quad (\text{E.1.6})$$

where  $Z_p$  is the projection of solute length, aving an orientation  $\Omega$  respect the director, along the axis of solvent tube. The first term in the equation E.1.6 is proportional to the lateral surface of a cylinder aving the same length of solute, while the second one is proportional to the area of the cylinder. Differently to the C model, in this model is described the solute–solvent interaction considering both cylindrical symmetry of solute and the orienting mean field of the liquid crystal. The possibility that, between the short range interaction, the dominant forces are depending to the superficial interaction solute–mean field of liquid crystal, had leading to develop other two independent models that explicitly involve only a single surface parameter: those model are called I and N and will be treated follow.

### E.1.3 I model

The I model, proposed from Burnell and Lange [1998], is substantially an expansion of the second term of CZ model, where instead the utilize of the full molecular projection, the circonference  $C_z(\Omega)$  is calculated in various position along the principal molecular axis oriented in the  $Z$  direction. The potential is therefore calculated by the integral of circonfereces along the  $Z$  axis by relation:

$$U(\Omega) = \frac{1}{2}k_S \int_{Z_{min}}^{Z_{max}} C_z(\Omega)dZ. \quad (\text{E.1.7})$$

The ring area  $C_z(\Omega)dZ$  is defined as  $|\sin\theta_{\hat{n}}|$  times the superficial area of molecule with integation limit  $[C, Z]$ . The physical meaning of this model is related to the representation of anisotropy of the solute–mean field interaction generated from the effective solute surface. The potential term can be written:

$$U(\Omega) = -\frac{1}{2}k_S \int dS_{\hat{n}} |\sin\theta_{\hat{n}}| \quad (\text{E.1.8})$$

where  $\hat{n}$  is the unitary versor normal to the surface element  $dS_{\hat{n}}$  and  $\theta_{\hat{n}}$  is the angle between  $\hat{n}$  and the nematic director.

### E.1.4 CI model

This model with two parameters Burnell and Lange [1998] born from the combination of C and I models both containing one parameter. Respect the two preVIOUS model separately this model provide a better agreement with NMR experimental results of fitting order parameters values. Infact the C model tends to overestimate the order parameter, while at the opposite the I model thend to underestimate it. In the combination of C and I, is obtained a sort of error copensation that give a good reproduction of experimental results. The CI model is expressed from the follow relation:

$$U(\Omega) = \frac{1}{2}k(C(\Omega))^2 - \frac{1}{2}k_S \int_{Z_{min}}^{Z_{max}} C_z(\Omega)dZ. \quad (\text{E.1.9})$$

This potential describe both the elastic distortion of the nematic phase and the anisotropic surface interaction between the solute with the mean field.



### E.1.5 Surface tensor model

In the surface tensor model [Burnell and Lange 1998] the interaction between the normal surface of solute and the mean field of liquid crystal, in analogy with the surface potential in a nematic phase, change as  $P_2(\cos\theta_{\bar{n}})$ .

It is assumed that the normal  $\bar{s}$  to every infinitesimal molecular surface element of surface  $dS$  tends to align perpendicularly to the nematic director of phase  $\bar{n}$ . This behavior is due to the energy contribution

$$dU_{st} = \varepsilon P_2(\bar{s} \cdot \bar{n}) dS \quad (\text{E.1.10})$$

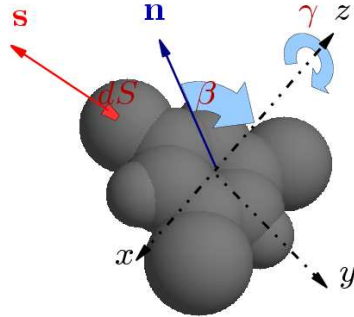
where  $P_2$  is the polynomial second order Legendre term and  $\xi$  is the intensity of interaction. Integrating over all the molecular surface  $S$  is obtained the global potential expression [Matteo et al. 2000]:

$$U_{st}^{(\beta)} = \xi \int_S P_2(\mathbf{n} \cdot \mathbf{s}) dS \quad (\text{E.1.11})$$

$$= -\xi \int_{\gamma} \int_S \left[ \frac{3}{2} (\mathbf{n} \cdot \mathbf{s})^2 - \frac{1}{2} \right] dS d\gamma. \quad (\text{E.1.12})$$

For uniaxial molecules the orientational dependence of the potential  $U_{st}^{(\beta)}$  is generally parametrized using the angle  $\Omega$  between the director and the principal molecular axis of the solute inertial frame. From the formulas it is possible to notice that both the surface tensor and the potential  $U_{st}$  are proportional to the total molecular surface, therefore for small molecules it is predicted a small short range contribution in the orientational potential.

Figure E.1: Vector and angle definitions in the surface tensor model



### E.1.6 Terzis-Photinos's model

In this model [Burnell and Lange 1998] the potential expression is written as a function of solute shape  $g^H(\Theta, r)$ , where  $r$  and  $\Theta$  are the position vector and the molecular orientation respectively. The function  $g^H(\Theta, r)$  describe a rigid body and assume the values 0 and 1 depending if the solute and the solvent molecule are overlapped or not. The orienting potential, learned from the solute molecule, is proportional to the orientational mean of its excluded volume respect to the solvent molecule as:

$$\frac{U(\Omega)}{k_B T} = \frac{N}{V} \int dr d\Omega' \tilde{f}(\Omega') [1 - g^H(\Theta, r)] \quad (\text{E.1.13})$$

where  $\frac{N}{V}$  is the density of the solvent molecules and  $\tilde{f}(\Omega')$  is their orientational distribution function and  $\Omega'$  indicate the solvent orientation relative to the nematic director.

# Appendix F

## Substitution algorithm

In the Molecular Dynamics simulation one of the main difficult is to obtain an equilibrated sample intending a molecular configuration able to represent, for the specific simulation temperature, the real phase space of the sample. The risks, during the sample equilibration, are due to the simulation of undesired metastable states that have not a correct interpretation of the molecular properties. Working with a system, like liquid crystals, where the small intermolecular interaction are the answerable to the phase macroscopic properties, these risks are again more important. Thus, in our opinion, the best way to built a solution is inserting the solute molecules inside a well equilibrated configuration for that we are reasonably sure that after the long previous equilibration time (some mounths of computer time). On the other hand, this inserption is not banal because the molecular flexibility of solvent molecules do not permit a immediate substitution of a solvent molecule with a new solute ones. Our strategy was to select inside the LC sample the solvent molecules that in a given configuration is more similar to the molecular shape of the solvent to insert. In this way the cavity obtained from the remotion of the solvent molecule is similar to the new molecule to insert. After insertion (translating and rotating the molecule in the right manner) we have to perform a small equilibration of the sample to completely stabilize the solute position and orientation inside the sample.

We have preforme the substitution by calculation of inertia tensor of solute to be inserted andall the solvent molecules inside an equilibrated configuration. The inertia tensor are computed as

$$I_{jk} = \sum_i m_i (r_i^2 \delta_{jk} - x_{i,j} x_{i,k}) \quad (\text{F.0.1})$$

where  $m_i$  is the mass of atom  $i$ ,  $\delta_{jk}$  is the Kronecker delta,  $r_i$  is the distance from the center of mass given by  $r_i = \sqrt{x_{i,a}^2 + x_{i,b}^2 + x_{i,c}^2}$  where  $a$ ,  $b$  and  $c$  are the three components.

Comparing the inertia tensor we choose the most similar solvent molecule to the new molecule, follow we remove the solvent molecule selected (generating the cavity inside the sample) and we put the new molecule here with its center of mass overlapped to the center of mass of the old solvent and rotate until the inertial axes became superimpose. After this crude insertion we performe a Monte Carlo simulation, fixing the intramolecular structure and the surrounding solvent molecules position and orientation, to obtain the energetically best position and orientation of the solute inside the solvent cavity.

This approach can be also performed if the solute dimension is similar to the solvent. If the solute is a bit bigger than solvent is possible slightly reduce (about some %) the length of internal bonds and so the molecular dimension.