

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE CHIMICHE

Ciclo XXV

Settore Concorsuale di afferenza: 03/A2 MODELLI E METODOLOGIE
PER LE SCIENZE CHIMICHE

Settore Scientifico disciplinare: CHIM/02 CHIMICA FISICA

MODELING AND SIMULATIONS OF SOME
ANISOTROPIC SOFT-MATTER SYSTEMS:
FROM BIAXIAL TO CHIRAL MATERIALS

Presentata da: LARA QUERCIAGROSSA

Coordinatore Dottorato:

Prof. ADRIANA BIGI

Relatore:

Prof. CLAUDIO ZANNONI

Correlatori:

Dr. ROBERTO BERARDI

Dr. MATTEO RICCI

Esame finale anno 2013

Contents

Abstract	3
1 Introduction	7
1.1 Aim of the work	7
1.2 Liquid Crystals	8
1.2.1 Thermotropic liquid crystals	9
1.2.2 Lyotropic liquid crystal	12
1.3 Molecular dynamics simulations	14
2 Coarse-grain modeling of dipolar biaxial nematics	17
2.1 Summary	17
2.2 Biaxial nematic materials	18
2.3 Biaxial Gay–Berne potential	20
2.4 Modeling of point dipoles	21
2.5 Characterization of the simulation results	24
2.5.1 Phase assignment protocol	24
2.6 Validation of the model	28
2.6.1 Point dipole representation	28
2.6.2 Cutoff radius	35
2.6.3 Comparison with Monte Carlo	48
2.7 Discussion	53
2.7.1 Central dipole along one molecular axis	53
2.7.2 Tilted central dipoles	59
2.7.3 Off-center dipoles	63
2.8 Conclusions	69

2.9	Appendices	70
2.9.1	Dimensionless units	70
2.9.2	Root mean square displacement	72
2.9.3	Tilt angles for layered phases	94
2.9.4	Order of phase transition	99
3	Coarse-grain modeling of DNA	107
3.1	Summary	107
3.2	DNA	109
3.2.1	Structural properties	110
3.2.2	Mechanical properties	112
3.2.3	Bases	121
3.2.4	Sugar-phosphate backbone conformation	128
3.2.5	Other possible structures	134
3.3	DNA nanotechnology	139
3.3.1	DNA-materials science	139
3.3.2	DNA-based soft phases	141
3.4	Modeling state of the art	150
3.4.1	Atomistic models	150
3.4.2	Coarse-grained models	151
3.5	Three-sites coarse-grained model	153
3.6	Validation of the model	154
3.6.1	Fitting of coarse-grained particles	154
3.6.2	Finding quaternions	164
3.6.3	Parameters: beads	178
3.6.4	Force field parametrization	180
3.7	Conclusions	205
3.8	Appendices	207
3.8.1	Structural analysis of DNA sequences used by Zanchetta group	207
3.8.2	DNA chirality index investigation	212
3.8.3	A possible application of DNA chirality index: mu- tations	221

4	Atomistic modeling of PBLG	257
4.1	Summary	257
4.2	PBLG and chiral discrimination	258
4.3	Modeling and validation	260
4.3.1	Physical and geometrical features	260
4.3.2	Parametrizing molecules	263
4.4	Results and discussion	272
4.4.1	Two poly(γ -benzyl L-glutamate) helices in dimethylformamide: parametrization and simulation	272
4.4.2	Parametrization and simulation of a system composed by a 144-residues long PBLG and heptyl butyrate in DMF	293
4.5	Conclusions	327
4.6	Appendices	328
4.6.1	CHARMM topology file	328
4.6.2	CHARMM parameter file	334
4.6.3	How to create initial sample	337
4.6.4	About NAMD input file	339
4.6.5	Chiral index analysis for various solutes	344
4.6.6	Behavior of a poly-glutamate helix in dimethylformamide	366
4.6.7	PBLG systems: a chirality index investigation	381
4.6.8	Helicoidal parameter	401
4.6.9	Docking	414
4.6.10	Docking of HEP starting from various rotated configuration	475

Abstract

We have modeled various soft-matter systems with molecular dynamics (MD) simulations.

The first topic concerns liquid crystal (LC) biaxial nematic (N_b) phases. These are LC phases that can be ordered along two directions, rather than just one as uniaxial LC, that are currently of great interest for their potentialities in realizing a new generation of fast displays. Unfortunately the synthesis of these materials has proved extremely difficult and there is a need for understanding the effects that changing in a rational way the features of candidate molecules has on the phase behaviour. Here we have investigated the phase organization of a relatively simple model where each molecule is represented by an attractive-repulsive biaxial Gay-Berne (GB) mesogen and we have considered the effects of the orientation, strength and position of a molecular dipole on the phase behavior. We have observed that for systems with a central dipole, nematic biaxial phases disappear when increasing dipole strength, while for systems characterized by an offset dipole, the N_b phase is stabilized at very low temperatures. This kind of results should be important for helping the design of new molecules showing N_b phases at room temperature.

In a second project, we are developing a DNA coarse-grained (CG) model to simulate DNA LC phases, in which sugar and phosphate group are represented with Lennard-Jones spheres, while bases with GB ellipsoids. This is important in view of the current interest on DNA as a nanomaterial, that once prepared in a suitable sequence can be easily replicated using current PCR technology, and of the liquid crystal phases formed by short DNA segments. For all these applications the use of fully atomistic for the various nucleic acid sequences is unfeasible, thus the interest of many groups in developing a sufficiently realistic CG representation. For our model, we have obtained shape, position and orientation parameters for each type of bead, so as to be able to best reproduce the atomistic structure of a B-DNA

helix. Using atomistic simulations results, we have completed a first parametrization of the force field terms, accounting for bonded (bonds, angles and dihedrals) and non-bonded interactions (H-bond and stacking). We are currently validating the model, by investigating stability and melting temperature of the double strains obtained for various sequences.

Finally, in a third project, we aim to explain the mechanism of enantiomeric discrimination due to the presence of a chiral helix of a functionalized protein, poly(γ -benzyl L-glutamate) (PBLG), in solution of dimethylformamide (DMF), interacting with chiral or pro-chiral molecules (in practice we have chosen heptyl butyrate, HEP). We have first tuned an atomistic force field (AMBER), so as to properly describe our system, in conditions of temperature and concentration similar to the experimental ones. We have observed that DMF and HEP molecules solvate uniformly the PBLG helix, but the pro-chiral solute is on average found nearer to the helix with respect to the DMF. The solvent presents a faster isotropic diffusion, twice as HEP, also indicating a stronger interaction of the solute with the helix. We are currently extending the trajectories of these very demanding simulations (53333 atoms) to improve the statistics of the orientational and chiral observables.

Chapter 1

Introduction

1.1 Aim of the work

We modeled and simulated various soft-matter systems.

First of all, we studied the phase organization of dipolar biaxial Gay-Berne (GB) mesogens and considered the effects of orientation, strength and position of the molecular dipole on the phase behavior and particularly on the possibility of obtaining a biaxial nematic phases.

Next, we developed a three-sites per nucleotide model, that represents sugar and phosphate groups with sphere and bases with ellipsoids. After having parametrized these beads (shape and orientation), atomistic data have been analyzed to develop a force field. This model is currently under construction: further tests are needed.

The last topic concerns the atomistic study of a system composed by an α -helix of a functionalized protein that is able to discriminate enantiomers. Our aim is to explain the details of the atomistic mechanism of this process with MD simulations.

1.2 Liquid Crystals

Liquid crystals (LC) are fluid phases characterized by a partial orientational and positional ordering, classified as mesophases, between the isotropic liquid state and the crystalline packing [1] (Figure 1.1).

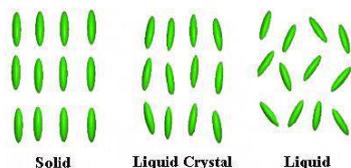


Figure 1.1: Comparison between solid, liquid crystal and liquid phases.

Hence, LC phases can flow like a liquid but their molecules are typically orientated in crystal-like fashions. Various types of LC phases have been experimentally described, thanks to their anisotropic optical properties. Liquid crystal phase can be divided as follow:

- thermotropic phases: inorganic and organic molecules system in which phase transitions are induced by temperature changes,
- lyotropic phases: molecular system in which phase transitions are induced by solvent concentration,
- metallotropic phases: organic and inorganic system in which transition depends on temperature, concentration and inorganic-organic ratio.

Besides their use in electronic displays, natural LC phases are found in many proteins system, cell membranes, soap and other detergents, some virus like the tobacco mosaic virus.

1.2.1 Thermotropic liquid crystals

1.2.1.1 Nematic phase

One of the most common liquid crystal phases is called nematic phase, where molecules have no positional order, but only long-range orientational order. The molecules can move and their center of mass are randomly distributed, as in a liquid, but all particles belonging to the same domain have the same direction. Usually nematic phases are uniaxial: it show a preferred axis of orientation, the major one, while the other two are equivalent. Such molecules can be represented with cylinders. Some biaxial nematic phases have also been observed: they are characterized by a secondary preferred direction of orientation [2].

The fluidity of nematic phases is similar to that of an isotropic liquid, but they can easily be aligned with an magnetic or electric field. An aligned nematic system has the same properties of an uniaxial crystal. Figure 1.2 shows the organization of a nematic phase where mesogens are directed represented as ellipsoids directed along their preferential alignment direction, arranged in any particular order.



Figure 1.2: Nematic liquid crystal.

Hence, these systems can be aligned along two mutually orthogonal directions of preferred alignment, called “directors”, keeping an uniform distribution of centers of mass. The principal director, \mathbf{n} , is typical of uniaxial nematic, while the secondary one, \mathbf{m} is specific for biaxial system, behaving like trirefrangent materials (i.e. system with three different refraction in-

dexes) [3]. Biaxial nematics are liquid crystals spatially homogeneous with three distinct optical axes; on the contrary, a uniaxial nematics has only one preferred directional axes, around which is rotationally symmetric.

1.2.1.2 Smectic phase

The smectic phase (Figure 1.3) is found at lower temperatures with respect to that of nematics and is formed of planes that can slide on each other. These phases have positional order along a preferred direction, besides directional order. The smectic phase A comprehend only molecules orientated perpendicularly to the other layers, while in the smectic C phase molecules are tilted with respect to the director [4, 5].

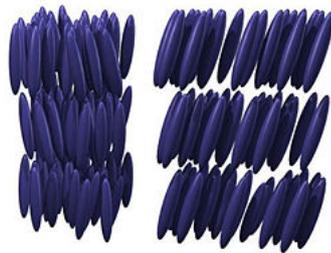


Figure 1.3: Smectic A (*left* and C *left* liquid crystal).

1.2.1.3 Chiral phase

Cholesteric phase (Figure 1.4) is a chiral phase, in which directors rotate along the phase. The twist angle between adjacent molecules is caused by the asymmetric packing, that leads to a chiral ordering. The smectic C^* , where $*$ denotes the chirality, shows molecules with a positional order and a layered structure, as normal smectic phases, but molecules are tilted of a finite angle with respect to the normal plane. A kind of spiral that rotate around molecules axes of the layer is formed [4, 6].

The chiral inclination, p , refers to the distance of a complete rotation of 360° . It is worth noting that the structure of a chiral nematic phase repeats every $p/2$, since directors at 0° and $\pm 180^\circ$ are equivalent. The p inclination



Figure 1.4: Chiral liquid crystal.

usually changes when temperature is altered or when other liquid crystal molecules are added (for example, if a chiral phase is doped with a chiral material the obtained liquid crystal will not be chiral). These systems have unique optical properties, like selective reflection.

1.2.1.4 Blue phases

The blue phases (Figure 1.5) are particular liquid crystal phases existing in a very small range of temperature, between the chiral nematic and isotropic liquid phases. Such phases have a regular three-dimensional cubic structures with defects on a periodic grid of hundred of nanometers. They exhibit a Bragg selective reflection in the range of visible light wavelength, corresponding to cubic grid.

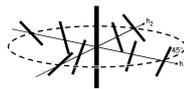


Figure 1.5: Blue phase.

1.2.1.5 Discotic phases

A disk-like mesogens can be oriented in layers, in a phase called discotic nematic (Figure 1.6). If groups of disks assume a columnar conformations, the phase will be called columnar discotic. The columns can arrange in rectangular or hexagonal ways. Chiral discotic phases also exist.



Figure 1.6: Discotic phase.

1.2.2 Lyotropic liquid crystal

A lyotropic liquid crystal has two components with liquid crystal properties only in determined ranges. In lyotropic phases, the solvent molecules fill the space around the compound to give fluidity to the system. On the contrary of thermotropic liquid crystals, the lyotropic one have a high degree of freedom for what concern the concentration, that give them the ability of form different phases.

A molecule composed by two immiscible parts, a hydrophobic and a hydrophilic one, is called amphiphilic. Most of these molecules show a sequence of lyotropic liquid crystal phases. These structures are formed through microsegregation of two incompatible nano-scale components. The soap is an example of everyday life of lyotropic liquid crystal. Water and other solvent content changes the way structure self-assembly. At very low amphiphilic concentration, the molecules scatter randomly without any order. At slightly higher concentrations, amphiphilic molecules will self-assembly in micelles and vesicles. This process hide hydrophobic tail of the amphiphilic molecules in the micelles, exhibiting only hydrophilic surface to aqueous solution. At high concentration, these clusters become ordered. The columnar hexagonal phase is typical of long cylindrical amphiphilic molecules, with a hydrophobic surface, arranging spontaneously in a hexagonal grid. This phase is called “half-soap”. At even higher concentration. a lamellar phase, called “clean soap” can be formed, where widespread sheets of amphiphilic are separated by thin water layers. For certain systems, the cubic phase can exist in hexagonal an lamellar phases,

in which a denser grid is formed. These spheres can be connected to each other, to arrange in a bicontinuous cubic phase.

Self-assembly objects are usually spherical, as for micelles, but can also be discotic (bicelles), stick-like and biaxial (micelles with three distinct axes). Some system, at high concentration, is characterized by inverse phases. For example, hexagonal inverse columnar phases (water column encapsulated by amphiphilic molecules) or inverse micellar phase (a liquid crystal mass with aqueous spherical cavities) can also exist. Usually, increasing the concentration of amphiphilic molecules the following phases are obtained:

- a discontinuous cubic phase (micellar cubic phase),
- hexagonal (columnar) phase,
- lamellar phase,
- bicontinuous cubic phase,
- inverse hexagonal phase,
- inverse hexagonal columnar phase,
- cubic (micellar) inverse phases.

Also in the same phase, the structures can vary depending on concentration: for example, in lamellar phases, the distance between layers increases with solvent volume. Since lyotropic liquid crystals rely on a counterbalance of intermolecular interaction, it is more difficult to analyze their properties and structures. Phases and similar characteristics are found in immiscible block copolymers.

1.3 Molecular dynamics simulations

Molecular dynamics is a simulation method based on solving Newton motion equations, predicting position of molecules depending on time and computing their average properties. First of all, initial conditions should be defined. Hence, forces are computed, using the force field. Solving Newton equation and controlling temperature and/or pressure, molecules trajectories are computed and saved. Thermodynamical and structural average properties can now be computed.

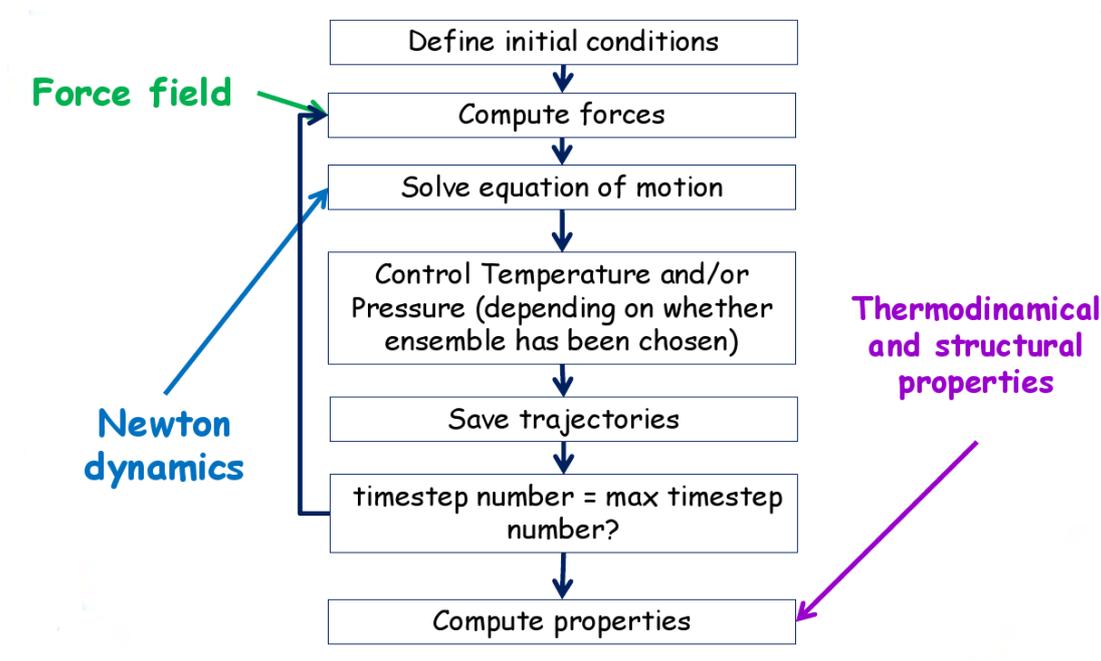


Figure 1.7: Molecular dynamics flow sheet.

Chapter 2

Coarse–grain modeling of dipolar biaxial nematics

2.1 Summary

We have investigated the phase organization of dipolar biaxial Gay–Berne (GB) mesogens and considered the effects of orientation, strength and position of the molecular dipole on the phase behavior and particularly on the possibility of obtaining a biaxial nematic (N_b). These mesophases are difficult to obtain experimentally but they are promising materials for faster switching time displays. We have mapped, using molecular dynamics (MD) simulations, the boundaries of the N_b phase for the coarse–grained models obtained adding a point dipole to an apolar biaxial mesogen that already yields a biaxial nematic, finding that the offset and the magnitude of the embedded point dipole are the parameters with the strongest influence on the mesogenic properties. For central dipoles, we find that the N_b organization is stable only for mesogens with relatively weak moments and that the N_b phase disappears if electrostatic interactions become comparable in magnitude with dispersion interactions. However, offset and skewed dipoles destabilize the smectic in favor of the N_b phase which becomes favored even at relatively high temperatures and for large dipole moments. These results show how specifically designed dipolar interactions can be used to influence the formation of N_b phases.

2.2 Biaxial nematic materials

Biaxial nematics (N_b) are a topic of currently great interest in liquid crystals (LC) [7, 8, 9, 10], since these anisotropic fluids have, with respect to the usual uniaxial variety, the characteristic of having two, rather than just one, directions of preferential alignment (directors) that can be independently controlled by external fields and surface treatments. A feature of particular interest for LC display technology is the expected significantly faster switching time of the secondary director [11]. Biaxial nematics should in principle form as easily as uniaxial ones and indeed their existence was predicted 40 years ago by Mean Field theory [12] and confirmed by lattice [13] and off-lattice [14, 15] computer simulations [10]. However, reality has revealed to be quite different [16, 9, 8]: while biaxial nematics have been found at an early stage in lyotropic [12], polymeric [17, 18, 19] and elastomeric LC [17], low molar mass thermotropic materials have defied synthetic chemistry until very recently, when bent core [20, 21], tetrapods [22] and similar mesogens have convincingly shown biaxial nematic phases.

Unfortunately, these important experimental findings are still not supported by systematic structure-phase property guidelines that would be very important in trying to bring the region of existence of these new LC materials within practically usable ranges of temperature, viscosity, etc.

Computer simulations can be particularly useful in this respect [10] as they allow a specific investigation of the effect that selected molecular features might have on the phase behavior. Here, we study, with this aim in mind, the effects of a molecular dipole positioned either in the center of the mesogen or off-set and at different orientations with respect to the molecular axes. The electric dipole is a particular interesting element in chemical design [23, 24, 25, 26, 27, 28] since, by suitable functional substitution, it can be placed at selected positions and orientations in an existing mesogen, controlling also, to some extent, its strength [29]. On the other hand, dipoles are also quite challenging and difficult to deal with in terms of predicting their collective effect [23], making common sense rather useless.

For instance, we have shown in the past that simply shifting a longitudinal molecular dipole from the central to a near-terminal position in an uniaxial mesogen can dramatically change the resulting smectic phase from SmA to striped $\text{Sm}\tilde{\text{A}}$ organization [23].

The chapter is organized as follow: first the biaxial dipolar Gay-Berne (GB) model is described and technical details of the MD simulations are given. Then in the following sections, the results are presented and discussed, while in the final section we draw some conclusion on the use of dipoles to tune phase biaxiality.

2.3 Biaxial Gay–Berne potential

We consider the mesogenic molecules as rigid ellipsoidal particles characterized by positional and orientational degrees of freedom (off-lattice) and interacting with a pair potential that is the sum of two terms: a Gay–Berne (GB) pair energy U_{GB} and an electrostatic contribution:

$$U_{12} = U_{12}^{GB} + U_{12}^{el}. \quad (2.1)$$

The first interaction term is the attractive–repulsive energy for a pair of rigid biaxial ellipsoids [30, 31, 10]:

$$U_{12}^{GB}(\mathbf{r}, \omega_1, \omega_2) = 4\epsilon_0 \epsilon(\mathbf{r}, \omega_1, \omega_2) [u^{12}(\mathbf{r}, \omega_1, \omega_2) - u^6(\mathbf{r}, \omega_1, \omega_2)]. \quad (2.2)$$

Here $\mathbf{r} \equiv \mathbf{r}_2 - \mathbf{r}_1$ is the center–center vector, whose modulus is the intermolecular distance r , while the orientations of the two interacting particles ω_i are given by a set of three Euler angles $(\alpha_i, \beta_i, \gamma_i)$ or equivalently by a quaternion (q_i) [10]. The function $u(\mathbf{r}, \omega_1, \omega_2) = \sigma_c / (r - \sigma(\mathbf{r}, \omega_1, \omega_2) + \sigma_c)$ contains the anisotropic contact term $\sigma(\mathbf{r}, \omega_1, \omega_2)$ which approximates the geometrical “*contact distance*” between two ellipsoids and depends on the axes lengths σ_x , σ_y and σ_z . The interaction term $\epsilon(\mathbf{r}, \omega_1, \omega_2)$ defines the angular dependent potential well depth and also depends on the interaction parameters ϵ_x , ϵ_y and ϵ_z , which are directly related to the potential well depths for two GB particles approaching with fixed parallel orientations along the three Cartesian directions. Using the standard notation, the molecular units of mass, length and energy are m_0 , σ_0 and ϵ_0 (not to be confused with the permittivity of vacuum).

2.4 Modeling of point dipoles

In each Gay–Berne particle, embedding two electric charges of same magnitude and opposite sign positioned at close distance (with respect to the molecular dimensions) allows to model the interaction effects due to a point dipole. For instance, for a central dipole the opposite $\pm q$ charges are symmetrically positioned at $\pm \mathbf{r}_q$ with respect to the center of mass and at a rather close distance $d = \|\mathbf{r}_+ - \mathbf{r}_-\| \ll \sigma_i$, ($d = 0.2 \sigma_0 = 2r_q$) thus approximating an electric point dipole $\mu = 2qd$. The electrostatic interaction between two particles 1, 2 is (2.1):

$$U_{12}^{el} = \sum_{a \in 1, b \in 2} \frac{q_a q_b}{r_{ab}} \quad (2.3)$$

where the sum runs over the charges q_a of particle 1 and q_b of particle 2 (at distance r_{ab}) and we use here a CGS notation for electrostatic energy, see reference [32] for further detail on the electric potential. For the computation of the essentially dipolar electrostatic interactions between mesogenic particles, we have relied on the reaction–field like method due to Tironi *et al.* [32]. The suitability of this approach for our type of samples has been checked by comparing the MD results of selected state points with those obtained independently from a full Ewald computation [23].

We are interested in studying effects of dipoles on biaxial mesogens that can give a biaxial nematic N_b phase to start with, so we have chosen as a reference the dipole–less model already studied in ref. [15], which was shown there to yield such a phase.

More specifically, we have chosen each biaxial ellipsoidal GB particle to have the following parametrization: unitary mass $m = m_0$, axes $\sigma_x = 1.4 \sigma_0$, $\sigma_y = 0.714 \sigma_0$, $\sigma_z = 3 \sigma_0$, $\sigma_c = 0.714 \sigma_0$ and interaction parameters $\epsilon_x = 1.7 \epsilon_0$, $\epsilon_y = 1 \epsilon_0$ and $\epsilon_z = 0.2 \epsilon_0$. All quantities have been employed in dimensionless form: e.g. temperature $T^* \equiv T/T_0 = T/(k_B^{-1} \epsilon_0)$, pressure $P^* \equiv P/P_0 = P/(\epsilon_0^{-1} \sigma_0^3)$, electrostatic charge $q^* \equiv q/q_0 = q/(\epsilon_0 \sigma_0)^{1/2}$ and dipole moment $\mu^* \equiv \mu/\mu_0 \equiv \mu/(\epsilon_0 \sigma_0^3)^{1/2}$, where μ_0 should not be confused with the permittivity of vacuum.

Using these definitions and taking as units values for length and energy $\sigma_0 = 0.5$ nm and $\varepsilon_0 = 1.381 \cdot 10^{-21}$ J, typical of a low molar mass mesogen such as 8CB, we have that $T^* = 1$ corresponds to $T = 100$ K, $q^* = 1$ corresponding to $q_0 = 8.767 \cdot 10^{-21}$ C and $\mu^* = 1$ to $\mu_0 = 1.314$ D = $4.38303096 \cdot 10^{-30}$ C m.

Table 4.24 summarizes positions, dipole modules and symmetry of the biaxial GB particles studied both with central and shifted dipoles.

Molecular dynamics simulations have been run with an in house written quaternion based code [33, 34], using a time step $\Delta t^* = 0.001$ at a chosen constant dimensionless pressure and temperature maintained by a Berendsen barostat and thermostat [35].

We have considered an orthorhombic sample box, endowed with periodic boundary conditions, whose sides were allowed to change independently, so as to accommodate more easily the smectic structures formed at the lowest temperatures. The systems considered here are composed of $N = 1024$ identical biaxial GB particles.

Each system has been studied at $P^* = 8$ and for reduced temperatures ranging from $T^* = 2.6$ to $T^* = 3.6$, starting with the lowest dipole module case. Each subsequent case with progressively increasing dipole moment has then been performed as a new cooling down sequence started from the equilibrated isotropic phase of the previous system.

position	Electric dipole			Electric charges		Symmetry
	label	alignment	orientation	position	magnitude (dipole)	
Centered	<i>cx</i>	$\boldsymbol{\mu} \parallel \mathbf{x}$, on-axis	$(0^\circ, 90^\circ)$	$\mathbf{r}_+ = -\mathbf{r}_- = (0.1, 0, 0)$	$ q = 2.5, 5, 7.5, 10, 15$	C_{2v}
	<i>cy</i>	$\boldsymbol{\mu} \parallel \mathbf{y}$, on-axis	$(90^\circ, 90^\circ)$	$\mathbf{r}_+ = -\mathbf{r}_- = (0, 0.1, 0)$		C_{2v}
	<i>cz</i>	$\boldsymbol{\mu} \parallel \mathbf{z}$, on-axis	$(0^\circ, 0^\circ)$	$\mathbf{r}_+ = -\mathbf{r}_- = (0, 0, 0.1)$		C_{2v}
	<i>ct1</i>	tilted #1, off-axis	$(0^\circ, 60^\circ)$	$\mathbf{r}_+ = -\mathbf{r}_- = (0.087, 0.000, 0.05)$	$(\mu = 0.5, 1, 1.5, 2, 3)$	C_s
	<i>ct2</i>	tilted #2, off-axis	$(30^\circ, 60^\circ)$	$\mathbf{r}_+ = -\mathbf{r}_- = (0.076, 0.044, 0.05)$		C_1
Offset	<i>ox</i>	$\boldsymbol{\mu} \parallel \mathbf{x}$, on-axis	$(0^\circ, 90^\circ)$	$\mathbf{r}_+ = (0.5, 0, 0)$ $\mathbf{r}_- = (0.3, 0, 0)$	$ q = 2.5, 5, 7.5, 10, 15$	C_s
	<i>oz</i>	$\boldsymbol{\mu} \parallel \mathbf{z}$, on-axis	$(0^\circ, 0^\circ)$	$\mathbf{r}_+ = (0, 0, 0.5)$ $\mathbf{r}_- = (0, 0, 0.3)$		C_s
	<i>ot</i>	tilted, off-axis	$(30^\circ, 50^\circ)$	$\mathbf{r}_+ = (0.091, 0.059, 0.007)$ $\mathbf{r}_- = (0.033, 0.009, 0.085)$	$(\mu = 0.5, 1, 1.5, 2, 3)$	C_1

Table 2.1: Details of the electric dipole moments embedded in the biaxial GB particles studied in this work. For every case, we give the positions (\mathbf{r}_+ , \mathbf{r}_-) and magnitude ($|q|$) of the two opposite charges (q_+ , q_-) with respect to the fixed molecular frame. The orientation (as spherical angles ϕ , θ), alignment and module (μ) of the dipole, and the overall symmetry of the dipolar GB particle are also given. Angles are measured in degrees, positions and distances in σ_0 units, charges and dipole moments in units of $q_0 = (\varepsilon_0\sigma_0)^{1/2}$ and $\mu_0 = (\varepsilon_0\sigma_0^3)^{1/2}$ units, respectively.

2.5 Characterization of the simulation results

2.5.1 Phase assignment protocol

The protocol we have used for assigning the resulting phase of each MD sample is based on a combined analysis of order parameters, positional correlation functions and visual inspection of the final equilibrated configuration at each temperature [36, 37].

The average second rank orientational order parameters, which are the first terms of the expansion of the single-particle orientational distribution in a basis set of symmetrized Wigner rotation matrices $R_{m,n}^2$, have been computed from the eigenvalues of Cartesian ordering matrices using now standard algorithm introduced in [38, 39]. In particular, we compute the standard order parameter:

$$\langle R_{0,0}^2 \rangle = \left\langle \frac{3}{2}(\mathbf{z} \cdot \mathbf{n})^2 - \frac{1}{2} \right\rangle, \quad (2.4)$$

which is the Maier-Saupe order parameter $\langle R_{00}^2 \rangle \equiv \langle P_2 \rangle$, ranging between 0 (isotropic) and 1 (completely aligned), typically used for characterizing uniaxial phases. In addition, the most telling order parameter for biaxial liquid crystals is [40, 38]

$$\langle R_{2,2}^2 \rangle = \left\langle \frac{1}{4} \left[(\mathbf{x} \cdot \mathbf{l})^2 - (\mathbf{x} \cdot \mathbf{m})^2 + (\mathbf{y} \cdot \mathbf{l})^2 - (\mathbf{y} \cdot \mathbf{m})^2 \right] \right\rangle, \quad (2.5)$$

where \mathbf{x} , \mathbf{y} and \mathbf{z} are the three molecular frame axes and \mathbf{l} , \mathbf{m} and \mathbf{n} are the overall mesophase directors, with \mathbf{n} the principal director, \mathbf{m} the transversal biaxial one perpendicular to \mathbf{n} and $\mathbf{l} = \mathbf{m} \times \mathbf{n}$, while $\langle \dots \rangle$ are ensemble averages. The order parameter $\langle R_{22}^2 \rangle$ unambiguously identifies biaxial phases [38], where also the \mathbf{x} and \mathbf{y} axes are macroscopically aligned, whenever it takes values larger than 0 (uniaxial) up to 1/2 (completely biaxial). In the framework of our MD simulations, with samples of relatively small size, we always observe uniformly ordered monodomains along some arbitrary direction, rather than polydomains. This is to some

extent advantageous, but the finite size of the samples also reduces our ability of observing first order transitions and the statistical fluctuations, proportional to $1/\sqrt{N}$, also affect the lowest values of the order parameters attainable in the isotropic phase. Taking into account these effects, we have arbitrarily identified the formation of a nematic from an isotropic sample if $\langle R_{0,0}^2 \rangle \geq 0.3$ (Figure 2.1–A) and, similarly, the spontaneous onset of biaxial phases with ordering along a direction \mathbf{m} perpendicular to \mathbf{n} , if $\langle R_{2,2}^2 \rangle \geq 0.1$ (Figure 2.1–B). It should be noted that even though the observation of these purely orientational order parameter being larger than the two thresholds is a necessary condition to assign uniaxial and biaxial phases, both nematic and smectic phases will match these criteria, so further tests are necessary.

To discriminate the nematic phases from the layered structures typical of smectic we have examined positional pair correlation functions. The first one we have computed is the standard radial pair correlation $g_0(r)$, giving the average probability of finding the center of mass of any two molecules separated by a distance r , using as reference the value expected from an uniform distribution

$$g_0(r) = \langle \delta(r - r_{12}) \rangle_{12} / (4\pi r^2 \rho), \quad (2.6)$$

where ρ is the number density of the sample and $\langle \delta(r - r_{12}) \rangle_{12}$ denotes an ensemble average with respect to the pair distribution function. In MD simulations this function is calculated as a discrete histogram, and smectic phases can be identified if a characteristic sequence of well defined maxima and minima in $g_0(r)$ reveals the presence of strong positional correlations extending over the first shell of neighboring molecules. As an example, Figure 2.2 shows the typical $g_0(r)$ profiles for a nematic and a smectic phase formed by our GB ellipsoids. The projection $g_{\parallel}(r)$ of the pair correlation along the principal director \mathbf{n} , and the second rank anisotropy $g_2^+(r)$ of the radial correlation function are also relevant for detecting the formation of layered structures in the MD simulation box (either as a monodomain or within local clusters). These two correlation functions are defined as:

$$g_{\parallel}(r) = \frac{1}{\pi R^2 \rho} \langle \delta(r - \mathbf{r}_{12} \cdot \mathbf{n}) \rangle_{12}, \quad (2.7)$$

$$g_2^+(r) = \frac{1}{4\pi r^2 \rho} \langle \delta(r - r_{12}) P_2(\mathbf{r}_{12} \cdot \mathbf{n}) \rangle_{12}, \quad (2.8)$$

where R is here the constant radius of a cylindrical sampling region perpendicular to the principal director \mathbf{n} .

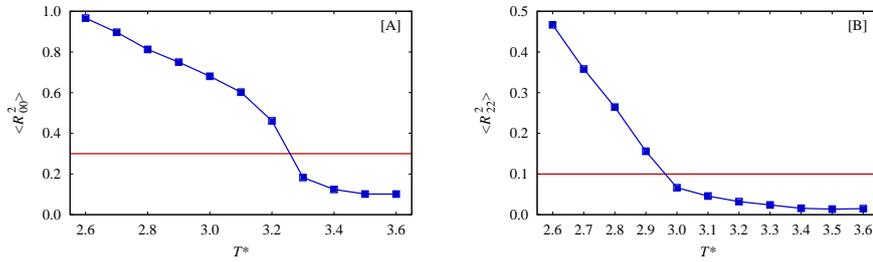


Figure 2.1: The average orientational uniaxial $\langle R_{00}^2 \rangle$ (plate A) and biaxial $\langle R_{22}^2 \rangle$ (plate B) order parameters for a bulk system of biaxial GB particles with an embedded electric dipole of dimensionless module $\mu^* = \mu/\mu_0 = \mu/(\epsilon_0 \sigma_0^3)^{1/2} = 0.5$ positioned at the ellipsoid center of mass and tilted with respect to the \mathbf{z} -axis (model *ct1*, see Table 4.24 for details). The thresholds used for identifying isotropic–nematic ($\langle R_{00}^2 \rangle < 0.3$) and uniaxial–biaxial ($\langle R_{22}^2 \rangle > 0.1$) phase transitions are shown as solid horizontal lines.

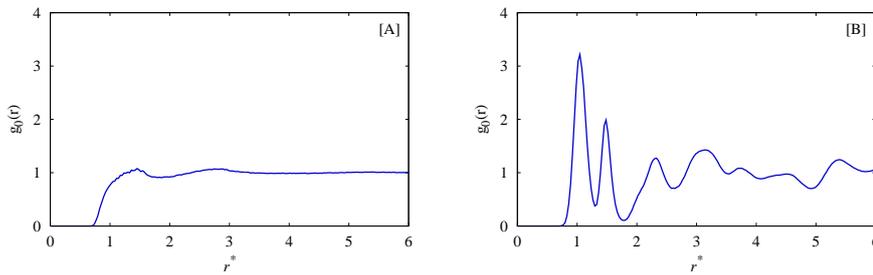


Figure 2.2: Comparison of the average radial correlation functions for a biaxial nematic at $T^* = 2.9$ (plate A) and a biaxial smectic at $T^* = 2.6$ (plate B) for the bulk system of Figure 2.1 (system *ct1*).

We have confirmed that the nematic phases are fluid-like and not glassy by checking, as in other cases [41, 42] from the time dependence of the mean square displacement that particles diffuse away from an arbitrarily chosen initial origin 2.9.2.

By computing velocity correlation functions and translational diffusion coefficients for the various nematic phases of our dipolar GB model, we have verified that in all cases the nematic organizations have diffusion coefficients which are roughly 20–30% lower than those measured for the corresponding isotropic liquid, while for the layered, lower temperature, organizations these values decrease by more than one order of magnitude [43]. The analyses of the correlation functions have also been supported with a direct visual inspection of the equilibrium molecular configurations (see for instance the snapshots of Figure 2.28, where each elongated GB particle is color coded according to its orientation with respect to the director [37]) to gain further insight into the structure of the various fluid phases. As a side remark we notice that the dipolar GB particles behaved in all cases as calamitic rod-like mesogens since the molecular axis giving the highest degree of orientational ordering was in all cases the \mathbf{z} one corresponding to the largest σ_i length. In addition, we did not observe columnar organizations.

2.6 Validation of the model

2.6.1 Point dipole representation

In order to verify if two charges placed close to each other onto a particle are able to emulate a point dipole, we place those charges at a small distance compared to the particle shortest axis (i.e. the smallest contact distance). Here we will call these new systems “nearer charges” and systems studied below (in “Results and discussion” section) “new code” (since simulations were carried using the last version of MDGB). The systems analyzed have central dipole aligned along the x axis (cx) with various dipole moments $\mu^* = 0.5$, $\mu^* = 1$ and $\mu^* = 2$.

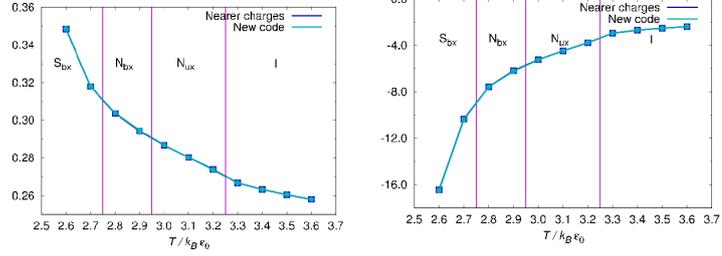
2.6.1.1 $\mu^* = 0.5$

This system is composed by molecules having two charges placed in positions shown in Table 2.2, with a dipole moment of $\mu^* = 0.5$.

Charge	x	y	z
+5.0	+0.05	0	0
-5.0	-0.05	0	0

Table 2.2: Positions of two charges acting as a central transversal dipole, $\mu^* = 0.5$, using charges placed nearer than systems studied.

The following graphs (Figure 2.3) showing a comparison between results obtained getting the charges closer and the charge distance we use for production runs.



(a) Number density.

(b) Potential energy.

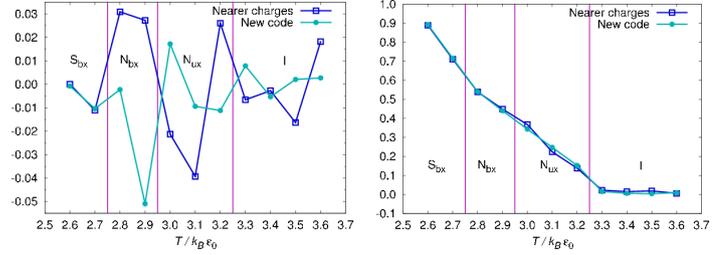
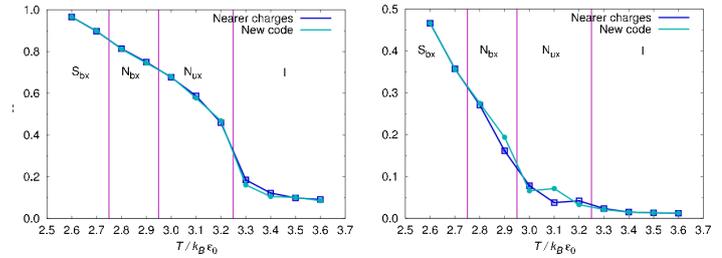
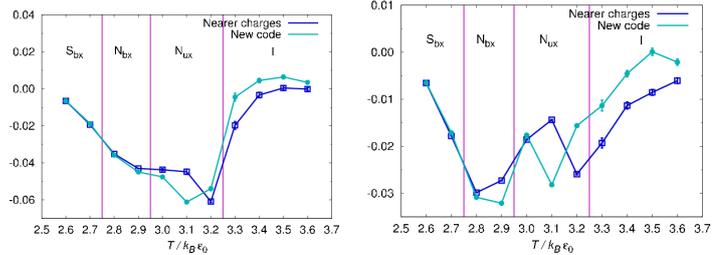
(c) $\langle P_1 \rangle$ (d) $\langle P_4 \rangle$ (e) $\langle R_{00}^2 \rangle$ (f) $\langle R_{22}^2 \rangle$ (g) $\langle R_{02}^2 \rangle$ (h) $\langle R_{20}^2 \rangle$

Figure 2.3: Physical quantities of a system with central transversal dipole ($\mu^* = 0.5$) on its molecules (charges placed nearer): Comparison with results obtained previously.

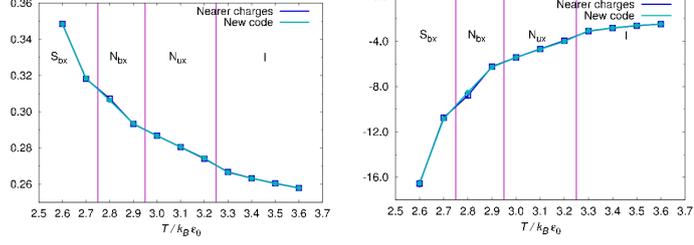
2.6.1.2 $\mu^* = 1$

This system is composed by molecules having two charges placed in positions shown in Table 2.3, with a dipole moment of $\mu^* = 1$.

Charge	x	y	z
+10.0	+0.05	0	0
-10.0	-0.05	0	0

Table 2.3: Positions of two charges acting as a central transversal dipole, $\mu^* = 1$, using charges placed nearer than the ones on previous system.

The following graphs (Figure 2.4) showing a comparison between results obtained getting the charges closer and the charge distance we use for production runs.



(a) Number density.

(b) Potential energy.

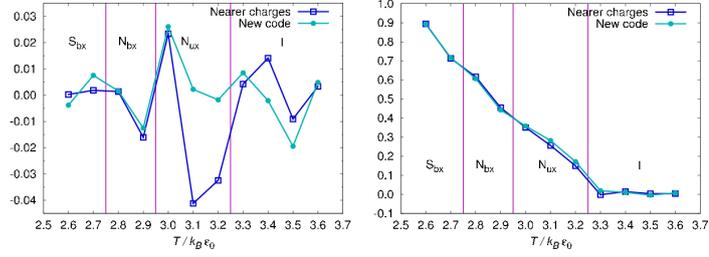
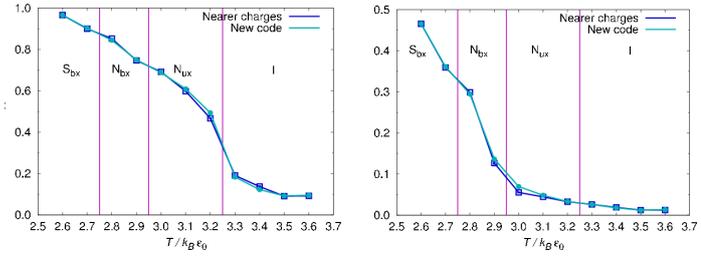
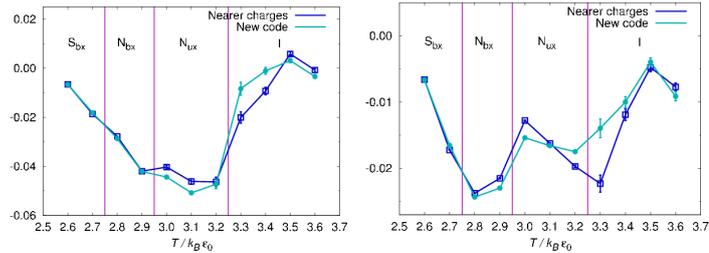
(c) $\langle P_1 \rangle$ (d) $\langle P_4 \rangle$ (e) $\langle R_{00}^2 \rangle$ (f) $\langle R_{22}^2 \rangle$ (g) $\langle R_{02}^2 \rangle$ (h) $\langle R_{20}^2 \rangle$

Figure 2.4: Physical quantities of a system with central transversal dipole ($\mu^* = 1$) on its molecules (charges placed nearer): Comparison with results obtained previously.

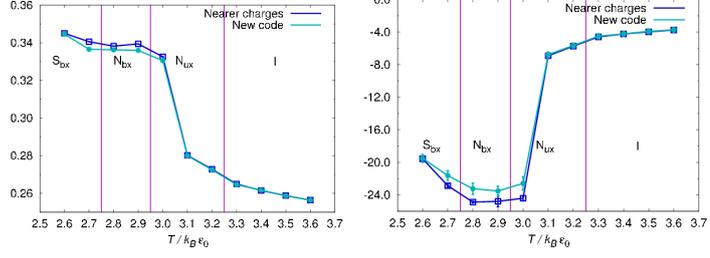
2.6.1.3 $\mu^* = 2$

This system is composed by molecules having two charges placed in positions shown in Table 2.4, with a dipole moment of $\mu^* = 2$.

Charge	x	y	z
+20.0	+0.05	0	0
-20.0	-0.05	0	0

Table 2.4: Positions of two charges acting as a central transversal dipole, $\mu^* = 2$, using charges placed nearer than the ones on previous system.

The following graphs (Figure 2.3) showing a comparison between results obtained getting the charges closer and the charge distance we use for production runs.



(a) Number density.

(b) Potential energy.

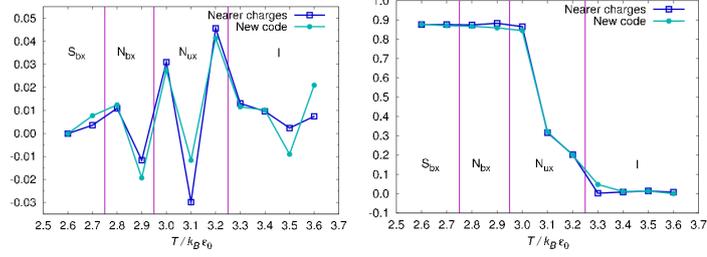
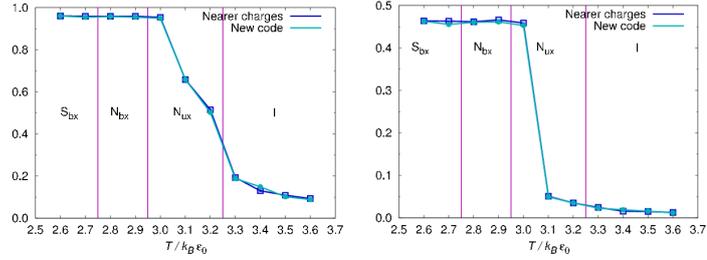
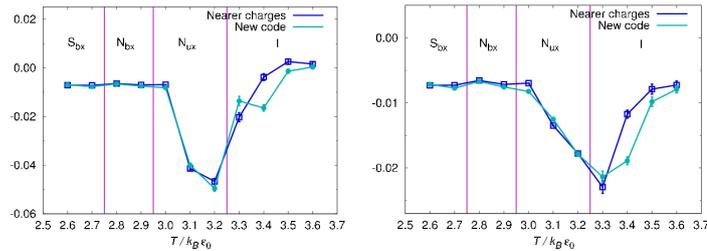
(c) $\langle P_1 \rangle$ (d) $\langle P_4 \rangle$ (e) $\langle R_{00}^2 \rangle$ (f) $\langle R_{22}^2 \rangle$ (g) $\langle R_{02}^2 \rangle$ (h) $\langle R_{20}^2 \rangle$

Figure 2.5: Physical quantities of a system with central transversal dipole ($\mu^* = 2$) on its molecules (charges placed nearer): Comparison with results obtained previously.

2.6.1.4 Conclusion

Table 2.5 shows it is clear that the phase sequence of the systematically studied models and this “validation model” with charges at a smaller distance are the same. Moreover, the average observables do not differ significantly from each other, so two point charges at not-so-close distance represent well a point dipole.

T^*	ox systems			Nearer charges ox systems		
	μ			μ		
	0.5	1	2	0.5	1	2
2.6	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}
2.7	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}
2.8	N_{bx}	N_{bx}	S_{bx}	N_{bx}	N_{bx}	S_{bx}
2.9	N_{bx}	N_{bx}	S_{bx}	N_{bx}	N_{bx}	S_{bx}
3.0	N_{ux}	N_{ux}	S_{bx}	N_{ux}	N_{ux}	S_{bx}
3.1	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}
3.2	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}
3.3	I	I	I	I	I	I
3.4	I	I	I	I	I	I
3.5	I	I	I	I	I	I
3.6	I	I	I	I	I	I

Table 2.5: Phases dependency on temperature in a system composed by molecules with a central transversal dipole: comparison with systems having charges placed at smaller distance.

2.6.2 Cutoff radius

In order to understand if charges cutoff radius is large enough to comprehend all contributes, we try to increase this value (6 Å). Moreover, from this analysis we could find out if the system is antiferromagnetic, so if it is correct to apply the reaction field method [44], since charges cutoff is indirectly related to how well adjacent dipoles shield each. If the values of physical quantities do not change a lot varying this parameter, it is possible to choose the smallest one, without losing precision.

The system analyzed is the one with central dipole aligned along the x axis (cx) of intensity $\mu^* = 1$, using the same parametrization a part from the variation of the charges' cutoff. All temperatures are tested, using the following cutoff values: 6 (the value used in all previous simulation), 8, 10 and 20 (Å).

2.6.2.1 $T^* = 2.6$

Figure 2.6 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

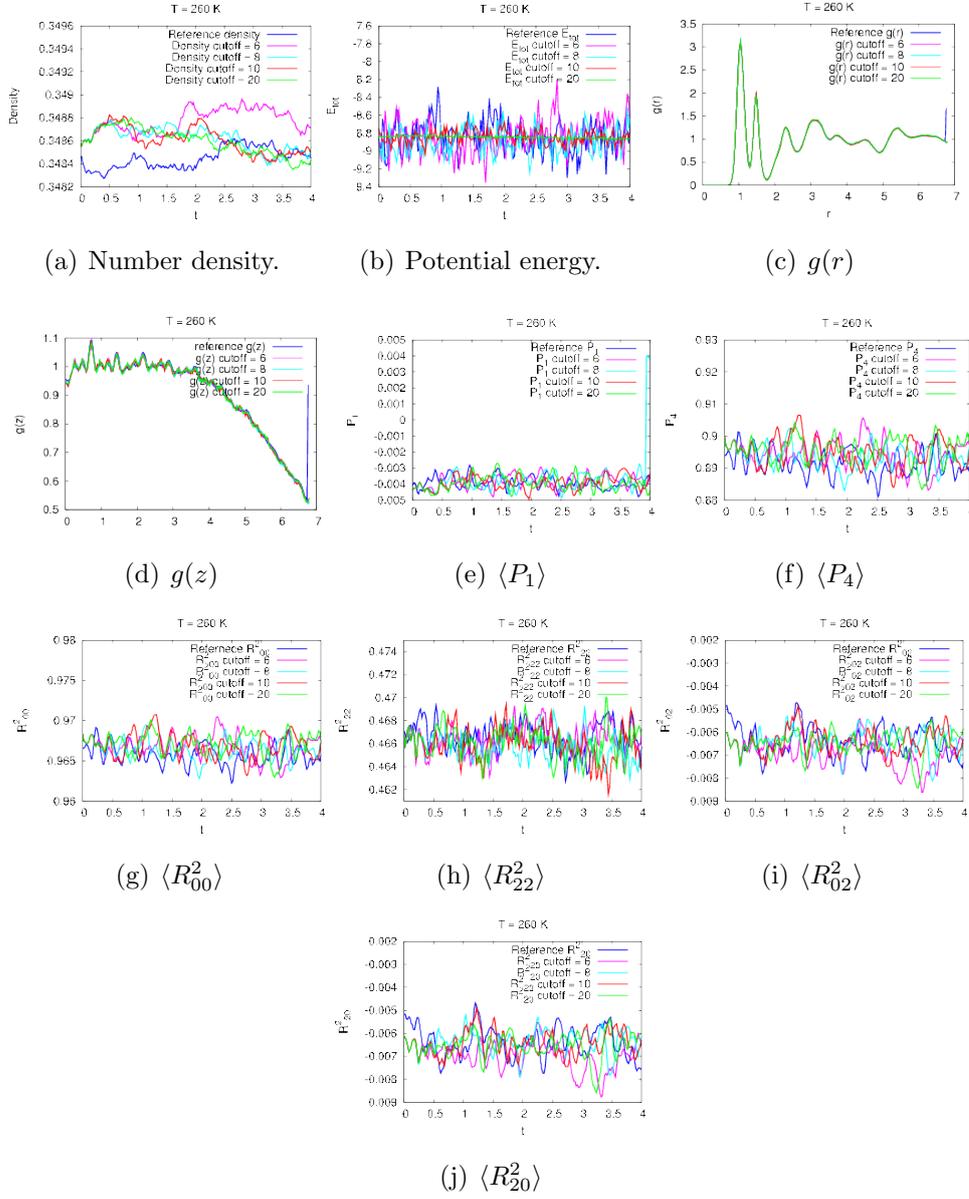


Figure 2.6: $T^* = 2.6$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.2 $T^* = 2.7$

Figure 2.7 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

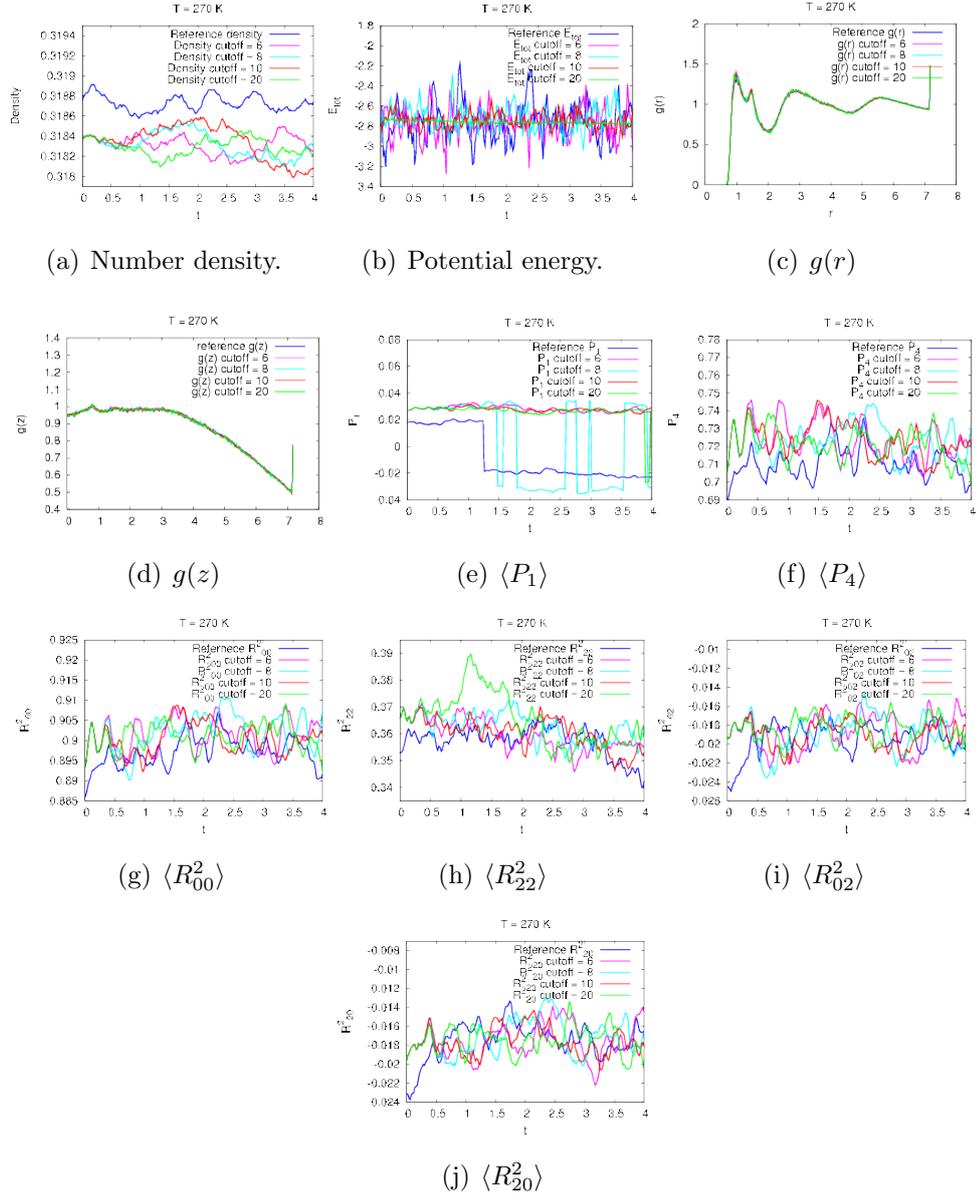


Figure 2.7: $T^* = 2.7$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.3 $T^* = 2.8$

Figure 2.8 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

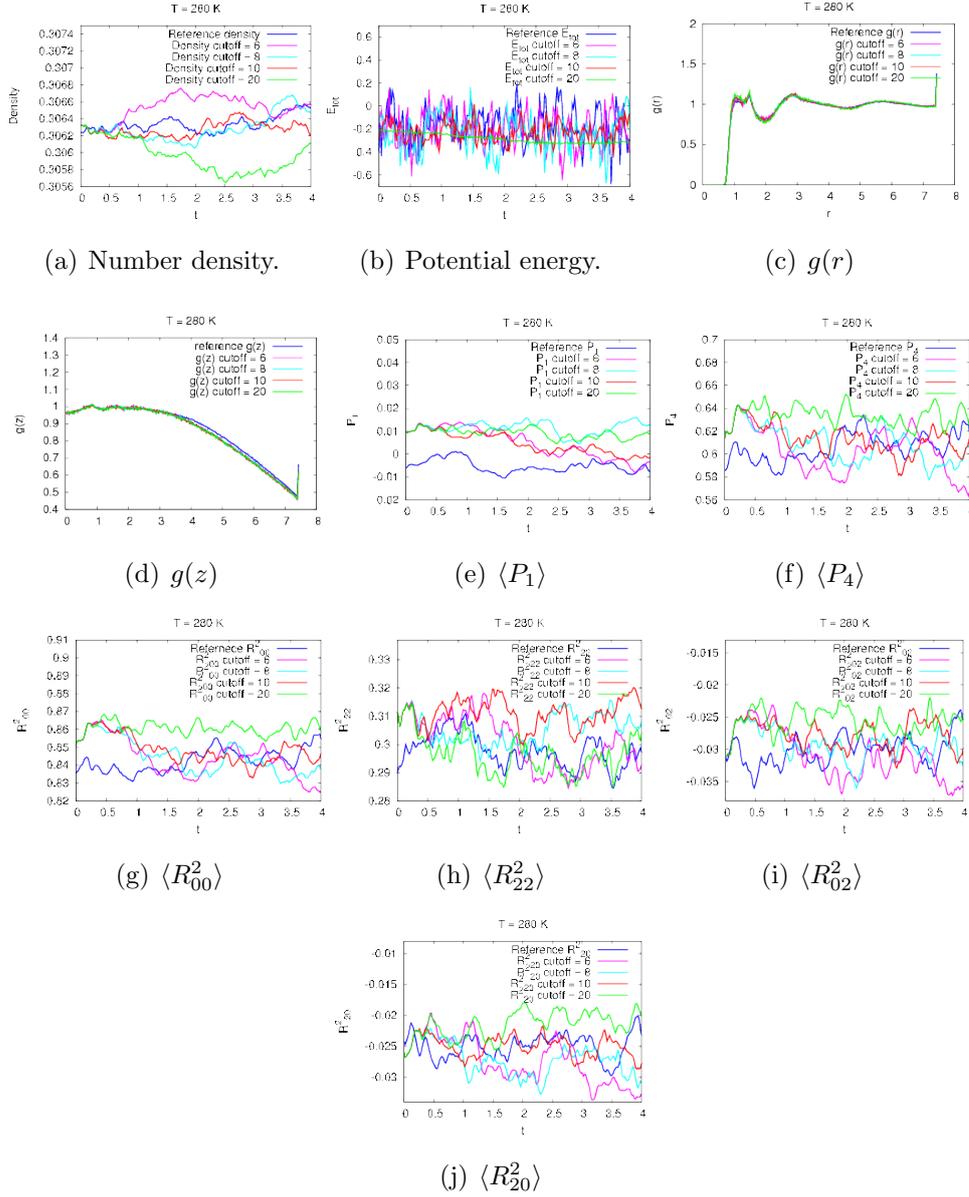


Figure 2.8: $T^* = 2.8$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.4 $T^* = 2.9$

Figure 2.9 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

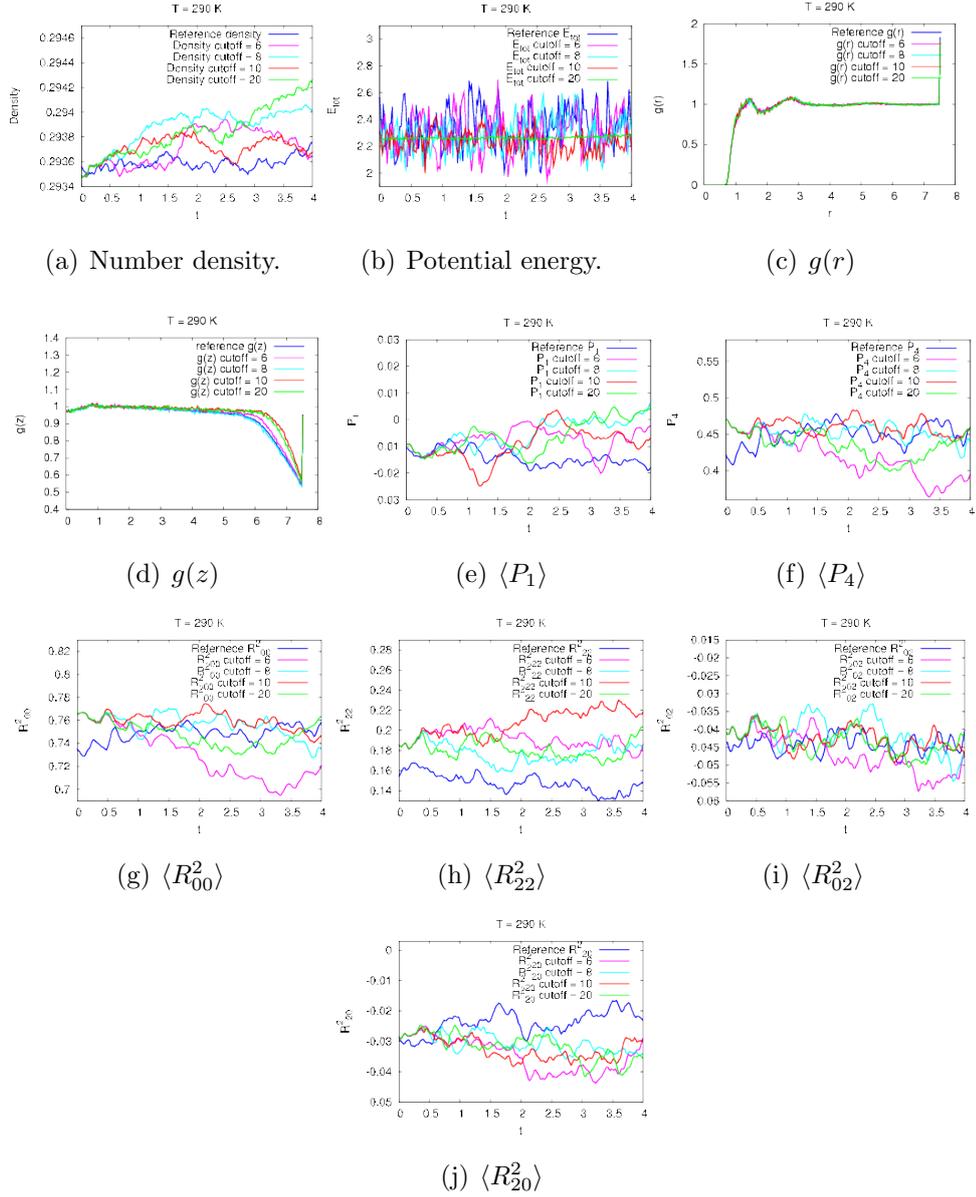


Figure 2.9: $PT^* = 2.9$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.5 $T^* = 3.0$

Figure 2.10 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

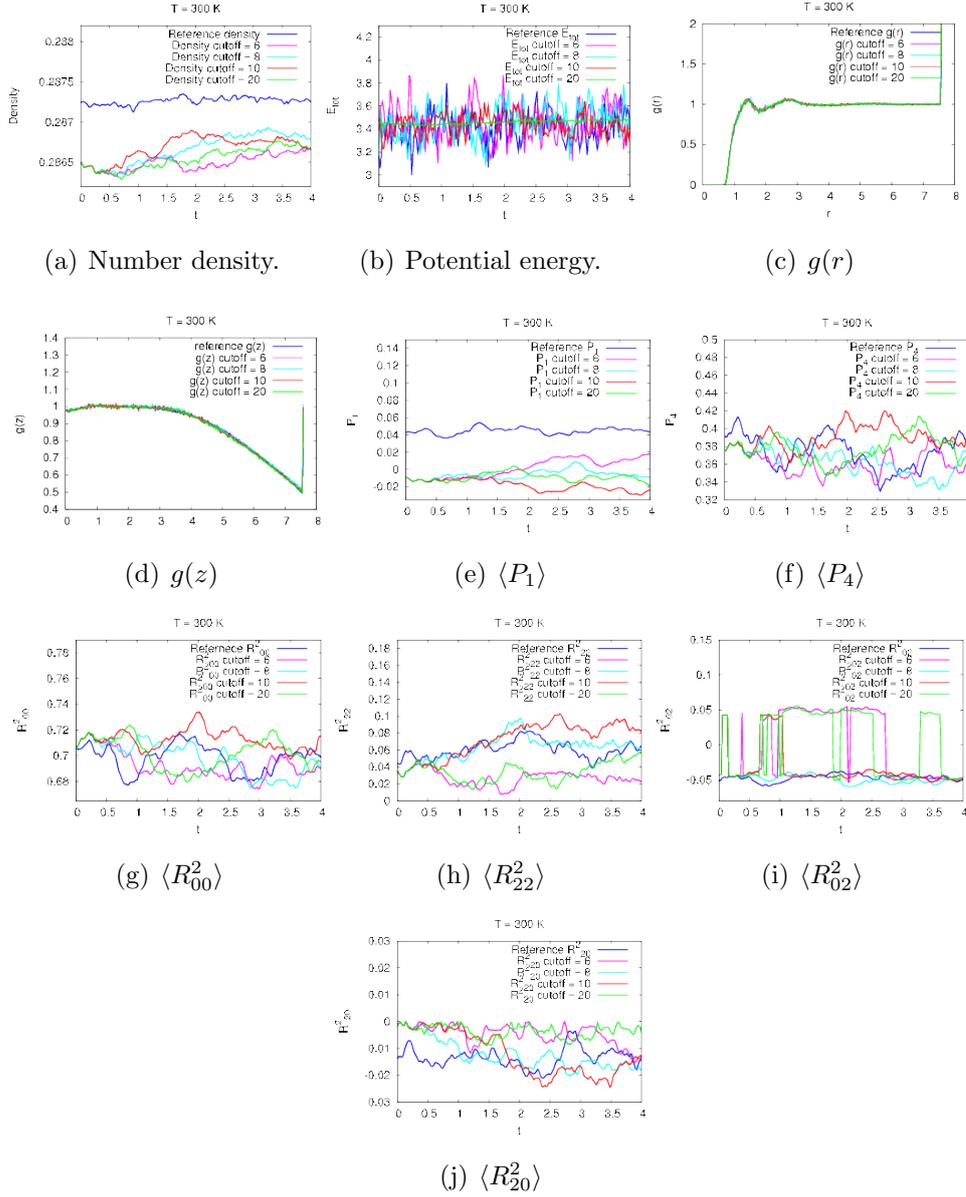


Figure 2.10: $T^* = 3.0$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.6 $T^* = 3.1$

Figure 2.11 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

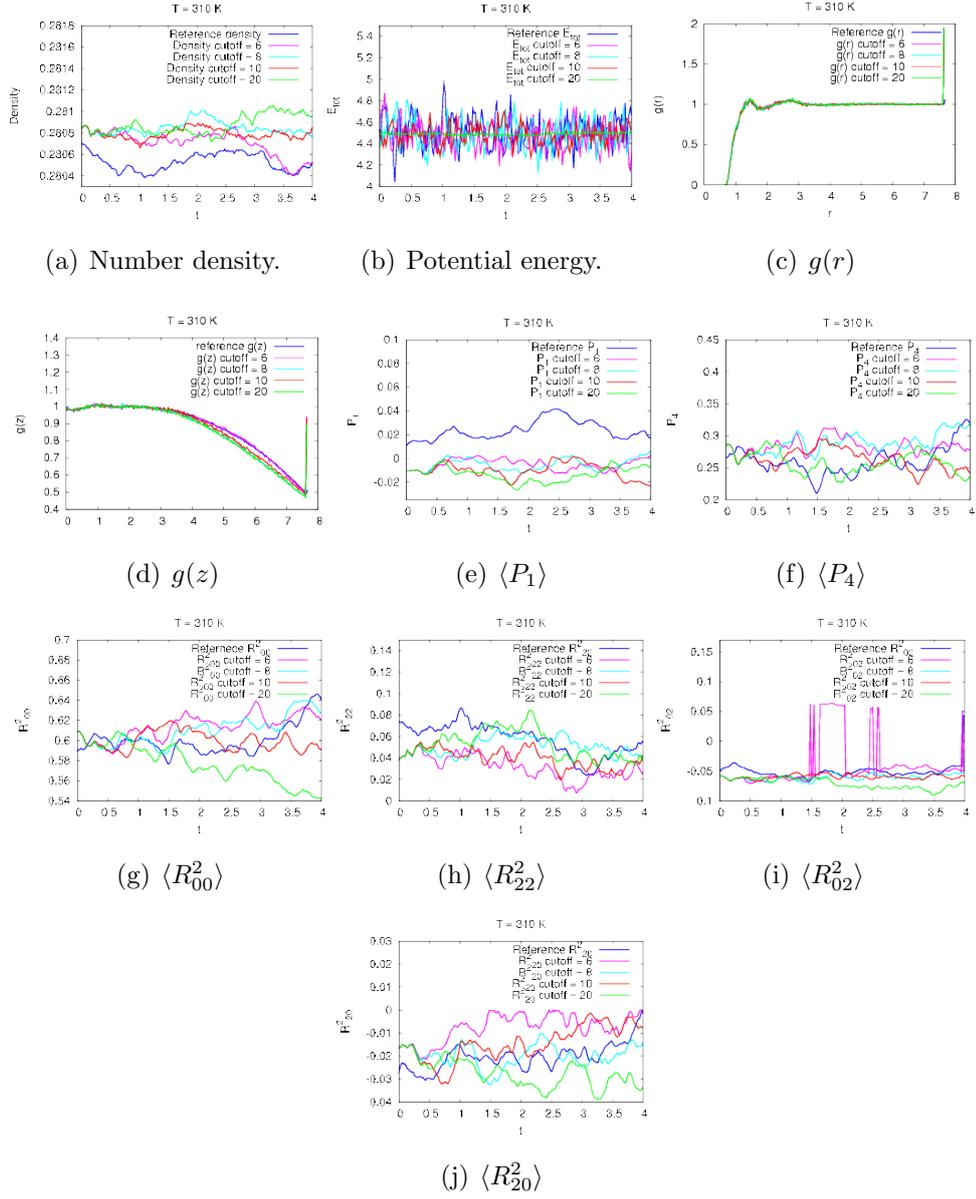


Figure 2.11: $T^* = 3.1$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.7 $T^* = 3.2$

Figure 2.12 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

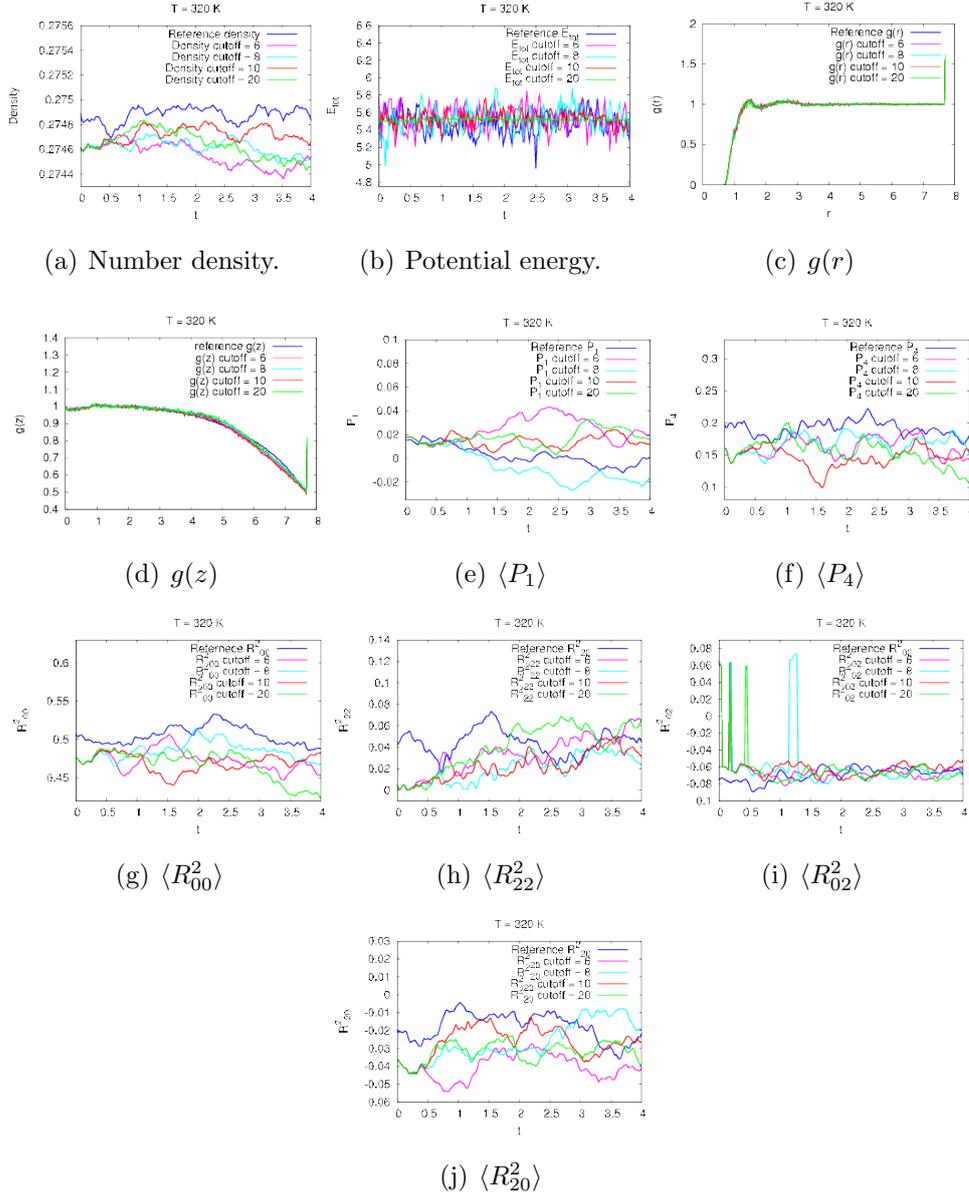


Figure 2.12: $T^* = 3.2$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.8 $T^* = 3.3$

Figure 2.13 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

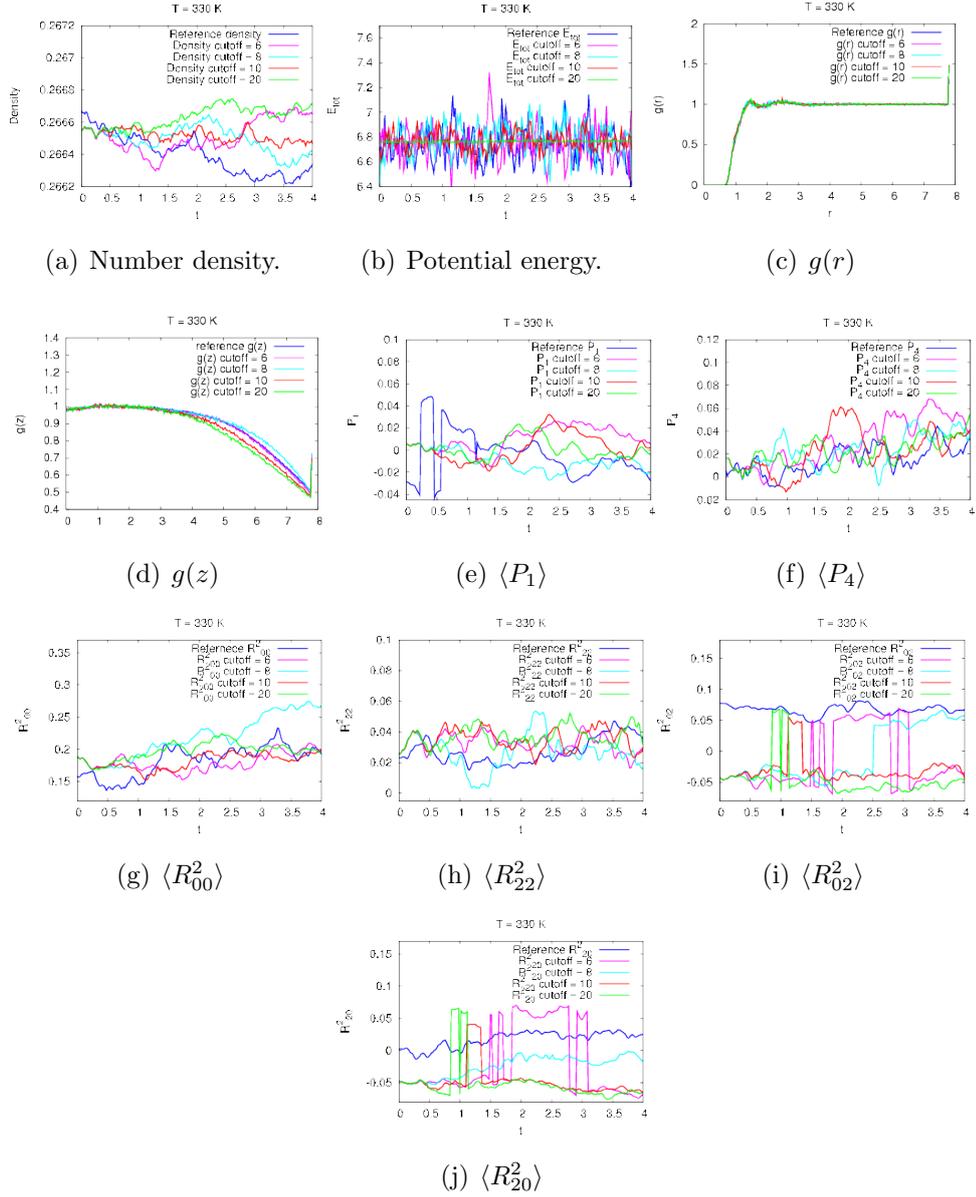


Figure 2.13: $T^* = 3.3$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.9 $T^* = 3.4$

Figure 2.14 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

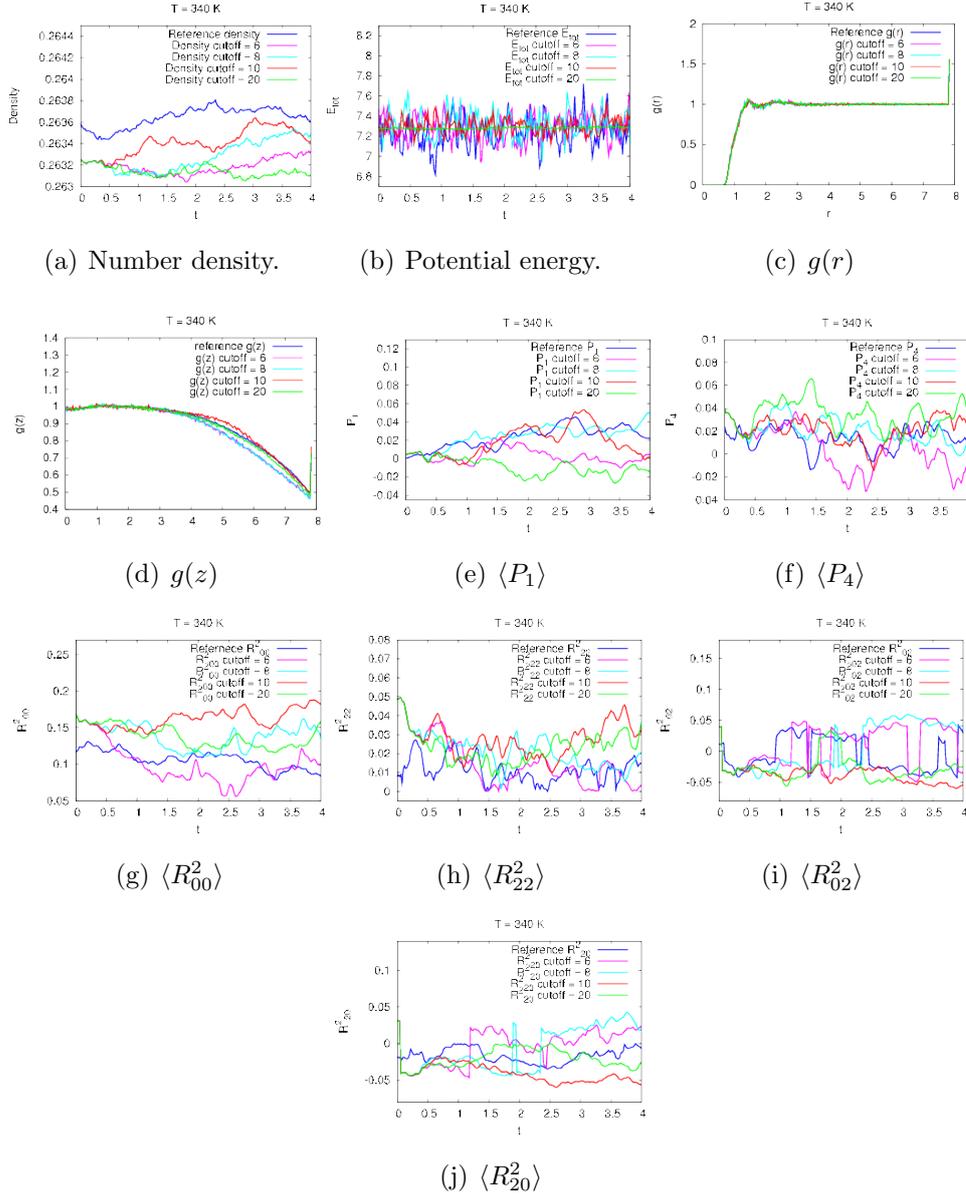


Figure 2.14: $T^* = 3.4$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.10 $T^* = 3.5$

Figure 2.15 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

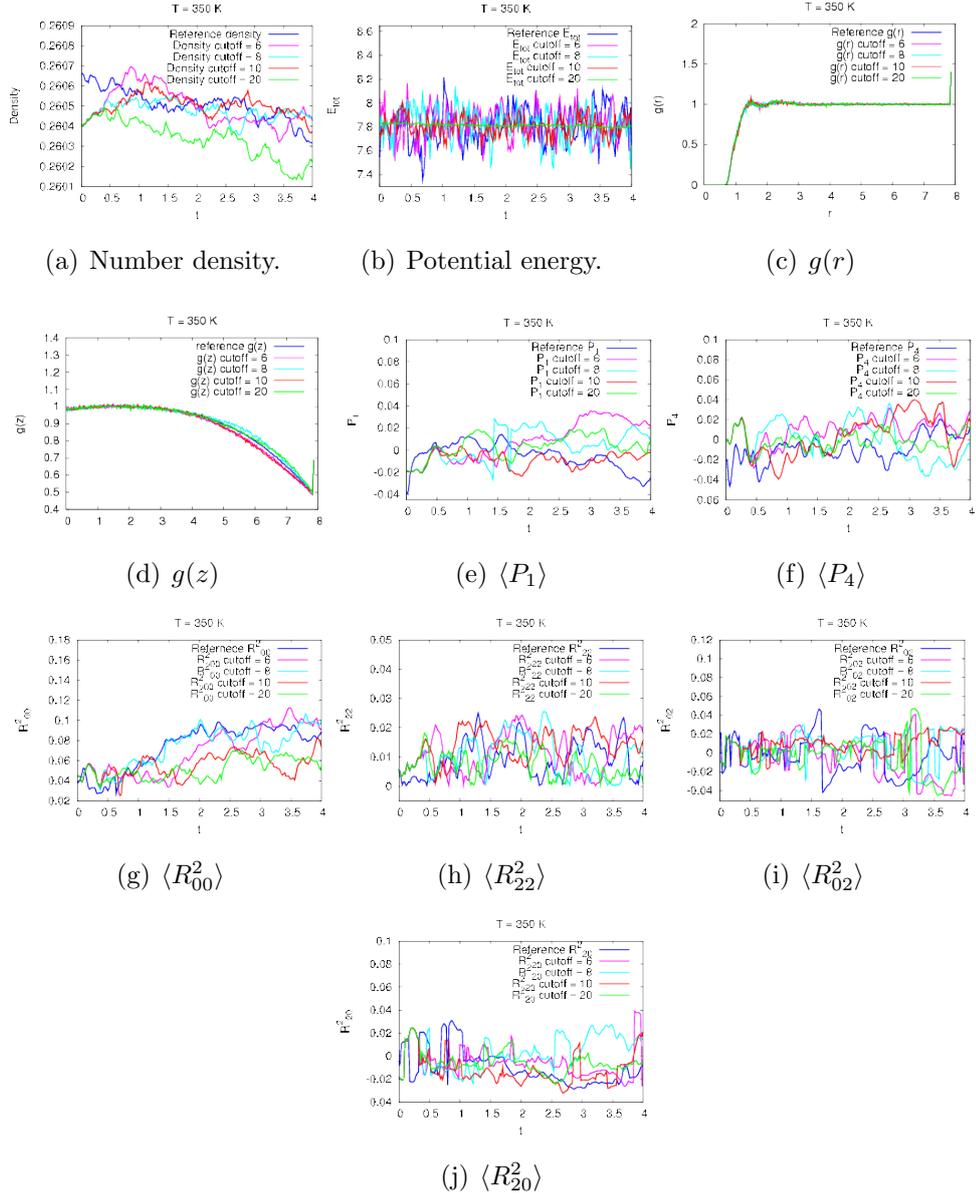


Figure 2.15: $T^* = 3.5$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.11 $T^* = 3.6$

Figure 2.16 shows the variation each physical quantity undergoes increasing the values of charges' cutoff.

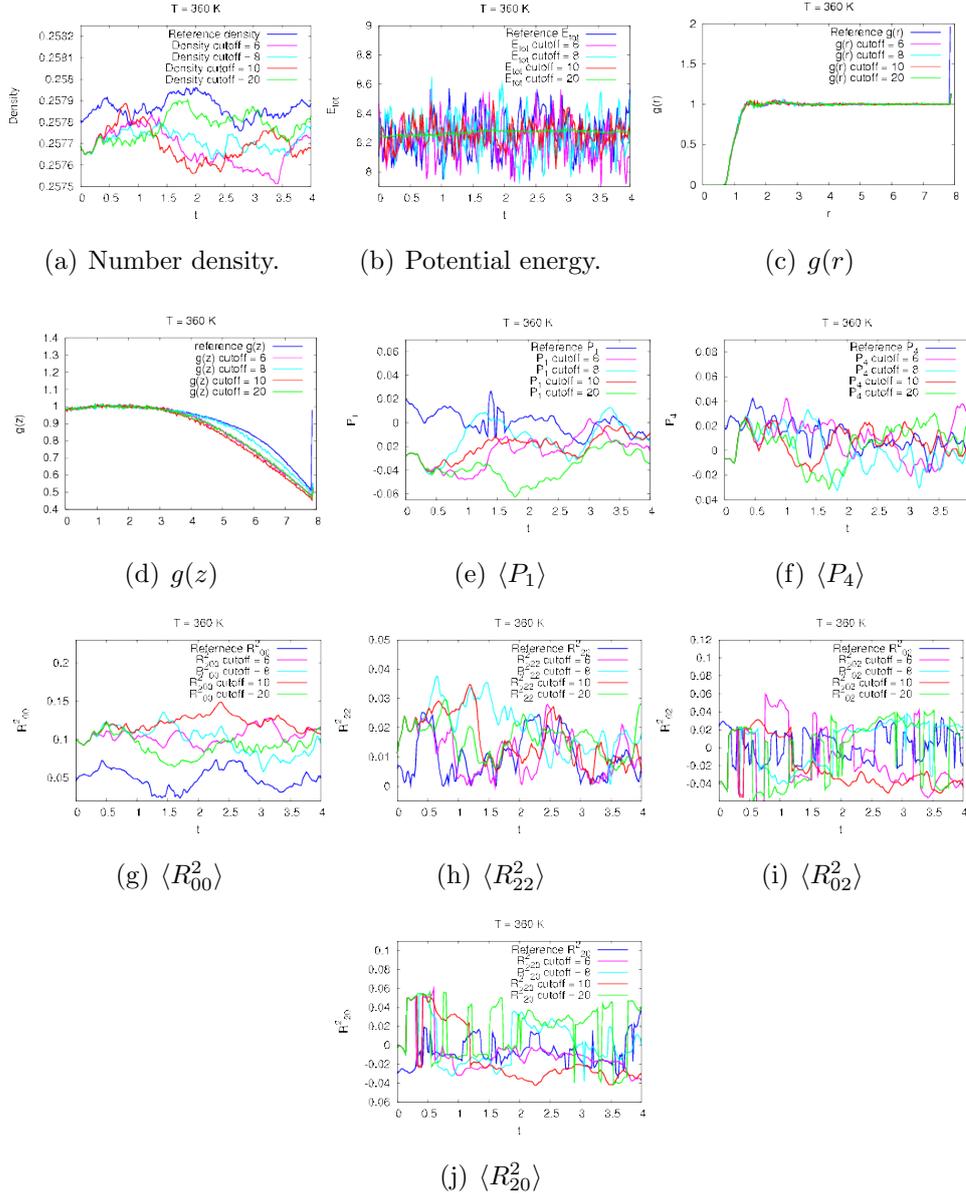


Figure 2.16: $T^* = 3.6$: Physical quantities of a system with central transversal dipole ($\mu^* = 1.0$) on its molecules (charges placed nearer): Comparison with results obtained using different cutoff values.

2.6.2.12 Conclusion

No difference in the analyzed characteristics is really important so it seems that charges' cutoff does not have an essential relevance. This is probably due to the fact that charges are screened enough: the reaction field method could finally be applied. Therefore, the value of this cutoff that should be used is the lowest possible, the same used for all other simulations, 6 Å.

2.6.3 Comparison with Monte Carlo

After having collected some results, we also compared a couple of systems obtained with Monte Carlo simulations. We chose to check **cz** system (system with a central dipole aligned along the z -axis) at $T^* = 2.8$ for $\mu^* = 0.0, 0.5$ and 2.0 .

Hence, we compare simulations obtained as explained in this chapter with molecular dynamics technique with these new Monte Carlo simulations. Monte Carlo technique does not take into account the dynamic of the system: it starts with a random configuration at a certain temperature, generates a new configuration (moving a molecule), compute energy and compare this value with that of the previous configuration. A “move” is accepted if the energy of the last configuration is lower, otherwise another random molecule is moved. The Reaction Field (RF) and the Wolf methods are used to simulate the effect of long range dipole-dipole interactions for simulations with periodic boundary conditions, as alternatives to the Ewald summation. In RF method [40], used also in the molecular dynamics simulations, around each molecule there is a “cavity” or sphere within which the Coulomb interactions are treated explicitly (molecule induce a polarization in this media). Over the cutoff radius a continuous fluid is assumed. On the other hand, Wolf method is based on the evidence that electrostatic Coulomb potential in condensed systems is short-ranged and that neutralization of the charge contained within the cutoff radius is crucial for potential stability [45].

2.6.3.1 Apolar system

Parameters	MDGB run	MC90 run
P^*	8.02 ± 0.05	8.01 ± 0.02
E_{tot}	0.71 ± 0.01	-7.42 ± 0.08
E_{gb}	-7.66 ± 0.03	-7.42 ± 0.08
E_{ch}	0.0 ± 0.0	0.0 ± 0.0
R_{00}^1	0.006 ± 0.003	-0.000 ± 0.001
R_{00}^2	0.831 ± 0.004	0.554 ± 0.009
R_{02}^2	0.0317 ± 0.0008	0.030 ± 0.002
R_{20}^2	0.0229 ± 0.0005	0.036 ± 0.001
R_{22}^2	0.254 ± 0.008	0.261 ± 0.004

Table 2.6: Neutral system.

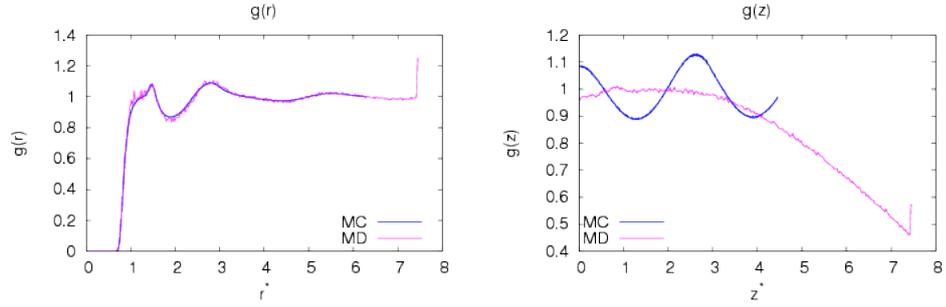


Figure 2.17: Apolar system.

2.6.3.2 cz system $\mu^* = 0.5$

Parameters	MDGB run	MC90 run (Wolf)	MC90 runs (Reaction Field)
P^*	7.99 ± 0.02	7.99 ± 0.01	8.07 ± 0.08
E_{tot}	1.161 ± 0.007	-6.7 ± 0.2	-6.80 ± 0.08
E_{gb}	-7.18 ± 0.05	-6.8 ± 0.2	-6.79 ± 0.08
E_{ch}	0.026 ± 0.002	0.0557 ± 0.0004	-0.0025 ± 0.0003
R_{00}^1	-0.018 ± 0.003	-0.000 ± 0.003	0.000 ± 0.002
R_{00}^2	0.797 ± 0.006	0.49 ± 0.03	0.45 ± 0.02
R_{02}^2	0.038 ± 0.002	0.033 ± 0.002	0.0299 ± 0.0002
R_{20}^2	0.031 ± 0.004	0.0403 ± 0.0009	0.0405 ± 0.0002
R_{22}^2	0.26 ± 0.01	0.23 ± 0.02	0.20 ± 0.01

Table 2.7: cz system, $\mu^* = 0.5$. Note that for results obtained using Reaction Field, E_{dip} is used to be compared with E_{ch} .

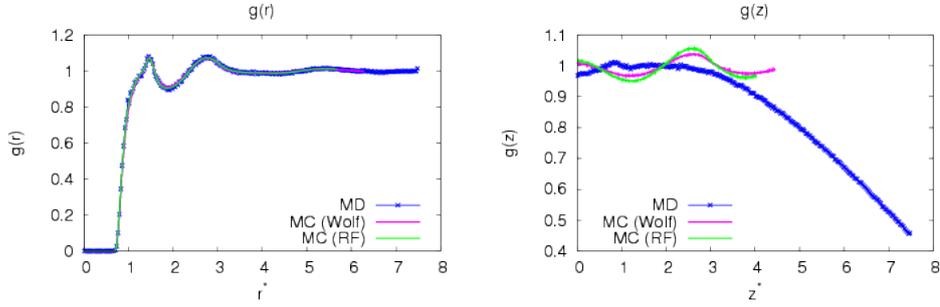
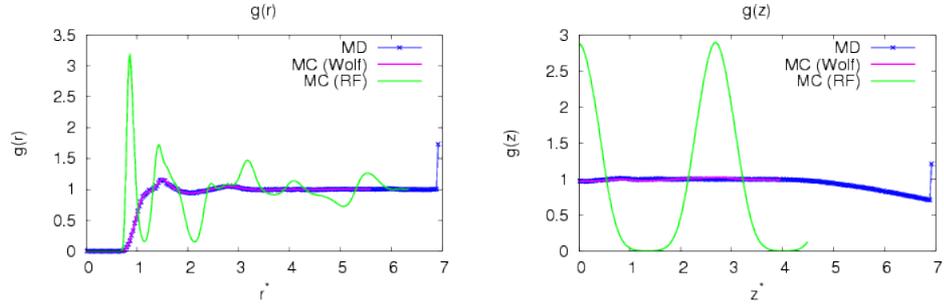


Figure 2.18: cz $\mu^* = 0.5$ system: $g(r)$ and $g(z)$.

2.6.3.3 cz system $\mu^* = 2$

Parameters	MDGB run	MC90 run (Wolf)	MC90 runs (Reaction Field)
P^*	8.00 ± 0.03	7.99 ± 0.01	5.14 ± 0.24
E_{tot}	-0.34 ± 0.02	-9.03 ± 0.05	-18.62 ± 0.07
E_{gb}	-6.77 ± 0.02	-7.10 ± 0.04	-14.15 ± 0.07
E_{ch}	-1.78 ± 0.02	-1.923 ± 0.007	-4.470 ± 0.005
R_{00}^1	-0.040 ± 0.003	0.000 ± 0.002	-0.0005 ± 0.0005
R_{00}^2	0.7934 ± 0.004	0.59 ± 0.02	0.923 ± 0.002
R_{02}^2	0.0292 ± 0.001	0.031 ± 0.002	0.007 ± 0.000
R_{20}^2	0.0370 ± 0.0007	0.0363 ± 0.0008	0.0072 ± 0.0001
R_{22}^2	0.244 ± 0.002	0.28 ± 0.01	0.462 ± 0.001

Table 2.8: cz system, $\mu^* = 2$.Figure 2.19: cz $\mu^* = 2$ system: $g(r)$ and $g(z)$.

2.6.3.4 Conclusions

Gay–Berne and electrostatic energies obtained with the three methods are actually quite different but this can be also due to the fact that the simulation types are completely different. Hence, increasing the value of dipole, results depend on the different techniques used to compute electrostatic interactions.

2.7 Discussion

The overall effect of the electrostatic interactions on the mesogenic properties of our biaxial GB ellipsoids is quite complex. However, we can identify a number of patterns in phase behavior arising from the specific values of the two model parameters with the strongest influence on the mesogenic properties, namely the dipole moment strength and its position with respect to the GB center of mass. In comparison with these two features, the orientation of the dipole with respect to the molecular frame (given as spherical angles ϕ and θ , e.g. see Table 4.24) plays a minor role, even though for an off-axis tilted dipole the lowered symmetry (from C_{2v} of the on-axis cases to C_s or even C_1 for the off-axis and offset models) helps into destabilizing layered structures in favor of nematic organizations. We now discuss the specific details of our MD results, following the classification of Table 4.24 into three groups of models with their respective simulations for the cases of a central dipole directed along one of the three molecular axes (Table 2.9), central off-axis tilted dipole (Table 2.10) and finally off centered dipole (Table 2.11).

2.7.1 Central dipole along one molecular axis

The first group of three models for which we discuss the MD results is that for the GB ellipsoids with a central dipole aligned with one of the molecular axis (i.e. see the models cx , cy and cz of Table 4.24). They show sequences of phases (see Table 2.9) reminding those observed for similar models [28, 27]: as the dipole moment increases the $I-N$ phase transition temperature is only marginally affected by the larger electrostatic interactions, while the transition from nematic to an orthogonal smectic/solid progressively shifts to higher temperatures, considerably reducing the stability ranges of the N_b and N organizations. Following this trend, the first nematic phase which disappears for $1.5 < \mu^* \leq 2$ is the N_b , and correspondingly to this we observe a direct transition from uniaxial nematic N

to a biaxial layered structure. Eventually, for $2 < \mu^* \leq 3$ also the uniaxial nematic is destabilized by the dipole-dipole interactions which strongly favor the *side-by-side* and *face-to-face* pair interactions. This results in the onset of fairly stable orthogonal layered structures and high temperature transitions from isotropic liquid to a biaxial smectic (or solid) phase. For these highly ordered systems we have not observed the spontaneous formation of polar cluster and neighboring dipoles adopt preferentially an antiferroelectric organization. Also the average $\langle P_1 \rangle$ order parameter is zero for all models at all temperatures.

The small differences between the *cx*, *cy* and *cz* models originate from the different closest distance that the two ellipsoidal particles can assume in a configuration with *side-by-side* antiparallel dipoles (and consequently from how important the electrostatic pair interaction can be). This is in turn mainly governed by geometrical, steric properties, namely the length of the GB ellipsoids axes determining the *so-called* contact distance, i.e. the distance $r_0 = \sigma(\mathbf{r}, \omega_1, \omega_2)$ for which $U_{GB}(r_0) = 0$. Thus, for the *cx* and *cz* models where this distance r_0 is $\sigma_y = 0.714 \sigma_0$ we obtain a similar mesogenic behavior, while for the *cy* model this antiparallel pair configuration can be observed only at larger separations ($r_0 = \sigma_x = 1.4 \sigma_0$), and the closest distance for the dipole-dipole interactions, $r_0 = \sigma_y = 0.714 \sigma_0$, is that for the *end-to-end* antiparallel dipoles, which gives an energy one half of the *side-by-side* at the same distance and orientation. The plots of the order parameters of Figure 2.23 and 2.24 show another interesting behavior of these central systems. Again, for small dipole moments (e.g. $\mu^* = 1$, Figure 2.23) the temperature dependencies of the uniaxial $\langle R_{00}^2 \rangle$ and biaxial $\langle R_{22}^2 \rangle$ order parameters are practically superimposable for all dipolar models and follow the trend observed for apolar system. On the other hand, for stronger dipole moments (e.g. $\mu^* = 2$, Figure 2.24) the $\langle R_{00}^2 \rangle$ order parameter profiles show small but well defined discontinuities across the nematic to orthogonal smectic (or solid) transitions. In spite of this, the $\langle R_{22}^2 \rangle$ plots for these central models are surprisingly not affected by the module and the orientation of the electric dipole. As before, the largest effects have been observed for the transversal *cx* (with $\boldsymbol{\mu} \parallel \mathbf{x}$) and

also longitudinal cz (with $\boldsymbol{\mu} \parallel \mathbf{z}$) cases (see Figure 2.24–B).

T^*	μ^*	orientation and magnitude μ^* of on-axis central dipole moments																	
		$\mu^* \parallel \mathbf{x} (cx)$						$\mu^* \parallel \mathbf{z} (cz)$						$\mu^* \parallel \mathbf{y} (cy)$					
	0	0.5	1	1.5	2	3	0.5	1	1.5	2	3	0.5	1	1.5	2	3			
2.6	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}			
2.7	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}			
2.8	N _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}			
2.9	N _{bx}	N _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}	N _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}			
3.0	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	S _{bx}			
3.1	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	S _{bx}			
3.2	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}			
3.3	I	I	I	I	I	N _{ux}	I	I	I	I	N _{ux}	I	I	I	I	I			
3.4	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I			
3.5	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I			
3.6	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I			

Table 2.9: The sequence of thermotropic phases obtained from the MD simulations of systems of $N = 1024$ elongated biaxial GB ellipsoids with an embedded electric dipole $\boldsymbol{\mu}$ of dimensionless magnitude μ^* , positioned in the GB center of mass, and aligned on-axis with either the \mathbf{x} , \mathbf{y} or \mathbf{z} molecular frame axes, model cx , cy and cz (see Table 4.24 for details). The phases observed at the various temperatures are isotropic (I), uniaxial nematic (N_{ux}), biaxial nematic (N_{bx}) and orthogonal biaxial smectic (S_{bx}).

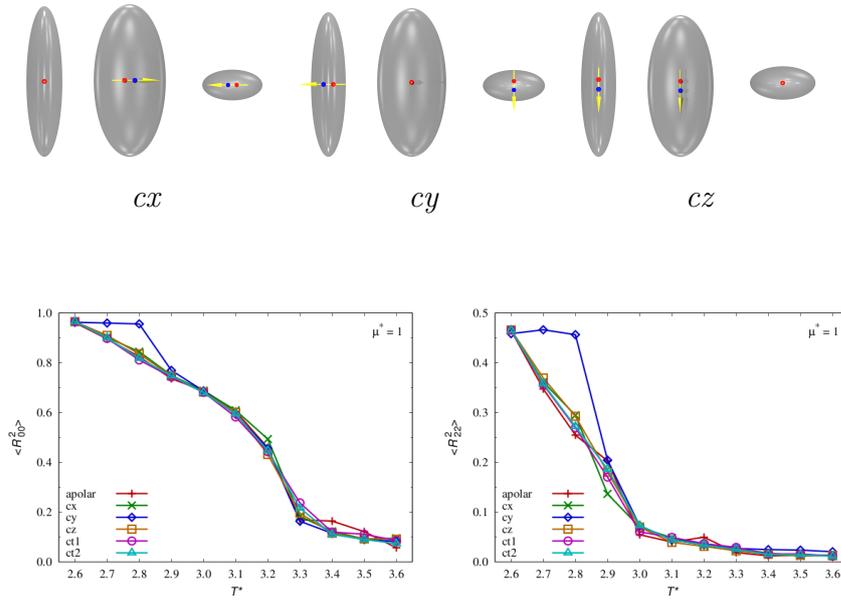


Figure 2.20: $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$ order parameter for system with $\mu^* = 1$, for model of cx , cy and cz of Table 4.24 (results shown also for model $ct1$ and $ct2$).

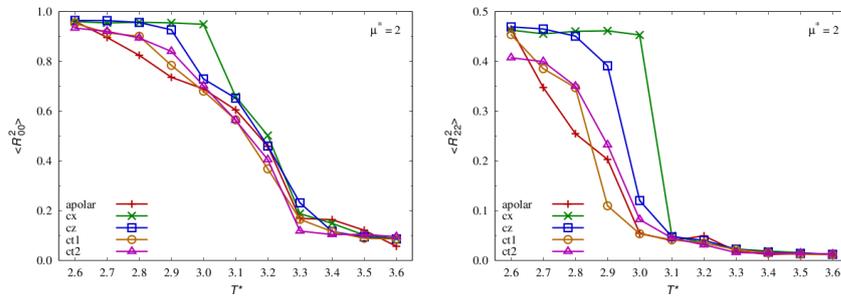


Figure 2.21: $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$ order parameter for system with $\mu^* = 2$, for model of cx , cy and cz of Table 4.24 (results shown also for model $ct1$ and $ct2$).

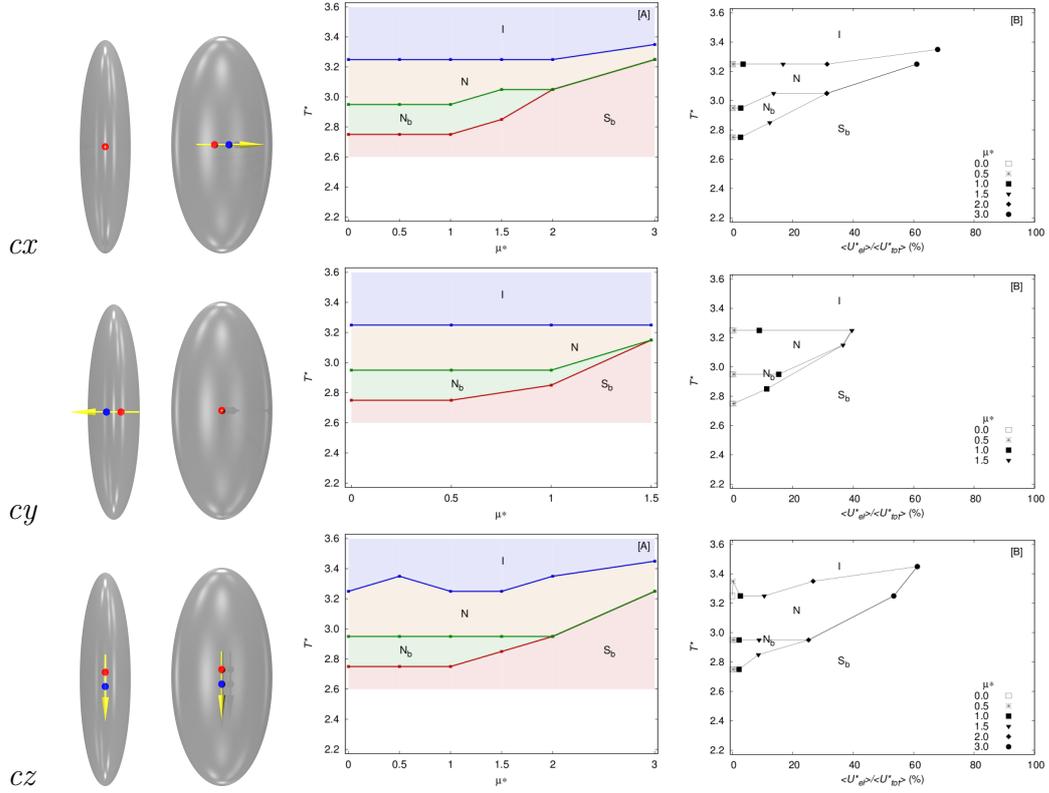


Figure 2.22: The transition temperatures between isotropic (I), nematic (N), biaxial nematic (N_b) and orthogonal biaxial smectic (S_b) for model *cx*, *cy* and *cz*, as a function of the dimensionless dipole modulus $\mu^* = 0.5, 1, 1.5, 2, 3$ (Plate A) and of the ratio between average electrostatic $\langle U_{el} \rangle$ and the total potential energy $\langle U \rangle = \langle U_{GB} \rangle + \langle U_{el} \rangle$ (Plate B). The straight lines are guides for the eye interpolating the MD results. The units for temperature and dipole modulus are $T^* = T/(k_B^{-1} \varepsilon_0)$ and $\mu^* = \mu/(\varepsilon_0 \sigma_0^3)^{1/2}$.

2.7.2 Tilted central dipoles

A second group of GB ellipsoids with a tilted central dipole (*ct1*, *ct2*) has been simulated (see Table 2.10) and also for these particles we observe for small dipole strength a little departure from the results of the reference apolar GB model.

In Figure 2.25 we plot the transition temperatures as a function of dipole module (plate 8–A) and the ratio $\langle U_{el} \rangle / \langle U \rangle$ between the average electrostatic energy and the total potential energy (plate 8–B). In particular, this second plot is revealing as it shows that as the dipole–dipole interactions become greater than $\approx 10\%$ of the total potential energy the layered phases start eroding the nematic temperature ranges, to eventually destroy the N_b phase for $\langle U_{el} \rangle / \langle U \rangle \approx 0.3$. Also the uniaxial nematic organization practically disappears when the electrostatic interactions account for $\approx 70\%$ of the total potential energy. This behavior is typical of highly dipolar symmetric particles, and it is consistent with previous simulation results [28, 27]. Regarding the symmetry of these mesogenic dipolar particles, we notice that, even though the *ct2* model does not have any symmetry plane, we have not seen any evidence of chiral organizations in our MD sample (containing only one enantiomeric form) possibly due to the relatively small number of particles and the periodic boundaries fluctuations large enough to overcome the small energy changes effects related to the formation of chiral clusters.

T^*	μ^*	orientation and magnitude μ^* of off-axis central dipole moments											
		$\phi = 0^\circ, \theta = 60^\circ$ (<i>ct1</i>)						$\phi = 30^\circ, \theta = 60^\circ$ (<i>ct2</i>)					
		0	0.5	1	1.5	2	3	0.5	1	1.5	2	3	
2.6		S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	
2.7		S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	S _{bx}	
2.8		N _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}	
2.9		N _{bx}	N _{bx}	N _{bx}	N _{bx}	N _{ux}	S _{bx}	N _{bx}	N _{bx}	S _{bx}	S _{bx}	S _{bx}	
3.0		N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	
3.1		N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	S _{bx}	
3.2		N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	N _{ux}	
3.3		I	I	I	N _{ux}	I	I	I	I	N _{ux}	I	I	
3.4		I	I	I	I	I	I	I	I	I	I	I	
3.5		I	I	I	I	I	I	I	I	I	I	I	
3.6		I	I	I	I	I	I	I	I	I	I	I	

Table 2.10: The sequence of thermotropic phases obtained from the MD simulations of systems of $N = 1024$ elongated biaxial GB ellipsoids with an embedded electric dipole μ of dimensionless magnitude μ^* , positioned in the GB center of mass, and with an off-axis tilt (given by the polar angles ϕ , θ) with respect to the molecular frame axes, model *ct1* and *ct2* (see Table 4.24 and Table 2.9 for details).

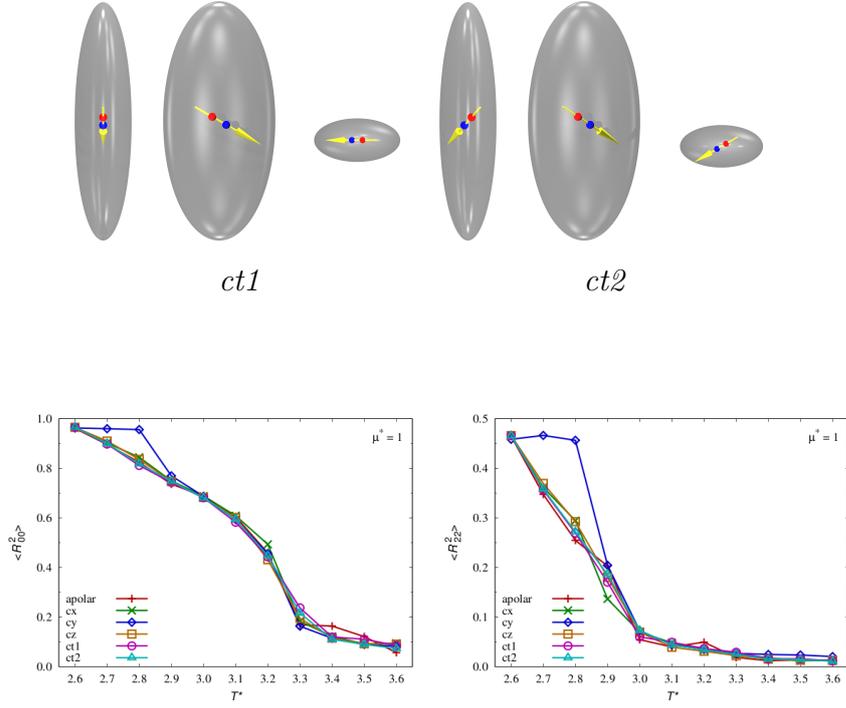


Figure 2.23: $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$ order parameter for system with $\mu^* = 1$, for model of *ct1* and *ct2* of Table 4.24 (results shown also for model *cx*, *cy* and *cz*).

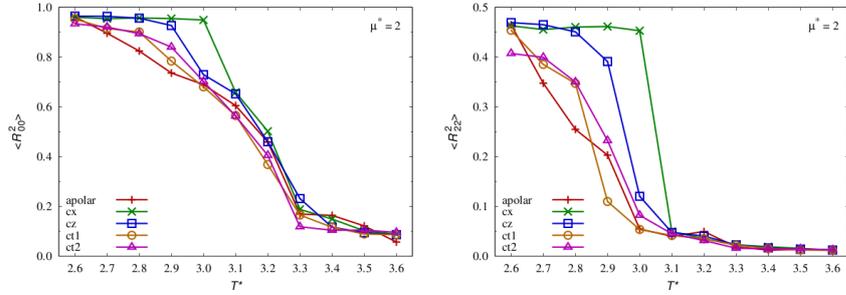


Figure 2.24: $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$ order parameter for system with $\mu^* = 2$, for model of *ct1* and *ct2* of Table 4.24 (results shown also for model *cx*, *cy* and *cz*).

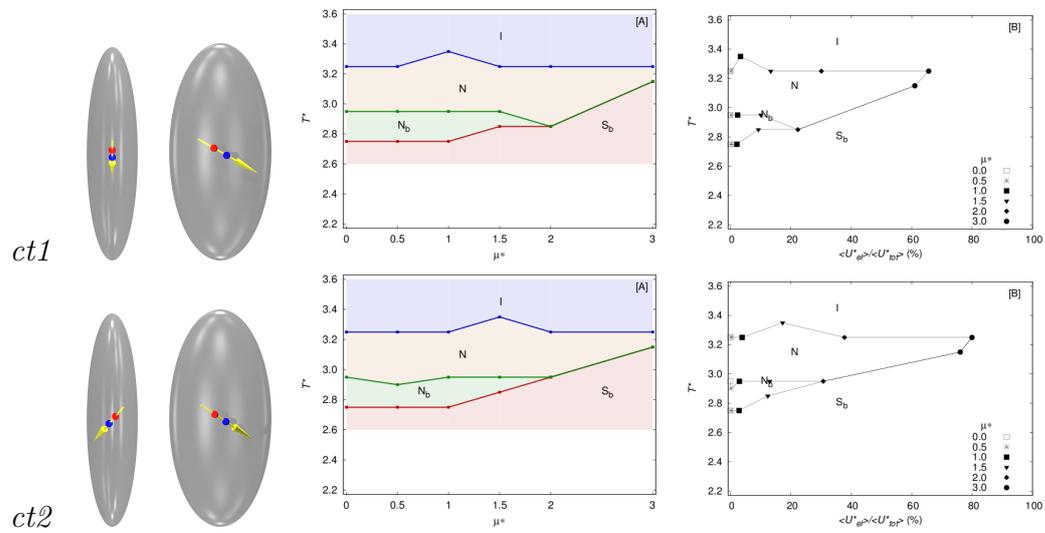


Figure 2.25: The transition temperatures for model *ct1* and *ct2*. For further details, see the caption of Figure 2.22

2.7.3 Off-center dipoles

The third kind of dipolar GB particles we have simulated are characterized by off-center positions of the point dipole, i.e. by a pair of electric charges not symmetrically placed with respect to the ellipsoid center of mass (see models *ox*, *oz* and *ot* in Table 4.24).

The sequences of thermotropic phases exhibited by these three models are reported in Table 2.11. We see that as the magnitude of the dipole moment increases, the electrostatic interactions determine significant differences in the phase diagrams with respect to the apolar GB reference system. This can also be seen in the order parameter plots of Figure 2.26 where we show the average values of $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$ for the models with dipole moment $\mu^* = 2$. The *ox* systems behave similarly to what observed for the GB particles with central dipole: for this intermediate value of μ^* the N_b organization has disappeared and the thermotropic phase transition from isotropic proceeds to an uniaxial nematic first and then directly to a biaxial layered phase. The most interesting results are those obtained from the MD simulations of the *oz* and *ot* models. The *oz* GB particles with an off-center dipole aligned with the molecular \mathbf{z} axis show a quite large enhancement of the temperature stability range for the N_b phase (see Figure 2.27) which is maximal for an intermediate value of the dipole moment $\mu^* = 1.5$ giving electrostatic interactions accounting for $\approx 10\%$ of the total energy. This behavior is different from what observed for the central dipole cases, where a destabilization of the N_b phase for increasing dipole moments was found instead. For the *oz* model we also observe a shift down in temperature for the transition to layered phases, while the $N-N_b$ transition temperature is essentially unaltered. As a consequence of this, the $\langle R_{22}^2 \rangle$ order parameter increases more slowly with decreasing temperature for the *oz* $\mu^* = 2$ model (see Figure 2.26) than for the *ox* system with same dipole moment. Another remarkable result is that the lowest N_b point has a fairly high $\langle R_{22}^2 \rangle = 0.36$, which for the dipolar GB models previously discussed was observed only in highly structured layered phases. A rationale for this behavior is that in the *oz* model the position

of the dipole does not lead to favorable electrostatic interactions when the GB centers of mass are aligned on parallel planar layers. The resulting competition between the dispersive and electrostatic interactions leads to a destabilization of the smectic phase. The *ot* model shows another peculiar phase behavior (see Table 2.11) and in this case, an increasing electric dipole modulus does not modify the sequence of thermotropic LC phases. Also the transition temperatures are marginally affected by large $\langle U_{el} \rangle / \langle U \rangle$ ratios and the order parameter plots for the $\mu^* = 2$ in Figures 2.26, *left*, and 2.26, *right*, closely follow those for the reference apolar GB system. However, in this case the $\langle U_{el} \rangle / \langle U \rangle$ ratio is well above 0.3; furthermore, in Figure 2.27 we see that the N_b phase can be found, albeit in a narrow temperature range, even for electrostatic interactions amounting to more than 80% of the total potential energy.

T^* / μ^*		orientation and magnitude μ^* of offset on- and off-axis dipole moments																	
		$\mu^* \parallel \mathbf{x} (ox)$						$\mu^* \parallel \mathbf{z} (oz)$						$\phi = 30^\circ, \theta = 50^\circ (ot)$					
		0	0.5	1	1.5	2	3	0	0.5	1	1.5	2	3	0	0.5	1	1.5	2	3
2.1											S_{bx}	S_{bx}							
2.2											S_{bx}	S_{bx}							
2.3											N_{bx}	S_{bx}							
2.4											N_{bx}	S_{bx}							
2.5											N_{bx}	N_{bx}							
2.6		S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}
2.7		S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	S_{bx}	N_{bx}	S_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}
2.8		N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}	N_{bx}
2.9		N_{ux}	N_{bx}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}
3.0		N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}
3.1		N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}
3.2		N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}	N_{ux}
3.3		I	N_{ux}	I	I	I	S_{bx}	I	I	N_{ux}	I	I	N_{ux}	I	I	I	I	I	I
3.4		I	I	I	I	I	S_{bx}	I	I	I	I	I	I	I	I	I	I	I	I
3.5		I	I	I	I	I	S_{bx}	I	I	I	I	I	I	I	I	I	I	I	I
3.6		I	I	I	I	I	S_{bx}	I	I	I	I	I	I	I	I	I	I	I	I

Table 2.11: The sequence of thermotropic phases obtained from the MD simulations of systems of $N = 1024$ elongated biaxial GB ellipsoids with an embedded electric dipole with a position offset with respect to the center of mass, and two on-axis and one off-axis orientations, models ox , oz and ot (see Table 4.24 and Table 2.9 and Table 2.10 for details).

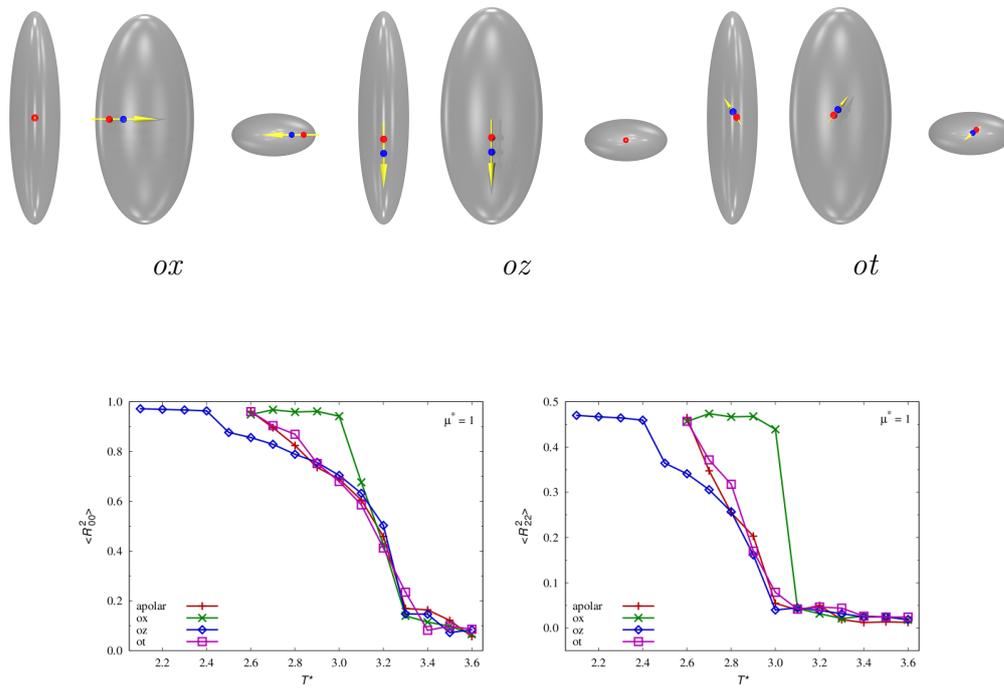


Figure 2.26: $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$ order parameter for system with $\mu^* = 1$ for models *ox*, *oz* and *ot* of Table 4.24.

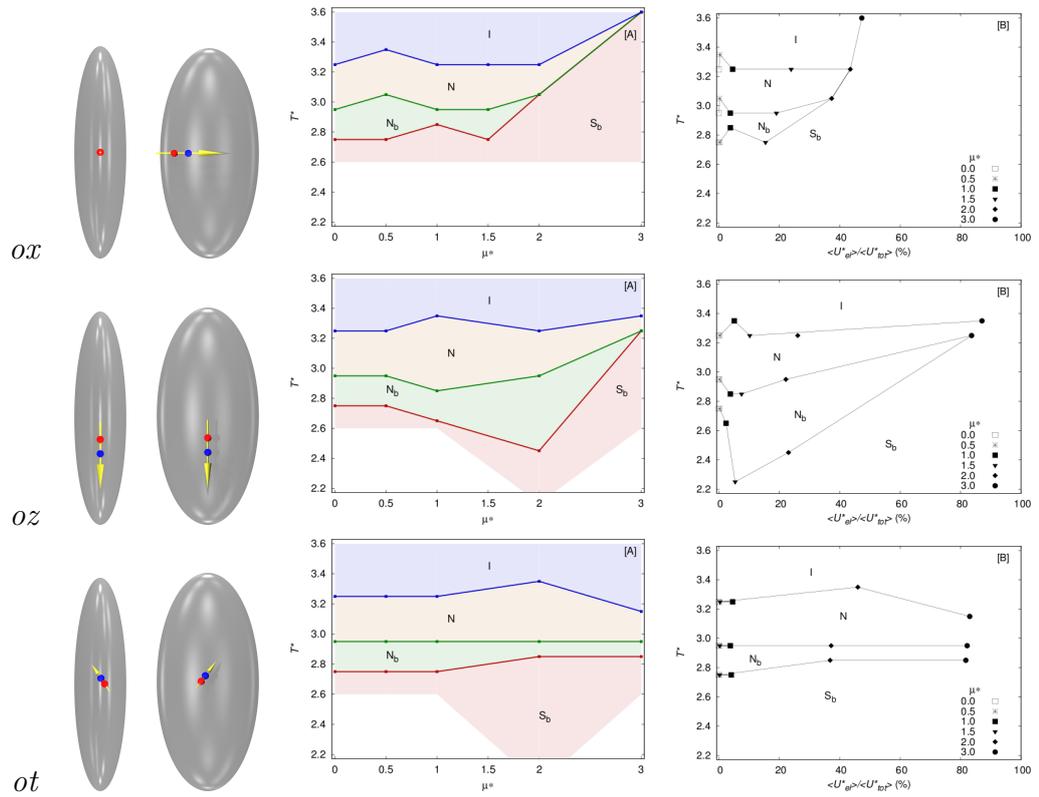


Figure 2.27: The transition temperatures for model *ox*, *oz* and *ot*. For further details, see the caption of Figure 2.22

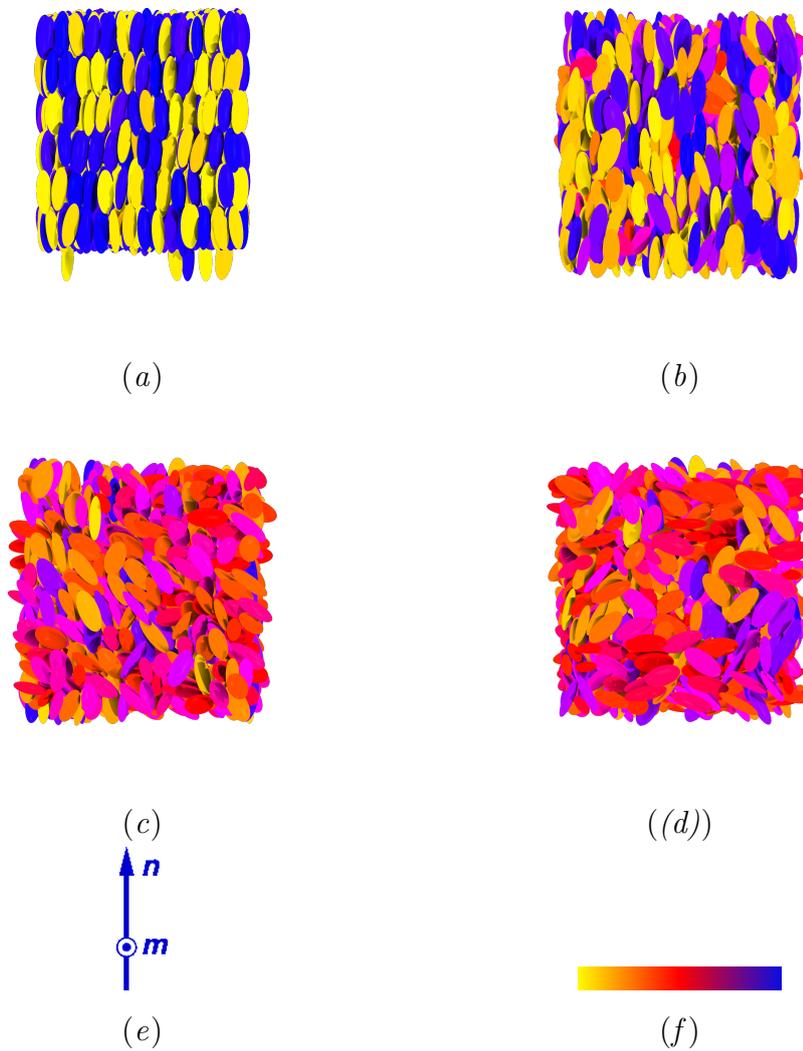


Figure 2.28: Representative snapshots of various condensed fluid phases formed by biaxial GB particles with an embedded off-axis (tilted with respect to the molecular long axis) electric dipole of dimensionless modulus $\mu^* = 1$ (see model *cx* of Table 4.24): (a) biaxial smectic at $T^* = 2.7$, (b) biaxial nematic at $T^* = 2.9$, (c) nematic at $T^* = 3.1$ and (d) isotropic at $T^* = 3.5$. We also show the reference frame with \mathbf{n} (principal) and \mathbf{m} (secondary) directors (e), and the palette (f) used to color code the orientation of the ellipsoids (ranging from yellow, for particles whose principal z -axis is parallel to the main director \mathbf{n} (i.e. they form an angle of 0°), to blue, for antiparallel orientations, at angle of 180°).

2.8 Conclusions

The main finding of our computer simulations is that for rigid D_{2h} particles, the electrostatic interactions between point dipoles of modulus comparable to the typical values found in thermotropic mesogens can be tuned, by a judicious positioning and orienting of the constituent electric charges, towards either destroying or enhancing the temperature stability for the N_b phase relying on a competition mechanism between the dispersive and the electrostatic interactions. By the same pathway, it is also possible to design dipolar particles whose mesogenic behavior and the stability of the N_b phase are not affected by dipole–dipole interactions. The first kind of behavior has been found for central symmetric dipoles: it is essentially independent of their orientation, and originates from the electrostatic interactions whenever they become relevant (usually $> 10\%$) within the total potential energy breakdown. The second and third cases are typical of models offset charges and by suitable tailoring the orientation of the dipole it is possible to either widen the temperature range of the N_b phase or either to stabilize it even in presence of very strong molecular dipoles and electrostatic interaction accounting for not less than 30% of the total potential energy. In all cases, when electrostatic interactions become dominant, any kind of nematic organization disappears and the cooling down of an isotropic fluid produces a highly structured biaxial smectic phase or crystalline layered structure.

2.9 Appendices

2.9.1 Dimensionless units

2.9.1.1 Electric charge

Also for charge we can find conversion factor, as explained below.

$$q^* = \frac{q}{\sqrt{4\pi\epsilon_0\epsilon_0\sigma_0}} \quad (2.9)$$

$$q^* = \frac{q}{q_0} \quad (2.10)$$

$$= \frac{1C}{\sqrt{1.113 \cdot 10^{-10} \frac{J}{V^2 m} \cdot 1.381 \cdot 10^{-21} J \cdot 5 \cdot 10^{-10} m}} \quad (2.11)$$

$$= \frac{1C}{\sqrt{7.6853 \cdot 10^{-41} \frac{J^2}{V^2}}} \quad (2.12)$$

$$= \frac{1C}{8.767 \cdot 10^{-21} \frac{J}{V}} = 1.141 \cdot 10^{20} \frac{C}{C} = 1.141 \cdot 10^{20} \quad (2.13)$$

$$1q_0 = \frac{1}{1.141 \cdot 10^{-20}} C = 8.767 \cdot 10^{-21} C \quad (2.14)$$

$$1q_0 = 8.767 \cdot 10^{-21} C \quad (2.15)$$

2.9.1.2 Electric dipole moment

Since $T^* = 1$ corresponds to $T = 100K$ ¹, we can find the value of ϵ_0 :

$$\epsilon_0 = \frac{T \cdot k_B}{T^*} = \frac{100K \cdot 1.381 \cdot 10^{-23} J/K}{1} = 1.381 \cdot 10^{-21} J \quad (2.16)$$

So for a system with $\sigma_0 = 5\text{\AA}$ and $\epsilon_0 = 1.381 \cdot 10^{-21} J$ we can find conversion factor for dipole moment. Useful values and units are:

¹This correspondence is true because we use 8CB as reference and $T_{N-I} \sim 350 K$ that for our 1-3 GB particles become $T_{N-I}^* = 3.50$

$$1D = 3.336 \cdot 10^{-30} \text{C m} \quad (2.17)$$

$$1e(\text{elementary charge}) = 1.602 \cdot 10^{-19} \text{C} = 1.602 \cdot 10^{-19} \frac{J}{V} \quad (2.18)$$

$$4\pi\epsilon_0 = 4\pi \cdot 8.854 \cdot 10^{-12} \frac{F}{m} = 1.113 \cdot 10^{-10} \frac{J}{V^2 m} \quad (2.19)$$

where ϵ_0 is permittivity.

The relation between dimensionless and real units dipole moment is:

$$\mu^* = \frac{\mu}{\sqrt{4\pi\epsilon_0\epsilon_0\sigma_0^3}} \quad (2.20)$$

Using the above relationships, we obtain that a dipole moment of 1D corresponds to a dimensionless value:

$$\mu^* = \frac{\mu}{\mu_0} \quad (2.21)$$

$$= \frac{3.336 \cdot 10^{-30} \text{Cm}}{\sqrt{1.113 \cdot 10^{-10} \frac{J}{V^2 m} 1.381 \cdot 10^{-21} J (5 \cdot 10^{-10})^3 m^3}} \quad (2.22)$$

$$= \frac{3.336 \cdot 10^{-30} \text{Cm}}{\sqrt{1.921 \cdot 10^{-59} \frac{J^2}{V^2 m^2}}} \quad (2.23)$$

$$= \frac{3.336 \cdot 10^{-30} \text{Cm}}{4.383 \cdot 10^{-30} \frac{J}{Vm}} = 0.761 \frac{\text{Cm}}{\text{Cm}} = 0.7611 \quad (2.24)$$

$$1\mu_0 = \frac{1}{0.7611} D = 1.314D \quad (2.25)$$

$$1\mu_0 = 1.314D \quad (2.26)$$

2.9.1.3 Example

In a cx model with two charges in position $r_+^* = -r_-^* = (0.1, 0, 0)$ with $|q^*| = 5$, we obtain in reduced units:

$$\mu^* = \sum_i q_i^* \mathbf{r}_i^* = 5 \cdot 0.1 + (-5) \cdot (-0.1) = 2 \cdot 5 \cdot 0.1 = 1 \quad (2.27)$$

Using S.I. units:

$$\mu = \sum_i q_i \mathbf{r}_i = (5 \cdot 8.767 \cdot 10^{-21} C) \cdot (0.1 \cdot 5 \cdot 10^{-10} m) + (-5 \cdot 8.767 \cdot 10^{-21} C) \cdot (-0.1 \cdot 5 \cdot 10^{-10} m) = \quad (2.28)$$

$$= 2 \cdot (5 \cdot 8.767 \cdot 10^{-21} C) \cdot (0.1 \cdot 5 \cdot 10^{-10} m) = 4.3835 \cdot 10^{-35} C \cdot m = \quad (2.29)$$

$$= 4.3835 \cdot 10^{-35} C \cdot m \cdot \frac{1D}{3.336 \cdot 10^{-30} C \cdot m} = 1.314D \quad (2.30)$$

Since:

$$1\mu_0 = 1.314D \quad (2.31)$$

the two values correspond.

2.9.2 Root mean square displacement

The mean square displacement in a simulation can be easily computed by its definition

$$MSD = \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle \quad (2.32)$$

where $\langle \dots \rangle$ denotes here averaging over all the particles. The MSD contains information on the atomic diffusivity. If the system is solid, MSD saturates to a finite value, while if the system is liquid, MSD grows linearly with time. In this case it is useful to characterize the system behavior in terms of the slope, which is the diffusion coefficient D :

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle \quad (2.33)$$

2.9.2.1 ox system $\mu^* = 0.5$

We have computed root mean square displacement for a system with offset transversal dipole of strength $\mu^* = 0.5$.

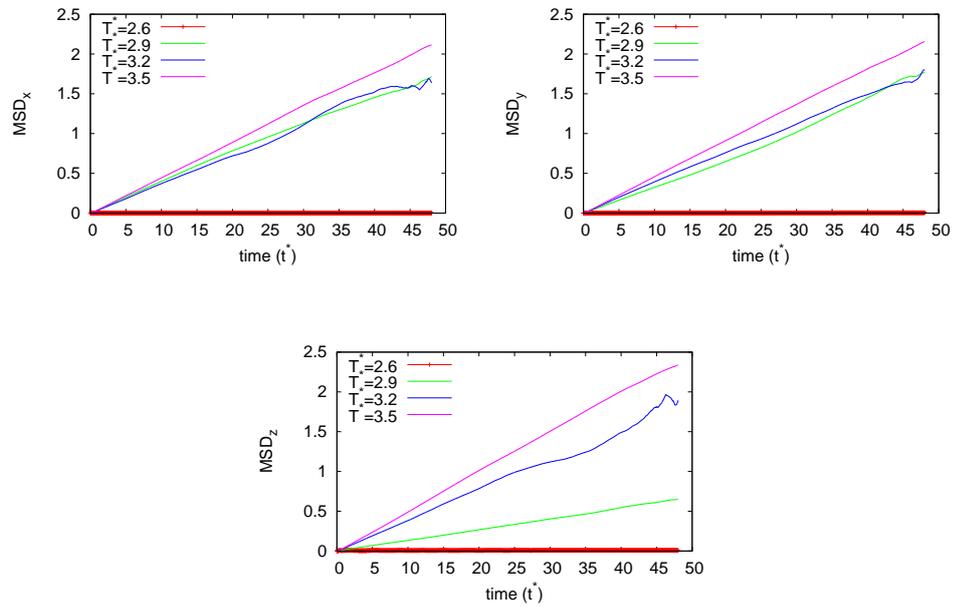


Figure 2.29: Root mean square displacement along x , y and z .

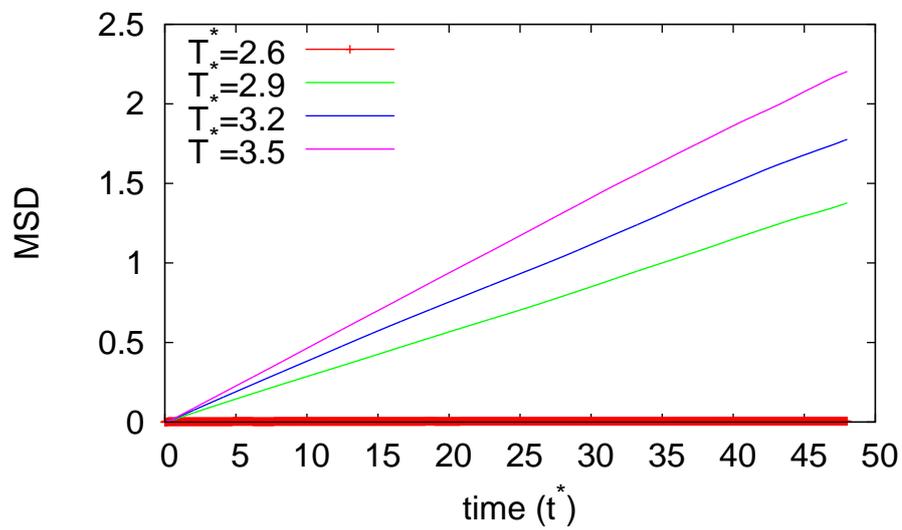


Figure 2.30: Root mean square displacement.

2.9.2.2 oz system $\mu^* = 2.0$

We have computed root mean square displacement for a system with offset longitudinal dipole of strength $\mu^* = 2.0$.

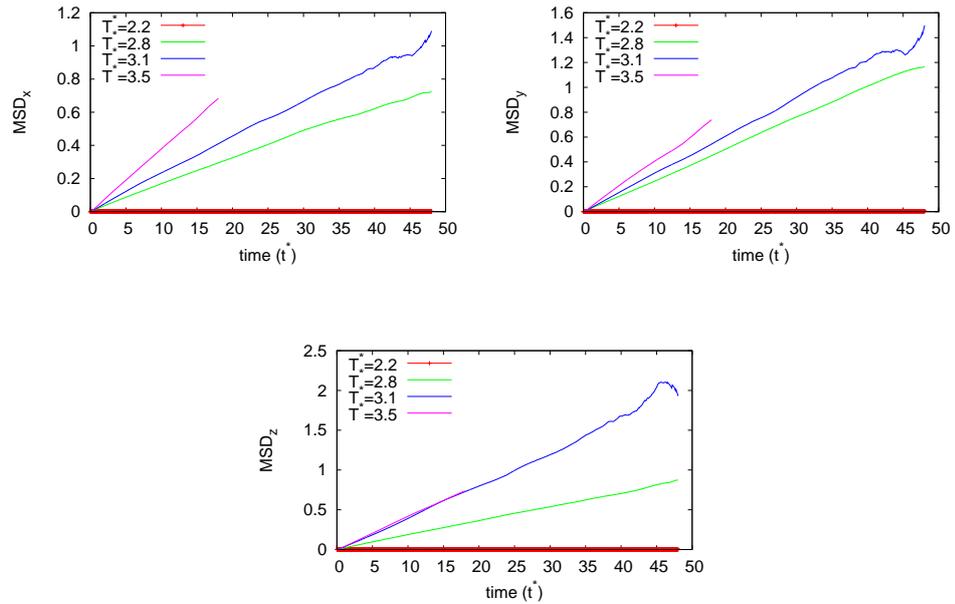


Figure 2.31: Root mean square displacement along x , y and z .

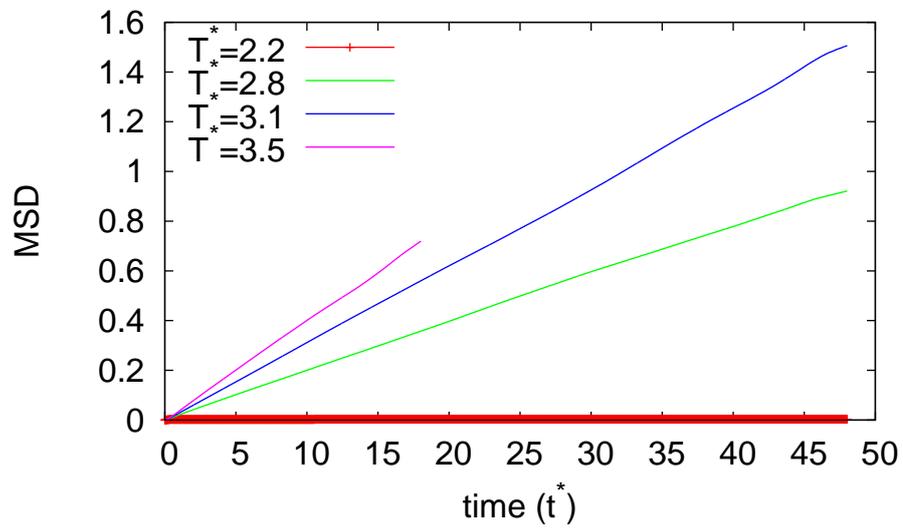


Figure 2.32: Root mean square displacement.

2.9.2.3 oz system $\mu^* = 3.0$

We have computed root mean square displacement for a system with offset longitudinal dipole of strength $\mu^* = 3.0$.

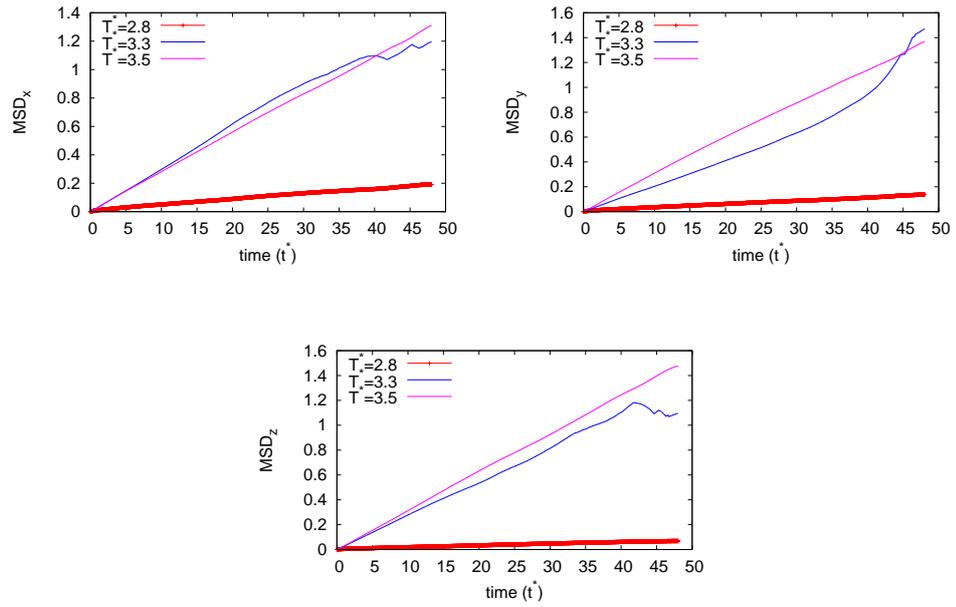


Figure 2.33: Root mean square displacement along x , y and z .

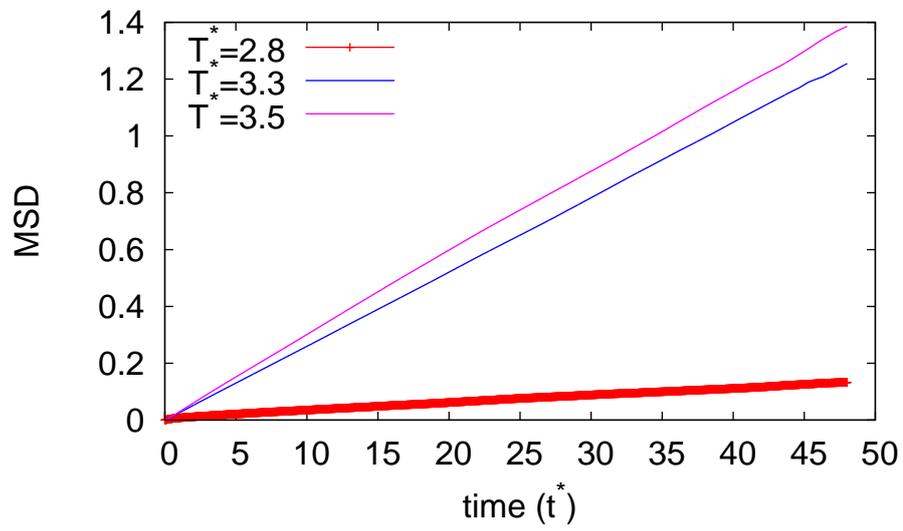


Figure 2.34: Root mean square displacement.

2.9.2.4 Mixed graphs

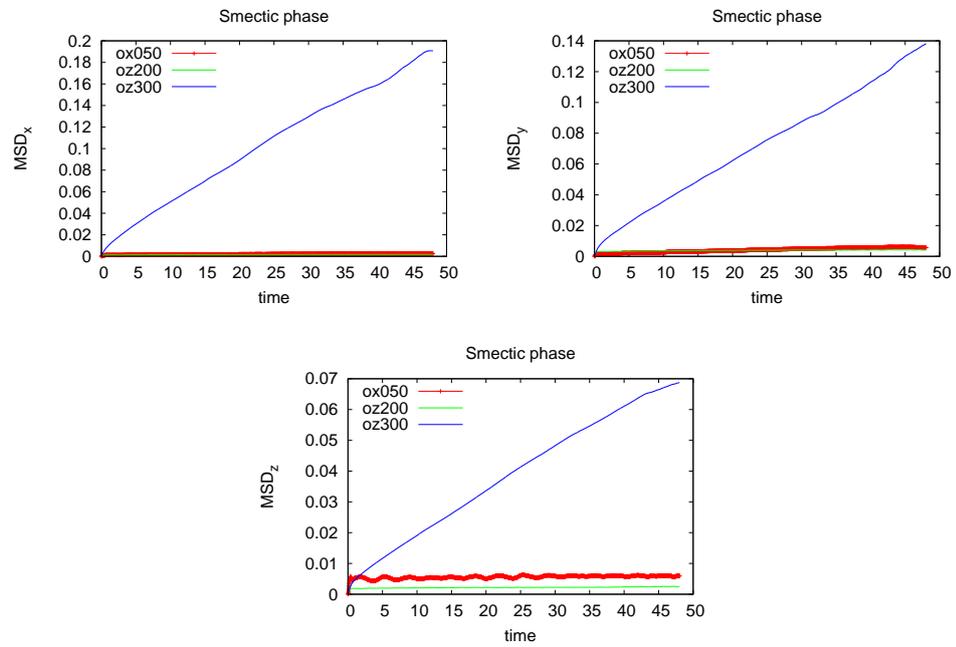


Figure 2.35: Root mean square displacement along x , y and z for smectic phase.

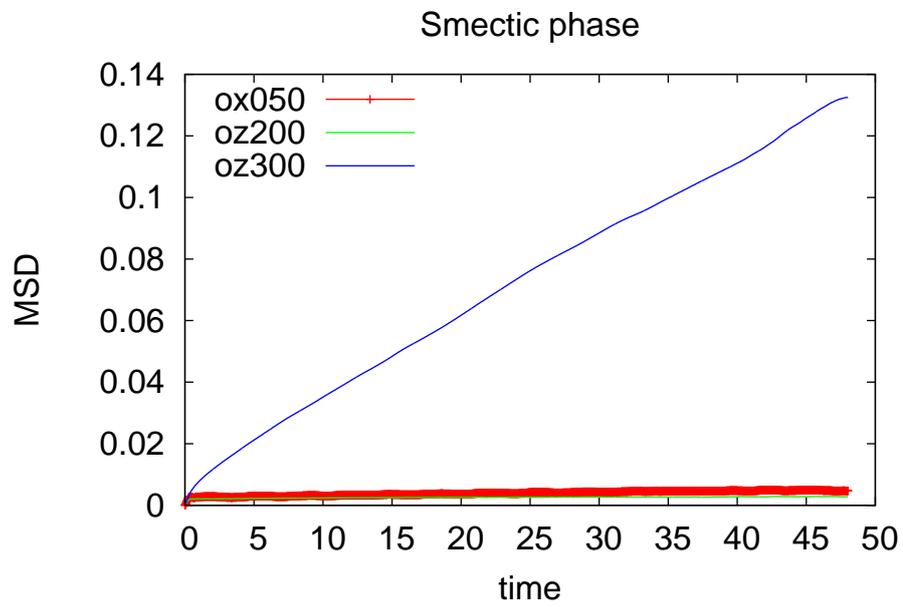


Figure 2.36: Root mean square displacement for smectic phase.

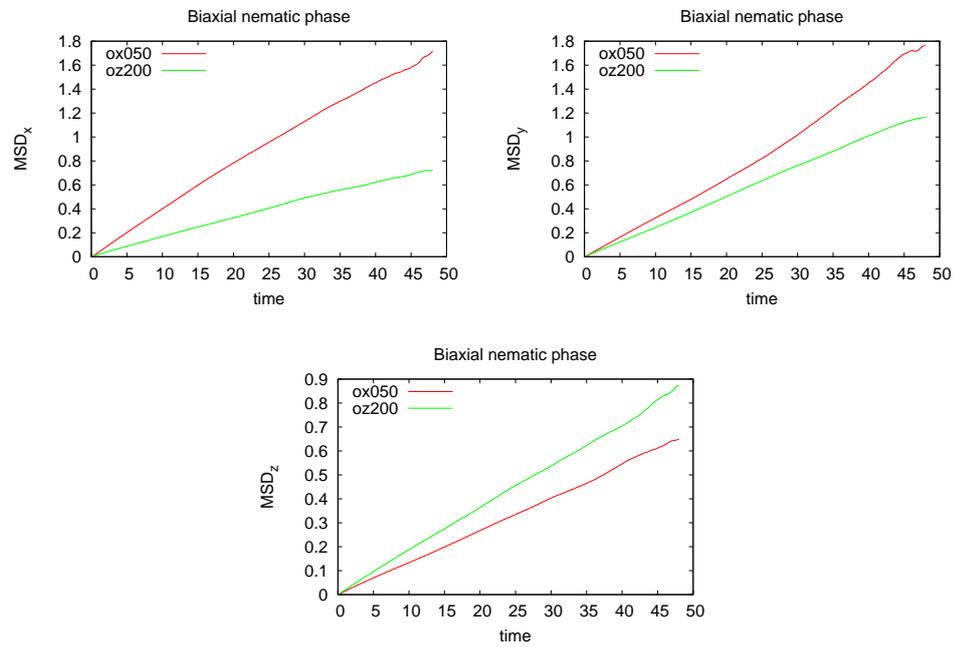


Figure 2.37: Root mean square displacement along x , y and z for biaxial nematic phase.

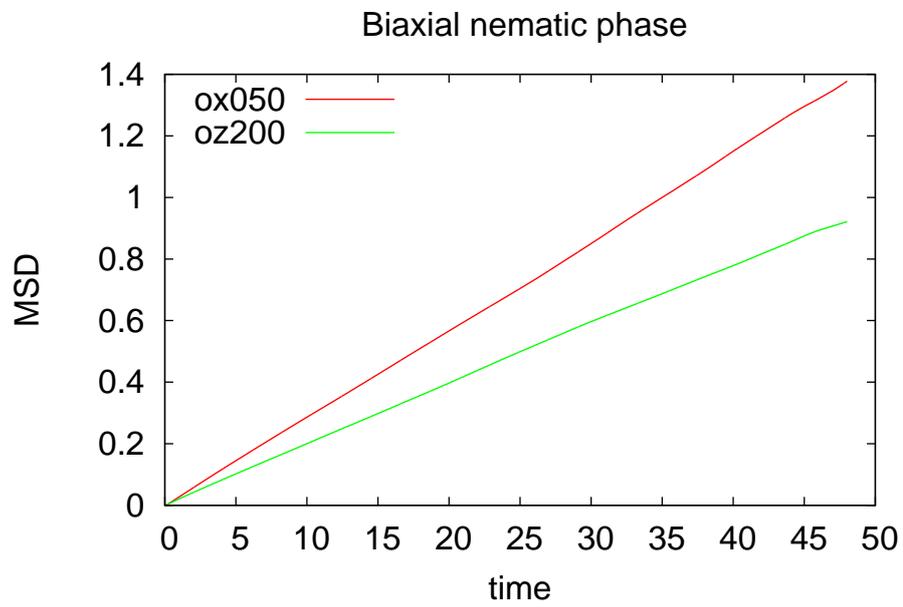


Figure 2.38: Root mean square displacement for biaxial nematic phase.

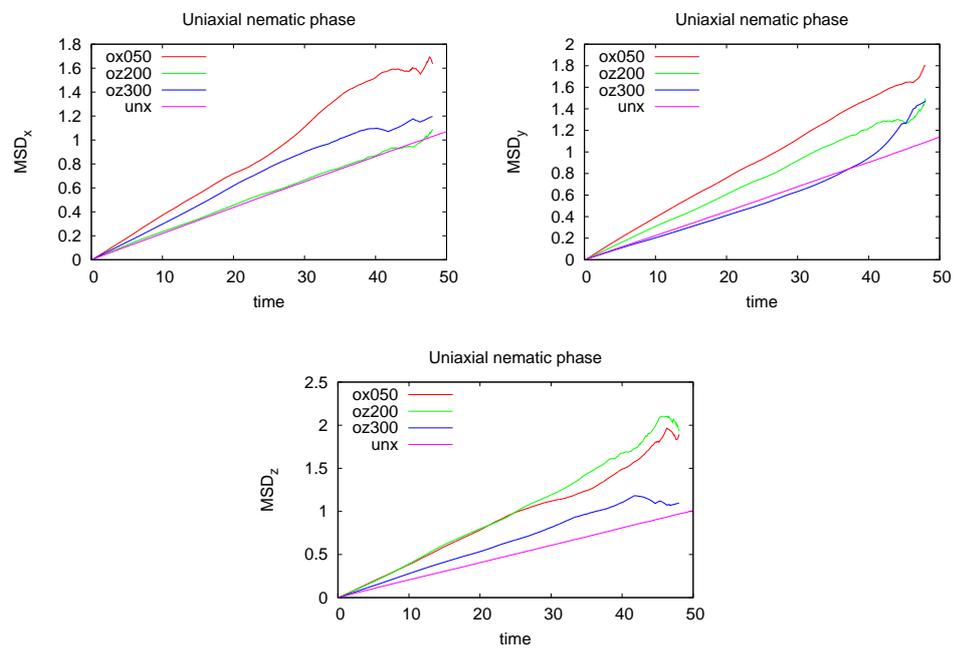


Figure 2.39: Root mean square displacement along x , y and z for uniaxial nematic phase.

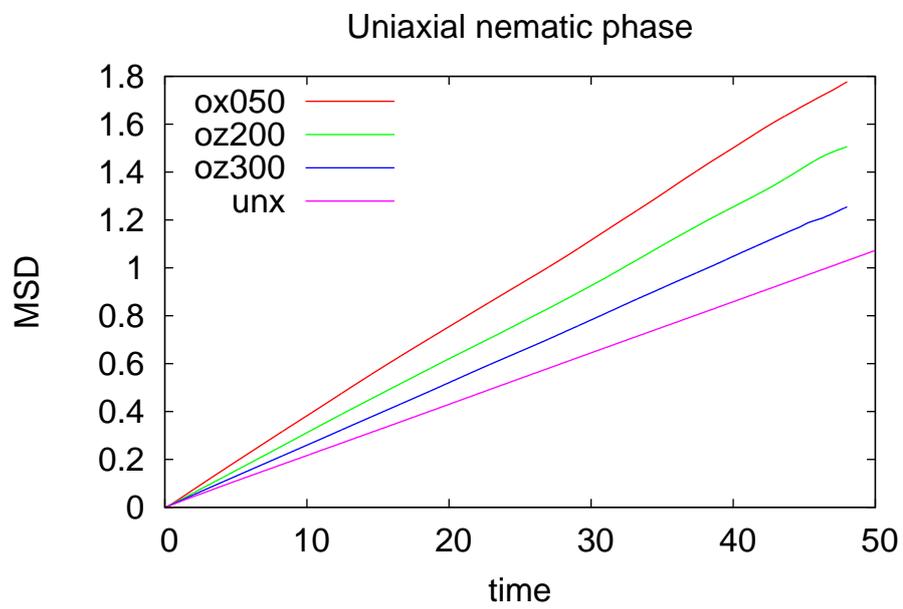


Figure 2.40: Root mean square displacement for uniaxial nematic phase.

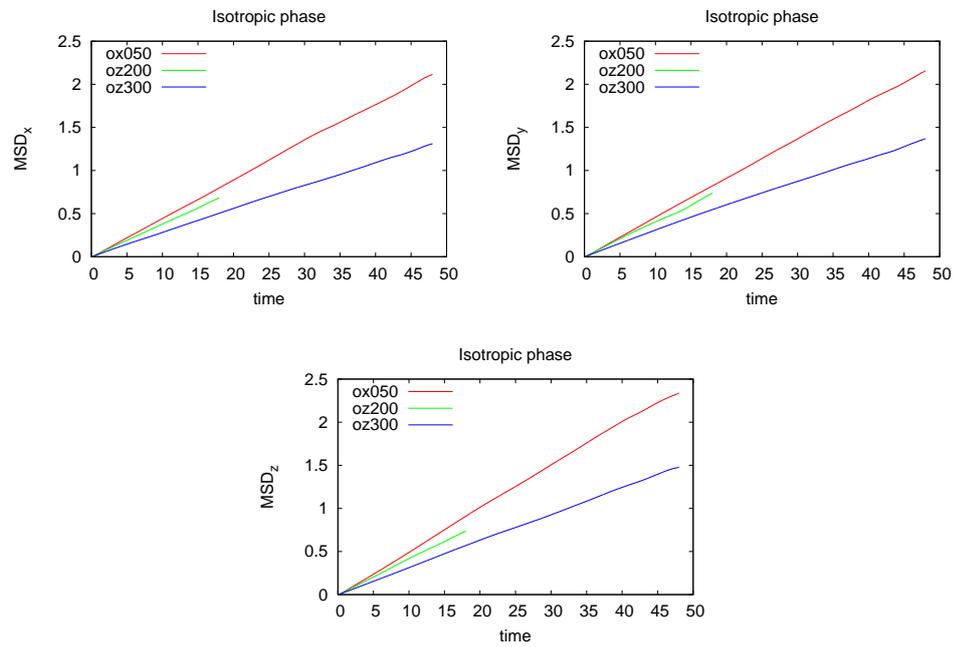


Figure 2.41: Root mean square displacement along x , y and z for isotropic phase.

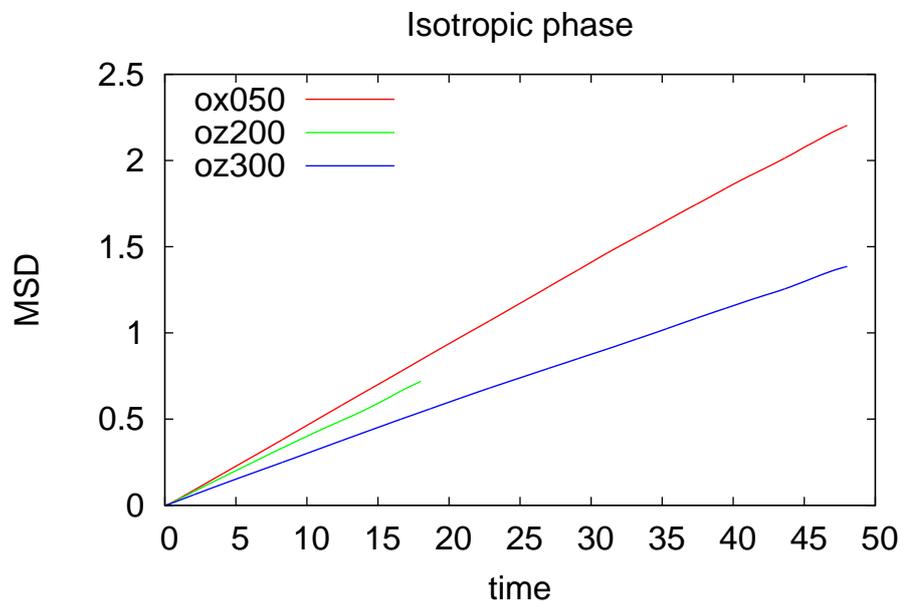


Figure 2.42: Root mean square displacement for isotropic phase.

2.9.2.5 Diffusion coefficient

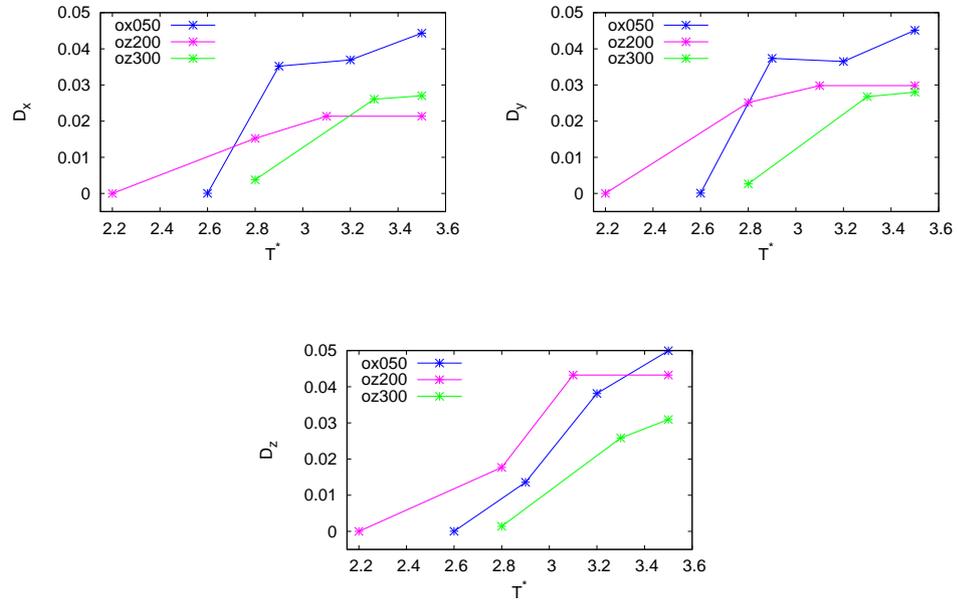


Figure 2.43: Diffusion coefficients for off-axis systems.

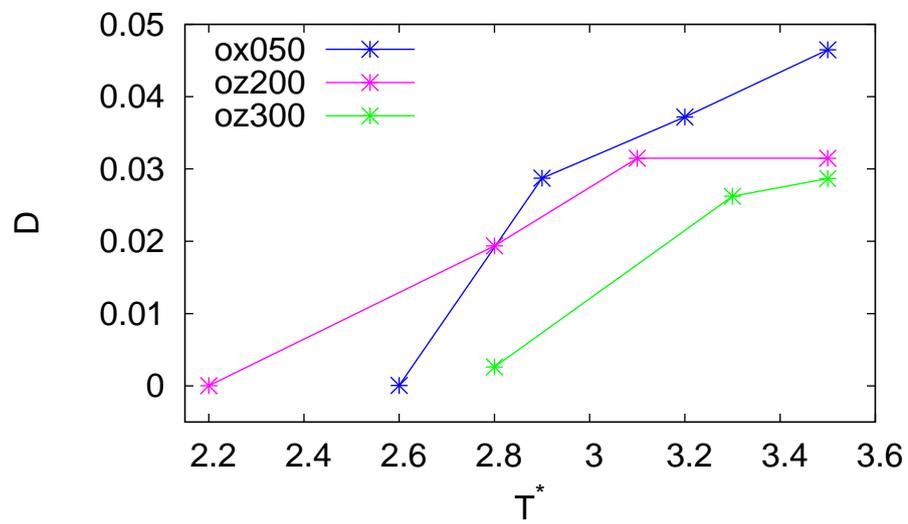


Figure 2.44: Diffusion coefficients for off-axis systems.

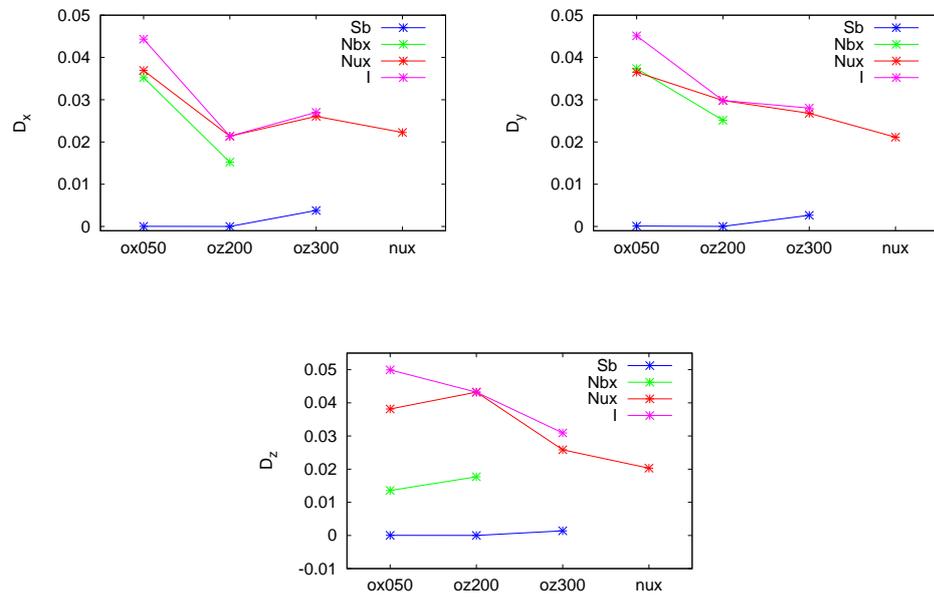


Figure 2.45: Diffusion coefficients for off-axis systems.

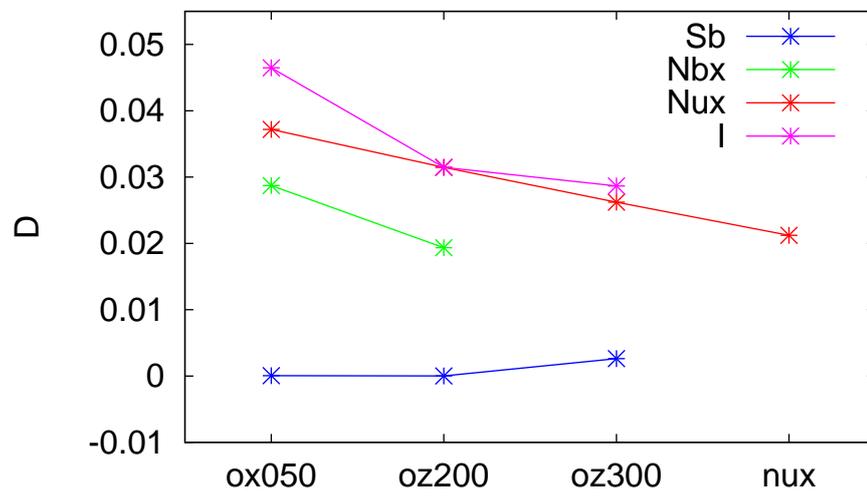


Figure 2.46: Diffusion coefficients for off-axis systems.

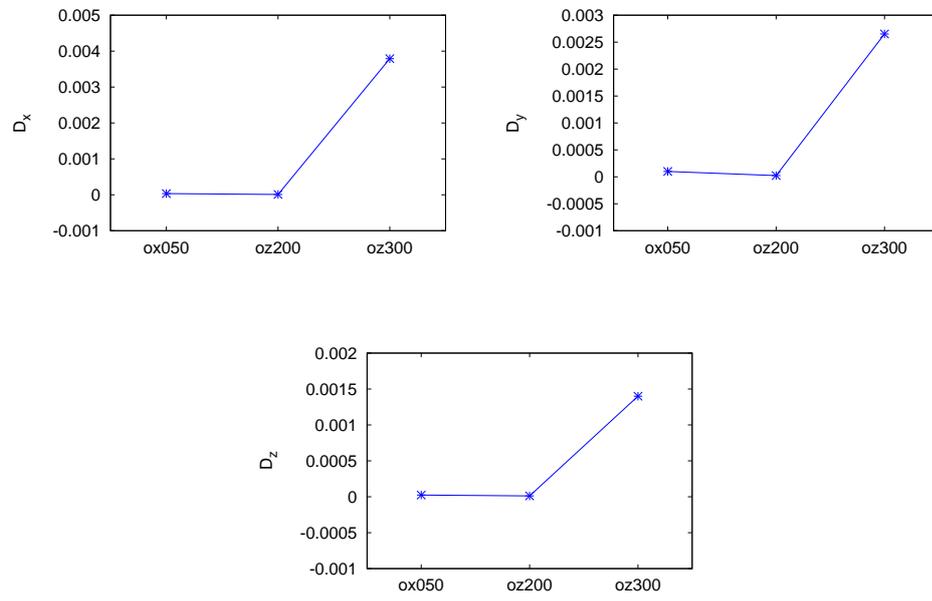


Figure 2.47: Diffusion coefficients for smectic phase of off-axis systems.

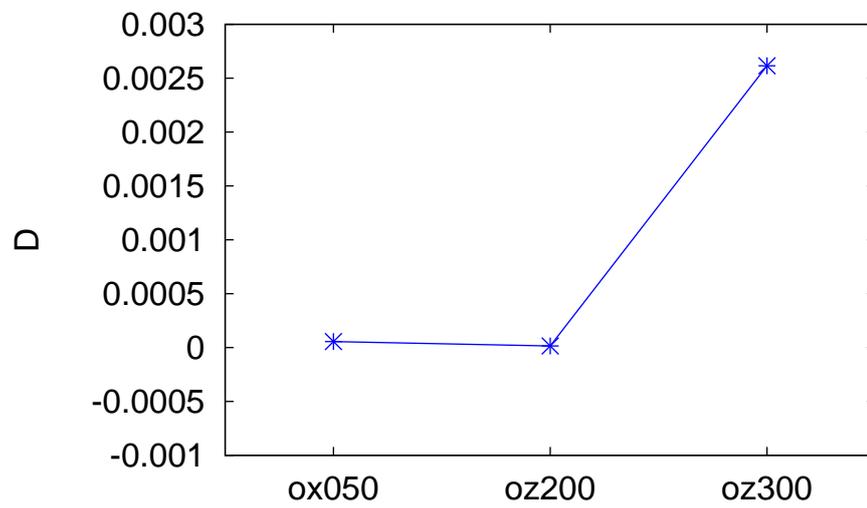


Figure 2.48: Diffusion coefficients for smectic phase of off-axis systems.

2.9.2.6 Conclusions

These results show that nematic phases are fluid-like and not glassy, since particles diffuse away from an arbitrarily chosen initial origin. The nematic phases have diffusion coefficients which are roughly 20–30% lower than those measured for the corresponding isotropic liquid, while for layered organizations these values are smaller by more than one order of magnitude.

2.9.3 Tilt angles for layered phases

We have computed the tilt angle between the director of the layer of a biaxial smectic (solid) phase and the average directions of GB particles, as the average orientation of molecules in each layer with respect to the perpendicular to the plane of the layer itself.

T/T_0	System	cx				
		$\mu^* = 0.5$	$\mu^* = 1$	$\mu^* = 1.5$	$\mu^* = 2$	$\mu^* = 3$
2.6		7.1	7.2	8.0	8.0	9.6
2.7		11.9	11.5	11.7	8.3	9.8
2.8				12.5	8.0	9.44
2.9					8.5	11.5
3.0					8.9	11.0
3.1						12.5
3.2						13.6

Table 2.12: Tilt angles for cx system.

T/T_0	System	cy		
		$\mu^* = 0.5$	$\mu^* = 1$	$\mu^* = 1.5$
2.6		7.4	7.3	7.3
2.7		10.9	8.1	8.3
2.8			8.5	7.4
2.9				10.9
3.0				7.6
3.1				8.4

Table 2.13: Tilt angles for cy system.

T/T_0 \ System	cz				
	$\mu^* = 0.5$	$\mu^* = 1$	$\mu^* = 1.5$	$\mu^* = 2$	$\mu^* = 3$
2.6	7.3	7.4	7.9	7.2	8.1
2.7	11.1	12.1	9.3	8.0	9.3
2.8			12.4	10.0	7.7
2.9				13.1	8.7
3.0					9.4
3.1					11.0
3.2					13.0
3.3					
3.4					
3.5					
3.6					

Table 2.14: Tilt angles for cz system.

T/T_0 \ System	ct1				
	$\mu^* = 0.5$	$\mu^* = 1$	$\mu^* = 1.5$	$\mu^* = 2$	$\mu^* = 3$
2.6	7.3	7.4	5.8	8.5	10.5
2.7	11.2	12.1	11.3	10.9	11.1
2.8			10.8	11.7	11.6
2.9				15.0	13.4
3.0					13.5
3.1					16.4

Table 2.15: Tilt angles for ct1 system.

T/T_0 \ System	ct2				
	$\mu^* = 0.5$	$\mu^* = 1$	$\mu^* = 1.5$	$\mu^* = 2$	$\mu^* = 3$
2.6	7.5	7.3	8.2	9.5	13.7
2.7	11.7	11.5	11.7	10.3	13.5
2.8			13.0	12.0	13.6
2.9				13.7	15.1
3.0					19.3
3.1					17.9

Table 2.16: Tilt angles for ct2 system.

T/T_0 \ System	ox				
	$\mu^* = 0.5$	$\mu^* = 1.0$	$\mu^* = 1.5$	$\mu^* = 2.0$	$\mu^* = 3.0$
2.6	7.1	7.4	7.7	7.7	23.1
2.7	12.4	7.9	8.1	7.3	22.8
2.8		8.3		7.8	19.9
2.9				7.8	22.1
3.0				8.7	31.2
3.1					25.9
3.2					31.5
3.3					31.0
3.4					30.2
3.5					*2
3.6					32.1

Table 2.17: Tilt angles for ox system.

System T/T_0		oz				
		$\mu^* = 0.5$	$\mu^* = 1.0$	$\mu^* = 1.5$	$\mu^* = 2.0$	$\mu^* = 3.0$
2.1	(-)	(-)	5.6	6.5	(-)	
2.2	(-)	(-)	7.4	6.7	(-)	
2.3	(-)	(-)		6.7	(-)	
2.4	(-)	(-)		8.1	(-)	
2.5	(-)	(-)			(-)	
2.6	7.7	8.6			8.6	
2.7	12.6				8.9	
2.8					11.6	
2.9					15.1	
3.0					17.1	
3.1					15.6	

Table 2.18: Tilt angles for oz system. Note that temperatures characterized by (-) have not been simulated.

System T/T_0		ot				
		$\mu^* = 0.5$	$\mu^* = 1.0$	$\mu^* = 1.5$	$\mu^* = 2.0$	$\mu^* = 3.0$
2.6	7.3	7.5	7.4	8.0	8.7	
2.7	12.6	11.7	12.0	11.4	8.7	
2.8				12.8	10.9	

Table 2.19: Tilt angles for ot system.

2.9.3.1 Conclusion

Usually, tilt angles increase with temperature, besides for cx system with $\mu^* = 2.0$ that shows almost the same tilt angle independently from the temperature. For systems with $\mu^* = 3.0$, these parameter values are higher than for systems with lower dipole intensity ($\theta < 14^\circ$). For all systems with the highest dipole strength the tilt angle are lower than 20° , besides the ox system ($\mu^* = 3.0$) for which angles range from 20° to 33° .

2.9.4 Order of phase transition

Using trends and histograms of total, electrostatic and Gay–Berne energies along with $\langle R_{00}^2 \rangle$ and $\langle R_{22}^2 \rangle$, we describe the order of the phase transitions of these systems. A transition of the first order is characterized by a discontinuous temperature profile of the derivatives of the free energy of the order parameters. Also the presence of double peaks in the histograms are typical of this kind of phase transition since they underline the coexistence of two different phases at the same temperature: increasing temperature these peaks does not modify their average values. For transition not of the first order, the trend will be continuous and the single peaks will shift with temperatures. We will take into account energy parameters for analyzing the transition between smectic and nematic phases, the $\langle R_{00}^2 \rangle$ for that between nematic and isotropic phase and $\langle R_{22}^2 \rangle$ for biaxial to uniaxial nematic (if there is). Figure 2.49 and Figure 2.50 show examples of the behavior of these parameters for transitions of the first and not of the first order (it is however possible that these transitions do not show a behavior typical of second order ones but they could be characterized by a transition that is a middle way between one of the first and one of the second order). Moreover, peaks in heat capacity are considered as a first–order transition behavior. Also moments and cumulants are computed. Following tables summarize order transition for systems studied.

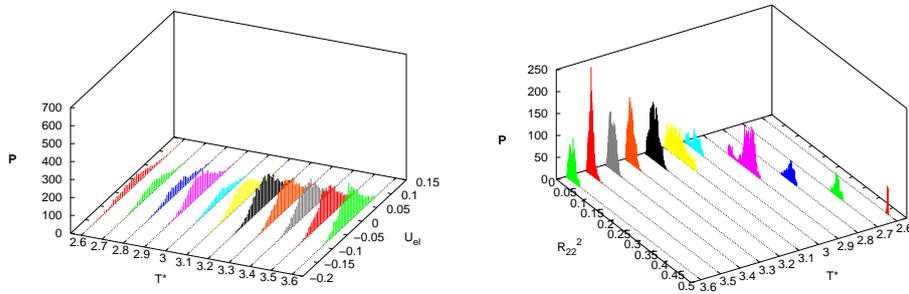


Figure 2.49: 3D histograms for electrostatic energy (for a transition not of the first order) and $\langle R_{22}^2 \rangle$ (for a transition of the first order). The example refers to a cz model with $\mu^* = 0.5$

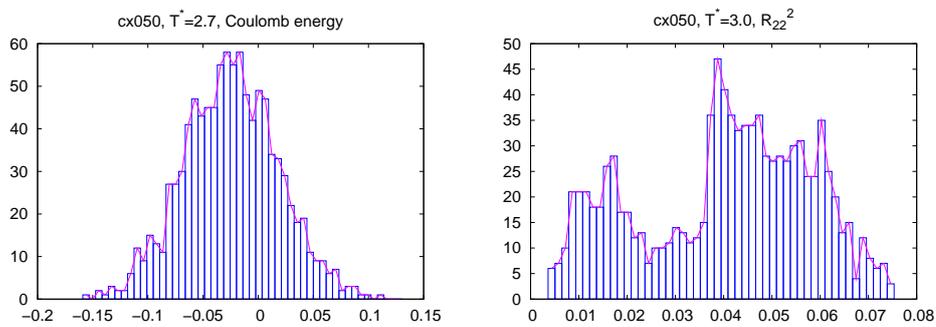


Figure 2.50: 3D histograms for electrostatic energy at $T^* = 2.7$ ($T_{S_b-N_b}$, a transition not of the first order) and $\langle R_{22}^2 \rangle$ at $T^* = 3.0$ (T_{N_b-N} , transition of the first order) for a cz system with $\mu^* = 0.5$.

2.9.4.1 Results

Transition \ μ/μ_0	cx				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	N	T_{S_b-N}	T_{S_b-N}
T_{N_b-N}	W	1	1	1	1
T_{N-I}	1	1	1	1	1

Table 2.20: Order of phase transitions for a cx system. “N” means that that transition is not of the first order, “W” that the transition is weakly of the first order and “1” the transition is of the first order.

Transition \ μ/μ_0	cy				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	W	W	T_{S_b-N}		
T_{N_b-N}	W	1	1		
T_{N-I}	W	1	N		

Table 2.21: Order of phase transitions for a cy system. See Table 2.20 for further information.

Transition \ μ/μ_0	cz				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	N	T_{S_b-N}	T_{S_b-N}
T_{N_b-N}	1	1	1	N	N
T_{N-I}	1	1	W	W	W

Table 2.22: Order of phase transitions for a cz system. See Table 2.20 for further information.

Transition \ μ/μ_0	ct1				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	N	T_{S_b-N}	T_{S_b-N}
T_{N_b-N}	1	1	1	N	N
T_{N-I}	1	1	1	1	1

Table 2.23: Order of phase transitions for a ct1 system. See Table 2.20 for further information.

Transition \ μ/μ_0	ct2				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	N	T_{S_b-N}	T_{S_b-N}
T_{N_b-N}	1	1	1	N	N
T_{N-I}	W	1	1	1	1

Table 2.24: Order of phase transitions for a ct2 system. See Table 2.20 for further information.

Transition \ μ/μ_0	ox				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	N	W	
T_{N_b-N}	1	1	1	1	
T_{N-I}	1	1	1	1	

Table 2.25: Order of phase transitions for a ox system. Note that the system with dipole intensity 3.0 is always S_b so there's no phase transition. See Table 2.20 for further information.

Transition \ μ/μ_0	oz				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	W	W	T_{S_b-N}
T_{N_b-N}	1	W	W	W	N
T_{N-I}	1	1	W	W	1

Table 2.26: Order of phase transitions for a oz system. Note that the system with dipole intensity 3.0 is always S_b so there's no phase transition. See Table 2.20 for further information.

Transition \ μ/μ_0	ot				
	0.5	1.0	1.5	2.0	3.0
$T_{S_b-N_b}$	N	N	N	N	N
T_{N_b-N}	1	1	1	1	1
T_{N-I}	1	1	1	1	1

Table 2.27: Order of phase transitions for a ot system. Note that the system with dipole intensity 3.0 is always S_b so there's no phase transition. See Table 2.20 for further information.

2.9.4.2 Conclusions

Usually, transitions between nematic and isotropic phases and between biaxial and uniaxial nematic phases are of the first order, while transitions between smectic and (biaxial or uniaxial) nematic phases are not.

Chapter 3

Coarse-grain modeling of DNA

3.1 Summary

We have been developing a coarse-grained three-sites per nucleotide model of DNA. Sugar and phosphate moieties have been considered as spherical interaction sites. To improve the published three-sites coarse-grained models which use only spheres, we chose GB ellipsoids to model bases, because of their shape closer to the atomistic structure than a sphere. The modeling procedure is as follows. First we have optimized the GB shape parameters of the various interaction sites to obtain the best approximation of the crystal structure of B-DNA. To determine the GB interaction parameters we have performed a preliminary MD simulation using an atomistic DNA model and the AMBER force field. The MD trajectory has been used to populate probability histograms and from those an effective energy profiles have been recovered via Boltzmann inversion. Note that each base will be described by an ellipsoid of different size. Anisotropic shapes also need an orientation to be specified: hence a quaternion has been associated to each base.

After collecting data on shape, position and orientation, the force field has been parametrized. First of all, we collect data from atomistic simulations of a 10 base pair long sequence and we compute histograms, using the Boltzmann inversion method. These data are now fitted using various

types of potential: the equations that fit better atomistic results are chosen and added to the MD code LAMMPS. We are currently checking the model obtained so far.

3.2 DNA

DNA is a double-stranded helical biopolymer composed of ordered sequence of repeating units, called nucleotides [46]. These monomeric units are made of three residues covalently bonded together: a phosphate group, a sugar and a base (Figure 3.1). The phosphate group give nucleic acids their acidity properties, as they are fully ionized at the physiological pH. The natural sugar in DNA is β -D-deoxyribose (a non-planar cyclic pentoses with a limited conformational flexibility, closely related to β -D-ribose in RNA). The bases are nitrogenous planar aromatic rings with lipophilic flat faces, and feature several hydrogen bond donors and acceptors along their edges.

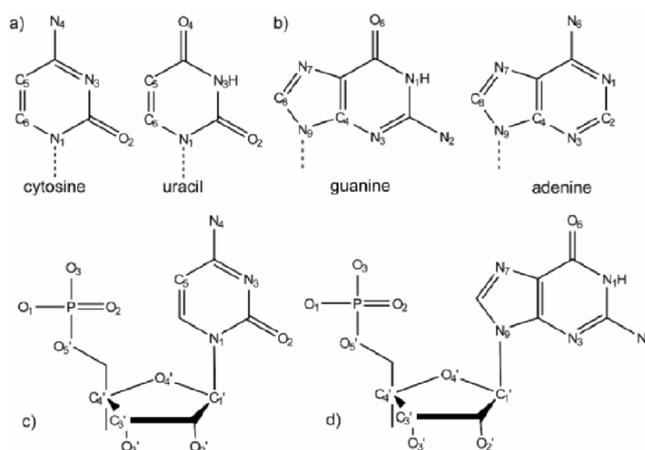


Figure 3.1: Chemical composition and nomenclature of nucleic acid components. a) Pyrimidines. Uracil occurs in RNA, DNA base thymine has a methyl group attached to C5. b) Purines. c) A pyrimidine nucleotide, cytosine-5'-phosphate. d) A purine nucleotide, guanine-5'-phosphate.

Nucleoside subunit is formed when a base and a sugar are linked by C-N glycosidic bond in the β -stereochemistry. These nucleosides are then linked through phosphate groups that are attached to the 3' carbon of one nucleotide and the 5' carbon of the other, hence the full repeating unit is a 3',5'-nucleotide. There are four types of nucleotides, which differ only by

the attached base and are divided in two classes. Adenosine-5'-phosphate (A) and guanosine-5'-phosphate (g) contain fused-ring purines (R), while cytosine-5'-phosphate (C) and thymine-5'-phosphate (T) are single-ring pyrimidines (Y). In RNA thymine is replaced by its demethylated form, uracil (U). Polynucleotide chains are numbered from the 5' end.

Bases can form pairs via hydrogen bonds between polar groups. Watson-Crick pairs are between G and C and between A and T/U. This is the most important pairing due to the fact that give high stabilization energies through hydrogen bonds, that the two pairs are almost isosteric, i.e. G-C and A-T have similar dimension along their long axes and that the minimal steric hindrance is furnished for a right-handed double helical conformations. There are various parameters to describe geometry of an isolated pair.

Secondary double helical structure is usually 22-26 Å wide and one nucleotide unit measure 3.3 Å [47]. The backbone of the two strands is composed by a sugar linked to two phosphate groups, with the C3' and C5' carbon atoms. This asymmetric bond cause each strand to have a proper direction: strands in a double-stranded DNA form has opposite directions. A sense strand is read in 5'-3' direction. Due to the Chargaff rules (that force C to form H-bonds only with G and T with A), the two antiparallel chains have complementary sequences.

3.2.1 Structural properties

Each DNA molecule contains the same information repeated twice, due to the complementarity of chains. This redundancy is biologically fundamental since it is an efficient model for replication of genetic information and if one of the two strands is damaged the other one is still conserved. Structural constrains to obtain the double helix form are:

- bilinearity: DNA is formed from two paired chains;
- complementarity: the chains are coupled with Chargaff rules;

- antiparallelism: the two chains are positioned in opposite sense: 5'–3' and 3'–5'.

The first two principles (bilinearity and complementarity) are related to the efficiency of duplication progress, while the last point is a structural constraint.

3.2.1.1 Symmetry properties

Nucleotide are intrinsically chiral molecules, since each sugar has at least three chiral carbons. Due to this property, a right- or left-handed Cartesian coordinate system can be defined. In particular, x axes associates with the direction of base chaining, y axes to the pairing direction and z axes is that from which the molecule is observed.

Four principles may be outlined [48].

- Principle of chain uniformity: each monomer has a “sense”, starting from the tail of a nucleotide and ending to the head of the next one, so that the chemical structure will be oriented in 5'–3' direction. Since also the strand has to maintain this verse, chaining can happen only if monomers has the same sense.
- Principle of complementarity: two bases can pair only with respect to the Chargaff rule.
- Mirror pairing: pairing direction is form the head of the monomer to that of paired monomer and every other monomer can pair only with one other monomer. The monomers belonging to a chain show the same chirality and, hence, the same orientation along the pairing direction: in such a way, the verses of the two paired chains are opposite. The advantage of this antiparallelism, with respect to the antiparallelism, is that you have only one reading verse.
- Free bilinear location: each molecule is free to arrange in each position of the chain, as long as the other principles are respected. So, the possibility that partial pairings happen is not excluded.

During the melting process (DNA replication), double-stranded DNA are denatured into two single stranded molecules: hence, pairing force is weaker than the chaining one. Moreover, it is worth noting that angle between phosphate group, sugar and next phosphate group is acute: so it is not possible that strands assume a bilinear configuration, only the antiparallel one is possible. In this arrangement, monomers are not exactly on the same plane since they are solvated, and a rotation of the backbone along x axes happens, causing the double-stranded form.

3.2.2 Mechanical properties

3.2.2.1 Geometry of DNA helices

Natural DNA double helices structures are A-DNA, B-DNA and Z-DNA. B-DNA is the form described by Watson and Crick [49] and seems to be preponderant in cells. A 10 base pair (bp) long isomer of this isomer is 23.7 Å wide and 34 Å long. The double helix pitch is 10.4/10.5 base pairs (number of bp per complete turn of the helix). This tilt frequency depends mostly on stacking forces that each base has on adjacent ones.

Also C, E, P, S, and the two enantiomeric forms D and L have been described [50, 51]. However, most of these structures have only been synthesized but they are not present in nature. Also triple-strand structures have been seen.

B-form of DNA is usually found in the aqueous environment of the cell, in presence of sodium ions. A-form is found only in dehydrated DNA samples, as those used for crystallographic experiments and maybe in hybrid pairs of DNA and RNA. Methylated DNA segments and protein-DNA complexes can adopt Z geometry, with a left-handed spiral.

3.2.2.2 DNA supercoil

B-isomer usually turns each 10.4/10.5 base pairs. However, many biological processes can induce a torsional stress and modify tertiary DNA

structure. An excess or a shortage of helical twisting is called respectively positive or negative supercoiling [52, 53] (Figure 3.2).

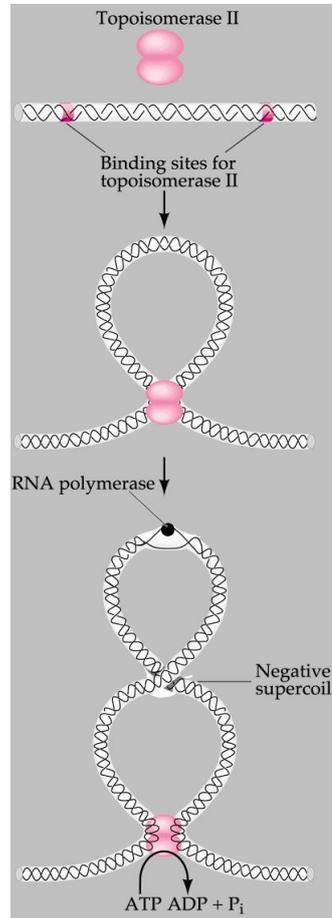


Figure 3.2: DNA supercoil.

3.2.2.3 Bending

DNA is a relatively stiff polymer, model as a worm-like chain. It has three significant degree of freedom (bending, twisting and compression), that limit the possible behavior of the DNA in the cell. Torsional and twisting stiffness is important for circularization of DNA and its direction when bounded to proteins. Axial and bending stiffness is also important for DNA wrapping. Extension and compression are not so important unless

in presence of high tension [54].

Persistence length and axial stiffness

In solution, DNA has not a unique firm structure but it change continuously due to thermal vibrations and collisions with water molecules. Hence, bending stiffness can be measured as persistence length, defined as length of DNA on which average time orientation of polymer becomes uncorrelated of factor e ". This value can be measured directly using an atomic force microscopy to obtain images of DNA molecules of different lengths. In aqueous solution, average persistence length is 46–50 nm or 140–150 base pairs¹ (Table 3.1), even if it can change significantly.

This parameter is referred to a section of DNA that somehow depends from the sequence. This can cause variation, due first of all to stacking energies and to residue lay in minor and major grooves [52, 53].

Sequence	Persistence length (base pairs)
Random	154 ± 10
$(CA)_{repeat}$	133 ± 10
$(CAG)_{repeat}$	124 ± 10
$(TATA)_{repeat}$	137 ± 10

Table 3.1: Example of B-DNA sequences and their persistence length.

Bending preferences

DNA molecules usually show a preferred bending direction (anisotropic bending), depending on bases in the sequence. A random sequence will not have any preferred bending direction (isotropic bending) [54].

Bending preferred direction is determined as stacking stability. If bases that stack together poorly are on the same strand of DNA helix, the molecule will bend so that to distance that direction. Increasing bending

¹DNA width is about 2 nm.

angle, also steric hindrance, the possibility to roll and paired residues (especially those laying in the minor groove) play an important role. Thymine and adenine residue will lay preferentially in minor groove in the internal part of bending. This effect is important in DNA–protein interaction, where a strong bond is induced.

DNA molecules with a high preferential bending can become intrinsically bent. An example is represented by sequences with 4–6 thymine and adenine residues separated by guanine– and cytosine–rich sequences with A and T laying in the minor groove on one side of the molecules. Intrinsically bended structure is induced from base pair propeller twist, that permits an unusual bifurcation of hydrogen bonds between base pairs. At high temperature, this structure and hence the intrinsic bending get lost [54]. All anisotropic DNA show a greater average persistence length and a greater axial stiffness. This increased stiffness is important to prevent random bending that will free the molecule to act isotropically.

3.2.2.4 Circular DNA

Circular DNA (Figure 3.3) depends on axial (bending) and torsional (rotational) stiffness of the molecule. In order to obtain a circular DNA, it has to be long enough to easily obtain a curvature to form a complete circle and it has to contain the correct number of base pair so that terminal bases will be in the correct rotation to permit bonding. Optimal length is around 400 base pairs (136 nm), with an integer number of helical turns, for example multiples of 10.4 base pairs. A molecule with 312 base pairs ($10.4 \cdot 30$) will form a circle structure a thousand time more quickly than a molecules with 317 ($10.4 \cdot 30.5$) base pairs.

DNA stretching

DNA long chains are entropically elastic and undergo tension. When DNA is in solution, various structural changes happen due to energy furnished by the solvent. This energy arises from thermal vibrations of molecules along with continuous collisions of water molecules. For entropic reasons, more compact and relaxes forms are thermally more accessible than stretched

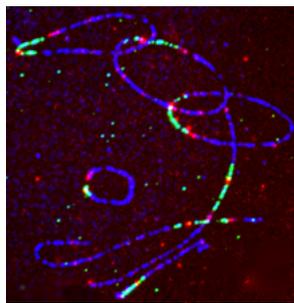


Figure 3.3: Circular DNA.

ones: hence DNA molecules are usually found in relaxed and tangled states. For this reason, a single DNA molecule will be stretched only by a force [54, 52, 53]. Under a sufficient tension and torsional force, DNA is believed to undergo a phase transition with the bases opening in a tilted way and phosphate moving to the center of the helix. Overstretched DNA structure is called P-form. Mechanical properties of compressed DNA have not yet been characterized due to experimental efforts to prevent polymer bending under compression force.

3.2.2.5 Superhelix DNA topology

Inside the cell, most of DNA is found in closed circles (as plasmids in prokaryotic) or as very long molecules whose diffusion coefficient produce actually closed fields (Figure 3.4). Linear sections of DNA are usually bounded to proteins or to physical structure as membrane so that they formed closed circles.

To analyze DNA topology [54] three parameters are used (Figure 3.4):

- *L – bond number*: bond number is the number of times a DNA strand roll around the other one. It is an integer number for a closed circle and is constant for a closed topological region. To compute this number, the molecule is flattened on a plane: it can not be modified twisting or wrapping molecules, unless DNA chains remains untouched.

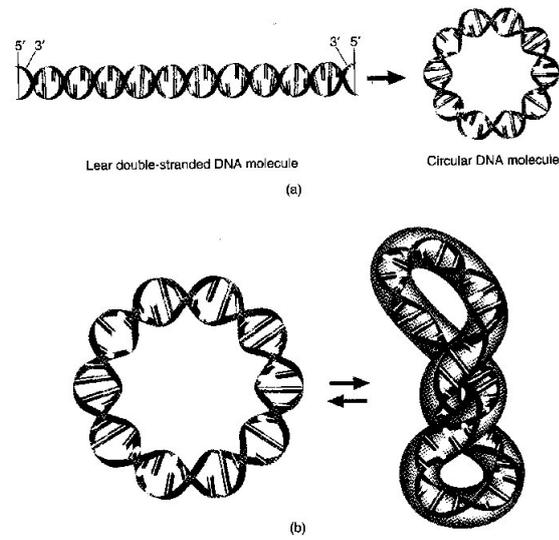


Figure 3.4: DNA topologies.

- *T – wrapping*: total number of helical turns that is usually equal to the number of turn of a DNA molecule in solution. It is the number of complete revolutions of a polynucleotidic chain around duplex axis, hence the number of bases divided by 10.4. Usually it shows positive value for right-handed duplexes.
- *W – torsion number*: number of turns that the axis of the double helix do around a superhelical axis. It is the measure of DNA supercoiling, that is zero when DNA is linear but circular, without superimposition.

Hence:

$$L = W + T \quad (3.1)$$

$$\Delta L = \Delta W + \Delta T \quad (3.2)$$

Each variation of T in a closed topological domain should be balanced by a change in W and vice versa: the resulting structure will have a higher

order. A closed molecule with zero torsion will be circular. If its twist increase or decrease with a supercoiling, torsion will be altered so that each molecule will a toroidal superhelicoidal coiling.

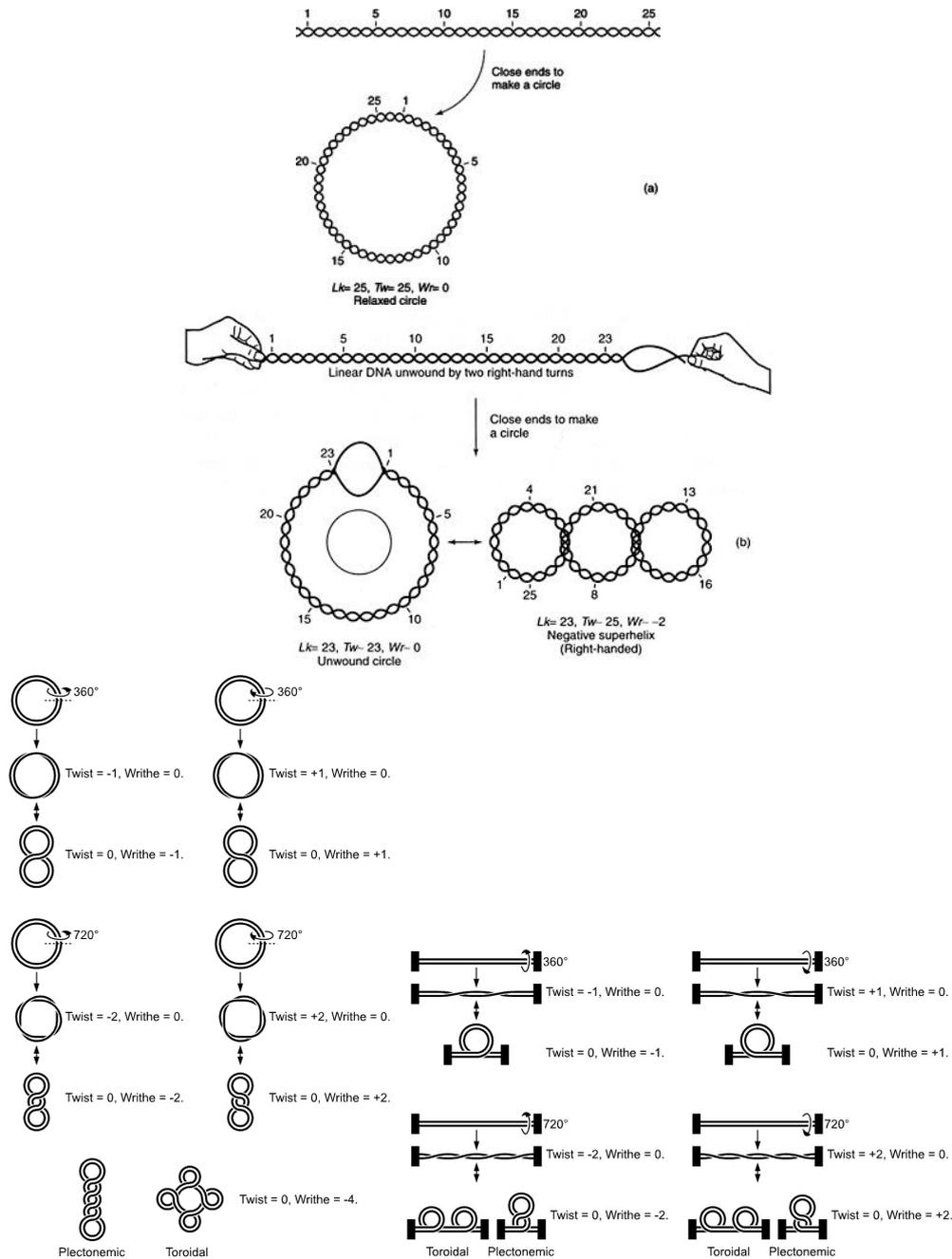


Figure 3.5: Example of analysis of various topologies.

When the circular form is obtained, strands are said to be topologically tangled. Topoisomerases unwind these structures to allow circular DNA replication and various linear DNA recombination.

3.2.2.6 DNA melting

During the DNA melting process [52, 53], interactions between strands of double helix are broken. These bonds are weak and easy to burst with a little warm up, due to specific enzymes or with physical forces. DNA melting occurs preferentially in certain points of DNA. An adenine– and thymine–rich region is melted more easily than a region containing a lot of guanine and adenine. Some step between base pairs are more efficiently broken in the melting process, in particular TA and TG. That the reason why many sequences as TATAA are used at the beginning of genes helping RNA polymerases during transcription (for which DNA melting is needed). For example, human DNA with high percentage of guanine and adenine (almost 50%) can melt at 70°, while *Streptomyces* bacterium, rich in guanine and cytosine, melts at 85°.

Melting temperature of DNA depends also on composition of solvent (Figure 3.7). A high ionic force, for example high NaCl concentration, is favorable to double–stranded helix state (increasing the melting temperature, T_m) due to the high concentration of sodium cations screening negative backbone phosphate charges [54].

DNA melting temperature depends also on the efficiency of base pairing. A synthetic DNA double–helix composed by some mismatched bases shows a lower melting temperature rather than that of a double–stranded DNA, with all base pairs precisely paired. This last property is an important feature for the use of DNA to detect similar sequences in DNA of another species. For example, the part of the DNA coding for a human cell enzyme can form double helices with mouse DNA sequences coding for the same enzyme; however, human–human and mouse–mouse duplexes melt at higher temperature with respect to mouse–human hybrid.

Bond break with a small heating (as PCR) is easy for molecule with less

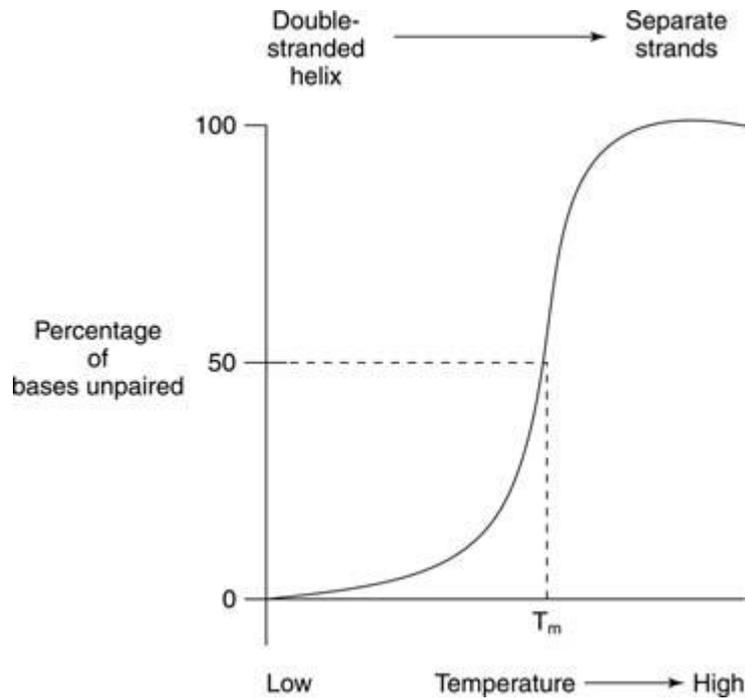


Figure 3.6: Melting temperature determined through percentage of denatured base pairs.

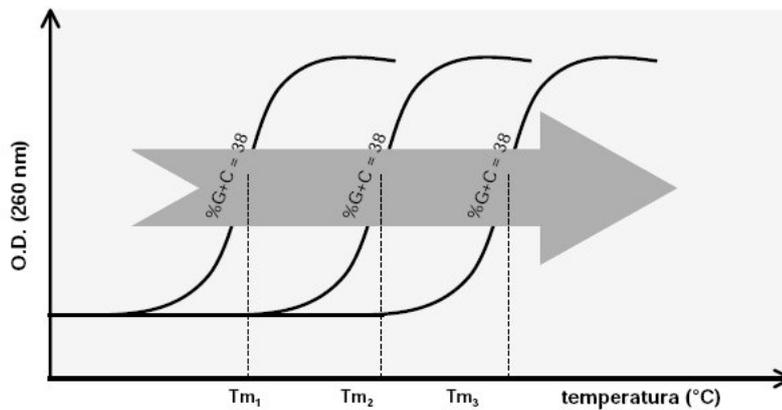


Figure 3.7: Effect of ionic force on DNA melting.

than 10 kbp. Bonds between the two strands of DNA make separation between long segments more difficult. The cell avoid this problem enzymes melting DNA, helicases, to work along with topoisomerases, that permit

the break of the backbone of one strand so that it can rotate around the other one. Helicases untangle strands to make easier the movements of enzymes reading sequence, as DNA-polymerases.

This process modify also qualitatively chemical–physical properties of DNA. The high viscosity of native DNA solution, due to the resistance to the deformation of its stiff sticks–like molecules, decrease significantly when DNA denatured in flexible single strands. In such condition, DNA UV absorption, depending almost totally on its aromatic bases, increase of the 40% due to the interactions between bases closed to each other. Evaluating the variation of light absorption at a specific wavelength (usually 260 nm) when increasing temperature a boost in light absorption happens. This behavior indicated that melting of DNA is a cooperative event where the collapse of a part of the structure destabilize the remaining part.

Hence UV light absorption can be used to analyze melting and helix formation. The bases that show stacking interactions screen each other from the light. As a consequence, the double helix DNA UV absorption, measured at 260 nm, is smaller than that of random coil DNA strands (Figure 3.8). This effect is called hypochromicity of double helix, i. e. “less color” relative to the decrease of absorbance of a material. It is possible to detect DNA melting also with fluorescence spectroscopy (Figure 3.9).

This process is a first order transition, for what concern thermodynamics. Melting temperature is usually identified as the point where the percentage of paired bases is %50. However, thermodynamically, this temperature is the point where the difference between the two energy minima is almost zero.

3.2.3 Bases

3.2.3.1 Base pairing geometry

In order to maintain Watson and crick geometry sugar groups should be bound to the respective bases in an asymmetrical way in the same side of base pairs. This asymmetric disposition defines mutual positions of two DNA strands: the atoms on the surface of sugar–phosphate backbone de-

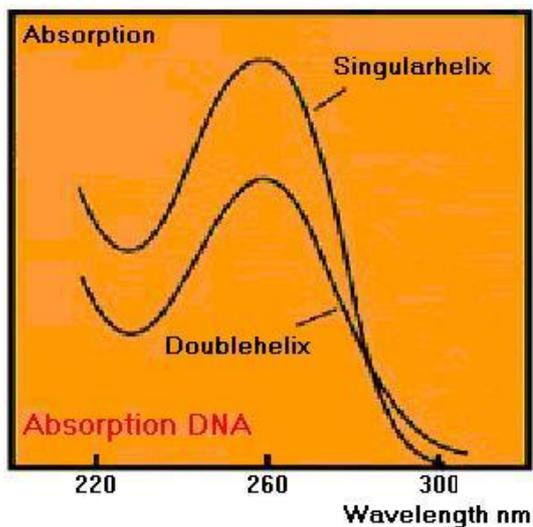


Figure 3.8: UV absorption melting DNA analysis.

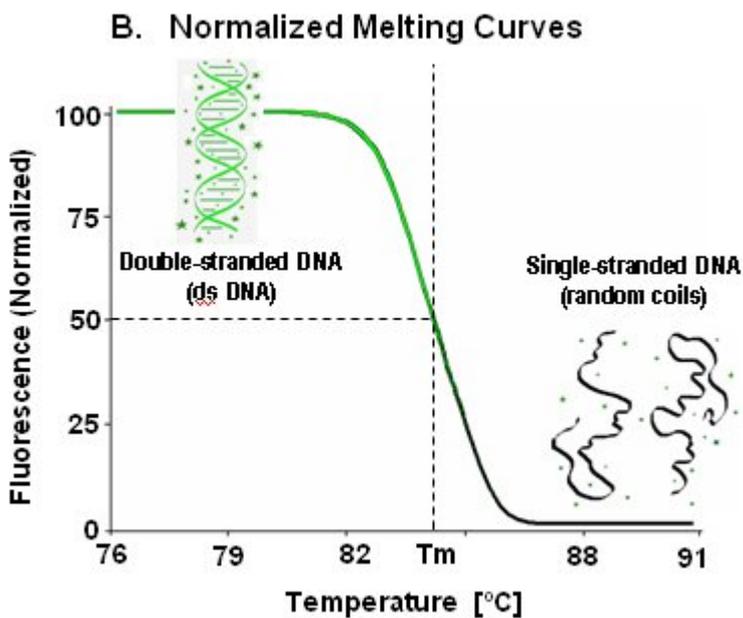


Figure 3.9: Fluorescent spectroscopy DNA melting analysis.

fine the major and minor grooves. The major groove is defined as that showing C6/N7/C8 purines atoms and their substituents or C4/C5/C6 pyrimidine atoms and their substituents, while minor groove is character-

ized by the presence of C2/N3 purines and C2 atom of pyrimidine and their substituents. The two base pairs should have almost identical dimensions in order to maintain this structure [52, 53].

Single bases are flat aromatic rings, but bases keep together by hydrogen bonds not stiff can show flexibility. Vertical disposition of bases and base pairs is flexible and kept almost completely by stacking interaction between π electrons of bases rings. This flexibility depends on base nature, partially on base pairs, but in particular from stacking environment. Hence, morphological description of bases has gained importance to describe and understand sequence-dependent properties and nucleic-acid deformations. These characteristics are usually considered necessary to comprehend what happen for what concern nucleotides, while for long-range effects like helix bending a more globally analysis should be done.

Many rotational and translational parameters are used to describe these geometrical relations between bases and base pairs, defined by 1989 Cambridge Agreement (Figure 3.10)

The *propeller twist* (w) [54], i.e. torsion between helix and the bases, is the dihedral angle between normals to bases view along the long axis of the base pairs. The angle usually shows a negative sign, with a clockwise rotation of the closest base viewed along the longest axis. This axis for a couple of purine-pyrimidine base is defined as the vector between C8 atom of the purine and pyrimidine C6. Analogous definitions can be applied to non standard pairs (purine-purine and pyrimidine-pyrimidine duplexes). The *buckle* (κ), i.e. the deviation from the idealized structure, is the rotation of one base with respect to the other in the same base pair (dihedral angle between bases), along the short axis, after helix torsion is set to 0° . The change of such deviation for subsequent steps is called “cup” and is useful to measure transformations happened along a sequence. The “cup” is defined as the difference of deviation at a certain step and that at the previous step.

The *inclination* (η) is the angle between base pair long axis and a plane perpendicular to helical axes. This angle is defined positive for a right-handed rotation around the vectors that goes from the helix towards major

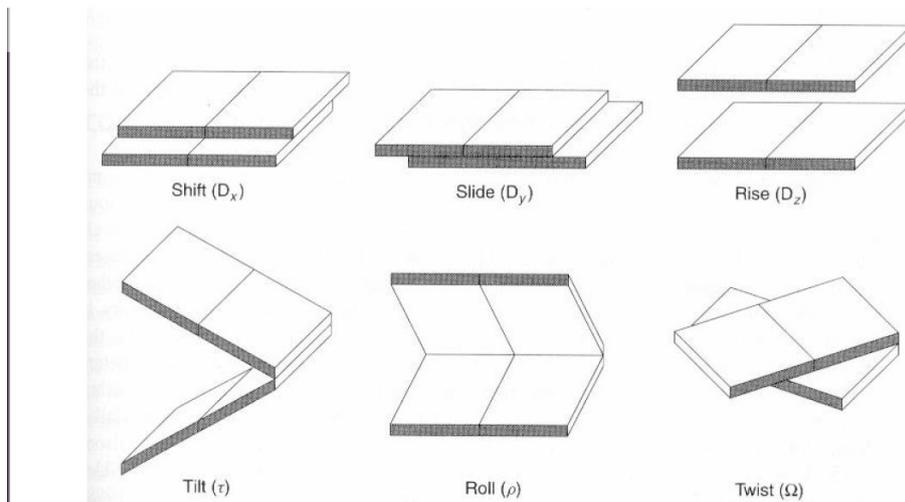
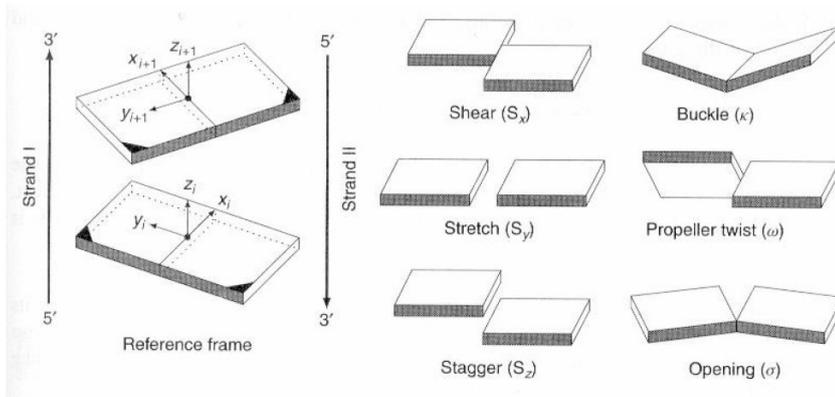


Figure 3.10: Parameters describing DNA helix.

groove.

The X and Y displacements [54], i.e. displacements along x and y axes,

define translation of base pairs in the plane plane in the middle with respect to central point of the long axis of the base pair starting from the helix axis. Displacement along x axis happen towards major groove when its value is positive. The displacement along y axis is orthogonal to the previous one and is positive if it aims to the first strand of the duplex.

The parameters for the other step of base pairs are three. The *helical twist* (Ω) [54], i.e. helical torsion, is the angle between consecutive bases measured as the variation of vectors C1'–C1' orientations, from one pair to the next one, projecting down the helical axis. To obtain a double helix repeating exactly, helical torsion should be $360^\circ/n$, where n is the repeating unit (nucleic acid).

The *roll* (ρ) corresponds to the dihedral angle for rotation of a base pair with respect to the base pair next to it, along the greatest axis of the base pair itself. A positive angle represents a “step” (a base pair) towards minor groove.

The *tilt* (τ) is the corresponding dihedral angle along the short axis (x axis) of the pair.

The *slide* [54] is the relative displacement of a base pair with respect to the other in the direction of the strand of the first nucleic acid (Y displacement), measured between the central points of C6–C8 axis along the base pair.

Geometry of a base pair can be completely characterized with six coordinates [54] (Figure 3.10): *rise*, *twist*, *slide*, *shift*, *tilt* and *roll*. The *shift* is the displacement along one axis in the plane perpendicular to that of base pairing, directed from minor to major groove. The *tilt* is the rotation around this axis. The *slide* is the shifting along an axis in plane on which base pairs are directed from one strand to the other; the *roll* is the rotation around this axis. The *rise* is the displacement along the helix axis and the *twist* the rotation around this axis.

Twist and *rise* determine the verse of helical rotation and helical pitch. The other coordinates can be zero. Usually *slide* and *shift* are small in B–DNA, while they have relatively high values in A– and Z–DNA. The *roll* and *tilt* force the next base pairs to be less parallel and have small values.

X ray analysis of B-DNA oligomers, it is clear that pyrimidine-purine interactions (C-G, A-T and C-A = T-G) usually follow the *roll* movement in opposite directions along their longest axis so that to open the side of the minor groove (“positive *roll*”), while a purine-pyrimidine interaction (G-C, A-T and A-C = G-T) usually undergoes a negative *roll*. The *tilt* value has also been used for the deviation of the axis of the base pairs between two strands, with respect to the helical axes. This parameter corresponds to a *slide* between subsequent base pairs; in the helix-based coordinates is better called “inclination”.

Note that local structure flexibility is a property of B-DNA depending also on its sequence.

3.2.3.2 Base pairing

Purines form hydrogen bonds with pyrimidines. A pairing between two pyrimidines will be energetically disadvantageous since the molecules will be too far apart from each other to create hydrogen bonds; on the other hand, the pairing between two purines is unfavorable since molecules will be too close and there will be an electrostatic repulsion. Since hydrogen bonds are not covalent bonds, they can be break and reformed quite easily. The two DNA strands of a double helix can be separated, both due to a mechanical force and to a high temperature.

Each pair if thymine and adenine base has two hydrogen bonds (Figure 3.11): each base act both as an electron donor and acceptor. The base pair guanine-cytosine comprehends three hydrogen bonds: guanine is acceptor for one of this bonds and donor for the other two. A DNA rich in guanine and cytosine is more stable than one with only a few of these bases, ma this is not due to the extra H-bond of the GC pair, but mainly to stacking interactions. This underline that the hydrogen bond furnish to bases the capability to pair selectively, but molecule does not gain any extra stability. Hence, the total length of the double helix of DNA and it GC content determine the association force between the two strands. DNA double helices with a high GC content show two strongly bounded strands,

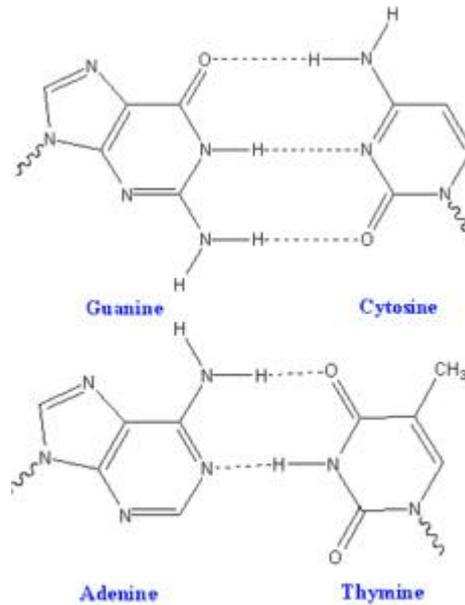


Figure 3.11: Base pairing.

while short double helices or regions of dsDNA with a high AT content are characterized by weakly bounded strands. This has a reflection in biology: when a double strands of DNA is needed to open easily, as for TATAAT Pribnow box of some promoters, it will be contain a high AT content.

This interaction can be measure founding the temperature needed to broke hydrogen bonds (i.e. the melting temperature). When all base pairs of the dsDNA melt, strands fall apart and exist in solution as two completely independent molecules. These molecules of single-stranded DNA does not have a unique structure, even if some configuration are stable than others. In conclusion, melting temperature depends on molecular length, GC content and number of mismatches. If there are a lot of guanine and cytosine in the dsDNA, the molecule will have high melting temperature: indeed it is not surprisingly that genome of extremophile organisms, as *Thermus thermophilus*, is particular rich in guanine and cytosine. On the contrary, regions of genomes that have to separate frequently, like promoters of gene that are often transcribed, are relatively poor in CG content [52, 53].

3.2.3.3 Base stacking

The base stacking interactions donate stability. It is due to the π stacking of aromatic rings belonging to the bases. The stacking interactions between guanine and cytosine with adjacent bases are the most favorable. Note that stacking interaction between GC bases with the next base is geometrically different from the CG interaction. These kinds of effects are important especially in RNA secondary structure: for example, RNA clover structures are stabilized by base stacking in long region [52, 53].

3.2.3.4 Base analogues and intercalation

Chemical analogues can substitute nucleotides and stabilize their non-canonical base pairs, leading to error (point mutations usually) in DNA replication and transcription. DNA intercalators are chemical molecules that take place easily in the space between adjacent bases on a single strand and force DNA-polymerase, during replication, to skip a nucleotide or to insert an extra one in the intercalation site [52, 53]. The most part of intercalators are composed by big polyaromatic systems and they are carcinogenic (like ethidium bromide).

3.2.4 Sugar-phosphate backbone conformation

The deoxyribose is not planar: this effect is called “puckering” [54]. The exact conformation of a sugar ring can be completely specified using five endocyclic torsion angles. The puckering of the ring derives from the effect of non-bonding interactions between the substituents of four carbon atoms: the most stable configuration shows all substituents as far as possible. This puckering can be described easily in terms of torsion angles inside the ring.

In principle, there is a continuum of interconvertible puckers, separated by energy barriers. These sugar conformations can be produced by systematically varying torsion angles of the ring. The pucker can be briefly described by P and τ_m parameters. The value of P , the pseudo-

tation phase angle, indicate the type of puckering, since P is defined using five torsion angles (τ_{0-4}):

$$\tan P = \frac{(\tau_4 + \tau_1) - (\tau_3 + \tau_0)}{2\tau_2(\sin 36^\circ + \sin 73^\circ)} \quad (3.3)$$

while τ_m refers to the maximum angle of the conformation and is specified by the follow equation:

$$\tau_m = \frac{\tau_2}{\cos P} \quad (3.4)$$

The pseudorotation phase angle can assume any values between 0° and 360° . If τ_2 has a negative values, the angle should be increased of 180° . This angle is usually represented as the pseudorotation of a wheel (Figure 3.12), that indicates the continuum of ring conformation. The τ_m values characterize the degree of puckering of the ring: experimental values obtained from crystallographic studies on mononucleosides are in the range between 25° and 35° . The five torsion angle are not independent, hence every single atom τ_j can be represented with only two variable:

$$\tau_j = \tau_m \cos[P + 0.8\pi(j - 2)] \quad (3.5)$$

A great number of distinct geometries for deoxyribose ring conformations has been experimentally observed with crystallographic or NMR techniques. When an atom in a ring is outside the plane of the other four, the type of puckering is so called letter-like. However it is more frequent that two atoms deviate from the plane of the other three, with those two lay in any part of the plane. Usually one of this two atom show a higher deviation from the plane, giving a stressed conformation. The atomic displacement direction from the plane is important: if the deviation is greater from the same side of the base and along a $C4'-C5'$ bond, the atom will be called *endo*, in the opposite case *exo*. Most common crystallographic conformations of isolated nucleosides and nucleotides are close both to $C2'-endo$ and $C3'-endo$. The $C2'-endo$ family show P values comprised between 140° and 185° , due to the position in the “puckering wheel” this

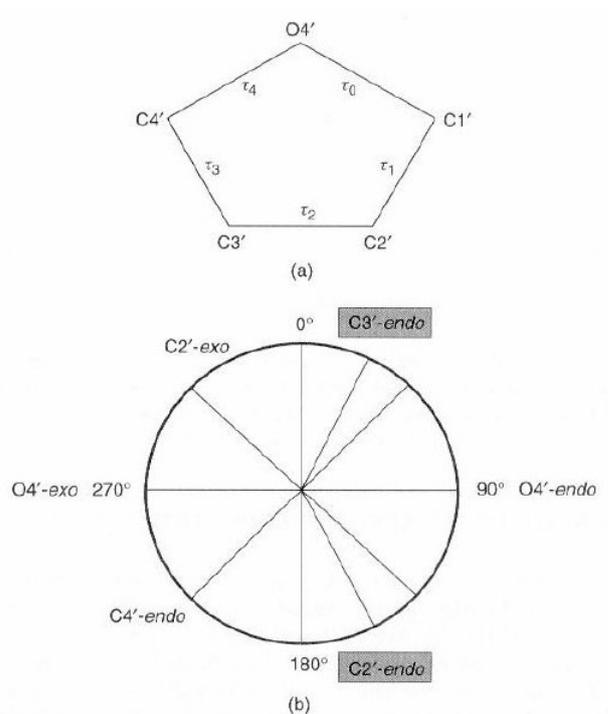


Figure 3.12: Figure (a): The five torsion angles inside a ribose ring. Figure (b): The pseudorotation wheel for a deoxyribose. The highlighted areas shows the most common pseudorotation angles..

structure is sometimes called S (south) conformation. The region $C3'$ -endo is characterized by P values comprised between -10° and 40° and it is called N (north) configuration. These structures are rarely observed, due to different ring substituents. As a consequence, the puckers are better described by torsion conformation. When the deviation outside the plane is greater in the *end* side, there will be a smaller deviation in opposite side (*eso*). To describe a twisted conformation of twisted deoxyribose is defining the deviation outside the greater plane followed by the minor one (for example, $C2'$ -end, $C3'$ -eso).

The pseudorotation wheel indicated that sugar conformations are inter-

changeable: hence there are barriers between the main conformations. The exact dimension of these barriers has been widely studied and now there's agreement on the fact that their height depends on the path around the pseudorotation wheel. For example, the preferred path for *C2'-endo* to *C3'-endo* interconversion goes through the *O4'-endo*, with a barrier of 2.5 kcal/mol, from some experimental data, or 1.5 kcal/mol from molecular dynamics studies. The first experimental value represent the total free energy for interconversion.

Puckers relative populations can be directly monitored with NMR spectroscopy measuring the ratio of coupling constants between protons H1'-H2' and H3'-H4'. On the contrary of freezed puckers present in solid structure of nucleosides and nucleotides, in solution a rapid interconversion happen. Relative population of most common conformations depends on the base type they are bounded to. The purine bases has a preferred conformation, the *C2'-endo*, while pyrimidines favor the *C3'-endo* one. The nucleoside with deoxyribose are usually (i 60%) in the *C2'-endo* conformation, while ribonucleosides in the *C3'-endo* one. The origin of this preferences lies in the non-bonded interactions between sugar ring substituents and somehow also in electrostatic properties. The ribose *C3'-endo* conformation will show hydroxyl groups in 2' and 3' positions further than in the *C2'-endo* pucker. The ribonucleosides have a lower mobility. Correlations between sugar conformation and backbone parameters have been with crystallographic and NMR studies, both in isolate nucleotides and in oligonucleotides. The sugar puckering are important factor for determine oligo- and polynucleotide conformation since it can alter C3', C4' and C1' substituents orientation, causing a big variation in the backbone and in the whole structure.

Glycosidic bond connect a deoxyribose sugar and a base with a C1'-N9 bond for purines and C1'-N10 for pyrimidines. χ torsion angle around this bond can assume a wide range of values, but structural constrains narrow it. Glycosidic torsion angles are defined with four atoms: O4'-C1'-N9'-C4 for purines and O4'-C1'-N1-C2 for pyrimidines.

Two main low-energy domains have been theoretically predicted for gly-

cosidic angle, according to experimental data. The *anti* conformation is characterized by N1,C2 for purines and C2,N3 for pyrimidines side turning from ring sugar so that hydrogen atoms bonded to purine C8 and pyrimidine C6 will be stretched on the sugar ring. Hence, Watson and Crick base pairs are not directed to to the sugar ring. These orientations show a *sin* conformation, with groups forming the H-bonds oriented towards O5' atom of the sugar. Analyzing many crystalline structure of pyridinic nucleosides, an hydrogen bond between sugar O5' and base N3 atoms stabilize this conformation. An exception to this rule is found in high guanosine content nucleoside that slightly prefer the *sin* conformation due to favorable electrostatic interactions between N2 of guanine amino group and phosphate group in 5' sense. For pyrimidinic nucleotide, *anti* conformation is preferred due to unfavorable contact between base O2 atom and 5' phosphate group. The resulting molecular mechanics minimization energies for all the four DNA nucleotides in *sin* and *anti* conformation (using AMBER force field) completely agree with this observations. Sterical preferred ranges for the two glycosidic angles domains are between -120° and 180° for the *anti* conformation and between 0° and 90° for the *sin* one. The χ value in region around -90° is often defined as "high *anti*". Correlations between sugar conformation and glycosidic angle exist: *sin* glycosidic angles are not found with *C3'-endo* conformation due to sterical hindrance between base and H3' atom, pointing at the bases in this form.

The backbone of a oligonucleotide has six torsion angles, called $\alpha, \beta, \gamma, \delta, \varepsilon$ and ζ , besides the five internal torsion angles of the sugar $\tau_0, \tau_1, \tau_2, \tau_3$ and τ_4 and glycosidic angle χ . As already shown, many of these parameters have highly correlated values. Hindrance reasons alone would not take into account limited ranges of angles. Also in this case, a wheel showing permitted values has been drawn (Figure 3.13).

Angles α, β, γ and ζ have three possible ranges each while ε has a wide range containing two separated regions: hence there is a high number of low energy conformations possible for nucleotides, in particular when glycosidic angles and sugar conformation are considered. Actually, only a few DNA oligonucleotides and some structural polynucleotides have been

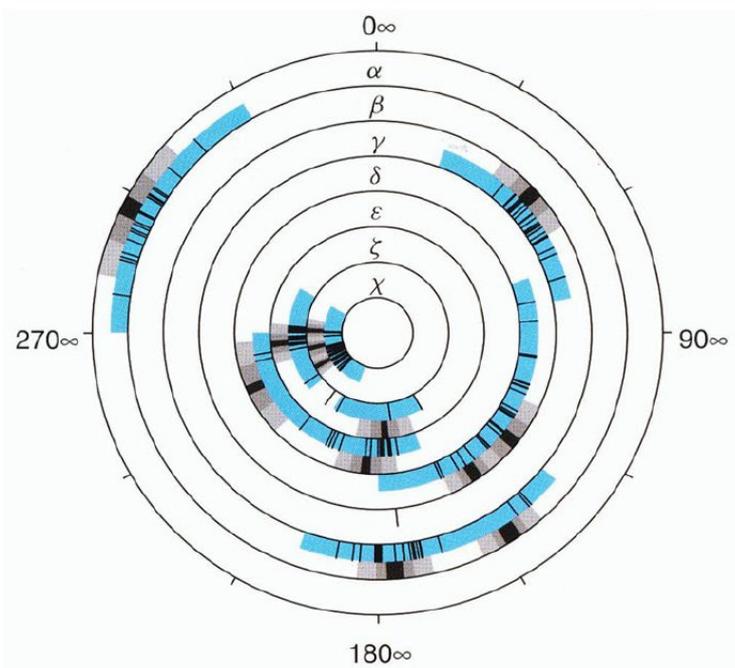


Figure 3.13: This conformational wheel shows torsion angles for BDL001.

experimentally observed, partially due to constraints imposed by Watson and Crick pairing on sugar conformation when two strands are bounded. On the contrary, crystallographic and NMR studies on a huge number of standard or modified mononucleosides showed a quite high conformational diversity. For mononucleotides, the conditions necessary for an efficient crystal packing can be sufficient to overcome the low energetic barrier between various values for a torsion angle. A wide range of base–base interactions characterized very big RNA molecules that can adopt various backbone conformations.

It is common to describe these backbone angles as *gauche*⁺ (g^+) for values of 60° , *gauche*[−] (g^-) for values of -60° and *trans* (t) for values of 180° . For example, α angles, around P–O5', and γ angle, the exocyclic angle around the C4'–C5' bond, may assume g^+ , g^- and t conformations. The two torsion angles around phosphate group, α and ζ , show a high flexibility level in various dinucleosidic crystalline structures, with tg^- , g^-g^-

and g^+g^+ . The A- and B-DNA forms adopt g^-g^- and g^-t , while Z-DNA g^+g^+ , tg^- and tg^+ . The β torsion angle around O5'-C5' bond is usually *trans*. All three possibilities are reported for γ angle in nucleosidic crystalline structures, even if g^+ conformation is the most common for right-handed double-stranded helices and polynucleotides. The δ torsion angle around C3'-C4' bond adopt values related to sugar conformation, since internal ring torsion angle τ_3 has value of 35° for *C2'-endo* and 40° for *C3'-endo*; σ is around 75° for *C3'-endo* and 150° for *C2'-endo*.

Some important correlations between backbone torsion angles, sugar conformations and glycosidic angles have been observed:

- Correlation between sugar conformations and glycosidic angle χ , especially for pyrimidinic nucleosides. The *C3' - -endo* conformation is usually associated to the average value of the *anti*-glycosidic angles, while the *C2' - -end* conformation is usually find with high *anti* angles. The *sin* glycosidic sugar angle prefer the *C2'-endo* sugar conformation.
- The scattering plots between α and χ angles show a clear distinction for A-, B- and Z-DNA classes. The same is observed for scattering plots between χ and ζ . Moreover, α, β, δ and χ angles characterize a fingerprint for nucleotide or for the entire structure able to predict the structure of double-stranded DNA.

3.2.5 Other possible structures

3.2.5.1 A-DNA

The B-DNA conformation when dehydrated undergoes a reversible conformation and become A-DNA [55], characterized by a right-handed helix wider and flatter, hence more compact than that of B-DNA form (Figure 3.14). An inverse relationship between crystal packing density and depth of major groove has been found. The A-DNA has 11 base pairs per turn and a pitch of 28 \AA , generating an axial hole. Hence this conformation has a higher number of base pairs each turn, due to a small rotational

angles. Moreover, the major groove is deeper and narrower, while minor groove is wider and less deep.

One of the most relevant A-DNA properties is that the plane on which the bases lie is tilted of 20° with respect to the helix axis. So, if B-DNA form is similar to a winding staircase, A form will resemble a winding staircase with steps tilted towards the center. These two conformations can be easily interconverted changing the degree of humidity. Hence, the process is quick and does not destroy ordered packing of fibers, since no separation of the two strands happens. In solution B-DNA conformation can be converted to A-DNA reducing water activity, for example adding a fraction of ethanol in aqueous solvent.

Since axis does not go through base pairs, this DNA structure has deep major groove and a slight minor groove, like a flat ribbon bounded around a cylindrical hole (6 Å diameter).

This conformation is found not only in dehydrated DNA samples as those used for crystallographic experiments, but also in hybrid DNA and RNA helices and in double-stranded RNA. Due to its flexibility, it is easy to found A-DNA in DNA-protein complexes: experimental evidences of a mixed A and B form for protein-DNA complex, bounded to TATA group.

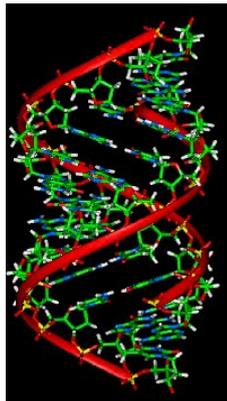


Figure 3.14: A-DNA structure.

3.2.5.2 C-DNA

This duplex form is characterized by 9.3 base pairs per turn, a little tilt of 6° from perpendicularity of the fiber and a slight base torsion. The C conformation, usually found in lithium salt at low humidity, is actually a variation of the B-DNA.

3.2.5.3 Z-DNA

This left-handed double helix structure, studied for the first time on the d(CGCGCG) crystalline structure, has 12 base pairs per turn, a pitch of 45 Å, a deep minor groove and a almost flat major groove. Its structure repeats every two base pairs, instead of one (Figure 3.15). The structures of Z type [56] show an alternation of cytosine and guanine in the first position.

Diffraction and NMR studies demonstrated that complementary polynucleotides with an alternation of purines and pyrimidines as poly-d(CG)·poly-d(GC) or poly-d(AC)·poly-d(GT) are usually in a Z conformation when there is a high salt concentration. Salt stabilizes Z-DNA with respect to B-DNA, since it reduces electrostatic repulsion between close phosphate groups but belonging to different chains (that distance 8 Å in Z-DNA, instead of 12 Å in B-DNA).

The problem is that in the Z conformation negative phosphate groups are close and so the left-handed helix is stable at extremely high ionic force. Nevertheless, a Z-DNA has been found in presence of polynucleotidic chains with C5 methylated cytosines. Evidence that this conformation is present also in the cell are cytosine methylating enzymes *Z anti-DNA* antibody able to bound to eukaryote chromosomes. The presence of B- and Z-DNA conformation in the same chain topologically permits an easier unwrapping of the double helix in biological important processes as replication or gene expression. It has been suggested that the reversible interconversion between the B and Z form may be a mechanism of gene expression regulation. However, it is difficult to demonstrate *in vivo* existence of Z-DNA, since it does not exist as a stable structure but only as a

transition state occasionally induced by biological activity.

This isomer seems to be promoted by an alternation of purine and pyrimidines, besides a negative supercoil of DNA, a low salt concentration and the presence of certain cations, at the physiological temperature (37° C) at pH 7.3/7.4.

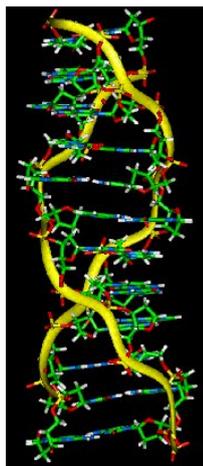


Figure 3.15: Z-DNA.

3.2.5.4 Comparison of different structures

Property	A-DNA	B-DNA	Z-DNA
Handedness	right-handed	right-handed	left-handed
Diameter	~ 26 Å	~ 20 Å	~ 18 Å
Bp per turn	11	10	12 (6 dimers)
Torsion per bp	33°	36°	60°
Helical pitch	28 Å	34 Å	45 Å
Helical raising per bp	2.6 Å	3.4 Å	3.7 Å
Base inclination (w.r.t. helix axis)	20°	6°	7°
Major groove	deep and narrow	wide and deep	flat
Minor groove	wide and not deep	deep and narrow	deep and narrow
Sugar pucker	<i>C3'-endo</i>	<i>C2'-endo</i>	<i>C2'-endo</i> (pyrimidines) and <i>C3'-endo</i> (purines)
Glycosidic bond	<i>anti</i>	<i>anti</i>	<i>anti</i> (pyrimidines) and <i>sin</i> (purines)

Table 3.2: Structural properties of A, B and Z types of DNA helices.

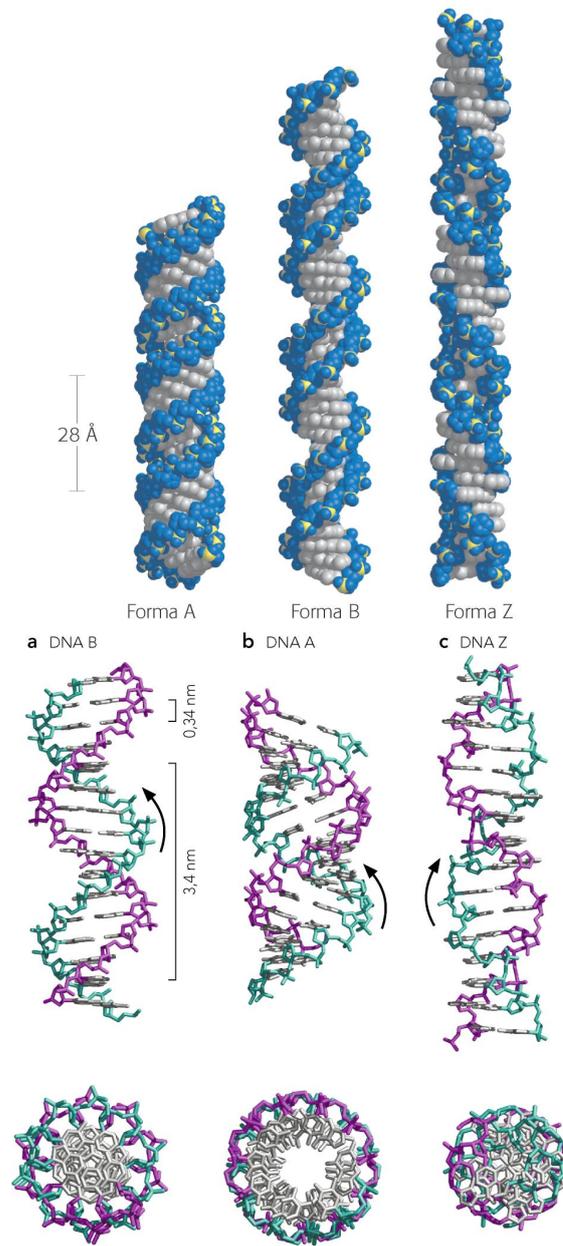


Figure 3.16: Various DNA conformations.

3.3 DNA nanotechnology

3.3.1 DNA–materials science

Nowadays, “DNA science” is expanding into various modern research fields [57], as shown in Figure 3.17.

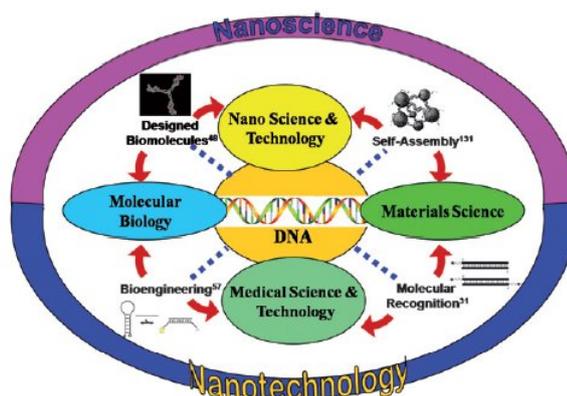


Figure 3.17: DNA science.

One of the main properties that make DNA a really useful materials to design structures and tools is its molecular and submolecular recognition capabilities. Many different architectures can be constructed *via* self-assembly. Molecular recognition properties of DNA have been used in the design of novel biosensors and diagnostic tools.

Optical properties of DNA and RNA can be used to observe intercalation of various molecules with DNA, using spectrometry techniques. Organic field-effect transistors (OFETs) and organic light-emitting diodes (OLEDs) has recently been constructed with DNA (Figure 3.18).

DNA is now widely used also for the preparation of inorganic nanoparticles. DNA can be easily synthesized and polymerase chain reaction (PCR) is widely used to duplicate DNA chains: hence it is really easy to obtain huge quantity of this material. Moreover, due to double helix DNA stiffness, it can be used as spacing between two functional end groups. Two possible approaches have been used to prepare DNA inorganic nanostructures: using it as a glue or as a template.

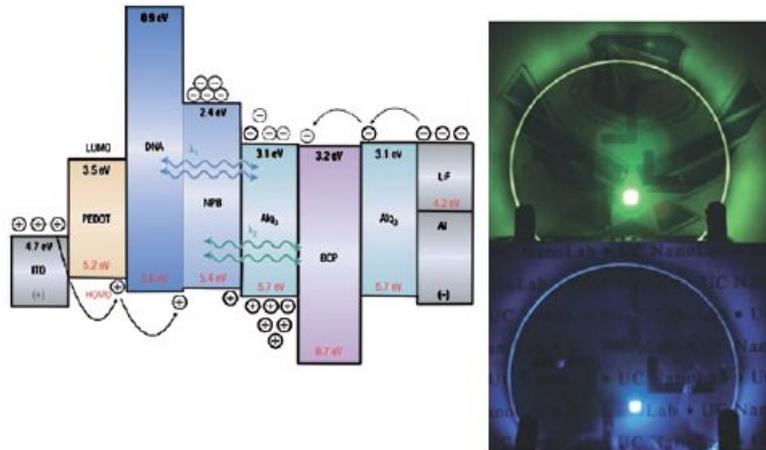


Figure 3.18: BioLED using DNA to increase light emission.

Also electrostatic and magnetic properties of DNA can be used. Also in this case this macromolecule is used as a spacing between two electrodes, field effect transistor (FET).

A DNA microarray [58] is an hybrid technology used in molecular biology and in medicine. Thousands of microscopic DNA oligonucleotide “points”, called “characteristics” contain each picomols of a specific DNA sequence. This sequence can be small gene region or other part of DNA used to hybridize a cDNA² or a cRNA (called target), under strict conditions. Target is hybridized through various probes.

DNA chips can be spatially arranged, as the gene chip (also called genome chip, DNA chip or gene array) or they can be specific DNA sequences marked to be independently identified in solutions. In standard biochips, the probes are bounds to a solid surface (glass or silica) through covalent bonds.

DNA microchips can be used to measure the changing of expression levels (gene expression profiling) or to detect single nucleotide polymorphism

²Complementary DNA (cDNA) is a double helix DNA synthesized from a mRNA sample.

(SNP). The gene expression profile can be used to identify genes whose expression undergoes some variation back to pathogens or other organisms, in order to obtain a comparison between gene expression in healthy and infected cells or tissues. Single nucleotide polymorphism detection can be used for forensic analysis, measure of disease inclination, lead identification, genetic mutations.

3.3.2 DNA-based soft phases

In this section we will outline systems in which DNA-mediated interactions promote the formation of “phases”, that is structures extending on length scales much larger than the building blocks. Their self-assembly typically involves a large number of interacting particles and often features hierarchical stages of structuring. Because of the possibility of fine-tuning the geometry and strength of the DNA-mediated interactions, these systems are characterized by a wide variety of patterns of self-assembly, ranging from amorphous, to liquid crystalline, to crystalline in one, two, or three dimensions [1].

3.3.2.1 All-DNA superstructures

The structures are designed to yield a two-stage interaction process. First, strongly bound aggregates are formed by exploiting WC pairing (usually standard linear double helices are formed, multi-strand aggregates can be obtained choosing carefully sequences). These aggregates mutually interact in many ways, comprehending steric, electrostatic, WC pairing and stacking interactions. Here we will report structure with stronger energies involved in the formation of the aggregates and weaker energies controlling their mutual interactions [1].

DNA liquid crystals

DNA LCs were first observed in vitro with long double strands; later they were recognized as the in vivo packing mechanism of some organisms and only quite recently they started being considered as a common ordering for DNA oligomers. The formation of these mesophases depends on the stacking of blunt ends and pairing and stacking of sticky ends [1].

Long DNA

LC order of DNA has been found in chromosomes, bacteria, viruses and sperm heads, suggesting a correlation between LC packing and biological activity, in particular with respect to protection from external stress or damage [1]. DNA double strands are obtained through enzymatic cut of nucleosomal DNA or through sonication, which yields average lengths between 10^2 and 10^4 bp. Two main mesophases were identified: a cholesteric or chiral nematic (N^*) phase and a columnar (COL) phases. The cholesteric phase is a positionally disordered fluid in which the constituent molecules align on average their axes along a common direction called the nematic director. The orientational order develops an additional macro-helical superstructure with the twist axis perpendicular to the local director. Hence, the phase consists of local nematic “layers” continuously twisted with respect to each other. This phase can be observed in polarized optical microscopy Figure 3.19. Nematic phase are bound at a concentration around 150 mg/mL at 100 mM monovalent salt conditions.

The columnar phase is characterized by parallel DNA helices align on a 2D lattice (Figure 3.20a) but remain free to slide relative to each other in the orthogonal direction. The continuous bending of the columns gives rise to the so-called developable domains, shown in Figure 3.20b. Columnar hexagonal phases are observed for concentration higher than 400 mg/mL. Experimental evidence was reported for the existence of other phases: a pre-cholesteric order in the form of a network of double-twisted cylinders, analogous to the thermotropic blue phases, a hexatic phase that replace the hexagonal columnar in very long DNA fragments, and a structure with

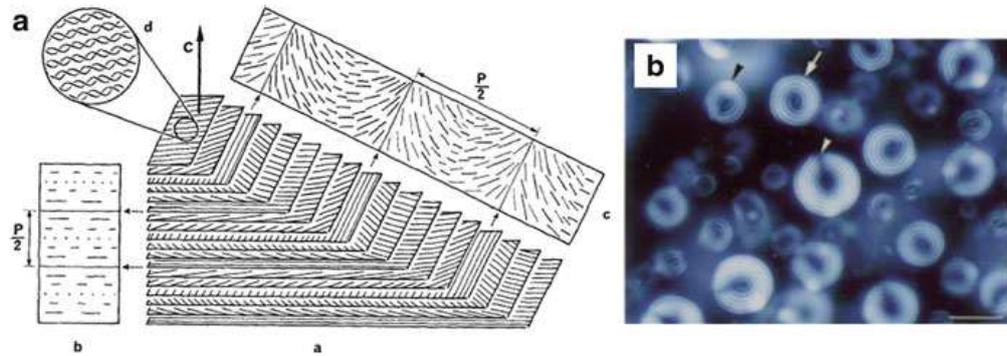


Figure 3.19: (a) Schematic representation of the structure of the chiral nematic phase of DNA, showing continuously twisting nematic layers, giving rise to a $p/2$ periodicity. (b) N^* droplets observed in polarized microscopy. The dark and bright stripes correspond to $p/2$ (size bar is $10 \mu m$) [1].

orthorhombic symmetry appearing in the transition to crystalline order. According to Onsager theory [1] based on excluded volume interactions, elongated semi-flexible repulsive rods (as dsDNA) at sufficiently high concentration are expected to develop orientational order. Such phases can be obtained with high concentration but also by adding in the solutions various condensing agents (for example alcohols), introducing effective attractive interhelical interactions. The Onsager theory predicts that helices longer than 100 bp display isotropic-to-nematic transition, while short DNA strands tend to collapse into longer aggregates (end-to-end stacking) [1].

Short DNA

According to the Onsager theory, rods with L/D^3 is the length of the double helix, while D is the effective diameter. ≥ 4 , i.e. sequence with less than 24 base pairs, do not form any LC phases. Surprisingly, in concentrated aqueous solutions of DNA self-complementary sequences with N (number of base pairs) as low as 6 bp (called nano-DNA, nDNA). The phase observed for nDNA are the same that for long DNA (N^* and COL)

³L

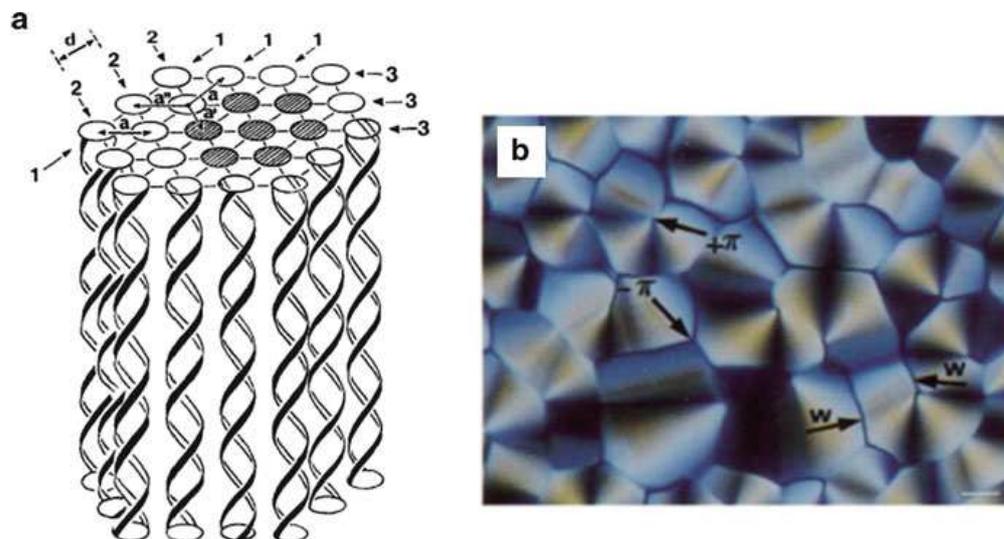


Figure 3.20: (a) Sketch of the structure of the hexagonal columnar phase of DNA, showing parallel molecules hexagonally packed in the plane perpendicular to their axis. “a” and “d” are the lattice parameters. (b) COL developable domains observed in polarized microscopy. “w” indicates defect walls between differently oriented domains, while $\pm\pi$ stands for point defect around which DNA molecules continuously bend (size bar is $10 \mu m$) [1].

but they exist at higher concentration. Pairing of complementary nDNA sequences yields to double helices (Figure 3.21a). Blunt-ended and sticky-ended helices then stack and form linear aggregates, that show N^* and COL LC phases (Figure 3.21b,c). On the contrary, helices with unpaired dangling ends can not aggregate, that segregate from mixture through the nucleation of LC domains (Figure 3.21b,d).

In nDNA, Zanchetta *et al.* studied sequence with 8–20 bp and found both right- and left-handed N^* phases. The handedness depends on oligonucleotide length (sequences with $N \geq 12$ show long DNA behavior, while shorter oligos can yield both right- and left-handed N^* phases), on the mode of terminal interaction (blunt-ended helices produce right-handed N^* phases, while sticky-ended helices behave like long DNA), on oligomer sequence and on concentration (higher concentration lead to right-handed

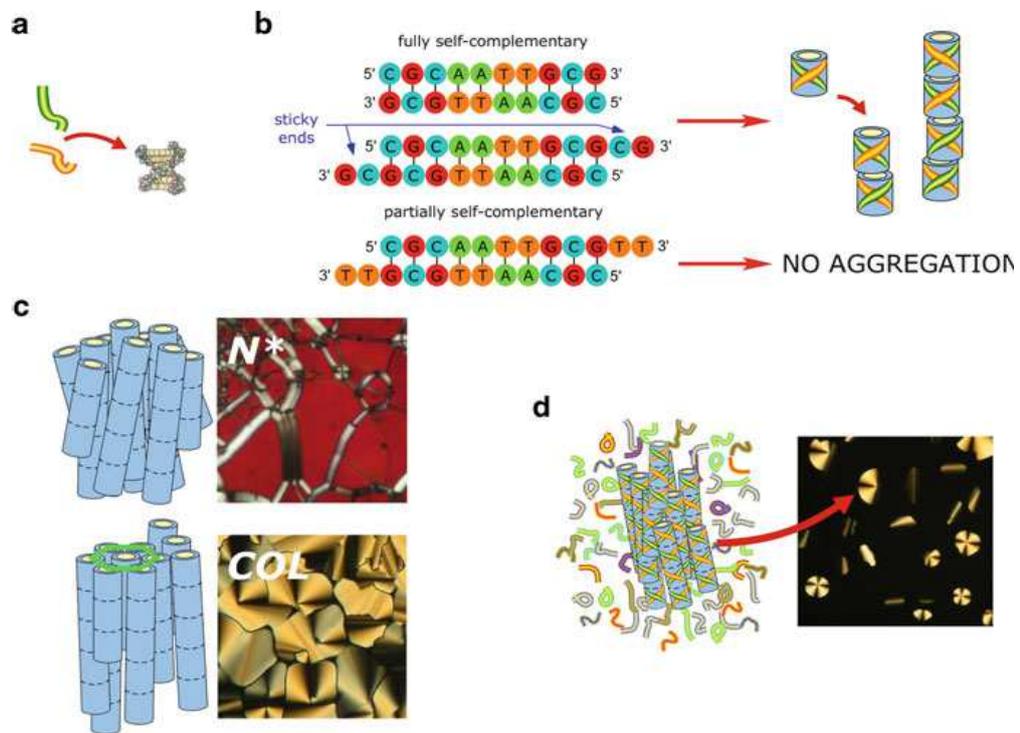


Figure 3.21: Schematic representation of short DNA self-assembly stages [1]. Details are explained in the text.

helices).

DNA crystals

The same inter-duplex interactions described above can lead also to crystalline symmetry or amorphous structure, thank to ad hoc chosen sequences. Examples of such structures are showed in Figure 3.22. These aggregates of DNA can be designed so as to seed the growth of crystals or other macroscopically organized phases [1].

Various two-dimensional crystals can be obtained with careful sequence design: they can be flexible (less than 8 arms) or strongly interlaced (polygons and double crossover⁴). These periodic aggregates are planar

⁴Double crossover is the linking of two or more parallel double helices

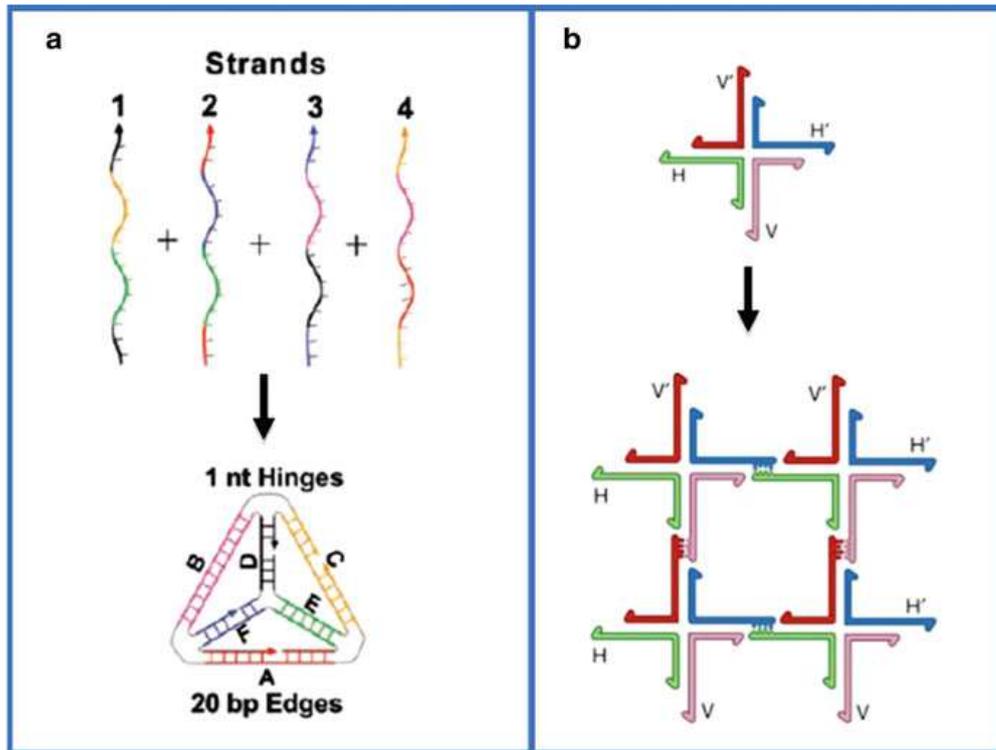


Figure 3.22: (a) Design principle for DNA tetrahedra. (b) Concept drawing of star-like junctions.

structures built assembling tiles in which helices can be at given angles (polygons) or all parallel (DX). Figure 3.22 show example of triangular interacting tiles and of double crossover (DX).

In the same way three-dimensional crystals can be obtained: for example layers of parallel helices are stacked together with a 120° rotation with respect to the one below. In these structures non-WC base pairing penalty is compensated by the benefit of stacking. This layers are bounded with unusual types of crossover arrangements. Such crystals have sizes up to a fraction of a millimeter: the pores are hence large enough to be permeated by proteins and can be used in molecular separation [1].

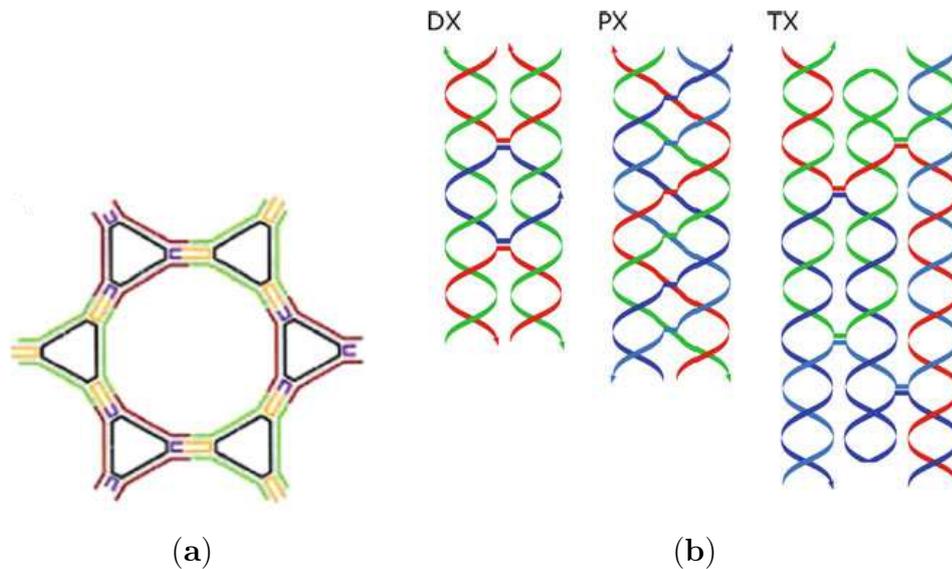


Table 3.3: (a) Interacting triangular tiles. (b) Double crossover structures formed by three (DX and PX variation) and four (TX) different sequences with complementary segments.

DNA origami

The interaction of short DNA to obtain long DNA is used to control nanofabrication, with an amazing control of size and shape of the structures. Nowadays, folding DNA strategy can lead to both 2D and 3D structures. An example of DNA origami is shown in Figure 3.23. The construction of three-dimensional smart materials is a possible application.

Other structures

Simple linear aggregates of DNA oligonucleotides (tubes) show tunable properties, which can be used in various applications: scaffold for various nanowires, guided controlled spacing of colloids and as drug nano-carriers. Hydrogels, with controlled stiffness, are amorphous cross-linked DNA networks that have a wide range of potential applications: drug delivery, cell culture, tissue engineering.

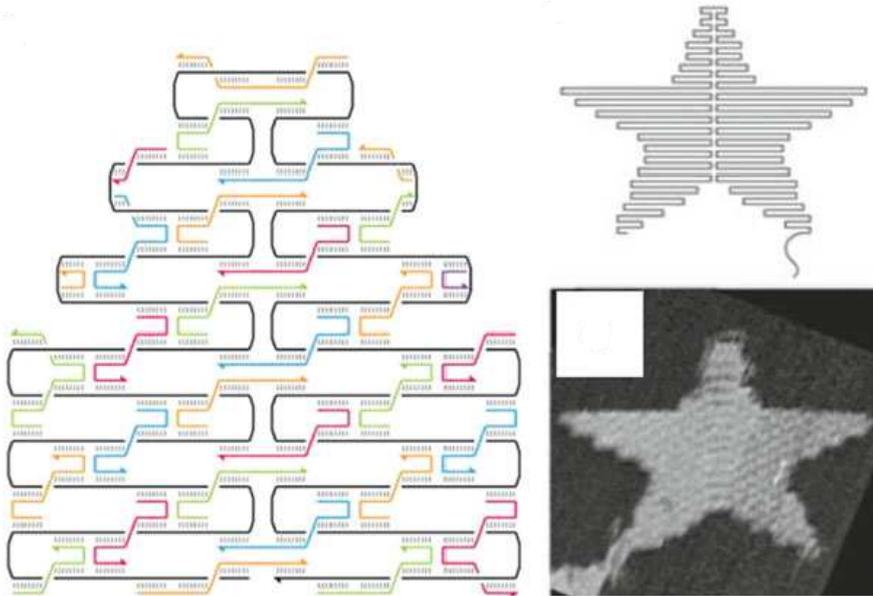


Figure 3.23: Sketch of the basic idea of DNA origami. A long strand composed by oligonucleotides complementary in specific regions is folded into designed shapes, such as a star. Also an AFM of such structure is obtained.

3.3.2.2 DNA-coated colloids

DNA-coated particles combine the collective nature of colloids (aggregation, flocculation) with the properties of DNA-based molecular interactions. Usually metallic nanoparticles, mainly gold, are directly covered with complementary sequences. An example of interactions between such nanoparticles is shown in Figure 3.24 [1].

The interactions between particles are of various types: an attractive part due to interparticle DNA hybridization, a steric and electrostatic repulsive term and a strong attractive van der Waals term at short distance.

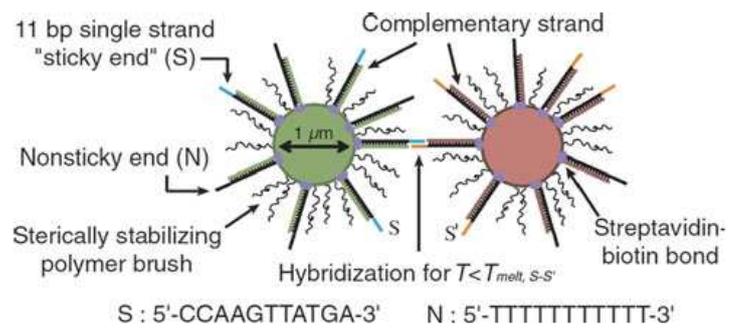


Figure 3.24: Hybridization of DNA functionalized nanoparticles.

3.4 Modeling state of the art

Molecular dynamics simulations, in particular, can describe well DNA structure and molecular dynamics motion. Nowadays, even better results can be obtained thanks to new molecular mechanics force fields, high computational abilities, parallelized simulation codes and increased long-range interactions interaction. The reliability of new models for studying processes on long time scale. Molecular dynamics simulation can be used to predict DNA behavior of single strand, duplex, triplex, quadruplex and various other structures as “zipper”, modified backbone and damaged DNA molecules.

Various level of description have been analyzed, ranging from atomistic representation to coarse-grained models. Atomistic models seem to be the most useful, since they furnish a more detailed description: however, when the number of sites increase, it becomes difficult to simulate due to high computational demand.

3.4.1 Atomistic models

The most detailed models are those representing each atom with a single site. Many force fields are available. Usually, the interactions between atoms are represented with the following potential energy (in particular for the CHARMM force field):

$$U = U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers} + U_{LJ} + U_{Coulombic} \quad (3.6)$$

Each term include various specific constants that refer to the particular atom involved. U_{bonds} deals with the bond between two atoms, separated by an equilibrium distance. U_{angle} describes the three-site interactions maintaining fixed angles between the two vectors. $U_{dihedrals}$ and $U_{impropers}$ are four-term interactions, keeping the right torsional equilibrium, chirality and planarity. U_{LJ} represents non-bonded and intermolecular interactions. The electrostatic dispersion energy are computed using Coulomb law.

3.4.2 Coarse-grained models

Many molecular models where bases and sugar are described by spheres have been developed. Atomistic models describe the interactions very well, but they have a high computational cost. Using atomistic model limits sample dimensions, the number of simulations and the simulated time, hence an appropriate error analysis is not possible. Moreover, the parametrization of systems containing other DNA molecules is difficult because protein-protein, protein-solvent and solvent-solvent interactions should be carefully described. To study bigger systems and reduce the number of parameters, various atoms are grouped in a single site. This problem is solved using coarse-grain models. The CHARMM force field, for example, show a set of united atom parameters that can be used with non polar hydrogen atoms and other heavy atoms. In particular, the Gō model is the most used [59]: it reduces the complexity of a residue to one, two or three interactions, limiting computational cost. Gō models are used to analyze protein folding, protein stretching and the influence of protein structure. Usually these models are in good qualitative agreement with experimental results. Gō models comprehend mainly three potentials:

$$U_{G\bar{o}} = U_{bb} + U_{nat} + U_{non} \quad (3.7)$$

The first term is referred to the backbone, the second to native contacts and the last one non-native energy contacts.

Another example of coarse-grained macromolecular models is represented by DNA-protein docking Poulain model [60]. Using a representative set of DNA-protein complexes, model can predict surface interactions in the bounded complex and DNA sequences and electrostatic influence on DNA and protein docking conformations. The direct electrostatic interaction between phosphate groups and lateral chains of amino acid strengthen DNA-protein complex. Moreover, this work demonstrate that a coarse-grained model can be useful to give a general and complete description and to understand the association of protein-DNA complex. The force field used are coarse-grained.

It is worth noting that there is another type of model that connect various resolution levels. These models, called “multiscale” models [61], include atomistic level of detail for important data and coarse-grained resolution for the part of the macromolecule that does intervene in the process we want to study.

3.4.2.1 Anisotropic models

Anisotropic models take into account the changing of properties depending on directions. In the last decades, experiments with single DNA molecules have revealed the complicated interaction network in supercoiled DNA or in local denaturation of double helix. Experimental results have been analyzed with two kinds of models: base pair transition and elastic and continuous worm-like chain variants. The first approach gives a more local resolution, describing relative position and orientation of base pairs. In particular, they give an interpretation of biological function of particular sequences. In the second type of approach, DNA models are represented with a worm-like chain (WLC) model. They use empiric parameters describing bending, twisting and stretching. Both models give a description in good agreement with experiments. WLC models are usually chosen to analyze biologically important processes, like supercoiling.

3.5 Three-sites coarse-grained model

Our aim is to parametrize a model that can be used to study DNA behavior in various biological systems, but also liquid crystalline phases. This model should predict well the key physical phenomena, like duplex stability, duplex denaturation, the effects of salt concentration. Hence, we need a model with low computational demand that can be used to perform long simulation of big samples.

We decided to represent each nucleotide with three sites. Sugar and phosphate group are simulated as spheres. To improve the other three-sites coarse-grained models that use only spheres, to symbolize bases we chose ellipsoids, because their shape seems to reproduce the atomistic structure better than sphere does. We fit the energy profiles with a Gay-Berne potential to obtain spheres and ellipsoids parameters. Note that each base will be described by an ellipsoid of different size. Anisotropic shapes also need to a orientation to be specified: a quaternion is associated to each base.

Next the force field have been parametrized. First of all, we collect data from atomistic simulations of a 10 base pair long sequences and we compute histograms, using the Boltzmann inversion method. These data are now fitted using various types of potential: the equations that fit better atomistic results are chosen and added to the MD code LAMMPS.

3.6 Validation of the model

3.6.1 Fitting of coarse-grained particles

We decide to fit sugar and phosphate groups with spheres, while bases are represented with ellipsoids. In order to obtain the best fitting particles, we found the atomistic energy profiles of groups (phosphate, sugar and bases) approaching themselves in different orientations. From these data, we find the σ_i and ϵ_i values.

3.6.1.1 Intramolecular potential

First of all, we have to decide which is the best force field for our system. There are mainly two possible force fields to be used for nucleic acid all-atom simulation: AMBER and CHARMM27.

CHARMM27

The functional form of the CHARMM force field is:

$$\begin{aligned}
 V = & \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] + \\
 & + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 + \\
 & + \sum_{\text{non-bonded}} \epsilon_{i,j} \left[\left(\frac{R_{\text{min}ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min}ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}} \quad (3.8)
 \end{aligned}$$

For DNA, RNA, and lipids, CHARMM27 [62] is used. Some force fields may be combined, for example CHARMM22 and CHARMM27 for the simulation of protein-DNA binding. Additionally, parameters for NAD⁺, sugars, fluorinated compounds, etc. may be downloaded. These force field version numbers refer to the CHARMM version where they first appeared, but may of course be used with subsequent versions of the CHARMM executable program. Likewise, these force fields may be used within other molecular dynamics programs that support them.

AMBER

The functional form of the AMBER force field is [63]:

$$\begin{aligned}
 V = & \sum_{bond} \frac{1}{2} k_b (l - l_0)^2 + \sum_{angle} k_a (\theta - \theta_0)^2 + \sum_{torsion} \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] + \\
 & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \left\{ \epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} \right\} \quad (3.9)
 \end{aligned}$$

The first term (summing over bonds) represents the energy between covalently bonded atoms. This harmonic (ideal spring) force is a good approximation near the equilibrium bond length, but becomes increasingly poor as atoms separate. The second term (summing over angles) represents the energy due to the geometry of electron orbitals involved in covalent bonding. The third term (summing over torsions) represents the energy for twisting a bond due to bond order (e.g. double bonds) and neighboring bonds or lone pairs of electrons. Note that a single bond may have more than one of these terms, such that the total torsional energy is expressed as a Fourier series. The last term (double summation over i and j) represents the non-bonded energy between all atom pairs, which can be decomposed into van der Waals (first term of summation) and electrostatic (second term of summation) energies.

The form of the van der Waals energy is evinced by the equilibrium distance (σ) and well depth (ϵ). The factor of 2 ensures that the equilibrium distance is σ .

To use the AMBER force field, it is necessary to have values for the parameters of the force field (e.g. force constants, equilibrium bond lengths and angles, charges). A fairly large number of these parameter sets exist. Each parameter set has a name and provides parameters for certain types of molecules.

- Peptide protein and nucleic acid parameters are provided by parameter sets with names beginning with "ff" and containing a two digit year number, for instance "ff99".

- GAFF (Generalized AMBER force field) provides parameters for small organic molecules to facilitate simulations of drugs and small molecule ligands in conjunction with biomolecules.
- The GLYCAM force fields have been developed by Rob Woods for simulating carbohydrates.

We decide to use this force field in order to obtain potential curves, since it seems to be one of the most reliable and well implemented force field for DNA bases. We use topology and parameter of the PARM94 version of the force field [63].

3.6.1.2 Fitting ellipsoids

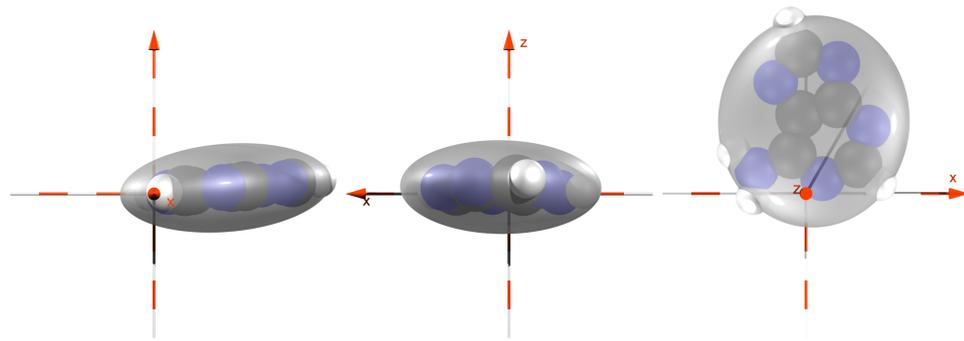
Using the AMBER force field and the homemade code “Outside”, we fit bases with ellipsoids. First of all, we use the “Biomolecular well depth” to compute the energy profile of two bases (actually they are the same) getting closer along the three axes. Averaging this approaching from the positive and negative part of the axis. Using the formula of the Gay–Berne potential, we fit σ_c , σ_x , σ_y , σ_z and ε_x , ε_y , ε_z . For example for σ_x and ε_x we use:

$$U_{GB}(x) = 4\varepsilon_x \left(\sigma_c / (x - \sigma_x + \sigma_c)^{12} - \sigma_c / (x - \sigma_x + \sigma_c)^6 \right) \quad (3.10)$$

For a first step of fit we obtain initial guess for σ_i and ε_i (with $i = x, y, z$), we average σ_i values to obtain σ_c . Then, the fit is done again to obtain correct values also for σ_i and ε_i . These values, along with snapshot of atomistic structures, are summarized below.

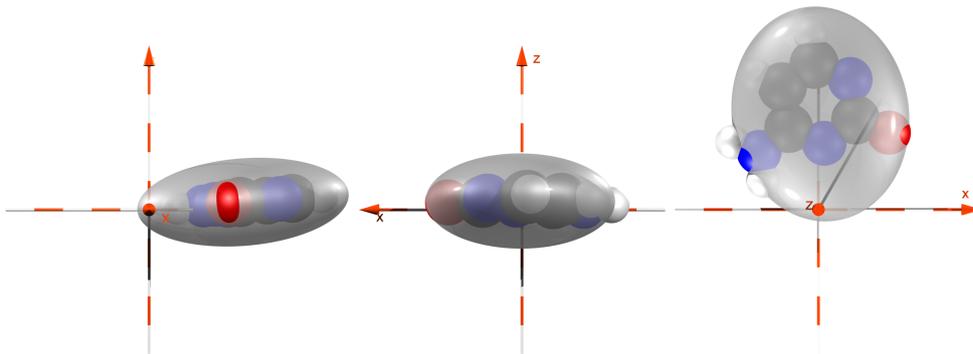
Adenine

Parameter	Value
σ_x	7.306
σ_y	6.647
σ_z	3.032
σ_c	3.308
ε_x	1.494
ε_y	1.579
ε_z	8.573

Table 3.4: Adenine: σ_i and ε_i .**Figure 3.25:** Adenine: crystallographic atomistic structure and best fitting ellipsoid. Note that ellipsoids are a bit “tilted” with respect to the molecular axes.

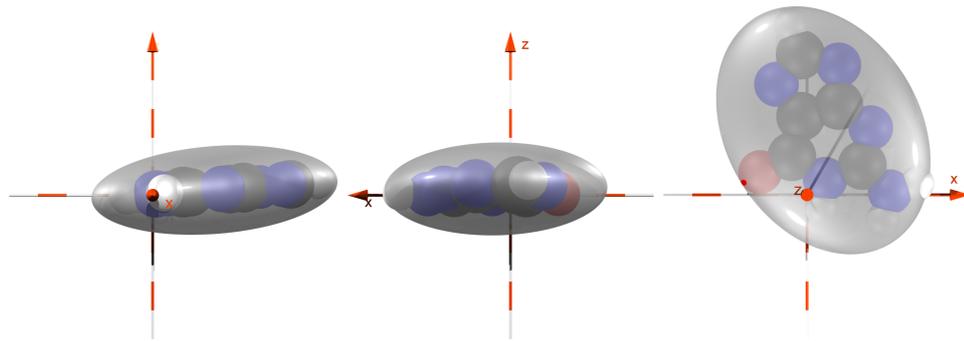
Cytosine

Parameter	Value
σ_x	7.410
σ_y	6.137
σ_z	3.023
σ_c	3.195
ε_x	0.855
ε_y	1.269
ε_z	6.085

Table 3.5: Cytosine: σ_i and ε_i .**Figure 3.26:** Cytosine: atomistic structure and best fitting ellipsoid.

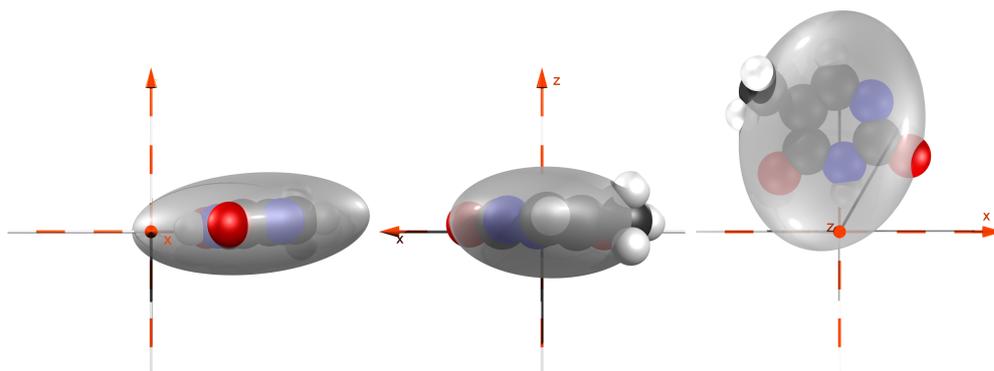
Guanine

Parameter	Value
σ_x	9.112
σ_y	6.766
σ_z	3.010
σ_c	3.258
ε_x	0.771
ε_y	1.517
ε_z	9.615

Table 3.6: Guanine: σ_i and ε_i .**Figure 3.27:** Guanine: atomistic structure and best fitting ellipsoid.

Thymine

Parameter	Value
σ_x	8.360
σ_y	6.370
σ_z	3.534
σ_c	3.005
ε_x	0.803
ε_y	1.289
ε_z	5.322

Table 3.7: Thymine: σ_i and ε_i .**Figure 3.28:** Thymine: atomistic structure and best fitting ellipsoid.

3.6.1.3 Fitting spheres

For spheres we follow the same procedure used above for fitting ellipsoids: at the end σ_i and ε_i were averaged to obtain only one value each.

Sugar

Fitted σ is 4.472, while ε 1.164.

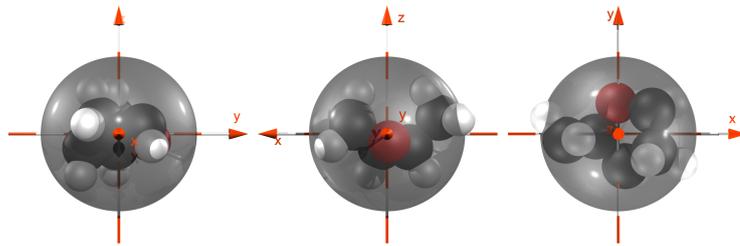


Figure 3.29: Sugar: atomistic structure and best fitting sphere.

Phosphate

Fitted σ is 5.508, while ε 1.566.

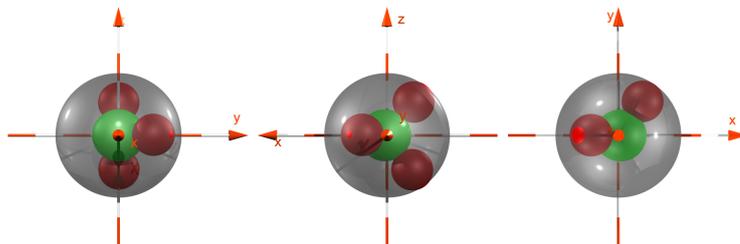


Figure 3.30: Phosphate: atomistic structure and best fitting sphere.

3.6.1.4 Beads parametrization

The following table sums up σ_i and ε_i values for spheres and ellipsoids.

	Adenine	Cytosine	Guanine	Thymine	Sugar	Phosphate
σ					5.508	4.472
σ_x	7.306	7.410	9.112	8.360		
σ_y	6.647	6.137	6.766	6.370		
σ_z	3.032	3.023	3.010	3.534		
σ_c	3.308	3.195	3.258	3.005		
ε					1.566	1.164
ε_x	1.494	0.855	0.771	0.803		
ε_y	1.579	1.269	1.517	1.289		
ε_z	8.573	6.085	9.615	5.322		

Table 3.8: Spheres and ellipsoids: σ_i and ε_i .

However, using these values of σ_x and σ_y for the ellipsoids, when we construct the double helix of B-DNA, bases bound by H-bonds overlap. So we decrease these values of the 30% in order to avoid this superimposition (Table 3.9 and following figures).

	Adenine	Cytosine	Guanine	Thymine	Sugar	Phosphate
σ					5.508	4.472
σ_x	5.114	5.187	6.378	5.852		
σ_y	4.653	4.296	4.736	4.459		
σ_z	3.032	3.023	3.010	3.534		

Table 3.9: Reduced ellipsoids and sphere.

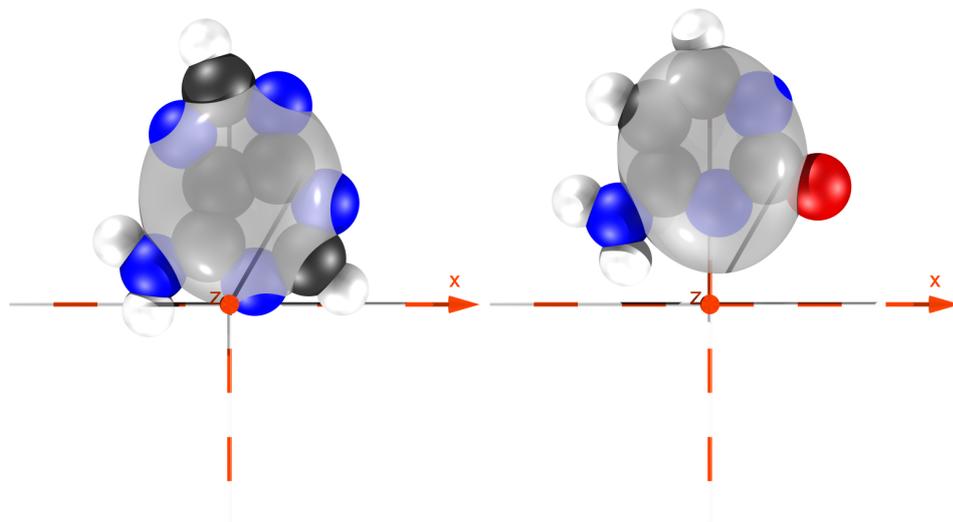


Figure 3.31: Adenine and cytosine: atomistic structure and reduced ellipsoid.

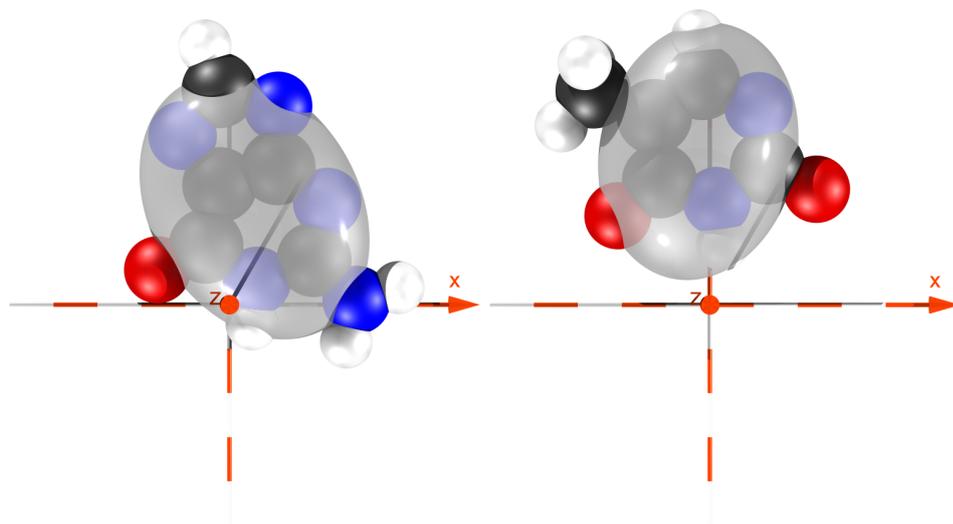


Figure 3.32: Guanine and thymine: atomistic structure and reduced ellipsoid.

3.6.2 Finding quaternions

Biaxial GB particles have an orientation which is specified via quaternions.

3.6.2.1 Protocol

The first step is to obtain from the crystallographic molecule the three inertial eigenvectors. Plotting the atomistic positions, the center of mass (COM) and the eigenvectors, obtained using the “inertial tensor”. We define the axes so that the molecule lays in the xy plane and the atoms involved H-bonds are in the positive part of the x and y axis. Once we have chosen the axes, we find an atom in a position which helps us assign the order of the axes: for example the projection of the atom on x (or y) axis should be very big and positive, projection on y (or x) axis should have an intermediate positive value and on z axis should be almost zero. Using these information we can assign eigenvectors and sign to the x and y axes, the z one will be computed with a cross product of the previous versors. From these versors, we are able to find out the associated quaternion, using a passive rotation since we want to describe the orientation of the molecule, leaving it in its crystallographic position, but using our reference system. Then, using LAMMPS, we want to check if the H-bond positions correspond to the atomistic one. We have first of all to translate the molecule in the COM and next to rotate it with an active rotation (we are not moving the reference frame but the object itself). Using these new coordinates and the quaternion obtained with the passive rotation in the LAMMPS input file we checked whether LAMMPS derived positions and atomistic H-bonds sites coincide.

Adenine

The following picture (Figure 3.33) shows adenine crystallographic configuration [64]. In the second one (Figure 3.34), the molecule in the “chosen” orientation and the eigenvectors are represented. Eigenvectors are shown as points⁵, besides the z -axis which is not appearing due to the fact that one point will cover the molecule and the other one will be underneath it. The yellow point is the center of mass of the molecules. Also H-bond sites are showed.

In this case, the red axes will be the y -axis and the green one the x -axis. We choose the H61 atom which will have a big and positive projection on the y -axis, an almost zero projection on the z -axis and an intermediate positive value on the x -axis.

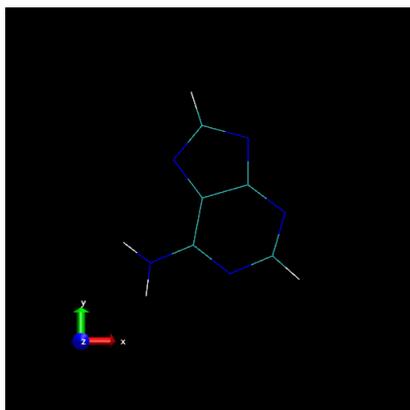


Figure 3.33: Adenine: crystallographic configuration.

⁵ $COM + eigenvectors * 10.0$ and $COM + eigenvectors * (-10)$

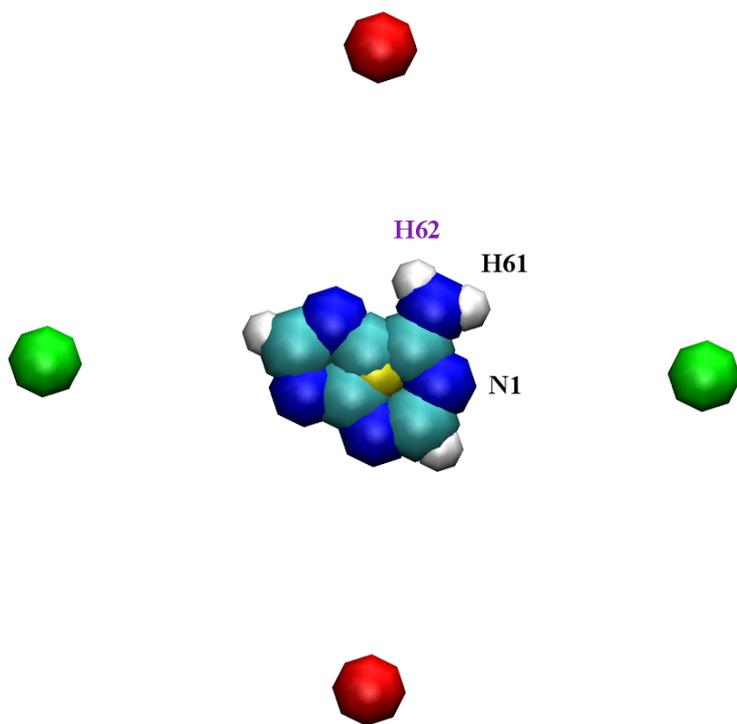


Figure 3.34: Adenine: the yellow atom represents the center of mass of adenine, while the red points and the green one two of the axes (the blue ones are not showed since one will appear on the adenine and the other one below it). The H61 and N1 are atoms involved in H-bonds.

Now we do dot product between eigenvectors and H62 atom position⁶ and associate the higher value to the y -axis and the intermediate one to x -axis, changing the sign whether the values are negative. Next we do a cross products between these two axes to obtain the z -axis.

	Eigenvector			Dot product
x	-0.0568	-0.9952	-0.0796	1.2044
y	-0.9983	0.0575	-0.0063	-2.7082
z	0.0108	0.0792	-0.9968	0.0003
q	0.0311	-0.6860	0.7265	0.0251
COM	0.2701	2.4699	0.2026	

Table 3.10: Adenine: eigenvectors, quaternion and center of mass.

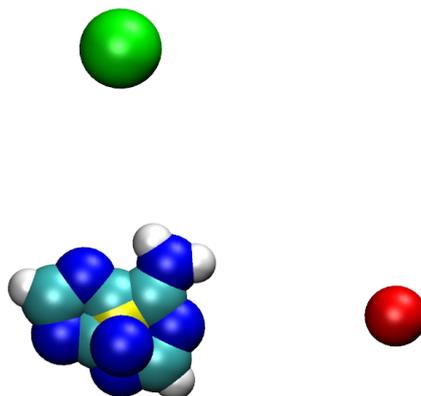


Figure 3.35: Adenine: x (red), y (green) and z -axes (blue).

⁶(position of H62 - COM)*eigenvectors

We obtain snapshots already shown.

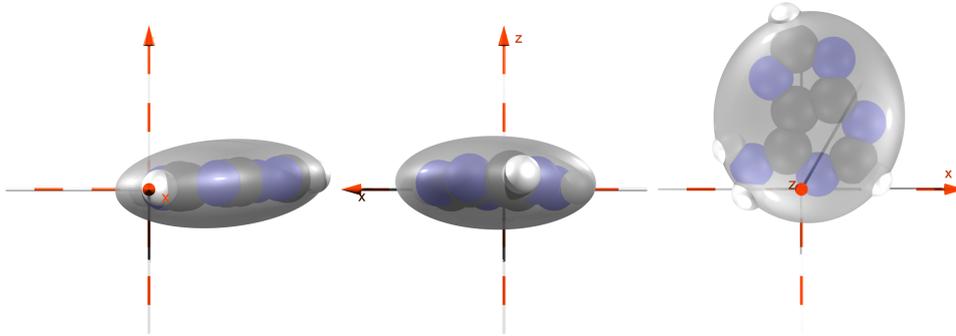


Figure 3.36: Adenine: atomistic structure and best fitting ellipsoid.

We just check if LAMMPS put the H-bond sites exactly in the atomistic positions: they superimpose exactly.

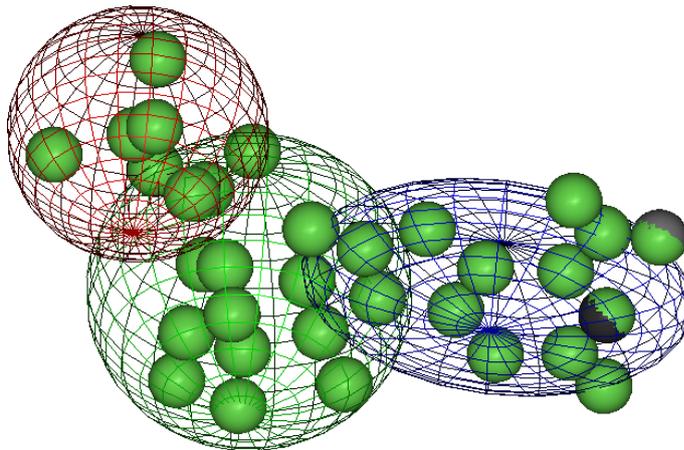


Figure 3.37: Adenine: LAMMPS. Coarse-grained and atomistic representations: the darker sphere are the position of the H-bonds position as computed by LAMMPS.

Cytosine

In this case, the red axes will be the y -axis and the green one the x -axis. We choose the H61 atom which will have a big and positive projection on the y -axis, an almost zero projection on the z -axis and an intermediate positive value on the x -axis.

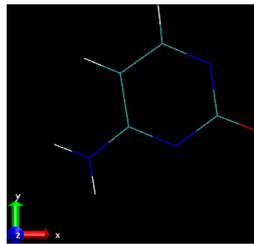


Figure 3.38: Cytosine: crystallographic configuration.

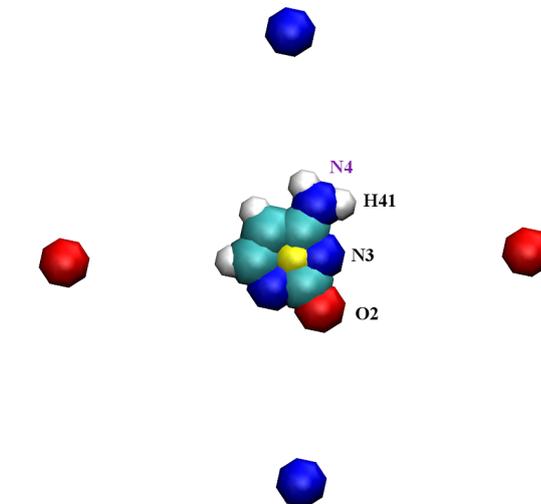


Figure 3.39: Cytosine: the yellow atom represents the center of mass of cytosine, while the red points and the green one two of the axes (the blue ones are not showed since one will appear on the cytosine and the other one below it). H41, O2 and N3 are atoms involved in H-bonds, while the N4 will be chosen to discriminate axes.

Now we do dot product between eigenvectors and N4 atom position⁷ and associate the higher value to the y -axis and the intermediate one to x -axis, changing the sign whether the values are negative. Next we do a cross products between these two axes to obtain the z -axis.

Eigenvector				Dot product
x	0.1559	-0.9848	-0.0765	0.9544
y	-0.9877	-0.1545	-0.0230	-2.3076
z	0.0109	0.0791	-0.9968	0.0000
q	-0.0336	0.7595	-0.6493	-0.02160
COM	0.0645	3.2736	0.2641	

Table 3.11: Cytosine: eigenvectors, quaternion and center of mass.

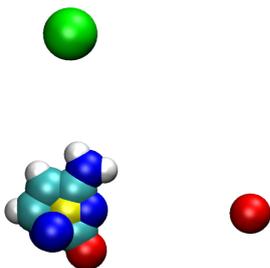


Figure 3.40: Cytosine: x (red), y (green) and z -axes (blue).

⁷(position of N4 - COM)*eigenvectors

We obtain snapshots already shown.

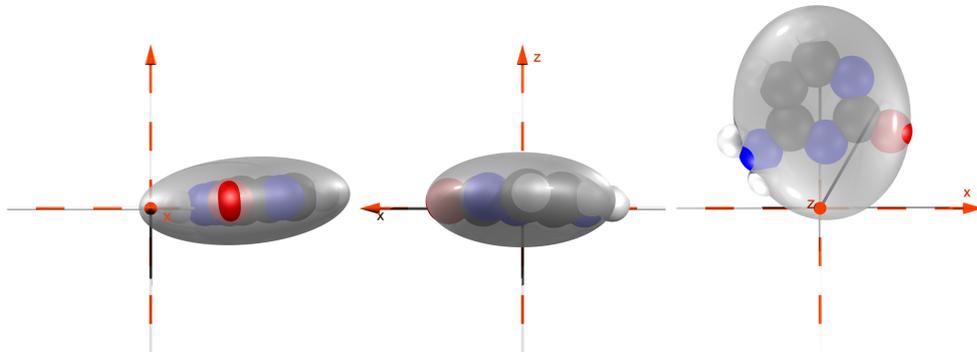


Figure 3.41: Cytosine: crystallographic positions and ellipsoids, x , y and z -view.

We just check if LAMMPS put the H-bond sites exactly in the atomistic positions: they superimpose exactly.

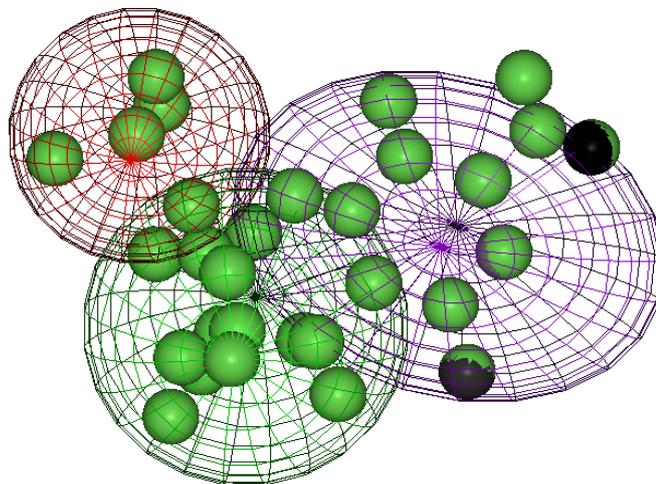


Figure 3.42: Cytosine: LAMMPS. Coarse-grained and atomistic representations: the darker sphere are the position of the H-bonds position as computed by LAMMPS.

Guanine

In this case, the red axes will be the y -axis and the green one the x -axis. We choose the H22 atom which will have a big and positive projection on the x -axis, an almost zero projection on the z -axis and an intermediate negative value on the y -axis.

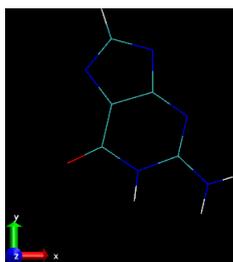


Figure 3.43: Guanine: crystallographic configuration.

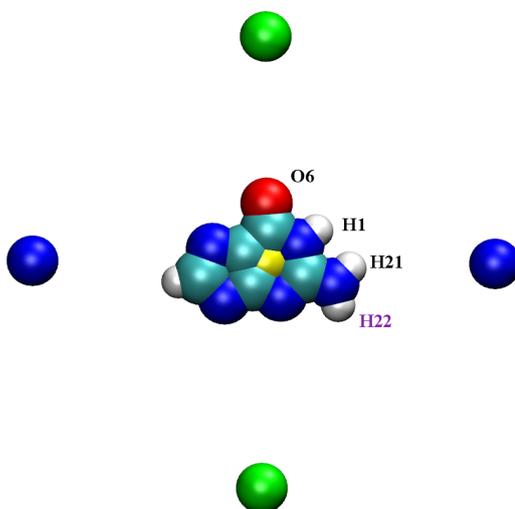


Figure 3.44: Guanine: the yellow atom represents the center of mass of guanine, while the red points and the green one two of the axes (the blue ones are not showed since one will appear on the guanine and the other one below it). H21, H1 and O6 are atoms involved in H-bonds, while the H22 will be chosen to discriminate axes.

Now we do dot product between eigenvectors and N4 atom position⁸ and associate the higher value to the y -axis and the intermediate one to x -axis, changing the sign whether the values are negative. Next we do a cross products between these two axes to obtain the z -axis.

	Eigenvector			Dot product
x	0.5100	-0.8579	-0.0626	-3.2297
y	-0.8601	-0.5077	-0.0499	-1.8788
z	0.0110	0.0793	-0.9968	-0.0011
q	-0.0372	0.8681	-0.4947	-0.0148
COM	0.5708	2.2047	0.1846	

Table 3.12: Guanine: eigenvectors, quaternion and center of mass.

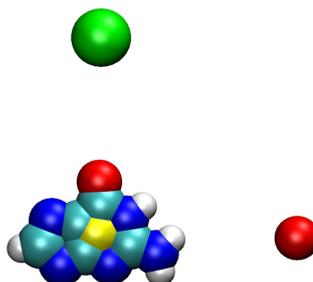


Figure 3.45: Guanine: x (red) and y (green). z -axes (blue) is beneath the molecule.

⁸(position of H22 - COM)*eigenvectors

We obtain snapshots already shown.

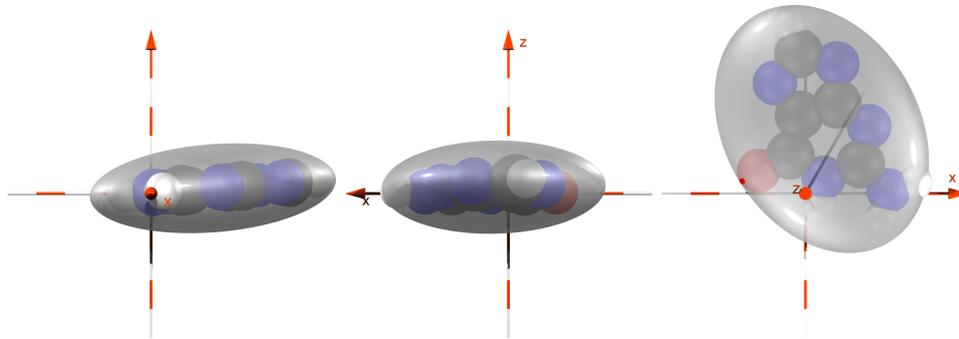


Figure 3.46: Guanine: crystallographic positions and ellipsoids, x , y and z -view.

We just check if LAMMPS put the H-bond sites exactly in the atomistic positions: they superimpose exactly.

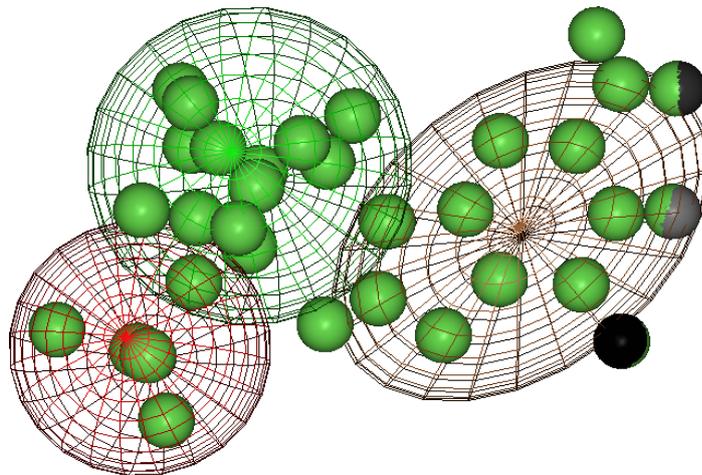


Figure 3.47: Guanine: LAMMPS. Coarse-grained and atomistic representations: the darker sphere are the position of the H-bonds position as computed by LAMMPS.

Thymine

In this case, the red axes will be the y -axis and the green one the x -axis. We choose the H51 atom which will have a big and positive projection on the y -axis, an almost zero projection on the z -axis and an intermediate positive value on the x -axis.

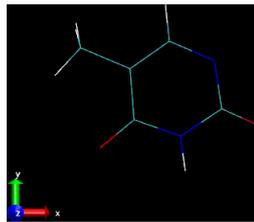


Figure 3.48: Thymine: crystallographic configuration.

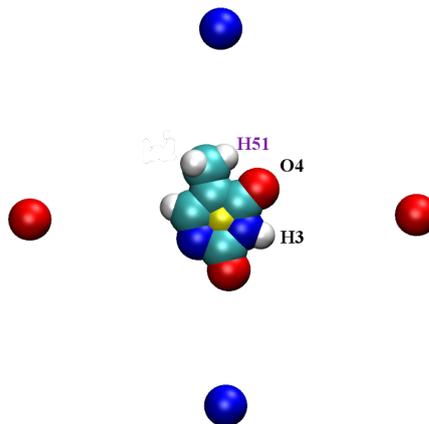


Figure 3.49: Thymine: the yellow atom represents the center of mass of thymine, while the red points and the green one two of the axes (the blue ones are not showed since one will appear on the thymine and the other one below it). H3 and O4 are atoms involved in H-bonds, while the H51 will be chosen to discriminate axes.

Now we do dot product between eigenvectors and N4 atom position⁹ and associate the higher value to the y -axis and the intermediate one to x -axis, changing the sign whether the values are negative. Next we do a cross products between these two axes to obtain the z -axis.

	Eigenvector			Dot product
x	-0.2140	-0.9736	-0.0797	0.0898
y	-0.9768	0.2142	0.0066	-3.1511
z	0.0107	0.0793	-0.9968	-0.0007
q	0.0290	-0.6262	0.7786	0.0276
COM	-0.26190	3.4545	0.2751	

Table 3.13: Thymine: eigenvectors, quaternion and center of mass.

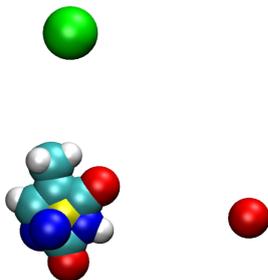


Figure 3.50: Thymine: x (red) and y (green). z -axes (blue) is beneath the molecule.

⁹(position of H51 - COM)*eigenvectors

We obtain snapshots already shown.

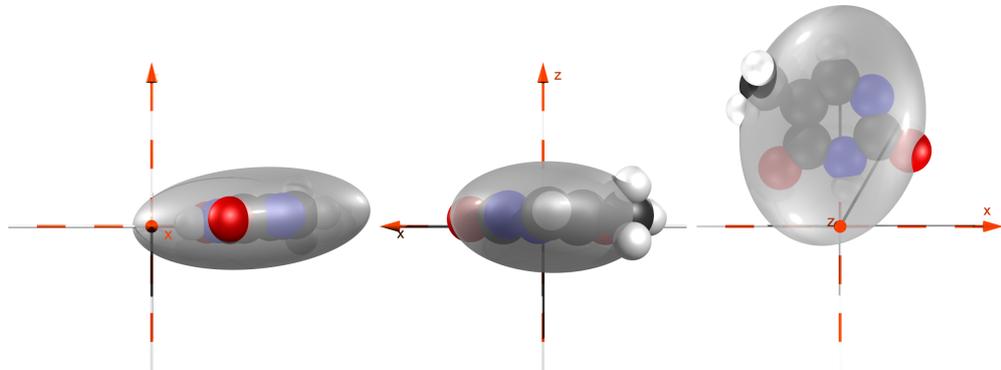


Figure 3.51: Thymine: crystallographic positions and ellipsoids, x , y and z -view.

We just check if LAMMPS put the H-bond sites exactly in the atomistic positions: they superimpose exactly.

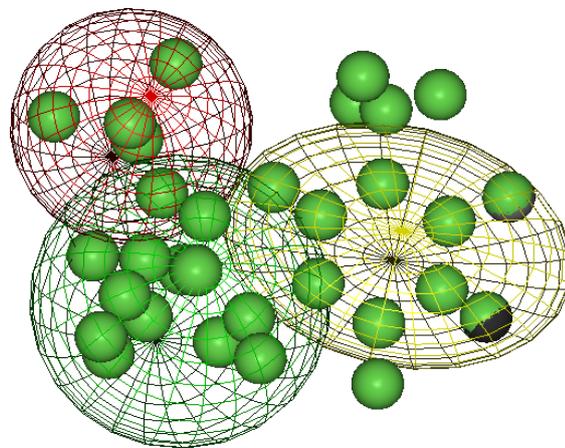


Figure 3.52: Thymine: LAMMPS. Coarse-grained and atomistic representations: the darker sphere are the position of the H-bonds position as computed by LAMMPS.

3.6.3 Parameters: beads

	σ			ϵ			
	σ_x	σ_y	σ_z	σ_c	ϵ_x	ϵ_y	
Sugar Phosphate	5.508				1.566		$COM = (2.3700, 6.5997, 1.2717)$
	4.472				1.164		$COM = (-0.6275, 8.8956, 2.1865)$
Adenine	σ_x	σ_y	σ_z	σ_c	ϵ_x	ϵ_y	ϵ_z
	5.114	4.653	3.032	3.308	1.494	1.579	8.573
Cytosine							$COM=(0.2701, 2.4699, 0.2026)$ $q=(0.0311, -0.6860, 0.7265, 0.0251)$
	5.187	4.296	3.023	3.195	0.855	1.269	6.085
Guanine							$COM=(0.0645, 3.2736, 0.2641)$ $q=(-0.0336, 0.7595, -0.6493, -0.02160)$
	6.378	4.736	3.010	3.258	0.771	1.517	9.615
Thymine							$COM=(0.5708, 2.2047, 0.1846)$ $q=(-0.0372, 0.8681, -0.4947, -0.0148)$
	5.852	4.459	3.534	3.005	0.803	1.289	5.322
							$COM=(-0.26190, 3.4545, 0.2751)$ $q=(0.0290, -0.6262, 0.7786, 0.0276)$

Table 3.14: Bead parametrization: spheres and ellipsoids.

3.6.4 Force field parametrization

For parametrizing the coarse-grained model we are developing, potential terms need to be characterized with energy constants. First of all, we carried out an atomistic simulation of a 10 base pairs long sequence (ACAA-GAACTA), at $T=302$ K for 40 ns. Energy constants are computed using the Boltzmann inversion method:

$$U_{eff}(x) = -k_B T \ln[p(\mathbf{x})] \quad (3.11)$$

where x is the order parameter of the potential, and $p(\mathbf{x})$ is the probability density of observing the system with the value of the order parameter being \mathbf{x} . We collect data for each parameter (i.e. for various type of bonds, angles, dihedrals, H-bonds, orientational angles, orientational dihedrals) and fitted them with various kind of potentials. Actually, we fit the plot of $p(\mathbf{x})$ and $-\ln p(\mathbf{x})$, since $-k_B T$ is a constant. We choose the best potential equation for each parameter and added it in LAMMPS. For each potential added in LAMMPS, various tests are carried out. First of all we check if the energy of this code and the one computed on atomistic simulations are the same. Once we are sure about the potential energy equation, we check also the force profile. Then, we run some NVE run for each type of parameter to check whether the energy is really constant during a simulation. Formula of potential energy and force are reported. Note that all potentials, besides electrostatic and Gay-Berne (which are not obtained from atomistic simulations) are multiplied by a factor $kt = 0.5997$ kcal/mol to report all data to the atomistic value.

3.6.4.1 Potential terms

Equations used for fitting data are shown below.

- harmonic:

$$E = k (r - r_0)^2 + s \quad (3.12)$$

- harmonic-like, including two different harmonic:

$$E = \sum_{i=1}^n k_i (r - r_0)^{i+1} + \sum_{j=1}^n k_j (r - r_1)^{j+1} + s \quad (3.13)$$

with $i = j = [1; 4]$;

- cosine–harmonic type:

$$E = k[1 - \cos(nr)] + s \quad (3.14)$$

- cosine type:

$$E = \sum_{i=1}^n k_n [1 - \cos(nr - r_n)] + s \quad (3.15)$$

- exponential type:

$$E = -\ln \left[e^{\frac{-k_1}{2}(r-r_1)^2} + A e^{\frac{-k_2}{2}(r-r_2)^2} \right] + s \quad (3.16)$$

- Morse type:

$$E = D \left[1 - e^{-\sqrt{\frac{K}{2D}}(r-r_0)} \right]^2 + s \quad (3.17)$$

- Lennard–Jones 6–12:

$$E = 4\varepsilon \left[\left(\frac{\sigma_o}{r} \right)^{12} - \left(\frac{\sigma_o}{r} \right)^6 \right] + s \quad (3.18)$$

- Lennard–Jones 10–12:

$$E = 4\varepsilon \left[\left(\frac{\sigma_o}{r} \right)^{12} - \left(\frac{\sigma_o}{r} \right)^{10} \right] + s \quad (3.19)$$

- Lennard–Jones 9–12:

$$E = 4\varepsilon \left[\left(\frac{\sigma_o}{r} \right)^{12} - \left(\frac{\sigma_o}{r} \right)^9 \right] + s \quad (3.20)$$

- Lennard–Jones 6–9:

$$E = 4\varepsilon \left[\left(\frac{\sigma_o}{r} \right)^9 - \left(\frac{\sigma_o}{r} \right)^6 \right] + s \quad (3.21)$$

- Lennard–Jones 4–6:

$$E = 4\varepsilon \left[\left(\frac{\sigma_o}{r} \right)^6 - \left(\frac{\sigma_o}{r} \right)^4 \right] + s \quad (3.22)$$

3.6.4.2 Bonds

For bonds two different potentials are used. For the S(5')-P backbone bond we use an “exponential” potential:

$$U(r) = -\ln\left[e^{-\frac{k_1}{2}(r-r_1)^2} + A e^{-\frac{k_2}{2}(r-r_2)^2}\right] + s \quad (3.23)$$

$$F(r) = \frac{-1}{Z + W A}[-Z k_1(r - r_1) - A k_2 W(r - r_2)] \quad (3.24)$$

with $Z = e^{-\frac{k_1}{2}(r-r_1)^2}$, $W = e^{-\frac{k_2}{2}(r-r_2)^2}$. For all other bonds we used an “harmonic-like” potential:

$$U(r) = k (r - r_0)^2 + s \quad (3.25)$$

$$F(r) = -2 k(r - r_0) \quad (3.26)$$

It follows the coefficient for each type of bonds, as used in the LAMMPS input files. Note that *fac* is a factor used to rescale the energy scale, since it is simply added to each potential.

```

bond_style hybrid harmonic_double exponential_dna
#Harmonic_double coeffs: k, r0, fac, kt
#Exponential_dna coeffs: r0, r1, k1, k2, prefac, fac, kt
bond_coeff 1 harmonic_double 35.3402 4.17089 0 ${kt}
bond_coeff 3 harmonic_double 102.39 5.12367 0 ${kt}
bond_coeff 4 harmonic_double 69.6751 4.6784 0 ${kt}
bond_coeff 5 harmonic_double 81.5909 4.68592 0 ${kt}
bond_coeff 6 harmonic_double 89.6242 5.17784 0 ${kt}
bond_coeff 2 exponential_dna 3.638 3.29 225.52 226.24 0.04 0 ${kt}

```

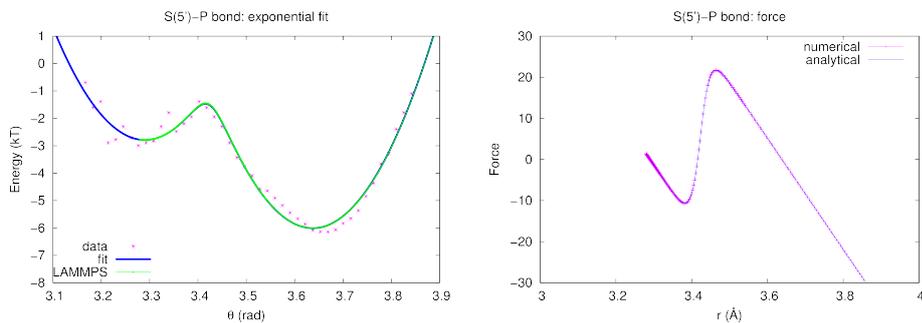


Figure 3.53: Bond, S(5')-P, type 2: potential energy and force.

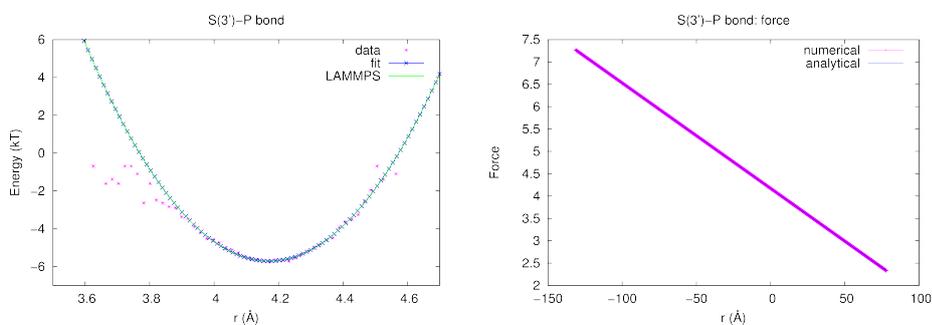


Figure 3.54: Bond, S(3')-P, type 1: potential energy and force.

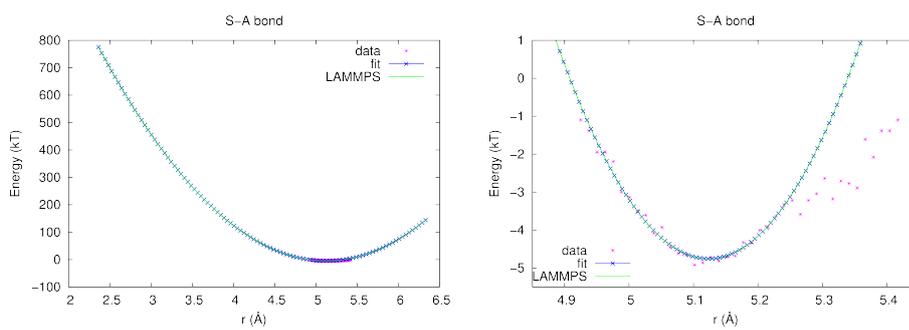


Figure 3.55: Bond, S-A, type 3: potential

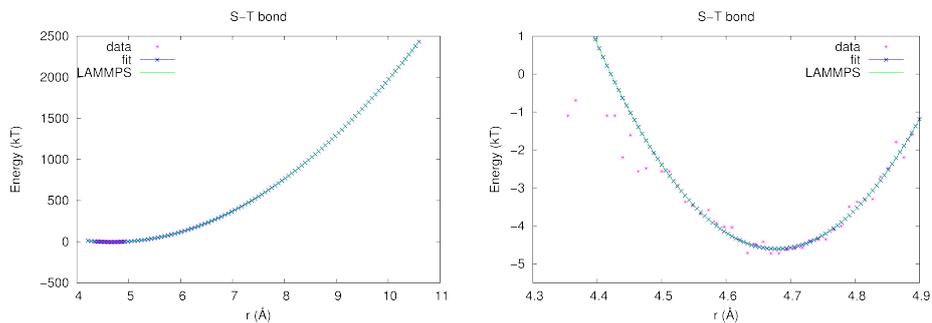


Figure 3.56: Bond, S-T, type 4: potential.

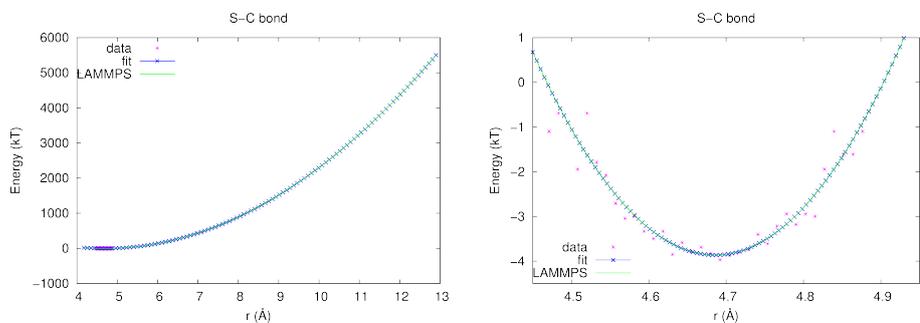


Figure 3.57: Bond, S-C, type 5: potential.

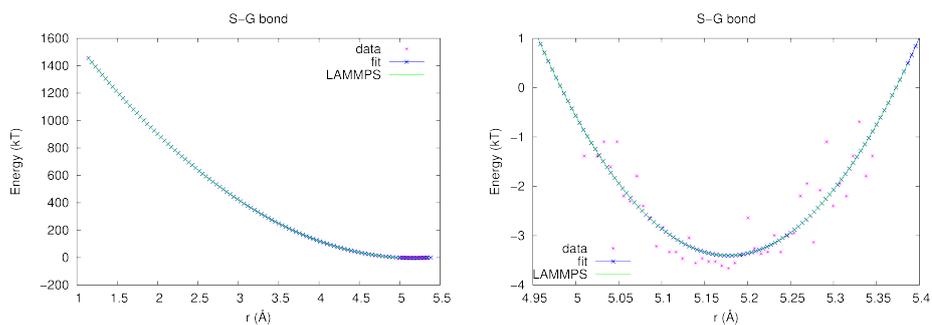


Figure 3.58: Bond, S-G, type 6: potential.

3.6.4.3 Angles

Also for angles, two different type of energy are used. The first one is an “harmonic-like” potential with the add of cosh:

$$U(\theta) = k(\theta - \theta_0)^2 + a \cosh(b\theta + c) + s \quad (3.27)$$

$$F(\theta) = -2k(\theta - \theta_0) - a b \sinh(b\theta + c) \quad (3.28)$$

For pyrimidine nucleotides (thymine and cytosine) we use an “exponential” energy term, always adding a cosh term:

$$U(\theta) = -\ln[e^{-\frac{k_1}{2}(\theta-\theta_1)^2} + A e^{-\frac{k_2}{2}(\theta-\theta_2)^2}] + a \cos(b\theta + c) + s \quad (3.29)$$

$$F(\theta) = \frac{-1}{Z + W A} [-Z k_1(\theta - \theta_1) - A k_2 W(\theta - \theta_2)] - a b \sinh(b\theta + c) \quad (3.30)$$

with $Z = e^{-\frac{k_1}{2}(\theta-\theta_1)^2}$ and $W = e^{-\frac{k_2}{2}(\theta-\theta_2)^2}$.

Note that we had to add a hyperbolic cosine in order to limit the “range of existence” of each angle type, so that outside this range the potential goes to infinite.

It follows the coefficient for each type of angles, as used in the LAMMPS input files.

```
angle_style hybrid harmonic_cosh exponential_cosh
#Harmonic_cosh coeffs: k, theta, fac, kk, r0, rc, kt
#Exponential_cosh coeffs: k1, theta1, prefac, k2, theta2, fac, kk, r0, rc, kt
angle_coeff 1 harmonic_cosh 54.6003 1.66965 0 3.25118e-22 19.978 -4.9906 ${kt}
angle_coeff 2 harmonic_cosh 15.501 2.05535 0 0.0322208 -7.47903 -14.9271 ${kt}
angle_coeff 3 harmonic_cosh 12.9177 1.65595 0 0.0958189 -7.76696 -13.0903 ${kt}
angle_coeff 4 harmonic_cosh 32.0463 2.11268 0 0.0165736 -10.5234 -22.5427 ${kt}
angle_coeff 9 harmonic_cosh 10.1243 1.73229 0 0.00297503 -11.6564 -21.1934 ${kt}
angle_coeff 10 harmonic_cosh 20.6649 2.11471 0 0.0377925 -9.54289 -21.3376 ${kt}
angle_coeff 5 exponential_cosh 92.6006 1.71071 3.77048 35.0268 2.27718 0 0.00208543 -10.3094 -21.6706 ${kt}
angle_coeff 6 exponential_cosh 457.525 1.3962 1.74063 59.5424 1.65626 0 0.0302943 -10.4557 -17.0901 ${kt}
angle_coeff 7 exponential_cosh 133.167 1.63911 0.924637 85.6403 1.9206 0 3.24126e-12 9.7508 -4.99046 ${kt}
angle_coeff 8 exponential_cosh 70.5017 2.32062 1.88889 180.842 2.0133 0 0.0267666 -10.3403 -23.0155 ${kt}
```

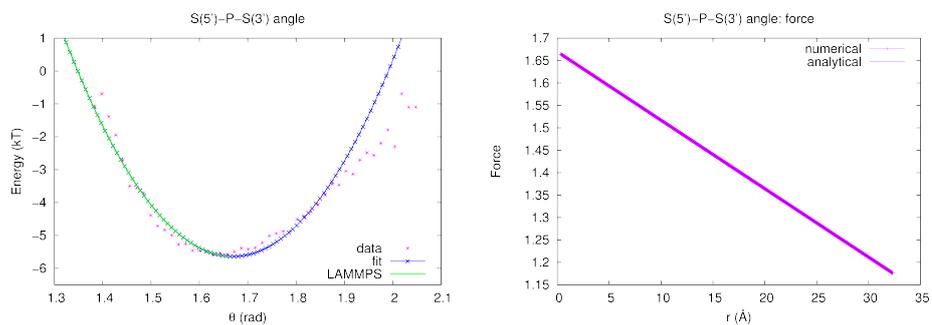


Figure 3.59: Angle, $S(5')\text{-P-S}(3')$, type 1: potential and force.

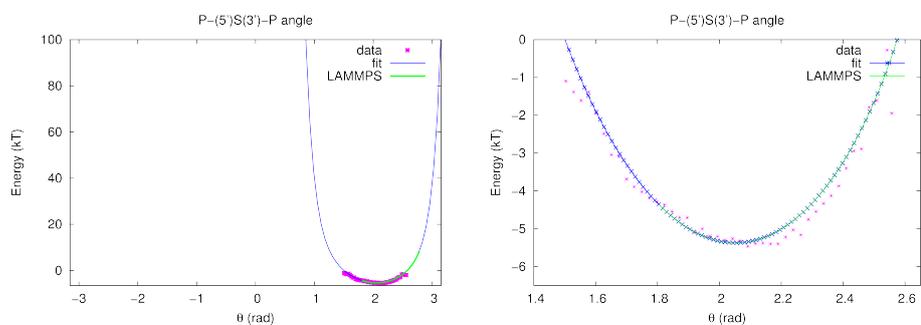


Figure 3.60: Angle, $P-(5')S(3')\text{-P}$, type 2: potential.

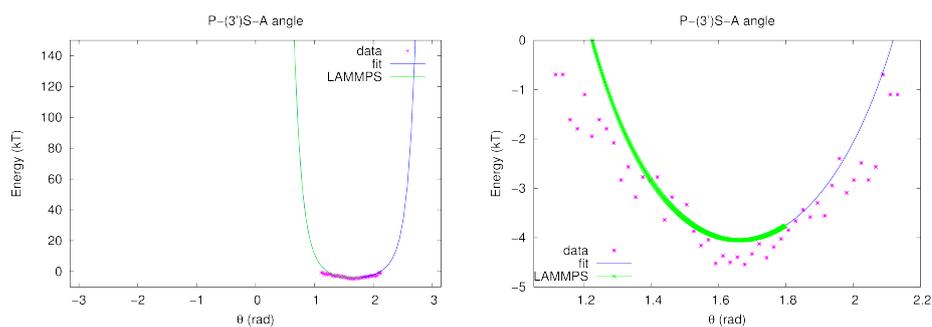


Figure 3.61: Angle, $P-(3')S\text{-A}$, type 3: potential.

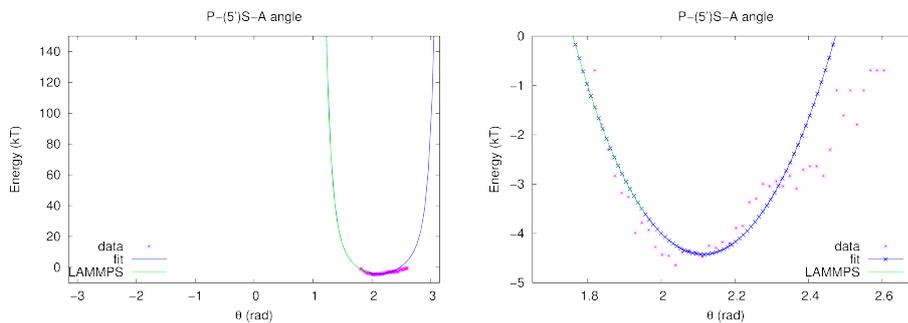


Figure 3.62: Angle, P-(5')S-A, type 4: potential.

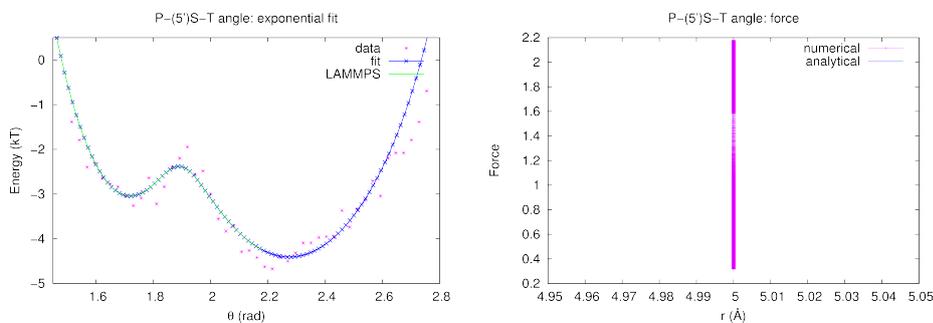


Figure 3.63: Angle, P-(5')S-T, type 5: potential and force.

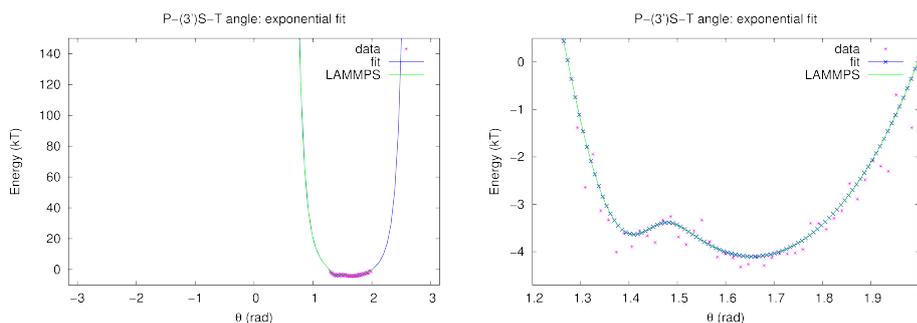


Figure 3.64: Angle, P-(3')S-T, type 6: potential.

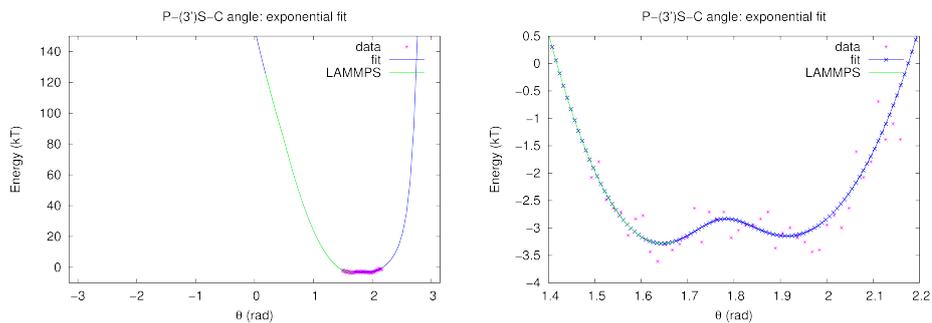


Figure 3.65: Angle, P-(3')S-C, type 7: potential.

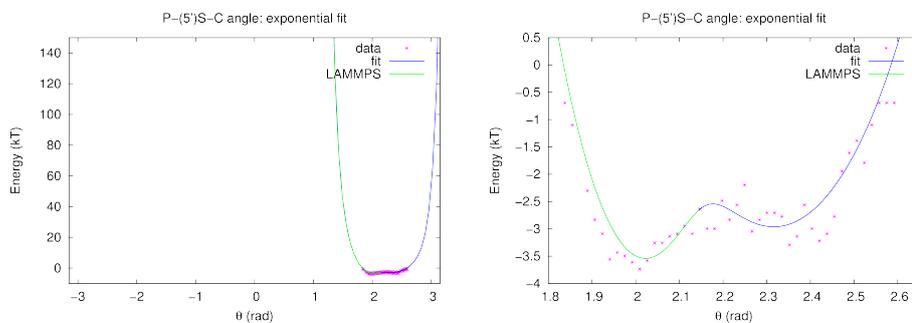


Figure 3.66: Angle, P-(5')S-C, type 8: potential.

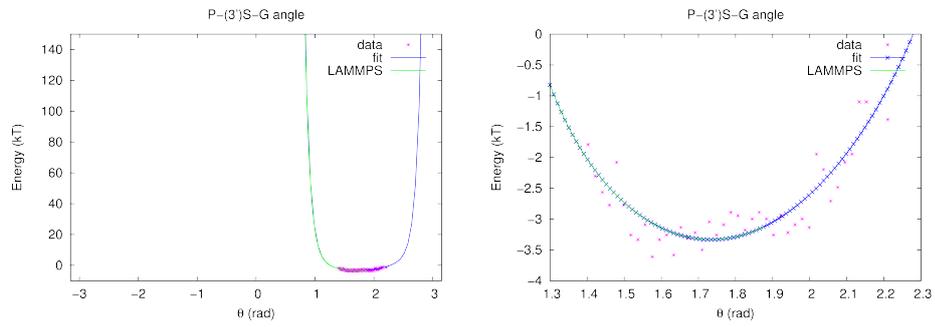


Figure 3.67: Angle, P-(3')S-G, type 9: potential.

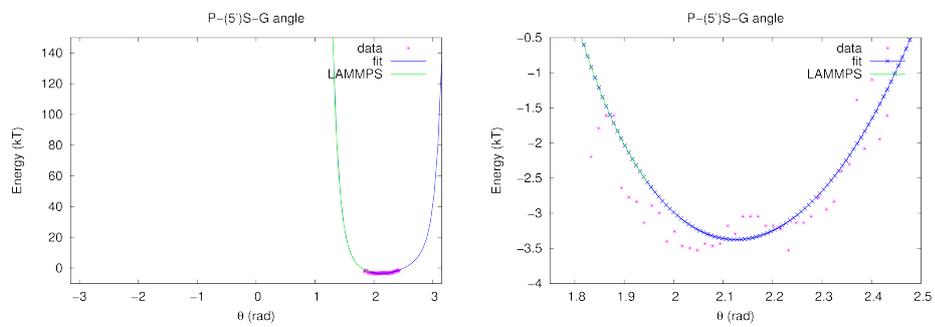


Figure 3.68: Angle, P-(5')S-G, type 10: potential

3.6.4.4 Dihedrals

For dihedrals type, we use only one potential type:

$$U(\phi) = k \cosh(r_0 \phi + r_c) + s + \sum_{i=1}^3 k_i [1 - \cos(i \phi - \phi_i)] \quad (3.31)$$

$$F(\phi) = -\frac{\partial U}{\partial \phi} = -k r_0 \sinh(r_0 \phi + r_c) - \sum_{i=1}^3 i k_i \sin(i \phi - \phi_i) \quad (3.32)$$

Note that in this case we also had to change how the dihedral angle and the derivative of this angle with respect to the various atom/particles involved in this angle are computed. We use the definition published in reference [65] multiplying all for a factor with takes into account the sign of the dihedral. So the definition of angle we use is:

$$\cos(\phi) = \text{sign}(\phi) \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}| |\mathbf{B}|} \quad (3.33)$$

$$\phi = \arccos \left[\text{sign}(\phi) \frac{\mathbf{A} \cdot \mathbf{B}}{|\mathbf{A}| |\mathbf{B}|} \right] \quad (3.34)$$

where

$$\mathbf{A} = \mathbf{F} \times \mathbf{G}, \quad (3.35)$$

$$\mathbf{B} = \mathbf{H} \times \mathbf{G}, \quad (3.36)$$

$$\text{sign}(\phi) = \begin{cases} +1 & \text{if } -\mathbf{G} [\mathbf{F} \times (-\mathbf{G})] \times [(-\mathbf{H}) \times (-\mathbf{G})] \geq 0 \\ -1 & \text{if } -\mathbf{G} [\mathbf{F} \times (-\mathbf{G})] \times [(-\mathbf{H}) \times (-\mathbf{G})] < 0 \end{cases} \quad (3.37)$$

and

$$\mathbf{F} = \mathbf{r}_i - \mathbf{r}_j, \quad (3.38)$$

$$\mathbf{G} = \mathbf{r}_j - \mathbf{r}_k, \quad (3.39)$$

$$\mathbf{H} = \mathbf{r}_l - \mathbf{r}_k. \quad (3.40)$$

The Figure 3.69 shows the order of atoms of a dihedral angle.

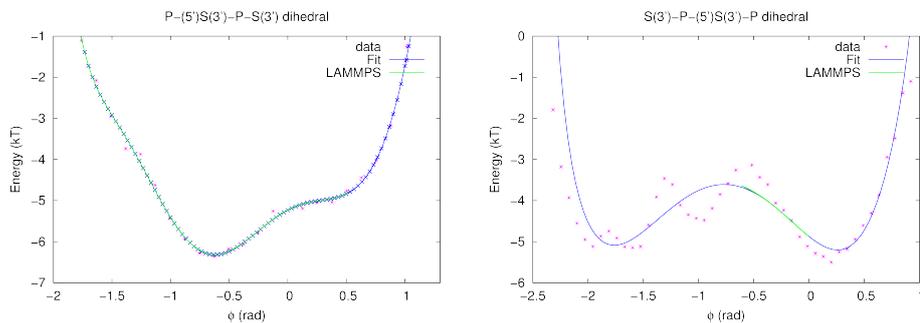


Figure 3.70: Dihedrals: P-(5')S(3')-P-(5')S (1st coefficient] and S(3')-P-(5')S(3')-P (2nd coefficient)].

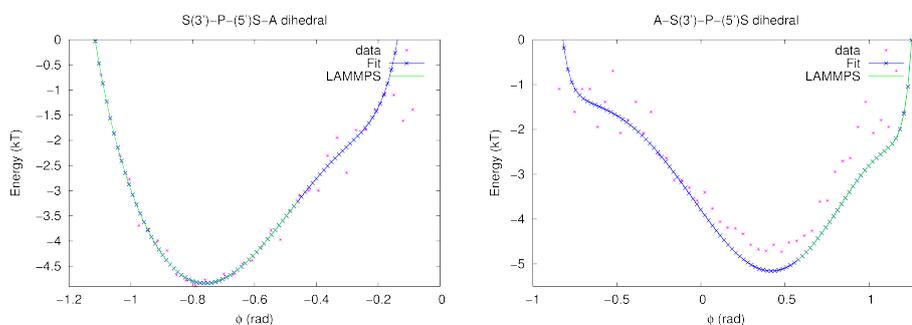


Figure 3.71: Dihedrals: S(3')-P-(5')S(3')-A (3rd coefficient] and A-S(3')-P-(5')S(3') (4th coefficient)].

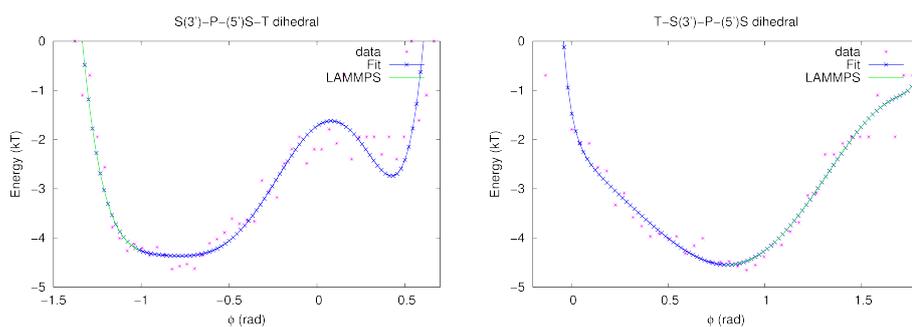


Figure 3.72: Dihedrals: S(3')-P-(5')S(3')-T (5th coefficient] and T-S(3')-P-(5')S(3') (6th coefficient)].

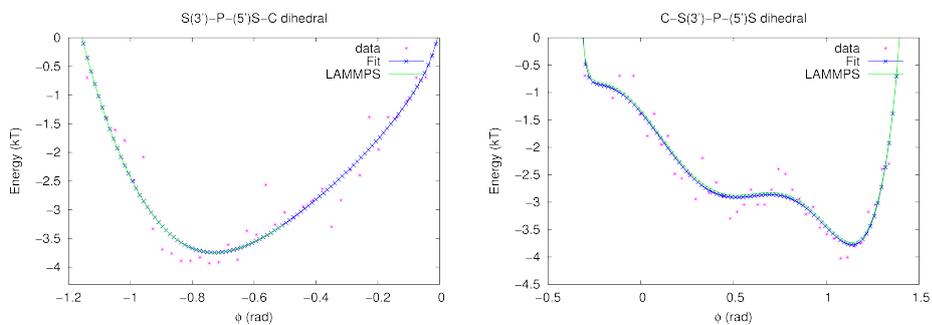


Figure 3.73: Dihedrals: S(3')-P-(5')S(3')-C (7th coefficient] and C-S(3')-P-(5')S(3') (8th coefficient].

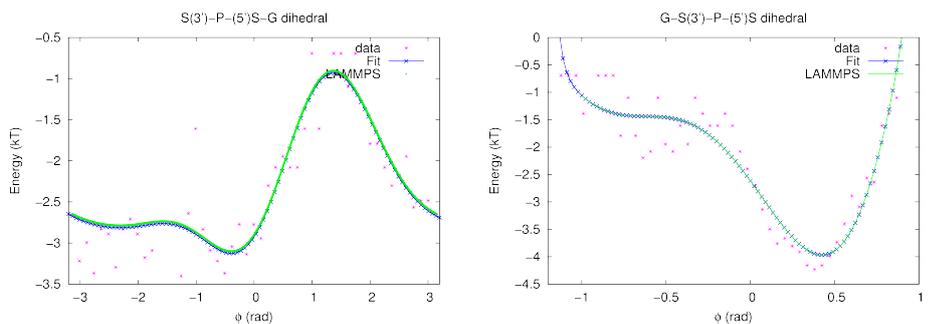


Figure 3.74: Dihedrals: S(3')-P-(5')S(3')-G (9th coefficient] and G-S(3')-P-(5')S(3') (10th coefficient].

3.6.4.5 H-bonds

To take into account the hydrogen bonds, a L-J 4-6 potential is used.

$$E(r) = 4\varepsilon [(\sigma_0/r)^6 - (\sigma_0/r)^4] + s \quad (3.46)$$

$$F(r) = -\frac{8\varepsilon\sigma_0^4}{r^7} \left(2r^2 - 3\sigma_0^2 \right) \quad (3.47)$$

It follows the coefficient for each type of H-bond, as used in the LAMMPS input files. Then cutoff and energy scale values have been modify by hand. The cutoff has been set to 2.5 Å, in order not to have discontinuous step in the potential

```
variable cutoffHbond equal 2.5

variable ntypes equal 4
variable nsitesADE equal 2
variable nsitesTHY equal 2
variable nsitesCYT equal 3
variable nsitesGUA equal 3
variable siteADEH61_x equal 2.737
variable siteADEH61_y equal 1.962
variable siteADEH61_z equal 0.000
variable shapeADEH61 equal 0.10
variable wellADEH61 equal 2.40
variable siteADEN1_x equal 1.939
variable siteADEN1_y equal -0.416
variable siteADEN1_z equal 0.000
variable shapeADEN1 equal 0.10
variable wellADEN1 equal 3.10
variable siteTHY04_x equal 1.800
variable siteTHY04_y equal 1.474
variable siteTHY04_z equal 0.000
variable shapeTHY04 equal 0.10
variable wellTHY04 equal 3.04
variable siteTHYH3_x equal 1.926
variable siteTHYH3_y equal -0.980
variable siteTHYH3_z equal 0.001
variable shapeTHYH3 equal 0.10
variable wellTHYH3 equal 2.40
variable siteCYTN3_x equal 0.998
variable siteCYTN3_y equal 0.021
variable siteCYTN3_z equal 0.000
variable shapeCYTN3 equal 0.10
variable wellCYTN3 equal 3.10
variable siteCYT02_x equal 0.986
variable siteCYT02_y equal -2.255
variable siteCYT02_z equal 0.000
variable shapeCYT02 equal 0.10
variable wellCYT02 equal 3.04
variable siteCYTH41_x equal 1.939
variable siteCYTH41_y equal 2.309
variable siteCYTH41_z equal 0.001
variable shapeCYTH41 equal 0.10
variable wellCYTH41 equal 2.40
variable siteGUAH1_x equal 2.247
variable siteGUAH1_y equal 1.401
```

```

variable siteGUAH1_z equal 0.000
variable shapeGUAH1 equal 0.10
variable wellGUAH1 equal 2.40
variable siteGUAH21_x equal 3.699
variable siteGUAH21_y equal -0.239
variable siteGUAH21_z equal 0.000
variable shapeGUAH21 equal 0.10
variable wellGUAH21 equal 2.40
variable siteGUA06_x equal 0.097
variable siteGUA06_y equal 2.622
variable siteGUA06_z equal 0.000
variable shapeGUA06 equal 0.10
variable wellGUA06 equal 3.04

variable sigmaADE_THY_H61_04 equal 1.529
variable epsilonADE_THY_H61_04 equal 7.40
variable sigmaADE_THY_N1_H3 equal 1.552
variable epsilonADE_THY_N1_H3 equal 18.88
variable sigmaCYT_GUA_N3_H1 equal 1.605
variable epsilonCYT_GUA_N3_H1 equal 21.60
variable sigmaCYT_GUA_O2_H21 equal 1.495
variable epsilonCYT_GUA_O2_H21 equal 10.54
variable sigmaCYT_GUA_H41_06 equal 1.653
variable epsilonCYT_GUA_H41_06 equal 8.14

pair_style hbondsite_dna ${ntypes} 3 ${nsitesADE} ${siteADEH61_x} ${siteADEH61_y} ${siteADEH61_z} ${siteADEN1_x} ${siteADEN1_y} ${siteADEN1_z} 4 $
#Hbondsite_dna coeffs: site, site, sigmaij, epsilonij, cutoff, facij, kt
pair_coeff * * hbondsite_dna 1 1 0.0 0.0 ${cutoffHbond} 0.0 ${kt}
pair_coeff * * hbondsite_dna 1 2 0.0 0.0 ${cutoffHbond} 0.0 ${kt}
pair_coeff * * hbondsite_dna 1 3 0.0 0.0 ${cutoffHbond} 0.0 ${kt}
pair_coeff * * hbondsite_dna 2 2 0.0 0.0 ${cutoffHbond} 0.0 ${kt}
pair_coeff * * hbondsite_dna 2 3 0.0 0.0 ${cutoffHbond} 0.0 ${kt}
pair_coeff * * hbondsite_dna 3 3 0.0 0.0 ${cutoffHbond} 0.0 ${kt}
pair_coeff 3 4 hbondsite_dna 1 1 ${sigmaADE_THY_H61_04} ${epsilonADE_THY_H61_04} ${cutoffHbond} 0.00000 ${kt}
pair_coeff 3 4 hbondsite_dna 2 2 ${sigmaADE_THY_N1_H3} ${epsilonADE_THY_N1_H3} ${cutoffHbond} 0.00000 ${kt}
pair_coeff 5 6 hbondsite_dna 1 1 ${sigmaCYT_GUA_N3_H1} ${epsilonCYT_GUA_N3_H1} ${cutoffHbond} 0.00000 ${kt}
pair_coeff 5 6 hbondsite_dna 2 2 ${sigmaCYT_GUA_O2_H21} ${epsilonCYT_GUA_O2_H21} ${cutoffHbond} 0.00000 ${kt}
pair_coeff 5 6 hbondsite_dna 3 3 ${sigmaCYT_GUA_H41_06} ${epsilonCYT_GUA_H41_06} ${cutoffHbond} 0.00000 ${kt}

```

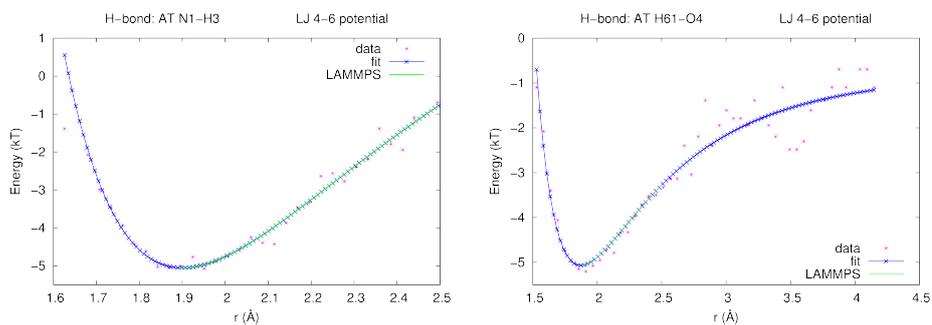


Figure 3.75: H-bonds, A-T: N1-H3 and H61-O4.

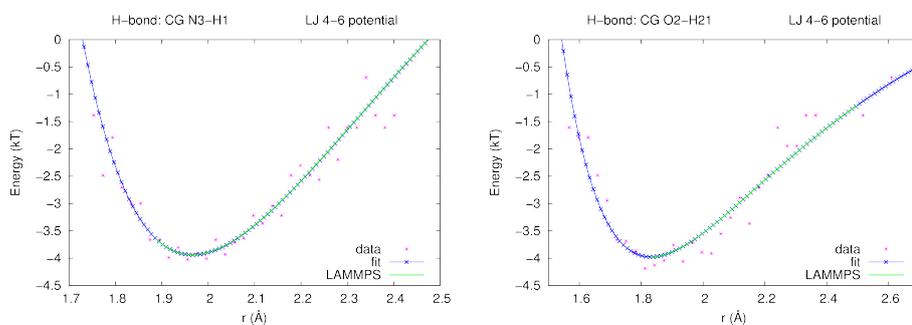


Figure 3.76: H-bonds, C-G: H1-N3 and H21-O2.

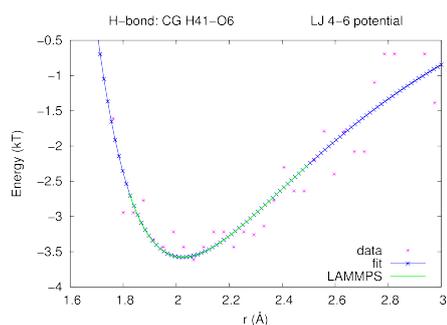


Figure 3.77: H-bonds, C-G: O6-H41.

3.6.4.6 Angle axis

Bases belonging to free nucleotides are able to rotate completely around the bond which connects the sugar and the base itself. On the contrary, when a nucleotide is part of a DNA molecule, due to steric interactions, H-bonds and stacking interactions, bases could not rotate freely. To avoid the full rotation, other two potentials are added to the force field that describes our coarse-grained DNA model: the first one concerns angles and the second one dihedral angles. We decide to compute angles (Figure 3.78, *left*) between the center of mass of the sugar, that of the base and each of the three eigenvectors of the inertia tensor of the base itself and dihedrals (Figure 3.78, *right*) between the center of mass of the phosphate group, that of the sugar, that of the base and each of the three eigenvectors. These two potentials force the base to maintain one of the biologically allowed positions.



Figure 3.78: Angle (*left*) and dihedral (*right*) between beads and eigenvectors.

Angles between beads and base axes (Figure 3.78, *left*) are computed using the following equation:

$$\alpha = 180.0 - \left[\arccos \left(\frac{\mathbf{A} \cdot \mathbf{B}}{d_A d_B} \right) \frac{180}{\pi} \right], \quad (3.48)$$

where $d_A = \sqrt{x_a^2 + y_a^2 + z_a^2}$, $\mathbf{A} = \mathbf{P1} - \mathbf{P2}$ and $\mathbf{B} = \mathbf{r}_i$ (\mathbf{r}_i are the three base axes).

On the other hands, dihedrals (Figure 3.78, *right*) between beads and base axes are computed using the following equation:

$$\beta = \frac{180}{\pi} \arctan 2(d_A \mathbf{C} \cdot [\mathbf{B} \times \mathbf{C}], [\mathbf{A} \times \mathbf{B}] \cdot [\mathbf{B} \times \mathbf{C}]), \quad (3.49)$$

where $\mathbf{A} = \mathbf{P1} - \mathbf{P2}$, $\mathbf{B} = \mathbf{P3} - \mathbf{P2}$ and $\mathbf{C} = \mathbf{r}_i$ (\mathbf{r}_i are the three base axes).

In first approximation we decide to analyze only the angles.

The potential and the force equations used are reported below:

$$U(\theta) = \sum_{i=1}^4 k_i \cos(i\theta) \quad (3.50)$$

$$F(\theta) = \sum_{i=1}^4 i k_i \sin(i\theta) \quad (3.51)$$

```
angle_style hybrid axis_cosines
#Axis_cosines coeffs: k1, k2, k3, k4, fac, kt
angle_coeff 11 axis_cosines -0.0939464 -34.4924 -0.523703 16.1368 0 ${kt}
angle_coeff 12 axis_cosines -0.40034 43.3321 -0.0621376 18.9803 0 ${kt}
angle_coeff 13 axis_cosines -4.14019 109.035 -1.17255 24.9779 0 ${kt}
angle_coeff 14 axis_cosines -0.39726 -12.9484 0.229668 11.5601 0 ${kt}
angle_coeff 15 axis_cosines -1.17129 20.299 -0.712862 16.1073 0 ${kt}
angle_coeff 16 axis_cosines -1.71566 45.1023 -0.560333 7.49234 0 ${kt}
angle_coeff 17 axis_cosines -0.411834 9.86552 -0.989607 14.264 0 ${kt}
angle_coeff 18 axis_cosines -1.02534 -9.86233 -1.13456 16.8377 0 ${kt}
angle_coeff 19 axis_cosines -2.55436 70.7296 -0.750511 12.7156 0 ${kt}
angle_coeff 20 axis_cosines -0.903887 12.0773 -0.656171 14.0762 0 ${kt}
angle_coeff 21 axis_cosines -0.37844 -11.6017 -0.906853 15.5343 0 ${kt}
angle_coeff 22 axis_cosines 4.8769 54.6031 1.93243 6.59921 0 ${kt}
```

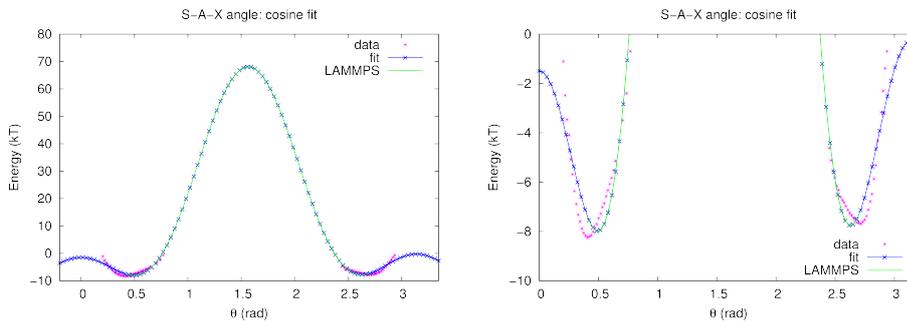


Figure 3.79: Angle axis, S-A-X

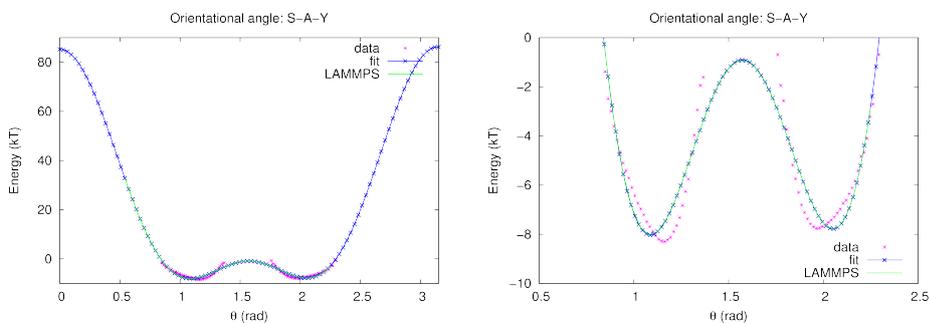


Figure 3.80: Angle axis, S-A-Y.

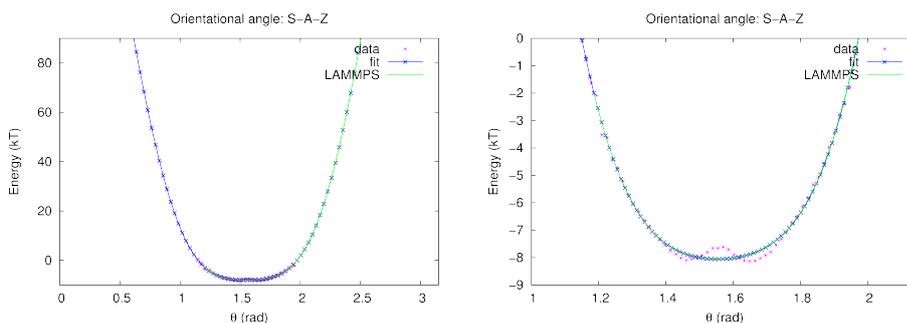


Figure 3.81: Angle axis, S-A-Z.

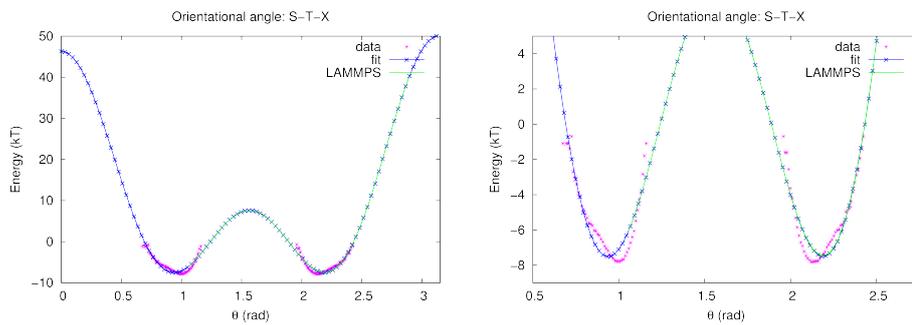


Figure 3.82: Angle axis, S-T-X.

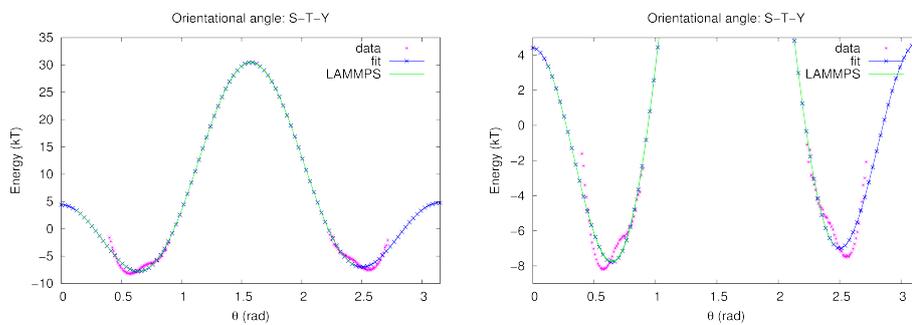


Figure 3.83: Angle axis, S-T-Y.

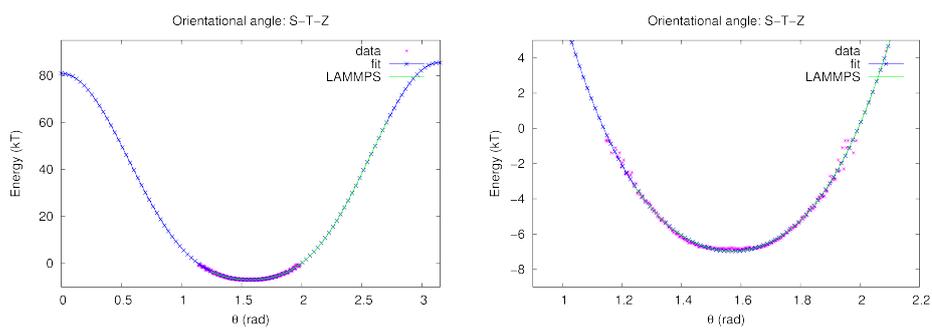
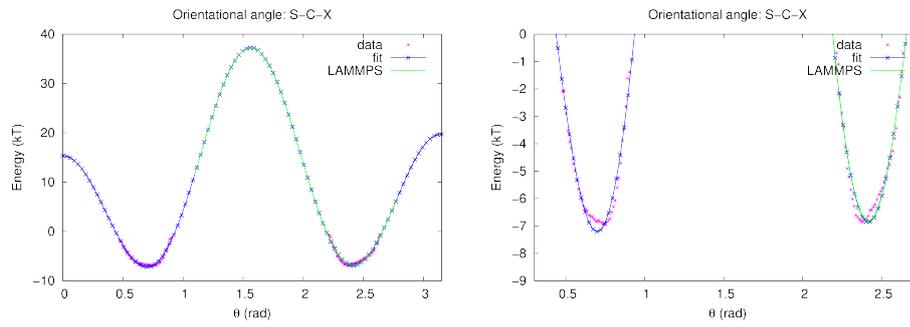
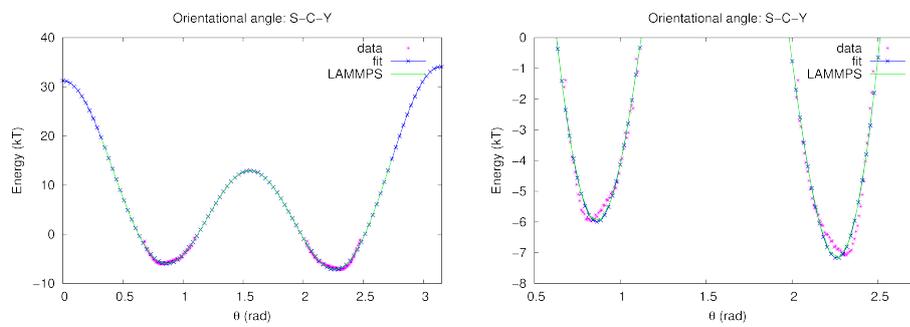
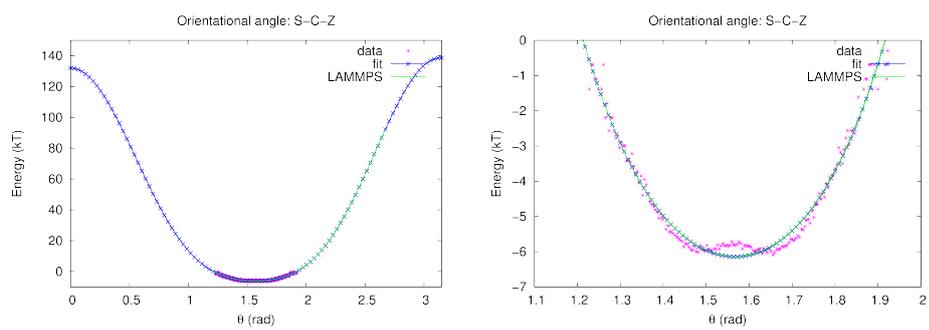


Figure 3.84: Angle axis, S-T-Z.

**Figure 3.85:** Angle axis, S-C-X.**Figure 3.86:** Angle axis, S-C-Y.**Figure 3.87:** Angle axis, S-C-Z.

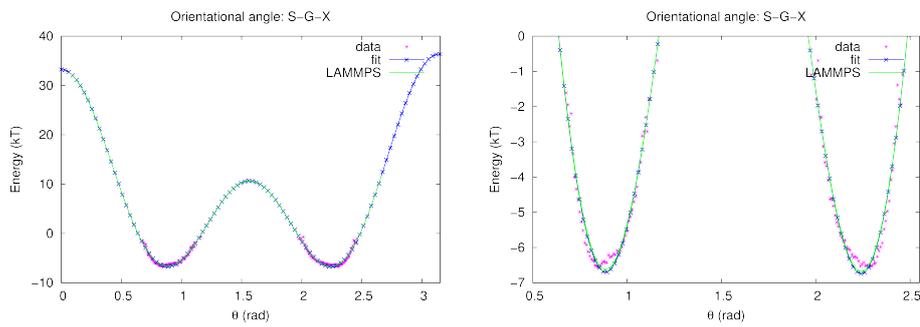


Figure 3.88: Angle axis, S-G-X.

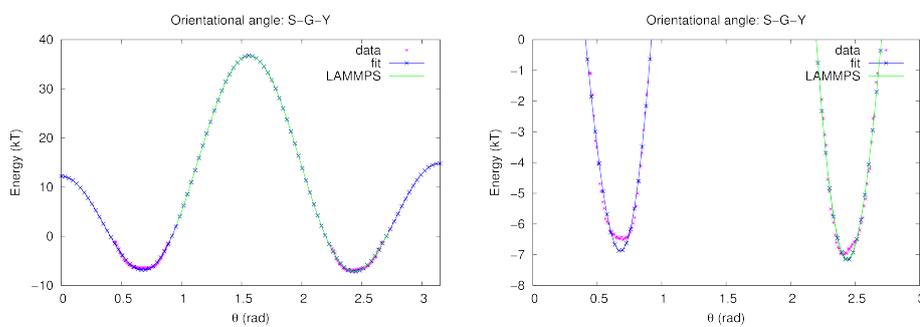


Figure 3.89: Angle axis, S-G-Y.

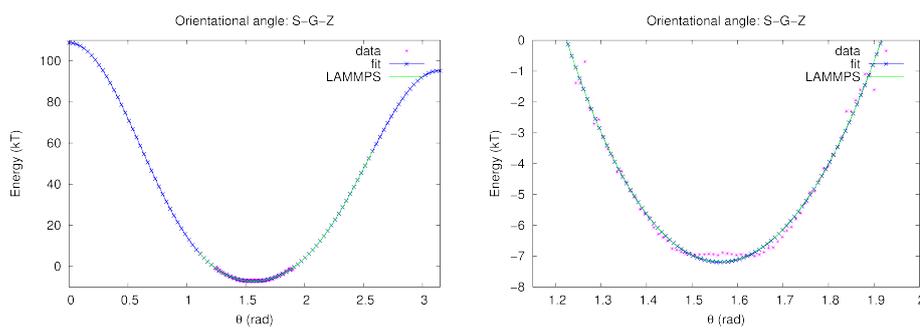


Figure 3.90: Angle axis, S-G-Z.

3.6.4.7 Stacking interactions

We prepare input configurations for each pair of bases, as they belong to the same strand of an helix and as they are contiguous.

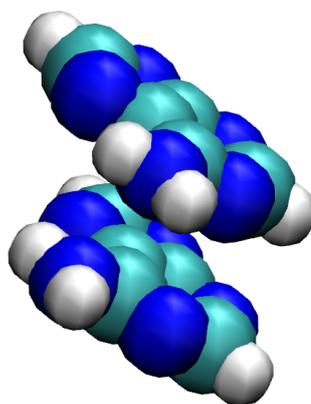


Figure 3.91: Example of two stacked adenine.

With the same protocol used to fit ellipsoids, we obtain σ_i and ε_i also for interaction between different bases. Table 3.15 summarizes values of σ_i (defining the shape of the ellipsoids), r and ϕ (defining how to construct the B-helix) and the mass of each bead.

	σ_x	σ_y	σ_z	r	ϕ	Mass (a.m.u.)
Sugar		5.508		7.0123	70.246	83.11
Phosphate		4.472		8.9177	94.035	94.97
Adenine	7.306	6.647	3.032	2.4846	83.759	134.1
Cytosine	7.410	6.137	3.023	3.2742	88.871	110.1
Guanine	9.112	6.766	3.010	2.2774	75.485	150.1
Thymine	8.360	6.370	3.534	3.4644	94.336	125.1

Table 3.15: Values of σ_i , r , ϕ and mass for bases [64].

It follows the interesting part of the LAMMPS input file, assuming that spheres do not interact with ellipsoids (since this stacking term does not

take into account part of the nucleotide that does not contain an aromatic base and that are not placed one above the other).

```

variable gb_cutoff equal 15.0
variable gb_mix equal 0.000000000001
variable sigmac_ps equal 3.498
variable sigmac_aa equal 3.308
variable sigmac_tt equal 3.00535
variable sigmac_cc equal 3.195070
variable sigmac_gg equal 3.258250
variable sigmac_at equal 3.283437
variable sigmac_ac equal 3.239883
variable sigmac_ag equal 3.359575
variable sigmac_cg equal 3.229553
variable sigmac_ct equal 3.061570
variable sigmac_gt equal 3.188250
variable aa_eps_x equal 1.49375
variable aa_eps_y equal 1.57871
variable aa_eps_z equal 8.57345
variable tt_eps_x equal 0.803368
variable tt_eps_y equal 1.289000
variable tt_eps_z equal 5.322470
variable cc_eps_x equal 0.855300
variable cc_eps_y equal 1.268740
variable cc_eps_z equal 6.085350
variable gg_eps_x equal 0.771314
variable gg_eps_y equal 1.516900
variable gg_eps_z equal 9.614650
variable ac_eps_x equal 0.311181
variable ac_eps_y equal 0.976414
variable ac_eps_z equal 6.396720
variable ag_eps_x equal 0.317917
variable ag_eps_y equal 0.924487
variable ag_eps_z equal 1.788460
variable at_eps_x equal 0.439928
variable at_eps_y equal 0.221046
variable at_eps_z equal 1.154980
variable cg_eps_x equal 0.815193
variable cg_eps_y equal 1.508990
variable cg_eps_z equal 7.514520
variable ct_eps_x equal 0.349480
variable ct_eps_y equal 0.025174
variable ct_eps_z equal 1.336630
variable gt_eps_x equal 0.347553
variable gt_eps_y equal 0.004132
variable gt_eps_z equal 1.865920

pair_style hybrid/overlay gayberne ${gamma_1_1} ${nuGB_1_1} ${muGB_1_1} ${gb_cutoff}
#Gayberne_dna coeffs: eps, sigma, eia, eib, eic, eja, ejb, ejc
pair_coeff * * gayberne ${ene_unitGB} ${sigmac_ps} ${gb_mix} ${gb_mix} ${gb_mix} ${gb_mix} ${gb_mix} ${gb_mix} ${gb_mix} ${gb_cutoff}
pair_coeff 3 3 gayberne ${ene_unitGB} ${sigmac_aa} ${aa_eps_x} ${aa_eps_y} ${aa_eps_z} ${aa_eps_x} ${aa_eps_y} ${aa_eps_z} ${gb_cutoff}
pair_coeff 3 4 gayberne ${ene_unitGB} ${sigmac_at} ${at_eps_x} ${at_eps_y} ${at_eps_z} ${at_eps_x} ${at_eps_y} ${at_eps_z} ${gb_cutoff}
pair_coeff 3 5 gayberne ${ene_unitGB} ${sigmac_ac} ${ac_eps_x} ${ac_eps_y} ${ac_eps_z} ${ac_eps_x} ${ac_eps_y} ${ac_eps_z} ${gb_cutoff}
pair_coeff 3 6 gayberne ${ene_unitGB} ${sigmac_ag} ${ag_eps_x} ${ag_eps_y} ${ag_eps_z} ${ag_eps_x} ${ag_eps_y} ${ag_eps_z} ${gb_cutoff}
pair_coeff 4 4 gayberne ${ene_unitGB} ${sigmac_tt} ${tt_eps_x} ${tt_eps_y} ${tt_eps_z} ${tt_eps_x} ${tt_eps_y} ${tt_eps_z} ${gb_cutoff}
pair_coeff 4 5 gayberne ${ene_unitGB} ${sigmac_ct} ${ct_eps_x} ${ct_eps_y} ${ct_eps_z} ${ct_eps_x} ${ct_eps_y} ${ct_eps_z} ${gb_cutoff}
pair_coeff 4 6 gayberne ${ene_unitGB} ${sigmac_gt} ${gt_eps_x} ${gt_eps_y} ${gt_eps_z} ${gt_eps_x} ${gt_eps_y} ${gt_eps_z} ${gb_cutoff}
pair_coeff 5 5 gayberne ${ene_unitGB} ${sigmac_cc} ${cc_eps_x} ${cc_eps_y} ${cc_eps_z} ${cc_eps_x} ${cc_eps_y} ${cc_eps_z} ${gb_cutoff}
pair_coeff 5 6 gayberne ${ene_unitGB} ${sigmac_cg} ${cg_eps_x} ${cg_eps_y} ${cg_eps_z} ${cg_eps_x} ${cg_eps_y} ${cg_eps_z} ${gb_cutoff}
pair_coeff 6 6 gayberne ${ene_unitGB} ${sigmac_gg} ${gg_eps_x} ${gg_eps_y} ${gg_eps_z} ${gg_eps_x} ${gg_eps_y} ${gg_eps_z} ${gb_cutoff}

```

3.7 Conclusions

We are developing a three-sites nucleotide. Sugar and phosphate group are simulated as spheres. To improve the other three-sites coarse-grained models that use only spheres, to symbolize bases we chose ellipsoids, because their shape seems to reproduce the atomistic structure better than sphere does. We fit the energy profiles with a Gay-Berne potential to obtain spheres and ellipsoids parameters. Note that each base will be described by an ellipsoid of different size. Anisotropic shapes also need to a orientation to be specified: a quaternion is associated to each base.

Now, the force field have been parametrized. First of all, we collect data from atomistic simulations of a 10 base pair long sequences and we compute histograms, using the Boltzmann inversion method. These data are now fitted using various types of potential: the equations that fit better atomistic results are chosen and added to the MD code LAMMPS.

We plan to check the force field parametrization and add a small number of charges to bases, by fitting the atomistic charge distribution, to have a better description of electrostatic interactions. We will next use melting temperatures of known DNA sequences to validate our model.

3.8 Appendices

3.8.1 Structural analysis of DNA sequences used by Zanchetta group

For sequences used by Zanchetta group aspect ratio values were calculated, finding a good agreement with the ones reported in literature [66]. It is worth underlining that these values do not depend on the sequence itself but only on sequence length.

3.8.1.1 Sequences

Zanchetta group works on systems containing short complementary B-DNA oligomers (6 to 20 base pairs in length). These systems show nematic and columnar liquid crystal phases [66]. They carried on a structural study which demonstrate that phases are due to end-to-end adhesion and consequent stacking of the duplex oligomers into polydisperse anisotropic rod-shaped aggregates, which can order into liquid crystal. The systems are initially composed by single strand sequences. Upon cooling, the complementary DNA sequences form duplexes, creating liquid crystal droplets and leaving the unpaired single strands in isotropic solution. Such ordering autocatalytic promotes polymerization of complementary oligomers. In this article [66], 13 sequences are used (Table 3.16).

Another work of the same group [67] is about other two sequences, both self-complementary and single-stranded: one is 10 base pairs long, CGCAATTGCG (filename: 10bp_ss) and the other one 20 base pair long (filename: 20bp_ss, CGCAATTGCGTTTTTTTTTTT).

File name	Length (bp)	Self-complementary	Sequence
6bp	6	yes	CGATCG
8bp	8	yes	CGCATGCC
10bp	10	yes	CGCAATTGCC
D12 ^a	12	yes	CGCGAATTGGCG
14bp	14	yes	ACCGGAATTGGCGT
16bp	16	yes	ACCGAGAATTCTGGCGT
20bp	20	yes	AACGCAAAAGATCTTTGGCGTT
12bp_1	12	yes	AACGCATGCCGTT
12bp_2	12	no	CCTCAAAAACCTCC
12bp_3	12	no	GGAGTTTTTGAGG
13bp	13	partially	CGCGAATTGGCGGT
14bp_1	14	partially	CGCGAATTGGCGGTT
22bp	22	partially	CGCGAATTGGCGGTTT'TT'TT'TT'

Table 3.16: Sequences used in [66].

^aDickerson dodecamer

3.8.1.2 Visualizing sequences

In this section, snapshots of analyzed sequence are collected.

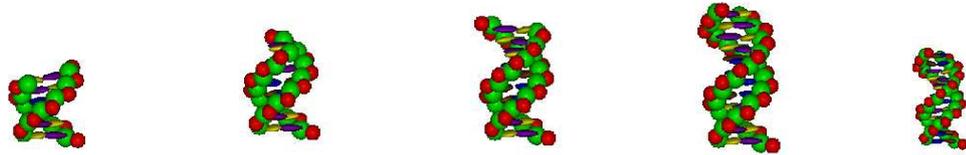


Figure 3.92: From left to right: 6bp, 8bp, 10bp, D12, 12bp_1

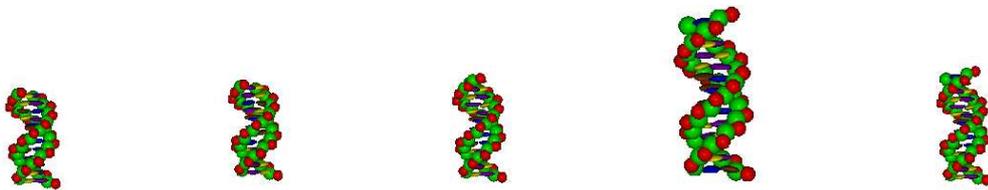


Figure 3.93: From left to right: 12bp_2, 12bp_3, 13bp, 14bp, 14bp_1

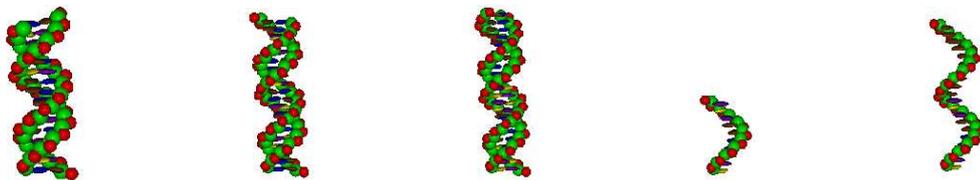


Figure 3.94: From left to right: 16bp, 20bp, 22bp, 10bp_ss, 20bp_ss

3.8.1.3 Results

Values of length and diameter of studied sequence are summed up in Table 3.17, while in Figure 4.69 aspect ratio values are collected.

Sequence	L (\AA)	D (\AA)
6bp	17.31	18.32
8bp	24.09	18.32
10bp	30.44	18.32
D12	37.13	18.32
14bp	44.31	18.32
16bp	51.25	18.32
20bp	64.31	18.32
12bp_1	37.34	18.32
12bp_2	37.24	18.32
12bp_3	37.24	18.32
13bp	40.79	18.32
14bp_1	44.31	18.32
22bp	71.03	18.32
10bp_ss	30.44	18.32
20bp_ss	62.26	18.32

Table 3.17: Data obtained evaluating length and diameter of studied sequences.

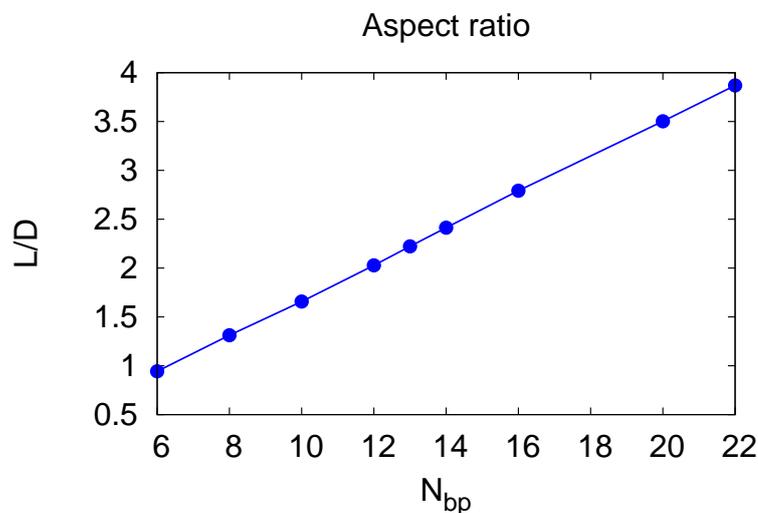


Figure 3.95: Aspect ratio dependence on length in base pairs.

3.8.1.4 Conclusion

These results are in good agreement with the ones reported in literature [66]. Sequence containing from 6 to 22 base pairs have an aspect ratio < 4.7 . Moreover, sequences of the same length show the same aspect ratio, i.e. this parameter does not depend on sequence itself. Finally, it is worth underlining that this value is not different for single stranded and double stranded DNA sequences of the same length (even if in the case of sequences 20 base pair long it seems to be a small difference of almost 1.5\AA). This disagreement is probably due to the way of calculating aspect ratio, taking the larger value of distance between the two bases at the sequence's end (belonging to the same or to the complementary strand).

3.8.2 DNA chirality index investigation

Chirality is an intrinsic property of DNA. Code studied for protein secondary structure investigation [68] had been modified in order to fit both atomistic and coarse-grained model. Atomistic model take into account ψ' and ϕ' angles to give the desired (characteristic of an α -helix) negative G^a , chirality index, value. Coarse-grained model had been analyzed with another version of the code which allows the index to be influenced by bases. The sequence seems to become an important part of this analysis, but results explicate that, when GC percentage is lower than 10%, the average value of G^a to which molecules tends to align is -0.04, i.e. the feature value of α helix. Note that when this percentage is 0%, 50% and 100%, G^a value behavior is symmetrical for each strand.

3.8.2.1 Introduction

DNA has an intrinsic chirality, so we applied to various sequences the chirality index investigation methodology [68] created for protein secondary structure to our DNA models.

3.8.2.2 Atomistic model

The code *chirality.f90* [68] had been changed in order to fit DNA atom type. Three set of atoms (see Figure 3.96) were analyzed: C5'–C4'–C3' (corresponding to ψ and ψ' angles), O5'–C5'–C4' (the correct ψ and ϕ angles) and C4'–C3'–O1P (angles ϕ' and ψ').

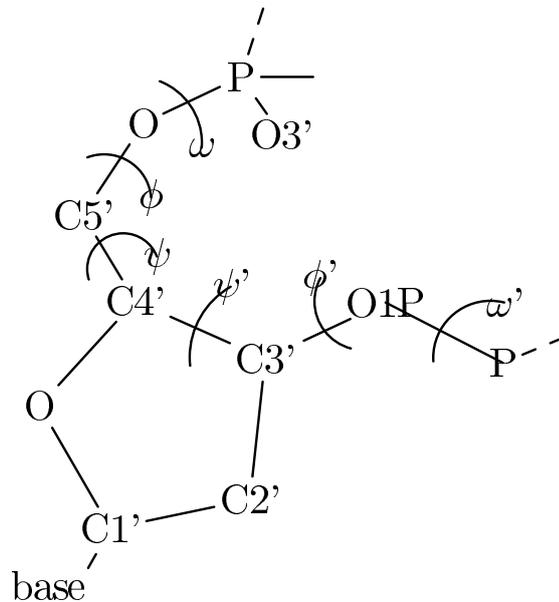


Figure 3.96: Atom names and angles for a DNA structure.

Values obtained for ϕ e ψ angles are slightly positive ($G^a = 0.005/0.017$), so this term has not been considered. The set C5'–C4'–C3' (ψ and ψ' angles) gives results comprised between $G^a = -0.045$ and $G^a = -0.03$: this break comprehends the standard G^a values for an α -helix ($-0.05/-0.04$), but takes into account only ψ angles. The last tern (C4'–C3'–O1P) corresponds to ϕ' and ψ' angles and gives G^a values in the interval $[-0.1; -0.08]$ for the DNA

helix. These values do not correspond properly to those characteristic of an α -helix ([68]) but are still negative, so we decided to use this set of atom. Note that, using this tern of atoms, make the G^a index independent from the base sequence (Figure 3.97). Note that all this values are obtained by multiplying for a factor of 10 each G^a obtained.

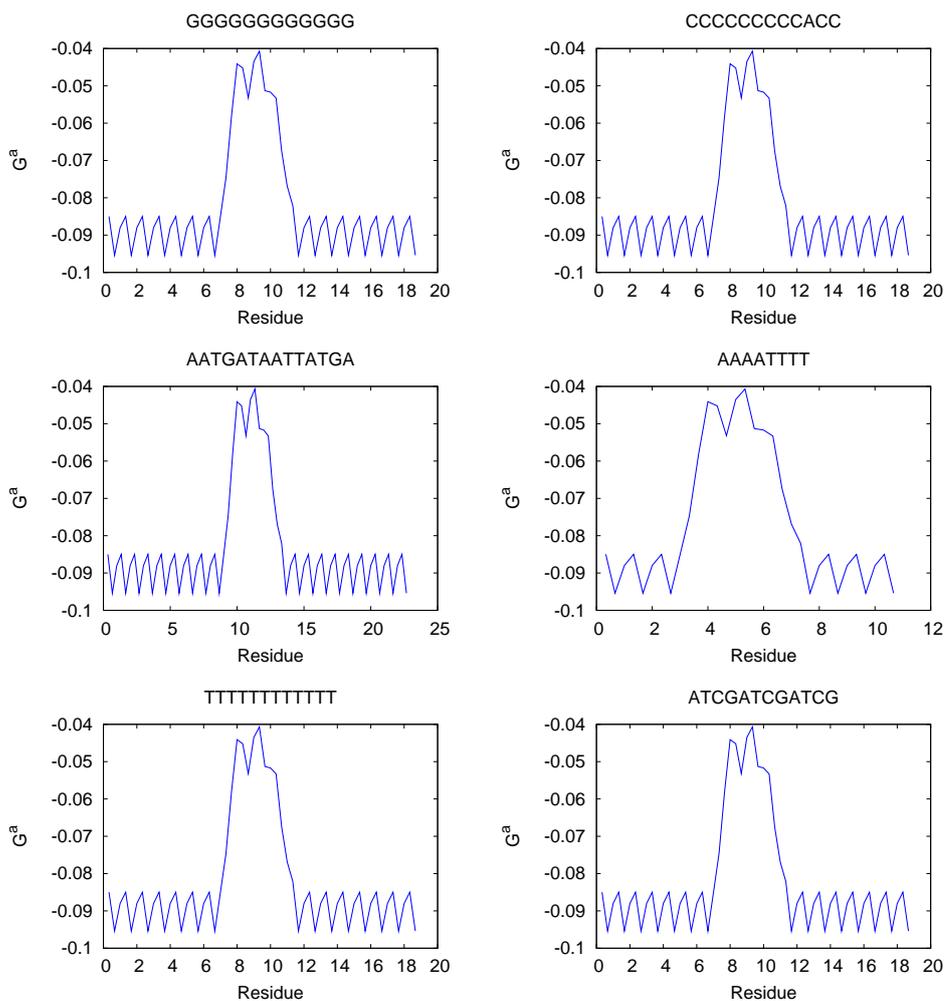


Figure 3.97: DNA chirality: G^a values for some atomistic sequences.

3.8.2.3 Coarse-grained model

Applying this methodology [68] to our new “under construction” coarse-grained model could be very interesting. The code has been slightly modified in order to achieve this goal. First of all, we try to analyze only the backbone (like for atomistic DNA and protein, considering a residue composed only by two beads, Phosphate and Sugar) but G^a values obtained were positive. This is not very surprising since angles between backbone’s beads could not be in the same place as for an atomistic model (Figure 3.98). So we applied the computation to all the three beads composing a residues: phosphate (recognized as “P” atom), sugar (“S”) and base (“A” or “C” or “T” or “G”). Calculating G^a this way, this value turned out to be base-dependent, at least at a first glance.

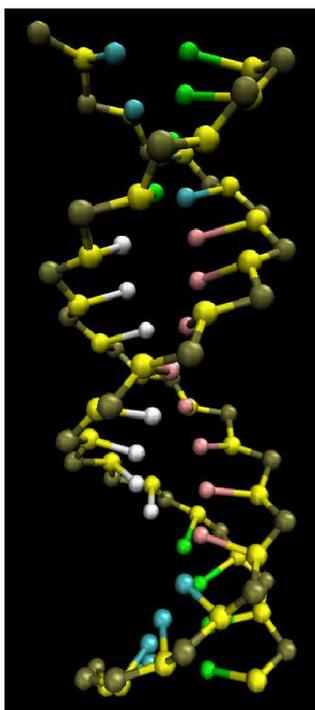


Figure 3.98: A snapshot of a coarse-grained DNA model (GGGGTTT-TAAAACCCC).

3.8.2.4 Results

This section sums up results obtained for a set of 17 sequences (Table 3.18, Figure 3.99, Figure 3.100 and Figure 3.101).

Sequence	Length (bp)	GC percentage (%)
ATCG	4	50
AAAAAAAAAAAA	12	0
CCCCCCCCCCCC	12	100
GGGGGGGGGGGG	12	100
TTTTTTTTTTTT	12	0
ATATCGCG	8	50
ATCGATCG	8	50
AATCGATAACCTGA	14	36
AATGATAATTATGA	14	14
ATCGATCGATCG	12	50
GGGGTTTTTAAAACCCC	16	50
CCCCGGGG	8	100
GGGGTTTT	8	50
AAAATTTT	8	0
CCCCAAAA	8	50
AAAAAAAAAAAAATA	13	0
CCCCCCCCCACC	12	83

Table 3.18: The 17 sequences analyzed.

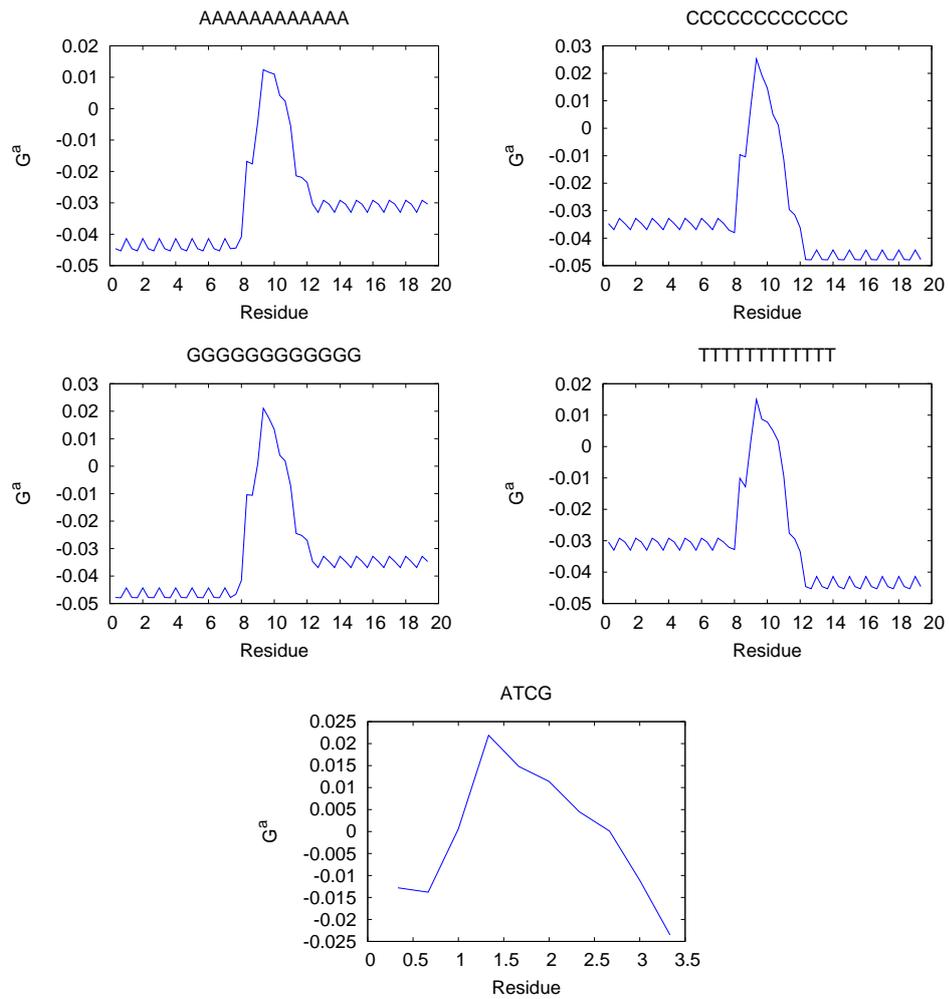


Figure 3.99: G^a for some sequences.

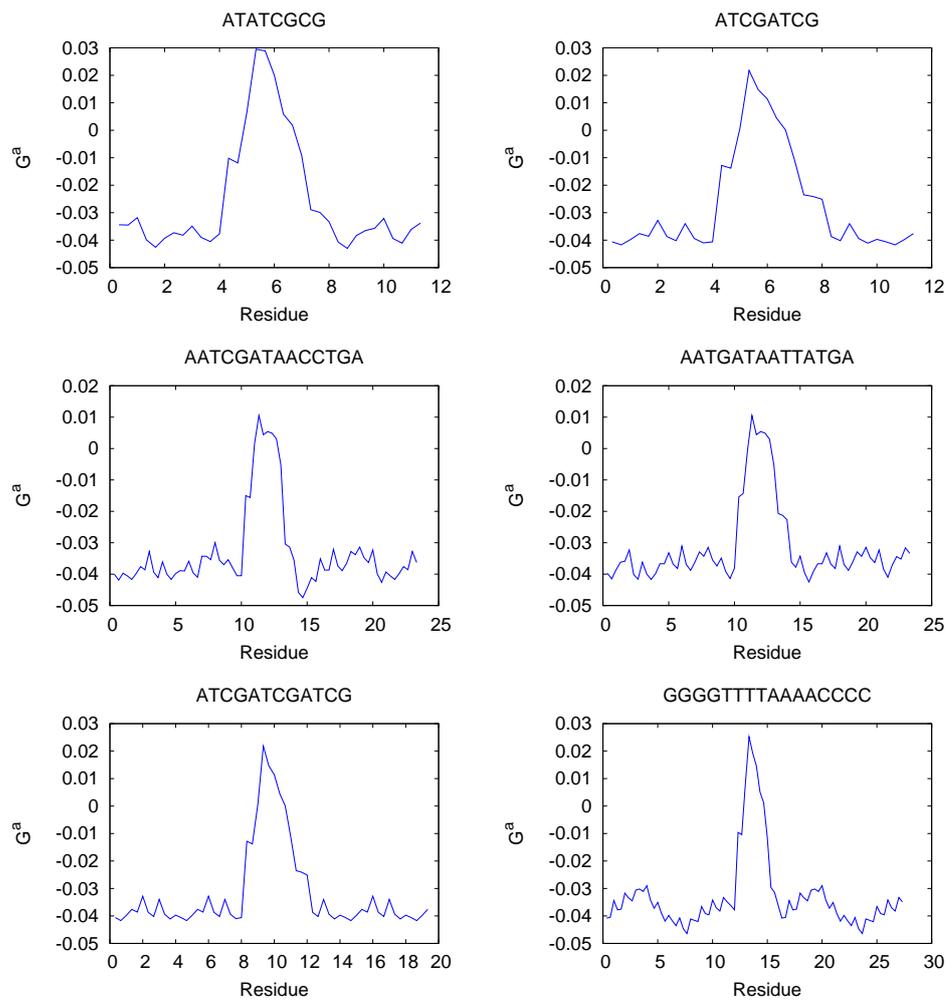


Figure 3.100: G^a for some sequences.

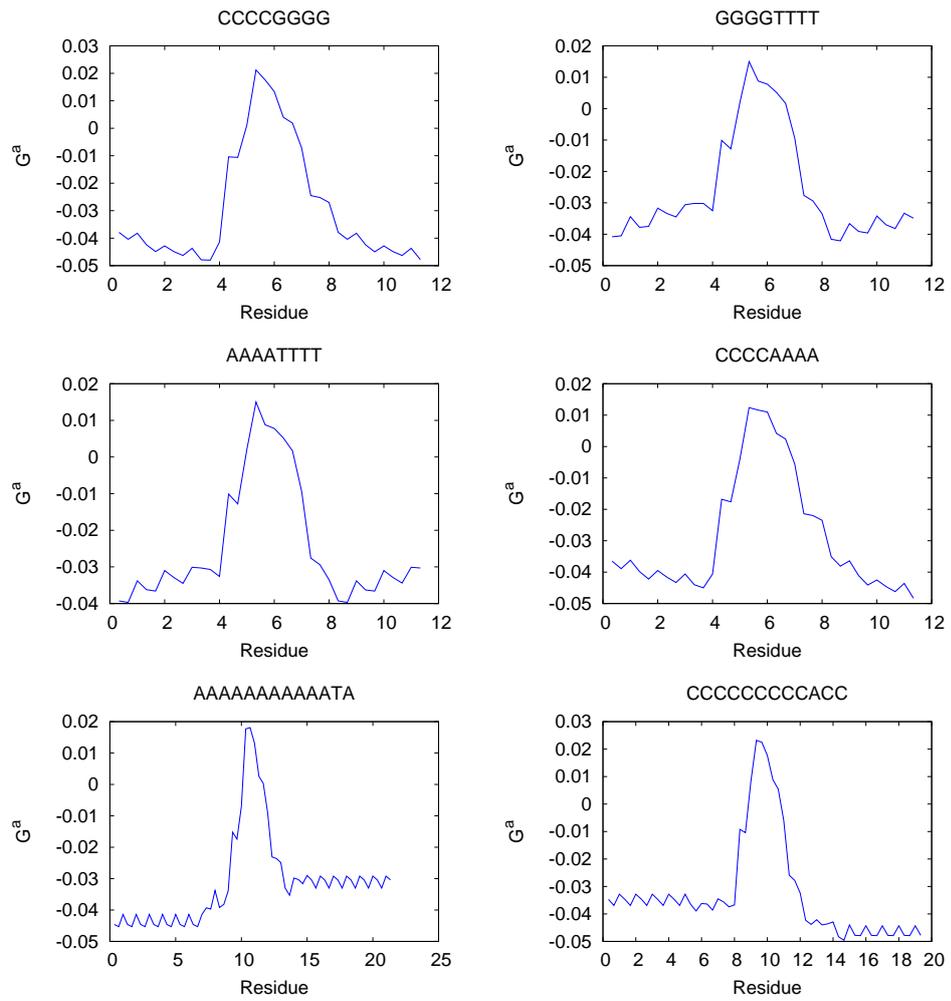


Figure 3.101: G^a for some sequences.

3.8.2.5 Conclusion

First of all, it is worth underlining that the four-base pair long sequence (ATCG, GC-content = $gc = 0.5$) does not show a stable plot, so only sequences longer than 4 bp has to be considered.

The four following plots, one for each 12-base pair sequence composed by only one base (AAAAAAAAAAAA, CCCCCCCCCCCC, GGGGGGGGGGGG and TTTTTTTTTTTT) show that each base have a proper value of G^a (note that each figure is ideally divided in two part, the first one for the sense strand and the second one for the antisense strand, explaining why residue number is almost twice as the base-pair length). Adenine base has chiral index values around $G^a = -0.045/-0.04$, cytosine $G^a = -0.04/-0.035$, guanine $G^a = -0.05/-0.045$ and thymine $G^a = -0.035/-0.03$.

Another feature recognizable from plots is that sequences having GC-ratio = 0.0, 0.5 or 1.0 are characterized by an almost symmetrical graph (the two strands shows almost the same behavior). When the GC-percentage is lower than 10%, G^a values aim at a value of chiral index typical of an α -helix (-0.04) and it is no more possible to recognize the specific contribution of each base.

3.8.3 A possible application of DNA chirality index: mutations

Mutations are changes in a genomic sequence, the DNA sequence of a cell's genome or the DNA or RNA sequence of a virus. Since mutations are responsible of lots of disease, we decided to analyze these genetic modifications, using the G^a computation's procedure [68] to wild-type and modified sequences. Results obtained from this study of different mutations' classes allow us to affirm that G^a could not be used to describe and trace nonsense point mutations, silent mutations, all kind of repeats and transitions. However, this parameter is a good tool for discriminate splice-site point mutations, frame-shift indels and transversions. Finally, a chiral index investigation on every sequences encoding for each amino acids was done, in order to find out if chemical properties could be recognized by G^a parameters. Analysis's results show that it is not possible to distinguish characteristic behaviors for each classes (basic, acidic, neutral polar and non-polar) of amino acids.

3.8.3.1 Introduction

Mutations are changes in a genomic sequence, the DNA sequence of a cell's genome or the DNA or RNA sequence of a virus. Mutations are caused by radiation, viruses, transposons and mutagenic chemicals, as well as errors that occur during meiosis or DNA replication [69, 70, 71]. They can also be induced by the organism itself. result in several different types of change in DNA sequences: these can either have no effect, alter the product of a gene, or prevent the gene from functioning properly or completely. Due to the damaging effects that mutations can have on cells, organisms have evolved mechanisms such as DNA repair to remove mutations [69]. Viruses that use RNA as their genetic material have rapid mutation rates [72], which can be an advantage since these viruses will evolve constantly and rapidly, and thus evade the defensive responses of e.g. the human immune system [73]. Since the importance of this type of processes, we tried to apply the chiral index investigation (i.e. the procedure of compute this index [68], modified for fit coarse-grained DNA), in order to discover if this parameter could describe DNA mutations. We applied the G^a computation's procedure [68] to wild-type and modified sequences.

3.8.3.2 Type of mutations

The following description is taken from an online biology textbook [74]. Figure 3.102 shows some examples of notable mutations.

Single-base substitutions

Single base substitutions are called point mutations. If a purine (A or G) or a pyrimidine (C or T) is replaced by the other, the substitution is called a transition, while if a purine is replaced by a pyrimidine or vice-versa, the point mutation is called a transversion. This kind of mutations could have different effects. With a *missense mutation*, the new nucleotide alters the codon so as to produce a different amino acid in the protein product. When the new nucleotide changes a codon that specified an amino acid to one of the STOP codons (TAA, TAG or TGA), the mutation is called

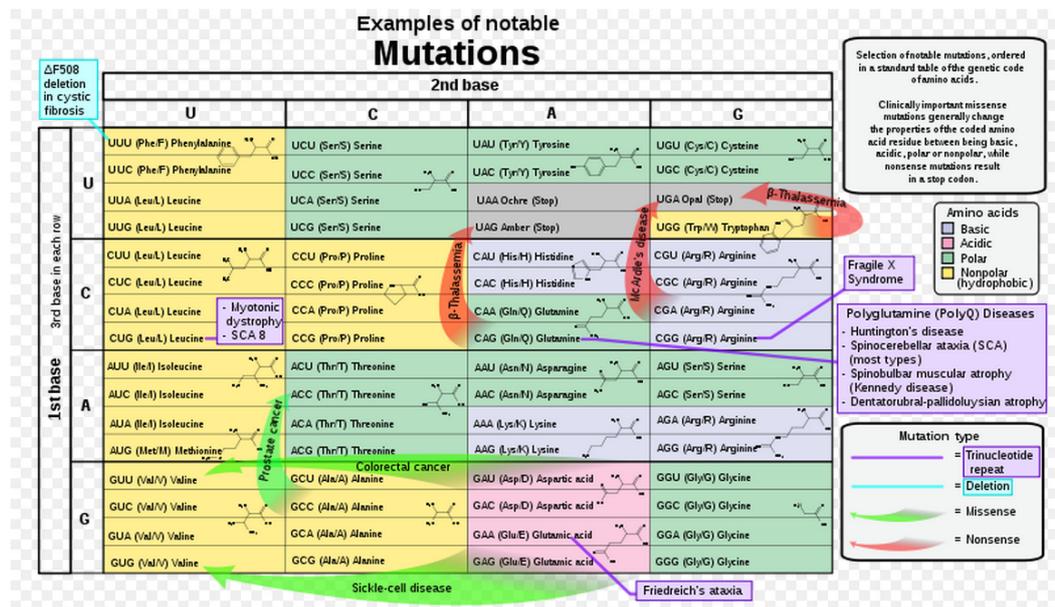


Figure 3.102: Example of notable mutations.

nonsense. The effect is that translation of the messenger RNA transcribed from this mutant gene will stop prematurely. The earlier in this gene occurs, the more truncated the protein product will be and the more likely that it will be unable to function. *Silent mutations* cause no change in their product and cannot be detected without sequencing the gene or its mRNA. The lack of effects is due to the fact that most amino acid are encoded by several different codons (triplets of nucleotides). For example, if the third base in the TCT codon for serine is changed to any one of other three bases, serine will still be encoded. The last type of single–base substitution is called *splice–site mutations*. The removal on an intron sequence¹⁰, as pre-mRNA¹¹ is being processed to form mRNA, must be done with great precision. Nucleotide signals at the splice sites (where exons are joined together) guide the enzymatic machinery. If a mutation

¹⁰An intron is a portion of a gene that is transcribed into RNA but is removed during the formation of the mature RNA molecule.

¹¹pre-mRNA is the first product of DNA transcription, which still have to undergo processing step to produce functional mRNA molecule.

alters one of these signals, then the intron is not removed and remains as part of the final RNA molecule. The translation of its sequence alters the sequence of protein product.

Indels

Indels is the collective name for insertions (extra base pairs addition) and deletions (extra base pair removal) from the DNA of a gene. The number can range from one to thousands. Indels involving one or two base pairs (or multiples thereof) can have devastation consequences to the gene because its translation results “frameshifted”. Figure 3.103 shows how, by shifting the reading frame one nucleotide to the right (deletion of the first nucleotide, G), the same sequence of nucleotides encodes a different sequence of amino acids. The mRNA is translated in new groups of three nucleotides and the protein specified by these new codons will be worthless. Frameshift often create new STOP codons and thus generate nonsense mutations. Indels of three nucleotides or multiples of three may be less serious because they preserve the reading frame. However, a number of inherited human disorders are caused by the insertion of many copies of the same triples nucleotides (e.g. Huntington’s disease and fragile X syndrome).

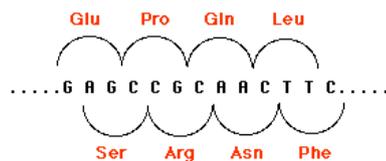


Figure 3.103: An example of frame-shift.

Duplications

Duplications are a doubling of a section of the genome. During meiosis, crossing over between sister chromatids¹² that are out of alignment can produce one chromatid with a duplicated gene and the other having

¹²A chromatid is one of the two identical copies of DNA making up a duplicated chromosome, which are joined at their centromeres, for the process of cell division (mitosis or meiosis).

two deletions. Note that unequal crossing over could create a gene which carries inappropriate promoters. If this promoter causes the gene to be expressed more strongly than the normal gene, the mutant gene will be dominant. Gene duplication has also been implicated in several human neurological disorders. However, gene duplication has occurred repeatedly during evolution of eukaryotes, which had been beneficial since duplicates can acquire different functions, provides redundancy and speciation.

Translocations

Translocations are the transfer of a piece of one chromosome to a non-homologous chromosome. This kind of mutation is often reciprocal (two nonhomologous swap segments). Translations can alter the phenotype in several ways. If the break occur within a gene, its function could be destroyed. Translocated genes may come under the influence of different promoters and enhancers so that their expression is altered. The breakpoint may occur within a gene, creating a hybrid gene. This may be transcribed and translated into a protein with an N-terminal of one normal cell protein coupled to the C-terminal of another.

3.8.3.3 Results

Point mutations: missense mutations

Sickle-cell disease

The replacement of a A by a T at the 17th nucleotide of the gene for the beta chain¹³ of hemoglobin changes the codon GAG (for glutamic acid) to GTG (which encodes valine). Thus the 6th amino acid in the chain becomes valine instead of glutamic acid.

Sequences analyzed are the wild-type one (CTGACTCCTGAGGAG) and the one having the 17th nucleotide turned into a T (CTGACTCCTGTGGAG).

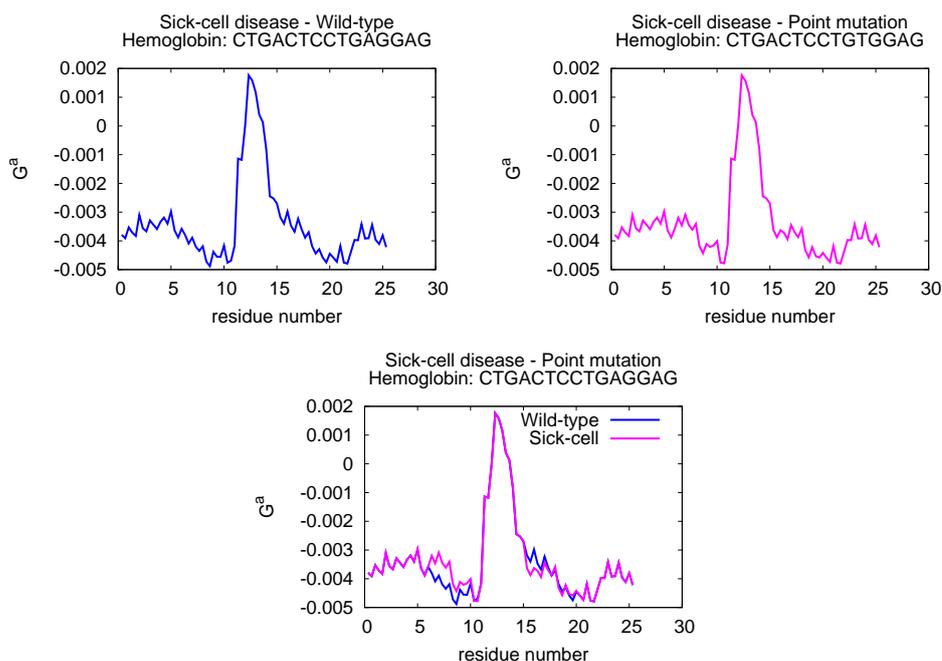


Figure 3.104: Sick-cell disease: hemoglobin point mutation.

Graphs (Figure 3.104) show a slight different behaviors of these two type of sequences. Even if these trends are not completely distinct, the wild-type

¹³Human haemoglobin A beta chain, GenBank: A01592.1

sequence and the modified one are distinguishable.

Cystic fibrosis

More than 1000 different mutations have been found in patients with cystic fibrosis. Each of these mutations occurs in a huge gene that encodes a protein of 1480 amino acids called the cystic fibrosis transmembrane conductance regulator (CFTR)¹⁴. The protein is responsible for transporting chloride ions through the plasma membrane. The gene encompasses over 6000 nucleotides spread over 27 exons on chromosome 7. Defects in this protein cause the various symptoms of the disease. Unlike sickle-cell disease, then, no single mutation is responsible for all cases of cystic fibrosis. People with cystic fibrosis inherit two mutant genes, but the mutations need not to be the same.

In this case, a missense point mutation at 482th nucleotide of the gene is studied. A G (GAGGAACGCTCTATC) is substituted by an A (GAGGAACACTCTATC), which cause 117th amino acid to become histidine (wild-type: arginine). Using G^a parameter these sequences could not be discriminated (Figure 3.106).

Point mutations: nonsense mutations

Cystic fibrosis

Another mutation in patient with cystic fibrosis is the substitution of a T for a C at nucleotide 1609, which converts a glutamine codon (CAG, in sequence TGTTCCTCAGTTTTCC) in a STOP codon (TAG, sequence now becomes TGTTCCTTAGTTTTCC). The protein produced by this patient has only the first 493 amino acids of the normal chain of 1480 and could not function.

¹⁴Homo sapiens cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7) (CFTR), mRNA, NCBI Reference Sequence: NM.000492.3

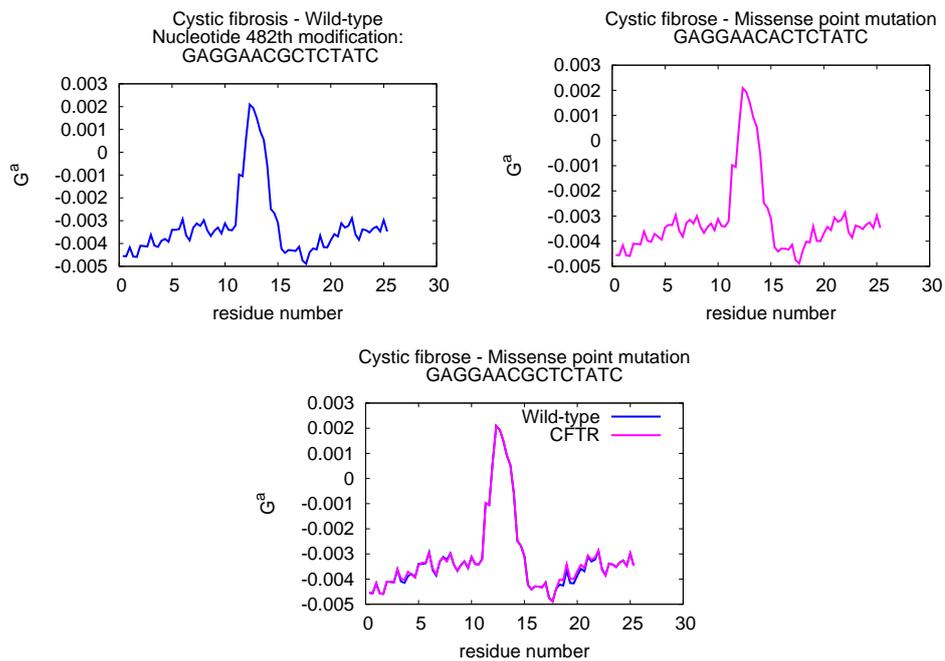


Figure 3.105: Cystic fibrose: CFTR point mutation.

Even in this case, G^a parameter is not able to recognize the mutation occurred.

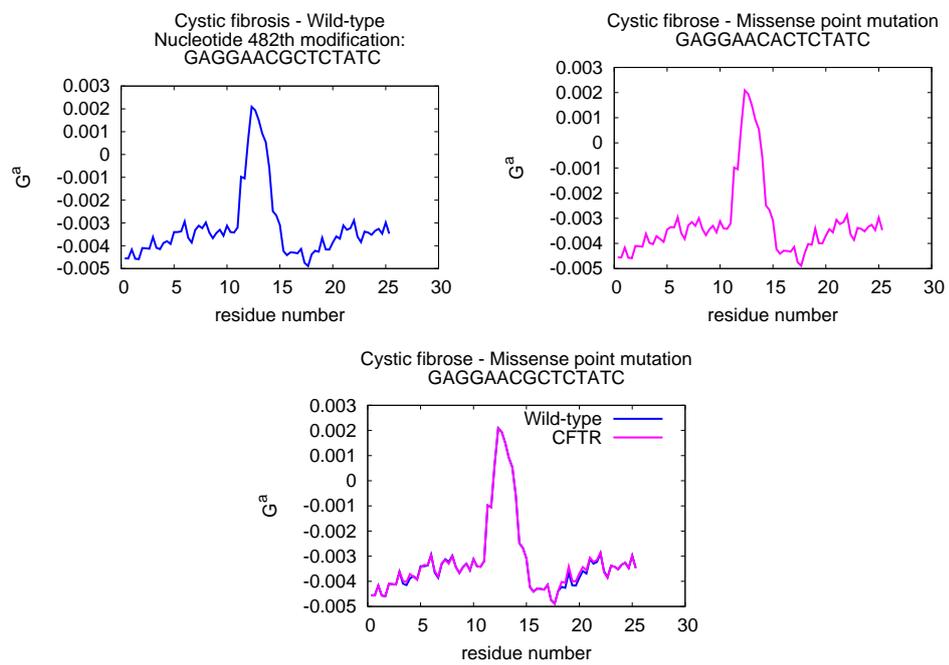


Figure 3.106: Cystic fibrose: CFTR point mutation.

TAG STOP codon

In this case, we want to analyze all possible mutation on the TAG STOP codon. The first, second and third nucleotide of this codon are subsequently changed (Table 3.19).

Nucleotide changed	Modified sequences			
1st	AAG	CAG	GAG	TAG
2nd	TAG	TCG	TGG	TTG
3rd	TAA	TAC	TAG	TAT

Table 3.19: All possible TAG point mutations.

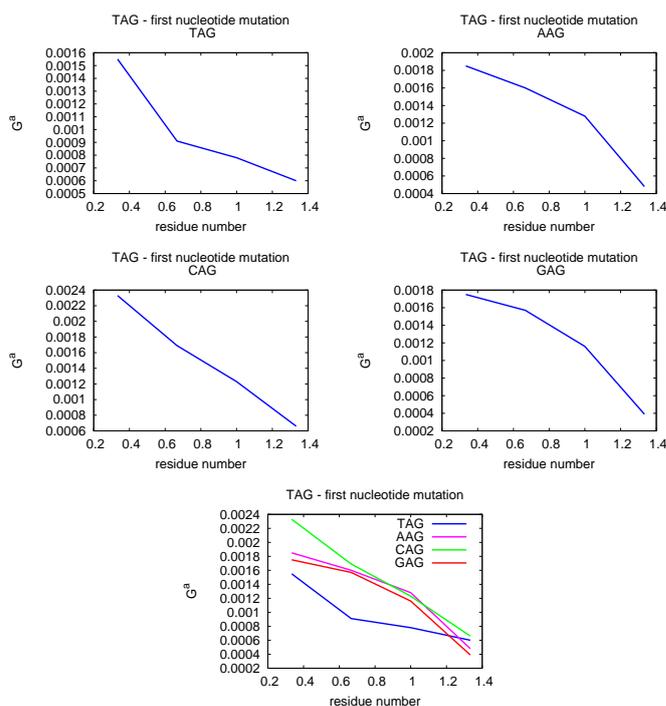


Figure 3.107: TAG STOP codon: first nucleotide modifications.

Note that these trends tends to zero: due to the briefness of these sequences, data are not reliable. So, even if TAG codon (except for the third nucleotide mutations) seems to posses the lower parameter values, it is not

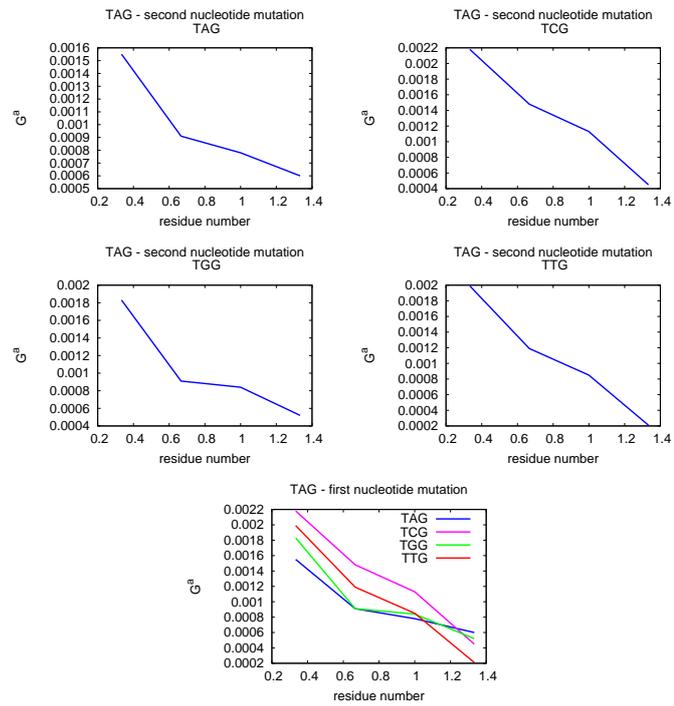


Figure 3.108: TAG STOP codon: second nucleotide modifications.

possible, using this chirality index, to discriminate between mutated and wild-type sequences.

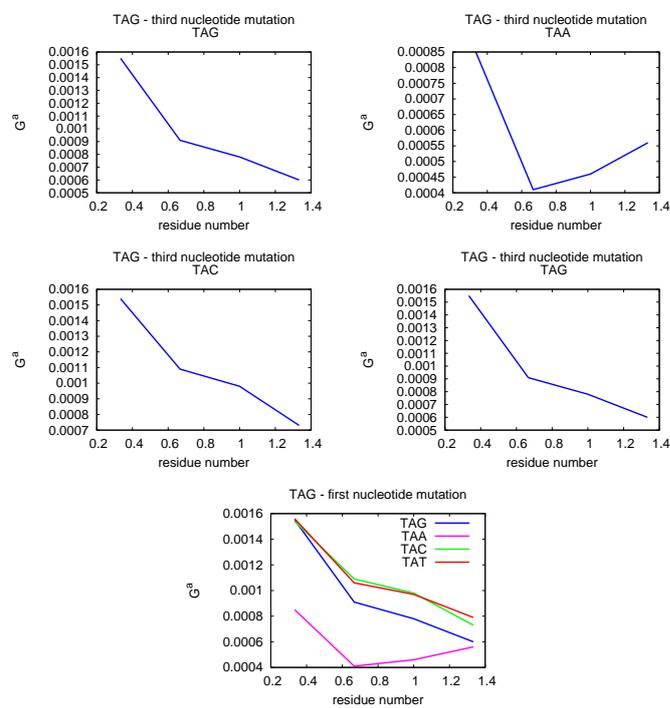


Figure 3.109: TAG STOP codon: third nucleotide modifications.

CFTR: a 478th nucleotide mutation

Also for a single-base substitution at 478th nucleotide of CFTR gene, chiral index was investigated. A G (GAGGAACGCTCTATC) has been replaced by a T (GAGTAACGCTCTATC): no glutamic acid is encoded and protein is only 115 amino acids long. Graphs show only small differences.

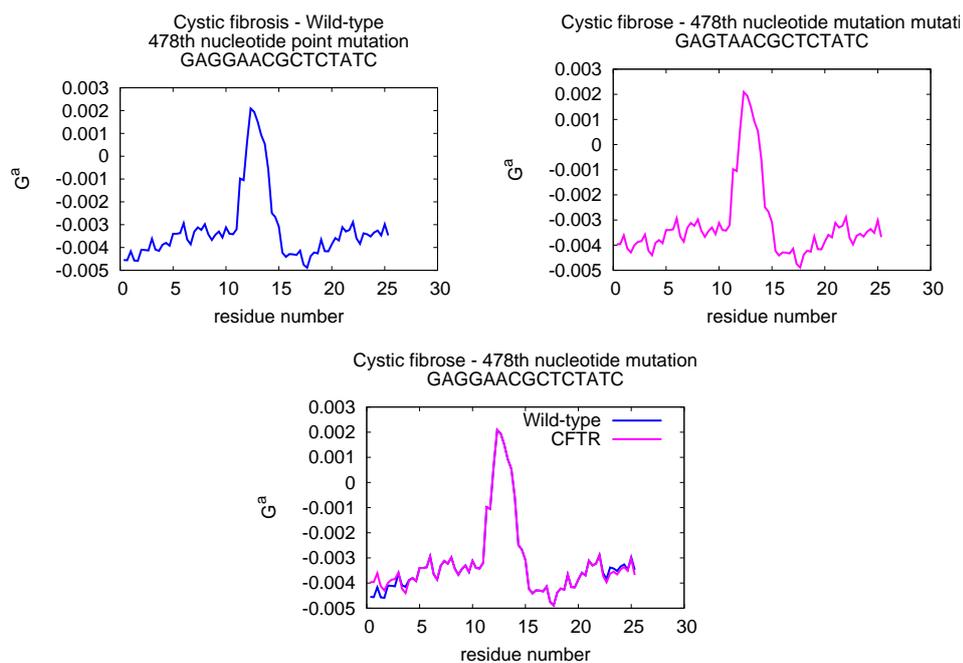


Figure 3.110: CFTR: 478th nonsense nucleotide modification.

Point mutations: silent mutations

We also analyze the effects of modifying the 1605th nucleotide in CFTR gene. TCT (TGTTCTCAGTTTTCC) triplet has been changed in TCA (TGTTTCACAGTTTTCC), TCC (TGTTCCCAGTTTTCC) and TCG (TGTTTCGCAGTTTTCC).

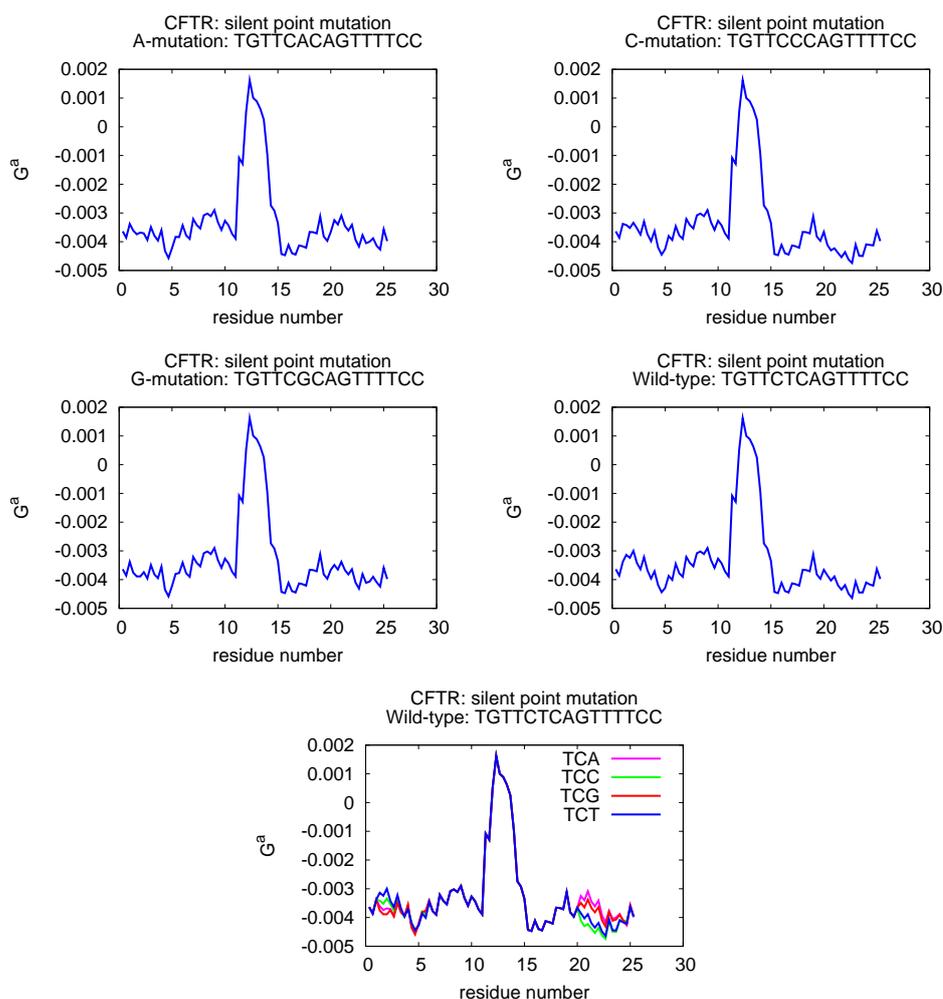


Figure 3.111: CFTR: 1605th nucleotide silent mutation.

Also in this case, it is not possible to determine substantial differences between wild-type sequence and the modified ones.

Point mutations: splice-site mutations

In reference [75], wild-type splicing patterns of various *Drosophila* genomes' introns were changed by mutations that opened the stem structure and restored by compensatory mutations that re-established the base-pairing potential, demonstrating that these secondary structures were indeed implicated in the splice site choice. We analyze mini-gene CG33298, that encodes for ATPase with phospholipid-translocation activity. Alternative donor usage during the splicing of its pre-mRNA is predicted to change the C-termini of the proteins. We modified the wild-type sequence (AGGTAAAAAGGCACAA) in order to view if there are some big differences (AGGTAAAAAGGCACAAA and GGTA AAAAGGCACAAA). Even if the average value of all sequences aim at the same value ($G^a \sim -0.004$), typical for DNA helical structure, there are slight differences between them.

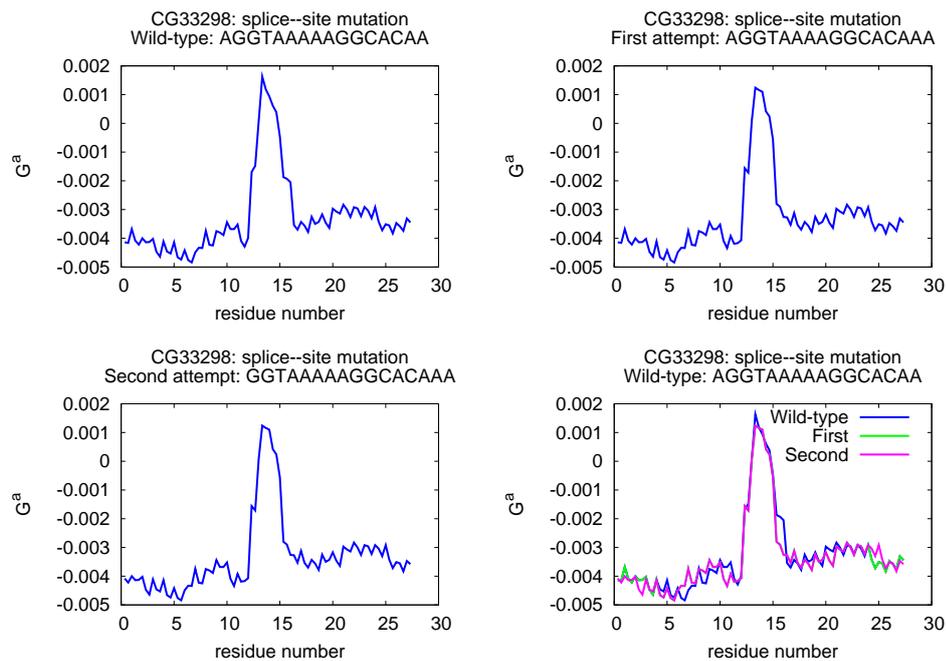


Figure 3.112: CG33298: splice-site mutations.

Indels: frame-shift

Here, we analyze the example described in online “Biology Page” [74] for frame-shift indels regarding the sequence GAGCCGCAACTT, which becomes Glu-Pro-Gln-Leu. The first G nucleotide has been deleted (AGCCGCAACTTC, which will be translated in this amino acid chain: Ser-Arg-Asn-Phe). We also tried to delete last three nucleotide (GAGCCGCAA, Glu-Pro-Gln).

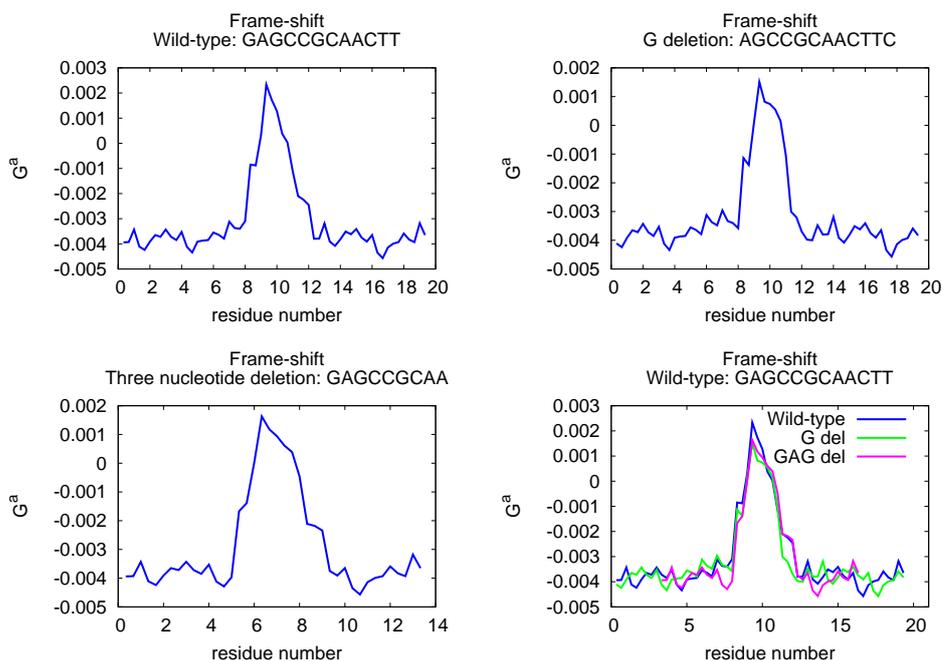


Figure 3.113: Frame-shift: one and three nucleotide deletions. Note that the cumulative graph has been made by shifting the plot of the three nucleotide deletion example in order to have peaks almost at the same position.

Differences in the behavior are great enough to discriminate between wild-type and mutated sequences.

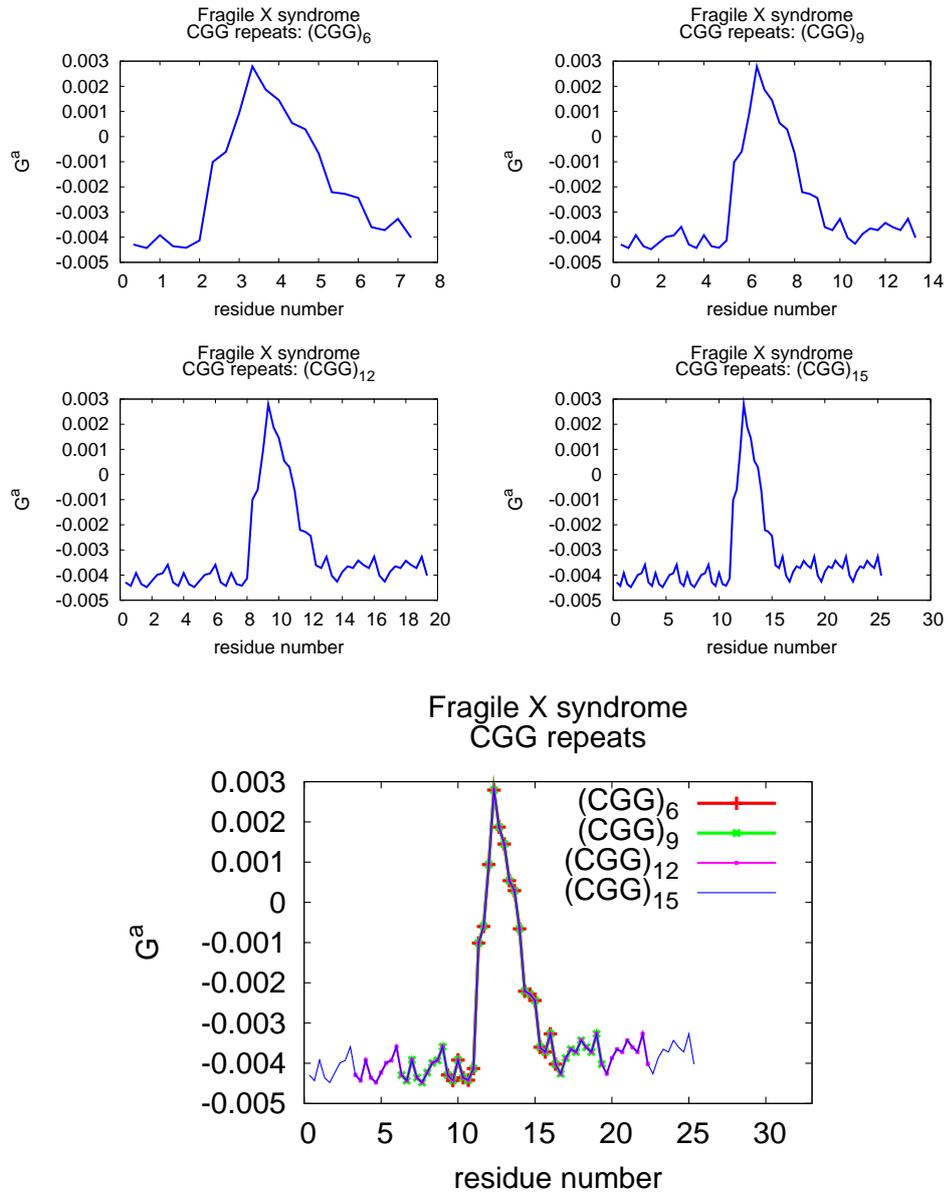


Figure 3.114: Fragile X syndrome: sequences $(CGG)_2$, $(CGG)_3$, $(CGG)_4$ and $(CGG)_5$. Note that the cumulative graph has been made by shifting plots in order to have peaks almost at the same position for each sequence.

Indels: fragile X syndrome

Several disorders in humans are caused by the inheritance of genes that have undergone insertions of a string of 3 or 4 nucleotides repeated over and over, as previously described. A locus on the human X chromosome contains such a stretch of nucleotides in which the triplet CGG is repeated. The number of CGGs may be as few as 5 or as 50 without causing a harmful phenotype (these repeated nucleotides are in a noncoding region of the gene). Even 100 repeats usually cause no harm. However, these longer repeats have a tendency to grow longer from one generation to the next (to as many as 4000 repeats). This causes a constriction in the X chromosome, which makes it quite fragile. Males who inherit such a chromosome (only from their mothers, of course) show a number of harmful phenotypic effects including mental retardation. Females who inherit a fragile X (also from their mothers; males with the syndrome seldom become fathers) are only mildly affected.

Sequences analyzed are CGGCGG, CGGCGGCGG, CGGCGGCGGCGG and CGGCGGCGGCGGCGG. No differences are recognizable (Figure 3.114).

Indels: Huntington's disease

Huntington's disease (HD) is a progressive neurodegenerative genetic disorder, which affects muscle coordination and leads to cognitive decline and dementia. In this disorder, the repeated trinucleotide is CAG, which adds a string of glutamines to the encoded protein (called huntingtin). The abnormal protein increases the level of the p53 protein¹⁵ in brain cells causing their death by apoptosis. Sequences under study are CAGCAG, CAGCAGCAG, CAGCAGCAGCAG and CAGCAGCAGCAGCAG. Also in this case, G^a values for repeats do not differ one from each other: our chiral index is not able to discriminate between sequences (Figure 3.115).

¹⁵p53 (also known as protein 53 or tumor protein 53) is a tumor suppressor protein that in humans is encoded by the TP53 gene [76, 77, 78]. p53 is important in multicellular organisms, where it regulates the cell cycle and, thus, functions as a tumor suppressor that is involved in preventing cancer. This protein plays a leading role in conserving stability by preventing genome mutation [79].

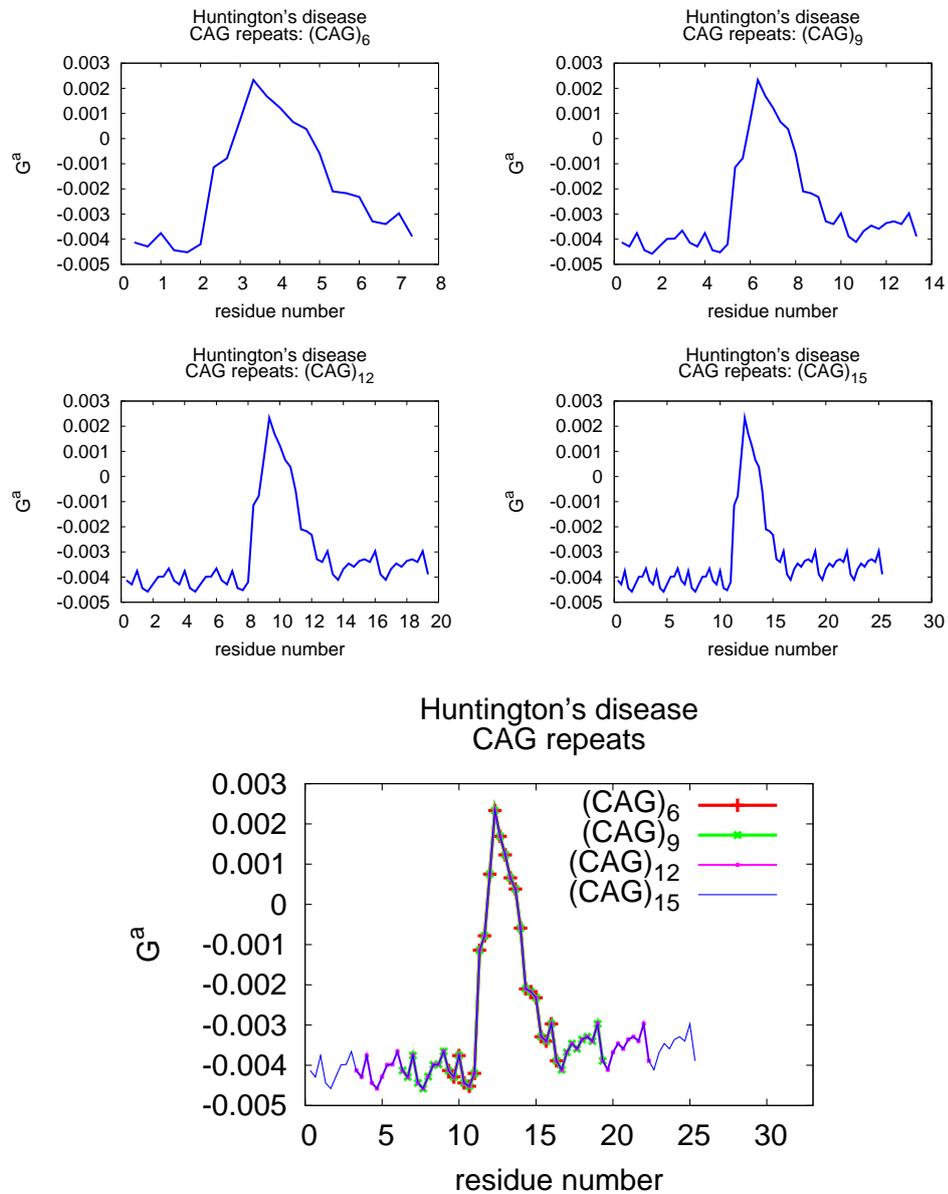


Figure 3.115: Huntington's disease: sequences (CAG)₂, (CAG)₃, (CAG)₄ and (CAG)₅. Note that the cumulative graph has been made by shifting plots in order to have peaks almost at the same position for each sequence.

Indels: muscular dystrophy

Some forms of muscular dystrophy that appear in adults are caused by tri- or tetranucleotide, e.g. $(CTG)_n$ and $(CCTG)_n$, repeats, where n may run into the thousands. The huge RNA transcripts that result interfere with alternative splicing of other transcripts in the nucleus.

 $(CTG)_n$

Here we analyze duplicates of CTG sequence. In particular repeats analyzed are $(CTG)_2$ (CTGCTG), $(CTG)_3$ (CTGCTGCTG), $(CTG)_4$ (CTGCTGCTGCTG) and $(CTG)_5$ (CTGCTGCTGCTGCTG). As for previously studied repeats, chiral index does not actually change, increasing the length of the repeated sequence.

 $(CCTG)_n$

In this section, we study $(CCTG)_n$: $(CCTG)_2$ (CCTGCCTG), $(CCTG)_3$ (CCTGCCTGCCTG), $(CCTG)_4$ (CCTGCCTGCCTGCCTG), $(CCTG)_5$ (CCTGCCTGCCTGCCTGCCTG) and $(CCTG)_{50}$. Even for the latter, longer sequence, no differences with the other ones could be found.

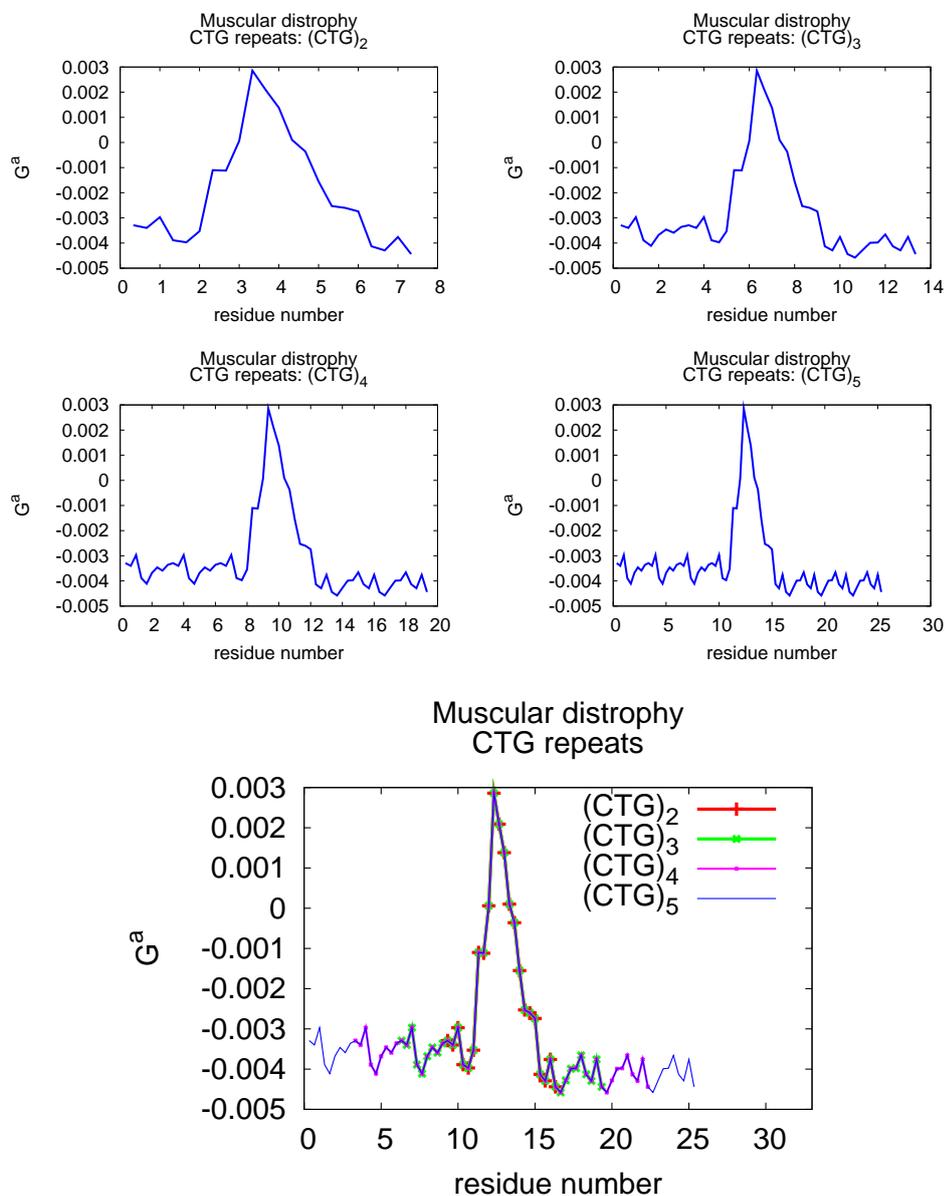


Figure 3.116: Muscular dystrophy. CTG repeats: sequences (CTG)₂, (CTG)₃, (CTG)₄ and (CTG)₅. Note that the cumulative graph has been made by shifting plots in order to have peaks almost at the same position for each sequence.

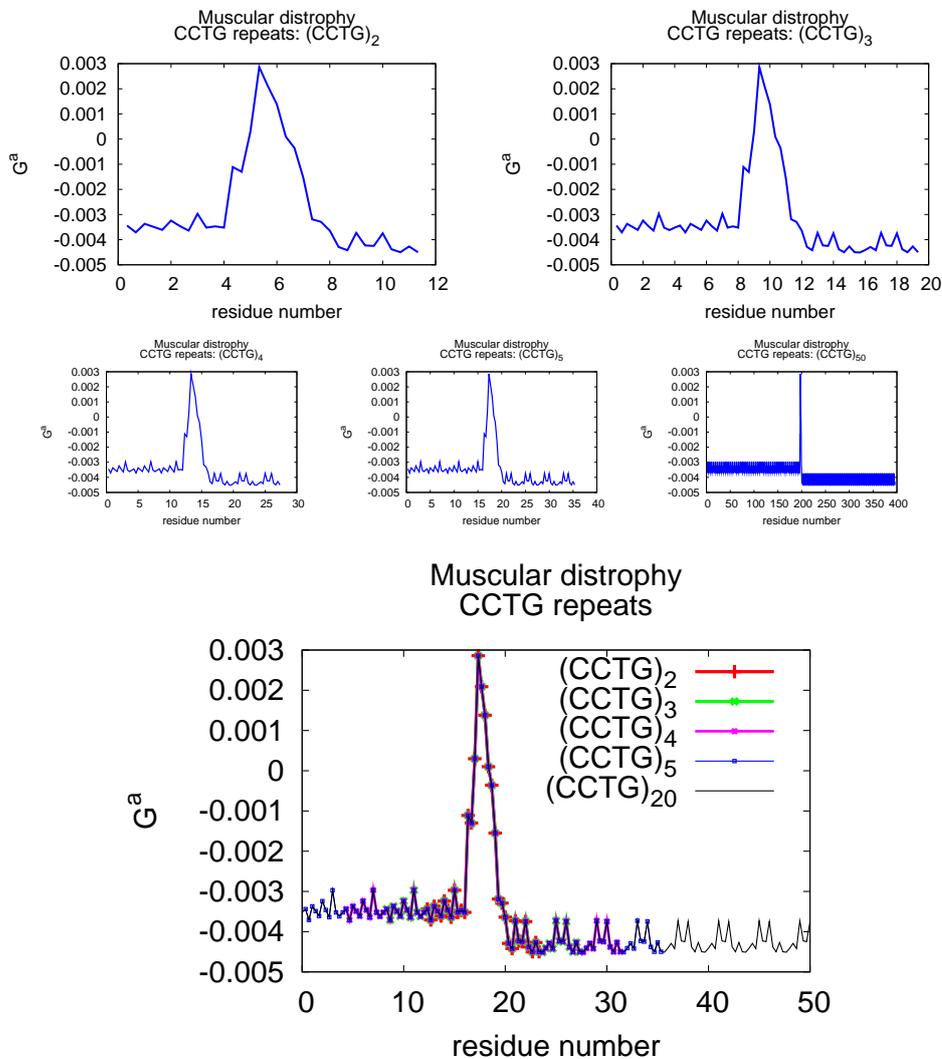


Figure 3.117: Muscular dystrophy. CCTG repeats: sequences $(\text{CCTG})_2$, $(\text{CCTG})_3$, $(\text{CCTG})_4$, $(\text{CCTG})_5$ and $(\text{CCTG})_{20}$. Note that the cumulative graph has been made by shifting plots in order to have peaks almost at the same position for each sequence.

Duplications

In this section, we try to analyze sequences and their doubles.

$(\text{CCTG})_4$ and $(\text{CCTG})_8$

Sequences studied are $(\text{CCTG})_4$ and $(\text{CCTG})_8$. As expected, it is not possible to obtain different chiral index behaviors when analyzing repeats.

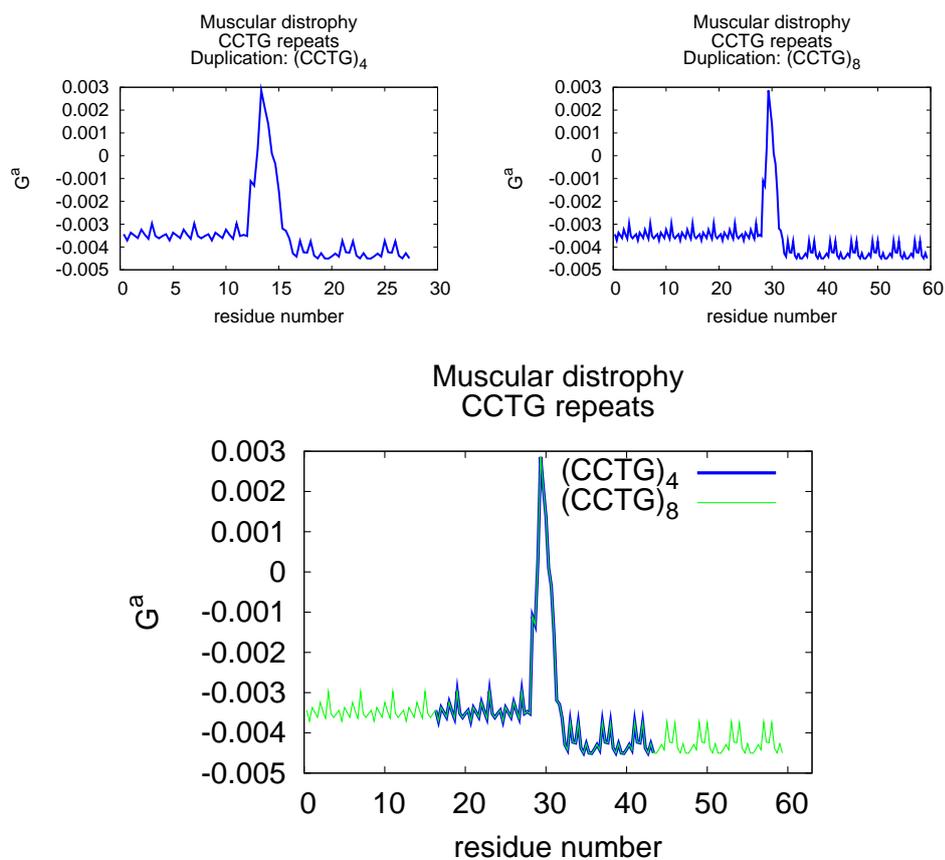


Figure 3.118: Muscular dystrophy. CCTG repeats duplication: sequences $(\text{CCTG})_4$ and $(\text{CCTG})_8$. Note that the cumulative graph has been made by shifting plots in order to have peaks almost at the same position for each sequence.

(CCTG)₅₀ and (CCTG)₁₀₀

We also try to analyze longer sequences: (CCTG)₅₀ and (CCTG)₁₀₀. Even in this case, little sequences repeated show the same trend for what concern G^a index.

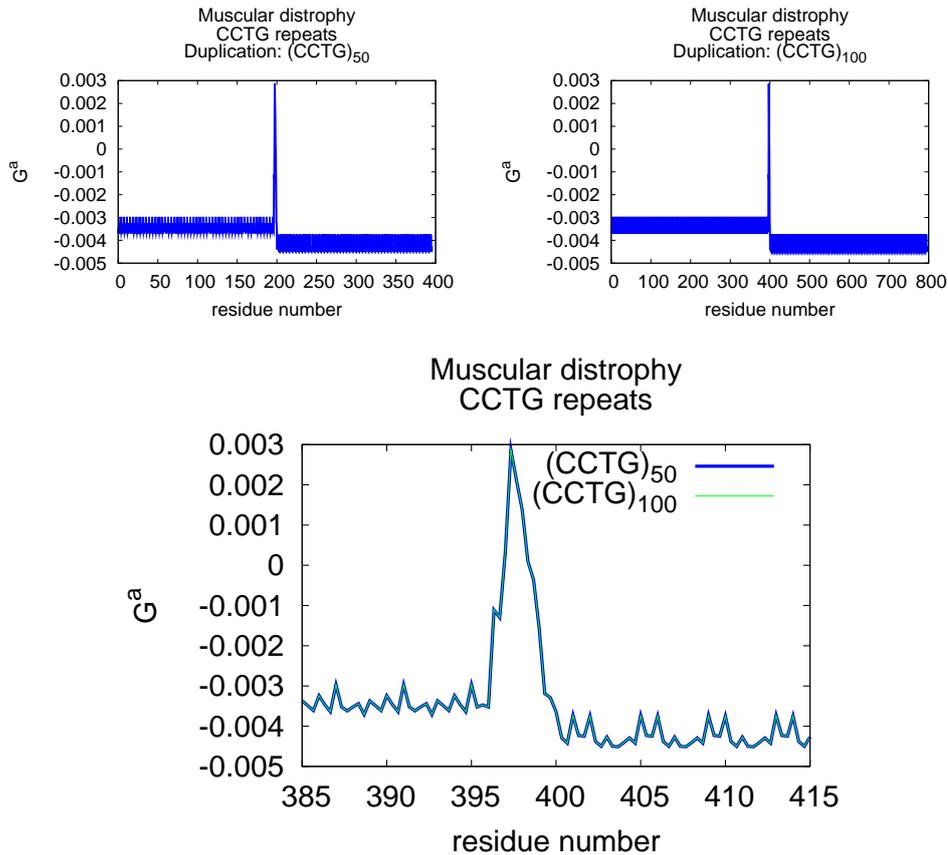


Figure 3.119: Muscular dystrophy. CCTG repeats duplication: sequences (CCTG)₅₀ and (CCTG)₁₀₀. Note that the cumulative graph has been made by shifting plots in order to have peaks almost at the same position for each sequence.

3.8.3.4 Discussions

From results collected, it is clear that chiral index G^a is not an appropriate parameter for analyzing all type of DNA mutations. In fact, nonsense point mutation, silent mutation and repeats are not recognizable by chiral index trend: wild-type and mutated sequences shows almost exactly the same behavior. On the contrary, splice-site point mutations and frame-shift indels seem to be easily detect from G^a trend. For what concern missense mutations, results are contradictory: in the sickle-cell disease example, mutation is pointed out clearly, while, for what concern the CFTR one, the behavior of chiral index does not change along with sequence mutation. Looking deeper into these examples, it is possible to note that the sickle-cell disease mutation is a transversion, while the CFTR one a transition. This feature worths an in-depth study.

3.8.3.5 Transition and transversion

Since the previous analysis on missense mutations gives contradictory results, other single-base substitutions are studied, on the sequence comprised between 7th and 21th nucleotide of the gene for the beta chain of hemoglobin (wild-type CTGACTCCTGAGGAG). Mutations are applied to 10th and 14th nucleotides.

Applying a transition on 10th nucleotide, sequence CTGGCTCCTGAGGAG is obtained; sequences undergone to transversion are CTGCCTCCTGAGGAG and CTGTCTCCTGAGGAG. Results are reported in Figure 3.120.

Transition from a C to a G, in the 14th nucleotide, makes the sequence become CTGACTCTTGAGGAG. Transversions are CTGACTCATGAGGAG and CTGACTCGTGAGGAG. Figure 3.121 shows results of this analysis. From these data, it is clear that G^a parameter detects transversions, but not transitions. It is also worth underlying that both possible transversions on a single nucleotide show the same behavior, which is of course different from the one typical of the wild-type. So this chiral index is a good tool to describe transversions.

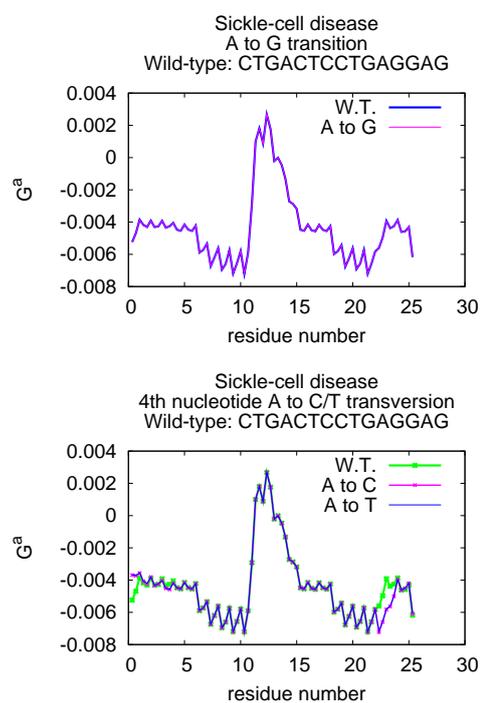


Figure 3.120: Gene for the beta chain of hemoglobin. 10th nucleotide point mutation: transition and transversion.

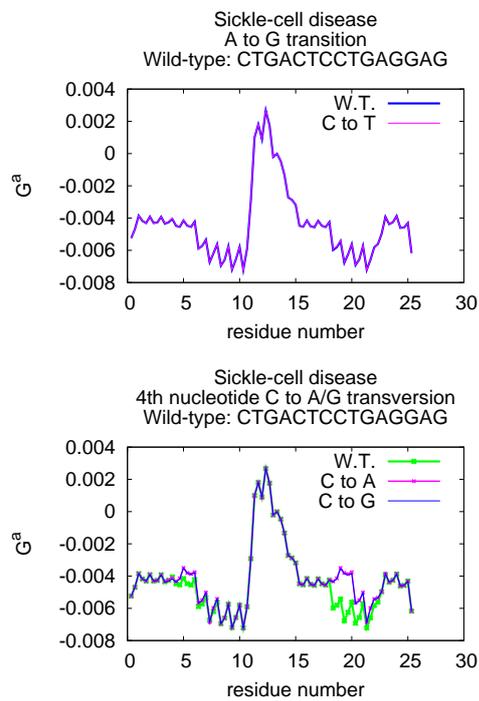


Figure 3.121: Gene for the beta chain of hemoglobin. 14th nucleotide point mutation: transition and transversion.

3.8.3.6 Conclusions

Results show clear that G^a could not be used to describe and trace nonsense point mutations, silent mutations, all kind of repeats and transitions. On the other hand, this parameter is a good tool for discriminate splice-site point mutations, frame-shift indels and transversions. Table 3.20 sums up which mutations are detectable or not using this chiral index parameter.

	Mutations	Examples of associated diseases	C^a detects these mutations?
Point mutations	Missense mutations: transitions	Cystic fibrosis	No
	Missense mutations: transversions	Sickle-cell disease	Yes
	Nonsense mutations	Cystic fibrosis	No
	Silent mutations	No change in protein product	No
	Splice-site mutations	Alteration of the protein product	Yes
Indels	Frame-shift	Alteration of the protein product	Yes
	Repeats	Fragile X syndrome, Huntington's disease Muscular dystrophy	No
Duplications	Repeats	Several neurological disorders could also be beneficial	No

Table 3.20: Types of mutation and chiral index adequacy for describing them.

3.8.3.7 Appendix: Analysis of correlation between codons and amino acid type

In this appendix, results about DNA sequences (triplets called “codons” transcribed in mRNA) and the amino acid for which they encode are collected. G^a trend. Figure 3.122 sums up the genetic code, linking triplets (mRNA and DNA) to their encoded amino acids.

		second base in codon					
		T	C	A	G		
T	first base in codon	TTT Phe	TCT Ser	TAT Tyr	TGT Cys	third base in codon	T
		TTC Phe	TCC Ser	TAC Tyr	TGC Cys		C
		TTA Leu	TCA Ser	TAA stop	TGA stop		A
		TTG Leu	TCG Ser	TAG stop	TGG Trp		G
C	CTT Leu	CCT Pro	CAT His	CGT Arg	T		
	CTC Leu	CCC Pro	CAC His	CGC Arg	C		
	CTA Leu	CCA Pro	CAA Gln	CGA Arg	A		
	CTG Leu	CCG Pro	CAG Gln	CGG Arg	G		
A	ATT Ile	ACT Thr	AAT Asn	AGT Ser	T		
	ATC Ile	ACC Thr	AAC Asn	AGC Ser	C		
	ATA Ile	ACA Thr	AAA Lys	AGA Arg	A		
	ATG Met	ACG Thr	AAG Lys	AGG Arg	G		
G	GTT Val	GCT Ala	GAT Asp	GGT Gly	T		
	GTC Val	GCC Ala	GAC Asp	GGC Gly	C		
	GTA Val	GCA Ala	GAA Glu	GGA Gly	A		
	GTG Val	GCG Ala	GAG Glu	GGG Gly	G		

		second base in codon					
		U	C	A	G		
U	first base in codon	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	third base in codon	U
		UUC Phe	UCC Ser	UAC Tyr	UGC Cys		C
		UUA Leu	UCA Ser	UAA stop	UGA stop		A
		UUG Leu	UCG Ser	UAG stop	UGG Trp		G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U		
	CUC Leu	CCC Pro	CAC His	CGC Arg	C		
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A		
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G		
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U		
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C		
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A		
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G		
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U		
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C		
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A		
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G		

Figure 3.122: DNA (*left*) and RNA (*right*) genetic code.

There are four classes of amino acids (after Timberlake [80]): the basic

ones (arginine, histidine, lysine), the acidic one (aspartic acid and glutamic acid), the neutral polar ones (asparagine, cysteine, glutamine, serine, threonine, tyrosine) and the neutral non-polar ones (alanine, glycine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, valine).

We analyze sequences encoding for each amino acids. Figure 3.123 sums up G^a behavior for all basic amino acids. Figure 3.124 contains the same information for acidic ones. Results for neutral non-polar and neutral polar amino acids are shown in Figure 3.125 and Figure 3.126, respectively. From these figures, it is clear that it is not possible to distinguish characteristic trends for each classes of amino acids.

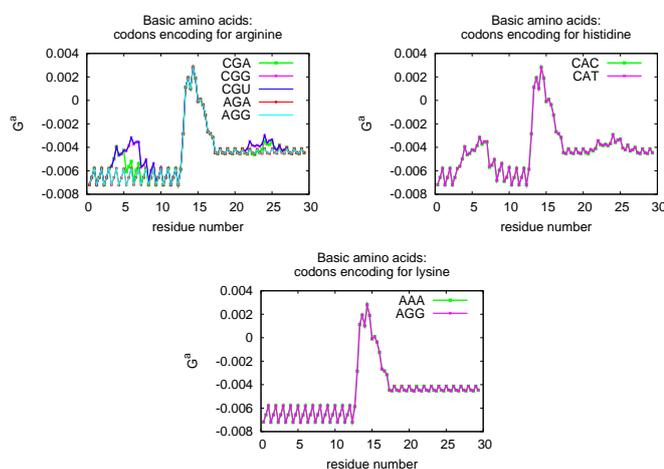


Figure 3.123: Basic amino acid chiral index investigation.

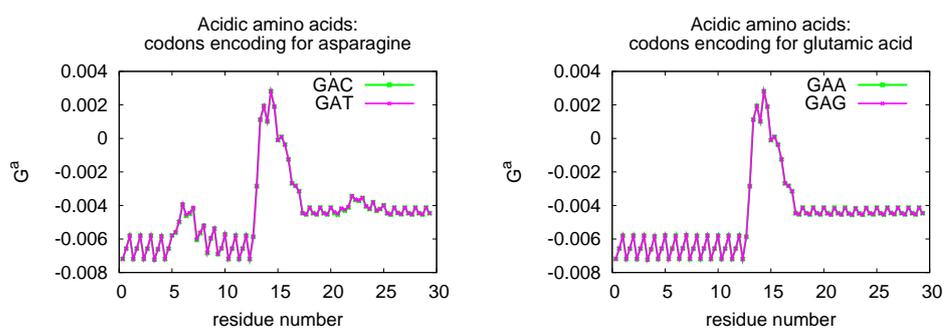


Figure 3.124: Acidic amino acid chiral index investigation.

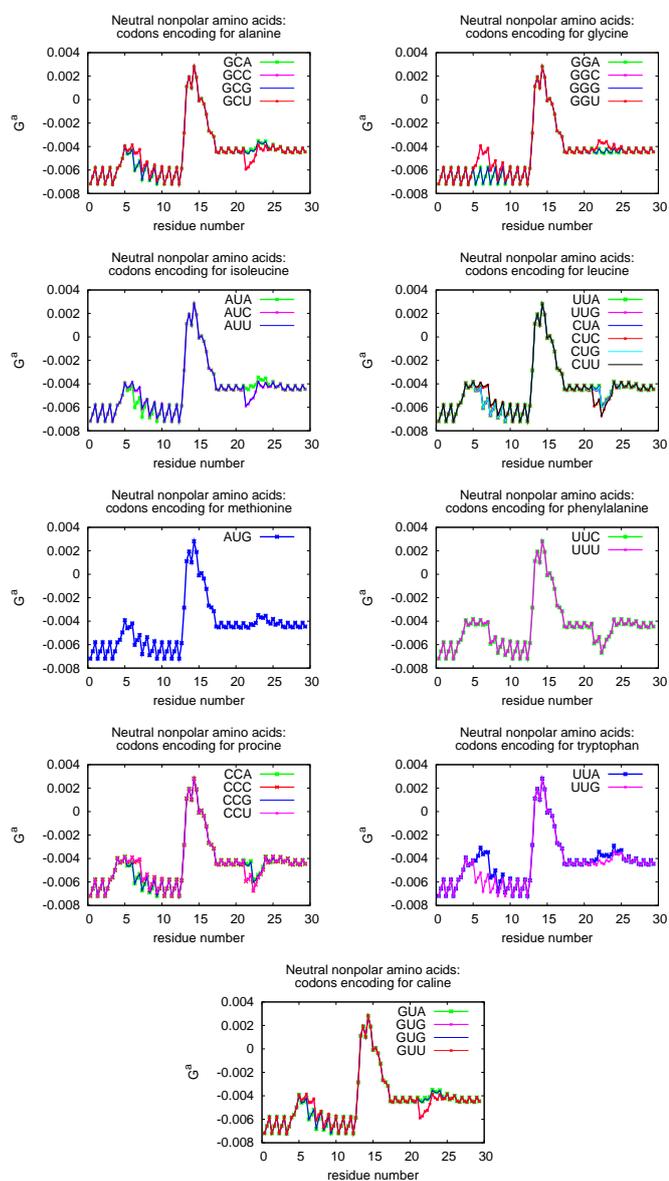


Figure 3.125: Neutral non-polar amino acid chiral index investigation.

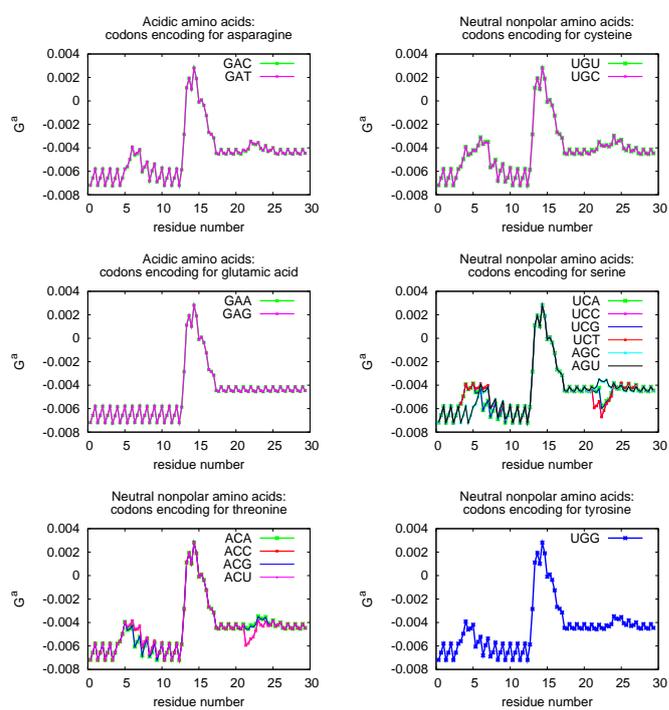


Figure 3.126: Neutral polar amino acid chiral index investigation.

Chapter 4

Atomistic modeling of PBLG

4.1 Summary

The interest in poly(γ -benzyl L-glutamate) (PBLG) arise from its chiral behavior, which can be quantitatively measured using residual quadrupolar splittings measured from ^{13}C and ^1H -Nuclear Magnetic Resonance (NMR) spectroscopy [81]. Organic solutions of this molecule show helicity [82], and are characterized by a large differential ordering effect (DOE) that can be exploited to discriminate enantiomers [83]. The molecular origin of this differentiation is still unknown. In order to comprehend how this discrimination might occur, we decided to run MD simulations with atomistic resolution.

After some preliminary studies, we have concentrated on a larger sample composed of a 144-residues long PBLG helix (18% w/w), 2912 molecules of dimethylformamide DMF solvent and 406 molecules of the pro-chiral solute heptyl butyrate (HEP), corresponding to a certain concentration (25% w/w). In order to examine the different interactions of the enantiomers with the helix, a chirality index previously developed [84, 85] in our group has been used. This HEP molecule has been chosen after having evaluated more than one hundred candidate compounds with our chiral index. High chirality solutes should interact better with the helix itself, making the study of the mechanism of the discrimination easier to follow and understand.

We have observed that DMF and HEP molecules solvate uniformly the

PBLG helix, but the pro-chiral solute is found closer to the helix backbone with respect to the DMF. The solvent presents a faster isotropic diffusion coefficient, twice as that of HEP, indicating a stronger interaction of the solute with the helix.

4.2 PBLG and chiral discrimination

PBLG chiral helices dissolved in a suitable organic solvent allow to discriminate enantiomers, by determining chemical shift anisotropy, residual quadrupolar splittings and differential ordering effect (DOE) measured by ^{13}C - and ^1H -NMR spectroscopy. These observables have been widely used in literature to measure the enantiomeric excess [83, 86, 81, 87]. On the other hand, the only computer simulations study is more than 15 years old and described only approximately the behavior of PBLG helices in DMF [88, 89]. This research is meant to analyze the physical mechanism used by such systems to induce a sufficient differential ordering effect to discriminate enantiomers [83]. The observables typically used for chiral discrimination are the residual dipolar couplings (RDC), that can provide both orientational and distance information. To be able to observe them, an anisotropic environment must be present. Such environment is achieved using as an orienting media an organic-solvent-based liquid crystal, the PBLG. In ref. [86], the orientational properties of an example compound are investigated. Marx et al. studied both diastereomorphous and enantiomeric combinations of PBG (L and D) with the two enantiomers of the solutes (isopinocampheol, (+)- and (-)-IPC). Enantiomeric mixtures, e.g. (+)-IPC/PBLG and (-)-IPC/PBDG, where PBDG is poly(γ -benzyl D-glutamate), do not give rise to different RDCs. On the contrary, diastereomorphous racemates show an effect on couplings. Two different behaviors can be recognized: in some cases the values of RDCs differ for the pure LC phases of PBLG and PBDG, while in other RDCs remain rather constant. In the first case, the RDCs show either linear dependence on the molar fraction of PBLG or a slightly s-shaped dependence. Since any “nonlinear effect” is detected, no diastereomorphous interaction is clearly

avored. Using the Euler angles, the differences between the two diastereomorphous orientations, e.g. (+)-IPC/PBDG and (+)-IPC/PBLG, are very small: only one angle (β) shows a linear dependence on the molar fraction of PBLG, while the other two remain constant (α , γ). Since the changes in beta angles are significant, a transition between the two diastereomorphous orientations happens. To estimate the mechanism of interaction, we follow a virtual experiment approach based on MD computer simulations and using an AMBER-23 force field [90] integrated with parametrizations specific for this system.

4.3 Modeling and validation

4.3.1 Physical and geometrical features

4.3.1.1 PBLG

Poly(γ -benzyl L-glutamate) is a polymer of functionalized glutamic acid. The chemical formula of the monomer is shown in Figure 4.1.

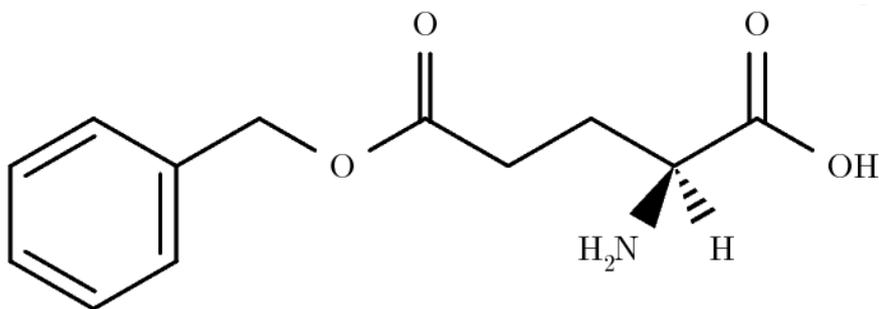


Figure 4.1: Chemical formula of benzyl L-glutamate.

This molecule polymerizes forming spontaneously an helix (Figure 4.2), which repeats every 18 residues. When the benzyl L-glutamate polymerizes, it forms an helix with a repeating unit composed as shown in Figure 4.3. This condensation implies the loss of a water molecule. A complete helix has 5 turns and the pitch is 27.04 Å long: so for each turn there are 3.6 residues (18/5) and the shifting angle is 100° (360/3.6). The axial rise per helical turn (27.04/5) is 5.4 Å; since there are 3.6 residues per turn (5.4/3.6), the axial rise per residue is 1.5 Å. So, when creating an helix, the next residue has to be placed in $(r, \phi+100^\circ, z + 1.5 \text{ \AA})$.

This polymer is well known for its ability to interact differently with enantiomers, allowing mixtures to be discriminated by NMR spectroscopy [81, 82, 83]. Moreover, PBLG has been used as thin films with polar order and piezo-pyroelectrical properties, particularly in photo-optical and electromagnetic applications and in the imaging of spatially resolved chemical libraries [91, 92, 93, 94, 95, 96, 97]. Self-assembly and chirality are the key feature for building these supramolecular constructs. These properties

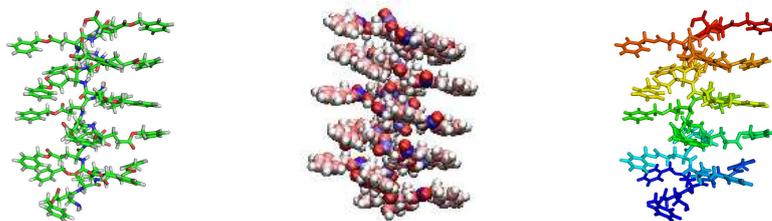


Figure 4.2: An 18₅ α -helix of PBLG. The snapshot in the center is color-coded with respect to charges computed with quantum chemistry method (Gaussian).

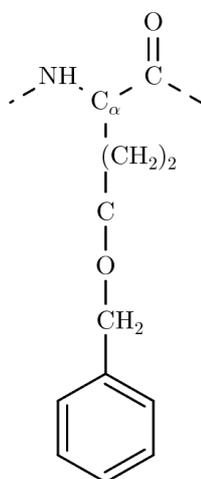


Figure 4.3: A repeating unit of benzyl L-glutamate.

make possible the in depth examination of the structure-function relationships, particularly those between optical characteristics, molecular order and the various types of birefringence. Finally, since this polymer shows a mimetic superstructure with biological supra-organization, it is expected to be biocompatible with collagen fibers [82].

4.3.1.2 DMF

The solvent, chosen following ref. [89], is dimethylformamide (Figure 4.4).

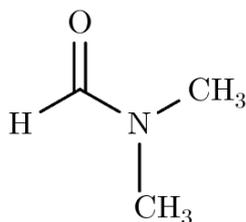


Figure 4.4: A single molecule of dimethylformamide.

This solvent is colorless and odorless, miscible with water and most of organic liquids. Dimethylformamide is a polar (hydrophilic) aprotic solvent with a high boiling point (425-427 K). It facilitates reactions that follow polar mechanisms, such as nucleophilic substitution (S_N2 reactions).

4.3.1.3 Heptyl butyrate

Heptyl butyrate is a colorless liquid, whose formula is shown in Figure 4.5 and has been chosen due to its quite high chiral index value (see appendix: “Chiral index analysis for various solutes”) and its relatively low steric effect (since no benzylic group is involved).

Heptyl butyrate is a food grade compound found abundantly in fresh apples and plums. Its odor resembles that of chamomile and sweet green tea. It is used as pesticide in residential and commercial areas in traps to control various species of yellowjackets and wasps. Based on the data reviewed by US Environmental Protection Agency (EPA), heptyl butyrate will not cause adverse effects to humans thanks to its low toxicity. The data submitted and reviewed showed that there is no reason to believe that any non-target organisms, including honeybees and other beneficial insects, would be attracted to or adversely affected by the use of heptyl butyrate in a wasp trap.

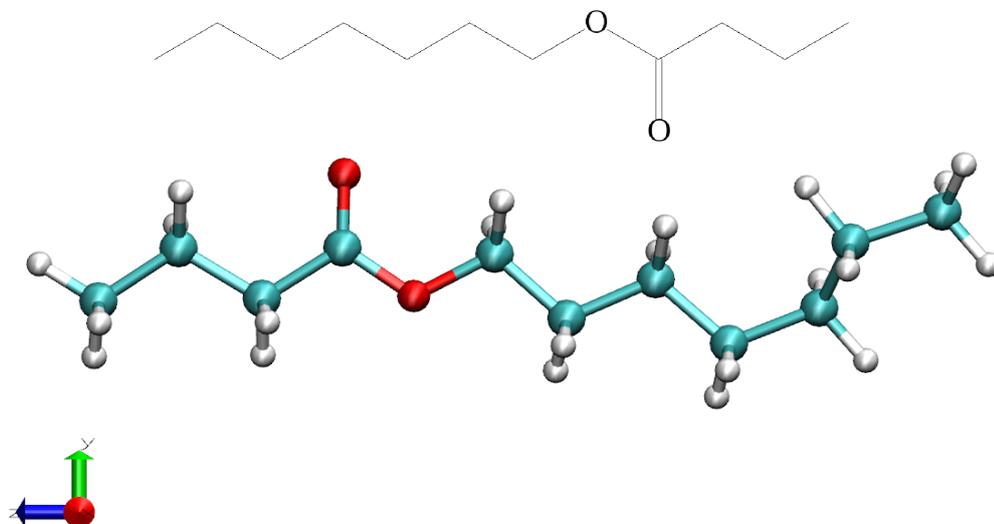


Figure 4.5: A molecule of heptyl butyrate.

4.3.2 Parametrizing molecules

4.3.2.1 PBLG

We start modeling an isolated molecule of benzyl L-glutamate (called BGL) using both Pymol [98] and Avogadro [99]. With Pymol [98], an α -helix of simple glutamate is built; then a benzylic group and all needed hydrogens are added for each residue (Figure 4.6). Then, we minimize this structure at its minimum energy geometry with Avogadro [99] using the Universal Force Field (UFF) [100].

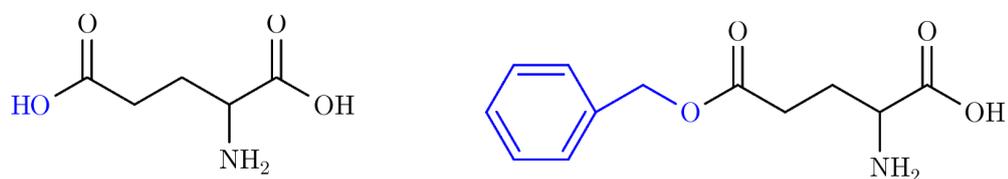


Figure 4.6: The chemical formula of glutamate molecule (*left*) and that of substituted benzyl glutamate (*right*).

With Gaussian [101], partial atomic charges on a 5-oligomer (i.e. a molecule

composed by five monomer units) are computed, using cc-pVTZ basis set and RESP (Restrained ElectroStatic Potential) technique [102].

These data are used to complete parameters of the topology modified file (obtained as a sum of the original glutamate AMBER topology file and this new parametrized part) and the parameter file.

However, to build the α_{18} -helix, a set of biomolecular programs (ISIS/Draw [103], Discovery Studio Visualizer [104], and Argus Lab [105]) is used. First a single residue (already deprived of the water molecule resultant from polymerization) is built and optimized. Then, this molecule is shifted of 1.5Å and rotated by 100° around z -axis (using “pdb.translation”) 17 times. These 18 residues obtained this way are then linked together and this last configuration is optimized using Avogadro [99].

4.3.2.2 DMF

This molecule is drawn with Pymol [98], minimized with Avogadro [99] and analyzed with Gaussian [101] as explained for the benzyl L-glutamate.

4.3.2.3 Heptyl butyrate

Also this molecule is drawn with Pymol [98], minimized with Avogadro [99] and analyzed with Gaussian [101] as explained for the benzyl L-glutamate. From results of such analysis we assign atom names and atomic charges follow (Table 4.1 and Figure 4.7):

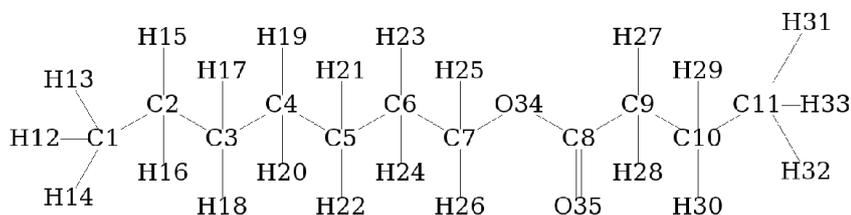


Figure 4.7: Atom names assigned to heptyl butyrate molecule.

Then, parameters contained in the file “par_amber_cornell.inp” are controlled to check if they are in good agreement with those typical of Gaus-

sian optimized structure. Distances and angles are in quite good agreement (Table 4.2).

Dihedral energy profile: O35–C8–C9–C10 parametrization

After obtaining information about the minimized structure and the atomic charges, we perform a minimization run using one molecule of heptyl butyrate in vacuum with NAMD in order to view if the simulation produces a different minimum state. As shown in Figure 4.8, the NAMD resultant configuration is quite different from the Gaussian one: hydrogens bounded to the α carbon with respect to the carbonyl group tend to align in *trans* to the oxygen (called O35) to get far away as possible from O35, on the contrary of what predicted by the Gaussian minimized structure. So a study of dihedral angle (O35–C8–C9–C10) profile is needed.

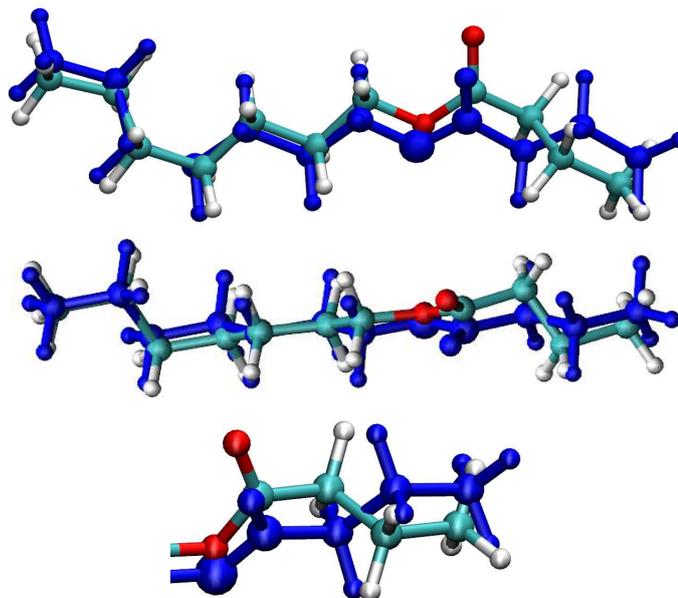


Figure 4.8: Comparison of heptyl butyrate minimized structures (the blue molecule is the one optimized with Gaussian) from different point of views. Note that the last figure is a zoom of the part of the chain that changes more.

The O35–C8–C9–C10 dihedral angle is minimized both using Gaussian and using NAMD. The *ab initio* calculation does not take into account

non-bonded term, van der Waals term and so on.

First of all, a Gaussian run is carried out (using *cc*-PVTZ basis set) in order to compute the energy profile related to the rotation of this dihedral angle. We obtained values of energy for $0^\circ < \phi < 180^\circ$ and construct the negative branch of the graph as the mirror reflection of the positive one. The values that have actually been fitted are the δE , i.e. the difference $E_i - E_{min}$, after having been converted in kcal/mol (from Hartree). The equation used is that of the AMBER force field [90]:

$$\begin{aligned} E_{\phi, AMBER} = & b_0 + (b_1) \cdot \cos x + b_2 \cdot \cos(2x) + b_3 \cdot \cos(3x) + b_4 \cdot \cos(4x) + \\ & + b_5 \cdot \cos(5x) + b_6 \cdot \cos(6x) + b_7 \cdot \cos(7x) + \\ & + b_8 \cdot \cos(8x) + b_9 \cdot \cos(9x) + b_{10} \cdot \cos(10x) \end{aligned} \quad (4.1)$$

Parameters obtained from this fit are reported in Table 4.3 and Figure 4.9 shows the trend of the energy profile.

Atom name	Atom type	Atomic charge
C1	CT	-0.4525
C2	CT	0.1930
C3	CT	-0.0491
C4	CT	-0.0830
C5	CT	-0.0761
C6	CT	-0.0514
C7	CT	0.2754
C8	C	0.8910
C9	CT	-0.5224
C10	CT	0.3359
C11	CT	-0.3799
H12	HC	0.1064
H13	HC	0.1064
H14	HC	0.1064
H15	HC	-0.0054
H16	HC	-0.0054
H17	HC	0.0228
H18	HC	0.0228
H19	HC	0.0278
H20	HC	0.0278
H21	HC	0.0210
H22	HC	0.0210
H23	HC	0.0428
H24	HC	0.0428
H25	HC	0.0120
H26	HC	0.0120
H27	HC	0.1364
H28	HC	0.1364
H29	HC	-0.0292
H30	HC	-0.0292
H31	HC	0.0882
H32	HC	0.0882
H33	HC	0.0882
O34	OS	-0.5315
O35	O2	-0.5901

Table 4.1: Heptyl butyrate parameters: atom names, atom types and atomic charges. Atom type are: CT for sp^3 aliphatic carbon, C for sp^3 C of a carbonyl group, HC for aliphatic hydrogen bonded to carbon without electronwithdrawal group, OS for ether and ester oxygen and O2 for carboxyl and phosphate group oxygen. Note that atomic charges are symmetrized, i.e. charges had been averaged for hydrogens bonded to the same carbon. Dipole moment components M_k , computed with Gaussian [101], are $M_x = 0.4143$, $M_y = -1.6703$, $M_z = -0.7712$ (with respect to the shape of molecule, where z is conventionally used for the long axis of the molecules and the y axis is directed as the oxygen atom, which is placed on the plane xy , see the lower image of Figure 4.5).

Distances			Angles		
Parameter	Gaussian optimized	Amber	Parameter	Gaussian optimized	Amber
C-CT	1.50	1.522	HC-CT-HC	106.4	109.50
CT-CT	1.52	1.526	CT-CT-HC	109.5	109.50
CT-HC	1.09	1.090	CT-CT-CT	113.3	109.50
OS-CT	1.43	1.410	CT-OS-C	115.9	117.00
OS-C	1.34	1.343	HC-CT-OS	109.0	108.50
O2-C	1.20	1.250	OS-C-O2	123.5	122.43
			O2-C-CT	125.5	117.00
			C-CT-HC	107.9	109.50

Table 4.2: Comparison between the parameters of the Gaussian minimized geometry computed with VMD [106] and those reported by the AMBER force field [107].

Constant	Value (kcal/mol)
b_0	0.725455
b_1	-0.512677
b_2	-0.0898118
b_3	-0.122593
b_4	-0.00507046
b_5	-0.000837269
b_6	0.00186292
b_7	-0.000788303
b_8	0.00220452
b_9	0.00247209
b_{10}	-0.000090763

Table 4.3: Values of the constants obtained by fitting with the equation for the AMBER torsional energy.

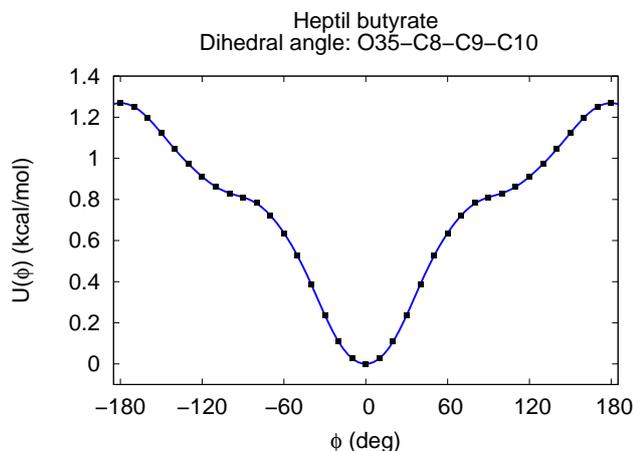


Figure 4.9: O35–C8–C9–C10 dihedral angle: Gaussian energy profile. Black points are data obtained from Gaussian output and the blue line is the fit with AMBER torsional energy.

Now, we determine the free energy differences along the reaction coordinate, ξ . This goal is achieved employing the adaptive biasing force (ABF) method [108] in its NAMD formulation and implementation [109]. ABF is based on the computation of the mean force along ξ , which is then canceled out by an equal and opposite biasing force, allowing the system overcome barriers and escape from minima of the free energy landscape. Ultimately, the dynamics of ξ corresponds to a random walk with zero mean force, and only the fluctuating part of the instantaneous force exerted along ξ remains. Virtual erasure of the roughness of the free energy landscape yields a uniform sampling along ξ . The ABF scheme assumes that the reaction coordinate, ξ , is fully unconstrained. This implies that in the course of the simulation, the complete reaction pathway discretized in small bins of width $\delta\xi$ will be explored in a continuous fashion. Sample of the instantaneous force acting along ξ are accrued in the different bins until a user-defined threshold is attained, beyond which the adaptive biasing force will be applied.

Before running this ABF simulation, we had to change the atom type of the C8 atom into CD, which is not present in our molecule so that we could modified values of dihedral angles which involves this atom without affecting other CT carbon atoms dihedrals. Then, we modify the topology in order to let this new atom type appear and create another .pdf file. The parameter file is also modified to take into account the CD atoms, copying bonds, angles, dihedral and impropers parameters involving such CD atoms. The new dihedral (“PHI” section of the parameter file) is “O2 CD CT CT”. Note that the set of dihedral angle “* CD CT *” could not be used since it comprehends 5 dihedrals while we want only that specific one (unitary multiplicity, $n = 1$). At the beginning, the dihedral angle is set to zero so that the bonded angle does not give any contribution to the energy. Since CHARMM takes into account only the first six terms, we want to obtain the force constants and minimum geometries of the dihedral angle (in degree) for such parameters.

$$\begin{aligned}
 E_{\phi,CHARMM} = & n_0 + n_1 \cdot \cos(x + p_1) + n_2 \cdot \cos(2x + p_2) + n_3 \cdot \cos(3x + p_3) + \\
 & + n_4 \cdot \cos(4x + p_4) + n_5 \cdot \cos(5x + p_5) + n_6 \cdot \cos(6x + p_6) + \\
 & + n_7 \cdot \cos(7x + p_7) + n_8 \cdot \cos(8x + p_8) + \\
 & + n_9 \cdot \cos(9x + p_9) + n_{10} \cdot \cos(10x + p_{10})
 \end{aligned} \tag{4.2}$$

The ABF simulation is repeated until the energy profile obtained from Gaussian and the one obtained with the NAMD superimpose. This convergence is achieved since parameters for our dihedral angle are fitted at every run and these new data are the starting point for the next NAMD simulation. When plots are almost the same, the parameters are:

02	CD	CT	CT	0.48175663	1	180.0
02	CD	CT	CT	0.83350102	2	180.0
02	CD	CT	CT	0.14097623	3	180.0
02	CD	CT	CT	0.10486943	4	180.0
02	CD	CT	CT	0.07681031	5	0.0
02	CD	CT	CT	0.02236823	6	180.0

describing the four atoms of the dihedral angle under study, the force constant, a kind of multiplicity (which is computed as n_i/n) and the minimum geometry of the angle (p_i). These data are necessary to run a proper simulation. To check if the parameters obtained are in good agreement with the Gaussian results, we run another minimization using NAMD. The minimized structure obtained in such way is in good agreement with the Gaussian one (Figure 4.10): so the new parameters describe well the behavior of such molecule.

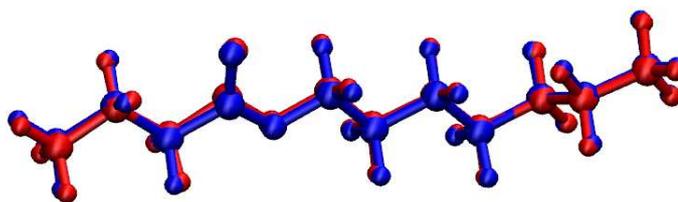


Figure 4.10: The last configuration of heptyl butyrate obtained with a NAMD minimization run (*blue*) and that obtained with Gaussian (*red*). These two structures almost superimpose, so the new parameters used for the NAMD simulation are correct.

4.4 Results and discussion

4.4.1 Two poly(γ -benzyl L-glutamate) helices in dimethylformamide: parametrization and simulation

These atomistic MD simulations confirm results obtained by Helfrich *et al.* [88, 89].

4.4.1.1 Setting up simulation

Since solvent and solute FF parametrization is completed, the two samples described in [89] could be generated. The first sample is made by two parallel segments (two 18_5 α -helices): these two helices are one the mirror image of the other (thus having different handedness and creating a racemic mixture). At the beginning, molecules are placed at a distance of 22 Å along x -axis with respect to the laboratory frame. In our model, the two helices are exactly the same: the second one is only translate of 22 Å along x -axis, with respect to the center of mass. Note that creating the racemic mixture is not really important, since PBLG state equations are well described by a simple excluded-volume model, which does not take into account helical nature of the molecule. Box sides are: $L_x = 94.9$ Å, $L_y = 60$ Å and $L_z = 27.04$ Å (the length of an 18 oligomer of PBLG), where L_x , L_y and L_z are the box lengths with respect to the laboratory frame. These values are chosen so that periodic images can interact. This sample contains 1144 DMF molecules.

The second sample comprehends two perpendicular helices, with an initial separation (between the center of mass) of 21 Å. The second helix is placed translating the first one of 21 Å and rotating it with an angle of $\pi/2$ around x -axis. Box sides are: $L_x = 63.8$ Å and $L_y = L_z = 54.08$ Å. In this case, there are no interactions between periodic images and DMF molecules are 1320.

Figure 4.11 shows snapshots of samples used as start-points for running simulations.

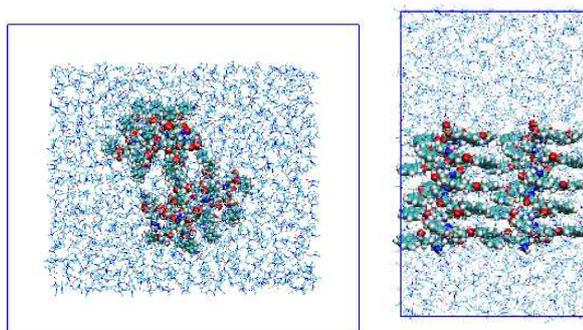


Figure 4.11: Systems with perpendicular (*left*) and parallel helices (*right*).

Simulations are run as explained in paragraph 4.6.4 switching off “fixed atoms” but keeping them “constrained”. Pressure is decreased to 1.01325 bar and the *zeroMomentum* is set *yes*. Simulations are characterized by the following features:

- timestep = 1 fs;
- total simulation time = 40 ns;
- T = 300 K, kept using rescale velocities algorithm;
- P = 1.01325 bar, using a Berendsen barostat;
- isobaric and isotropic ensemble.

Figure 4.12 shows a snapshot of sample after 40 ns (note that periodic boundary conditions split up helices).

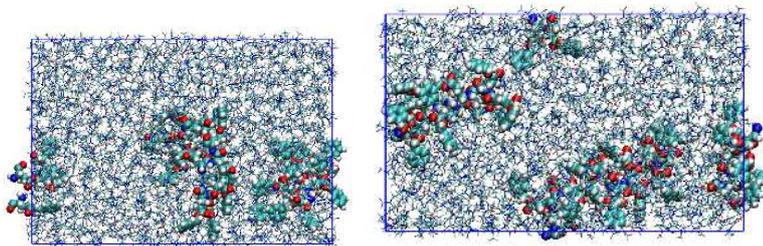


Figure 4.12: Systems with perpendicular helix (*left*) and with parallel one (*right*) at the end of the simulation (40 ns).

4.4.1.2 Results

The following table (Table 4.15) sums up average energy, temperature, volume and density values for both systems (with parallel and perpendicular helices). Note that energy values are not normalized, since they refer to the entire system.

Parameter (Units)	System with	
	parallel helices	perpendicular helices
Bond energy (kcal/mol)	4421.75	5047.30
Angle energy (kcal/mol)	6379.73	7299.74
Dihedral energy (kcal/mol)	1575.74	1780.82
Electrostatic energy (kcal/mol)	1758.65	2267.75
Van der Waals energy (kcal/mol)	-5767.28	-6587.82
Total energy (kcal/mol)	21760.89	25104.54
Temperature (K)	300.05	300.05
Volume (\AA^3)	168338.90	192604.71
Density (g/cm^3)	0.90	0.90

Table 4.4: Parameters: comparison between systems with parallel and perpendicular helices.

Since systems are different for dimensions and solvent molecules number, energy values are not comparable.

Also other parameters are analyzed:

- radial distribution functions: DMF-DMF, DMF-BGL and BGL-BGL $g(r)$ and $g(z)$;
- aspect ratio;
- orientational order parameters (P_2 and molecular biaxiality R_{02}^2) for BGL;
- time correlation functions;

- displacement analysis;
- solvation sphere;
- atom–atom distance trend.

Moreover, a comparison with a few data of the reference article [89] is done.

Radial distribution function and density distribution function

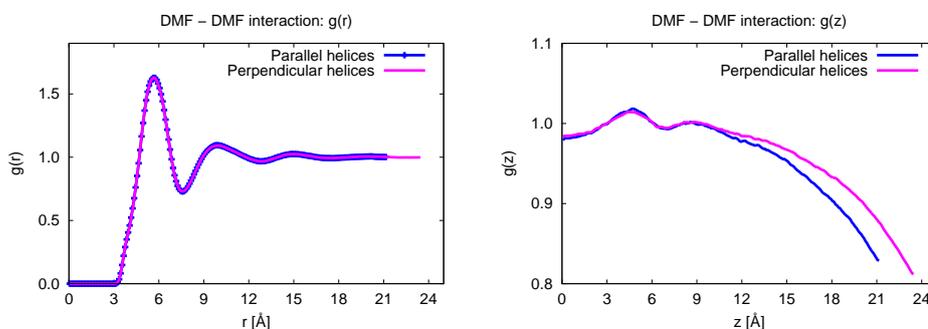


Figure 4.13: $g(r)$ and $g(z)$ computed with respect to two DMF molecules.

For DMF molecules, distribution functions show the typical trend of an isotropic solvent. $g(z)$ represents a density distribution function normalized for the total density along z -axis. The $g(r)$ trend shows two well-defined peaks, corresponding to two solvation shells.

From Figure 4.14 it is clear that DMF molecules solvates uniformly the helices.

Finally, $g(r)$ trends are more structured for the perpendicular system and $g(z)$ shows trend similar to the initial part of a wave, typical of helicoidal system. This last behavior depends on the density which is not constant, since helix has a well-structured shape.

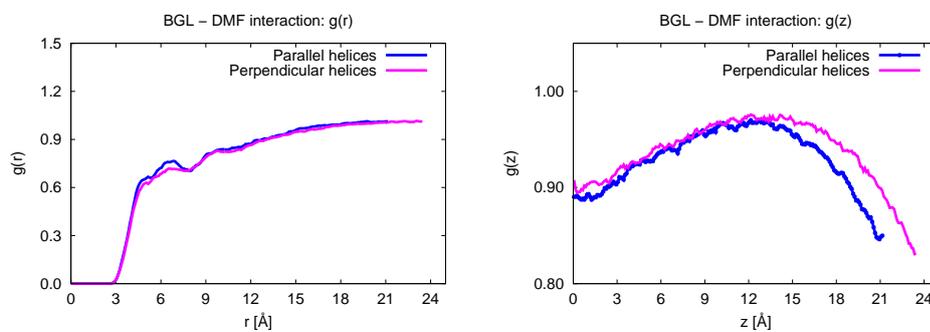


Figure 4.14: $g(r)$ and $g(z)$ of DMF molecules with respect to the BGL helices.

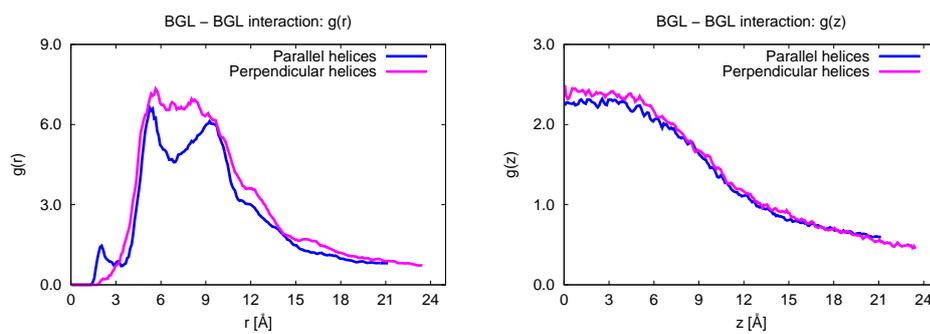


Figure 4.15: $g(r)$ and $g(z)$ typical of BGL molecules with respect to other BGL molecules.

Aspect ratio

The distribution of the aspect ratio becomes wider and the probability of having the same molecular length to breadth ratio decreases. So, helices probably undergo a distortion, in particular terminal residues. Since molecules are not stable, we need larger helices.

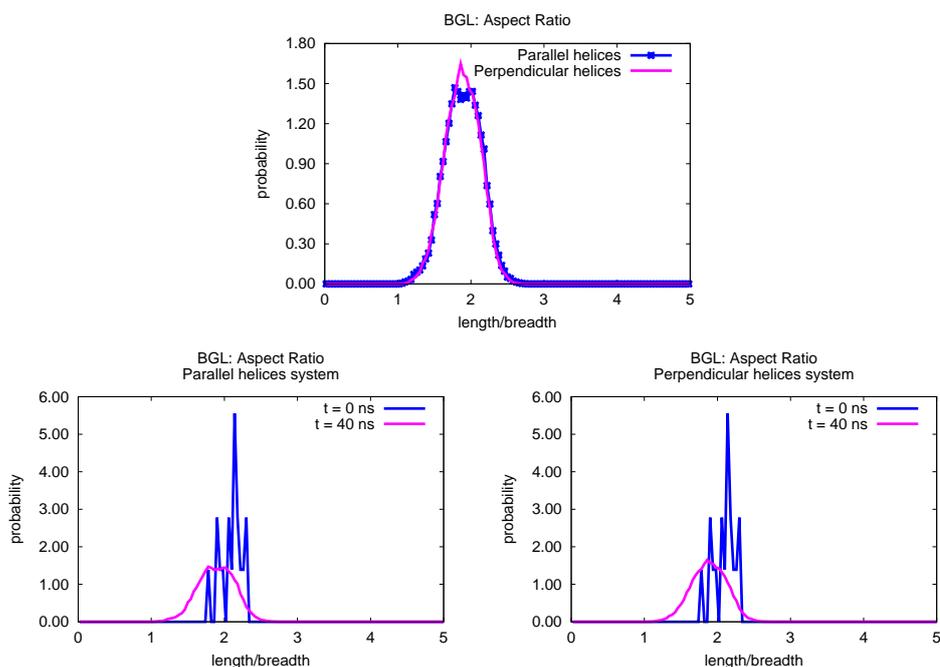


Figure 4.16: BGL aspect ratio. The first graph shows a comparison between initial sample for each system. The graphs below compare, for both systems (with parallel and perpendicular helices), initial samples and the same samples after a simulation time of 40 ns.

Order parameters

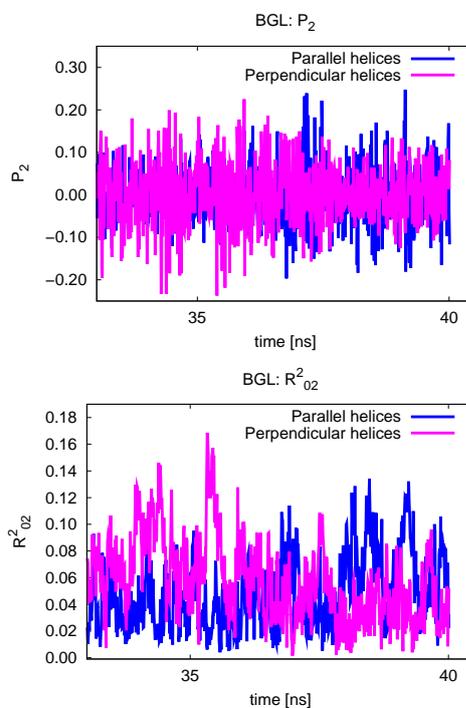


Figure 4.17: Order parameters computed with respect to BGL.

The orientational order parameter, P_2 , shows typical isotropic zero values. Even if helices are truly ordered, residues are arranged in a random way. Further analysis, in particular using a helicoidal order parameter, are needed. Molecular biaxiality R_{02}^2 values have high uncertainties but introduce an uniaxial behavior.

Time correlation functions

Since time correlation functions do not change, helices seem not to become spatially modified, but indeed they do, due to uncertain statistics and largely distorted extreme residues.

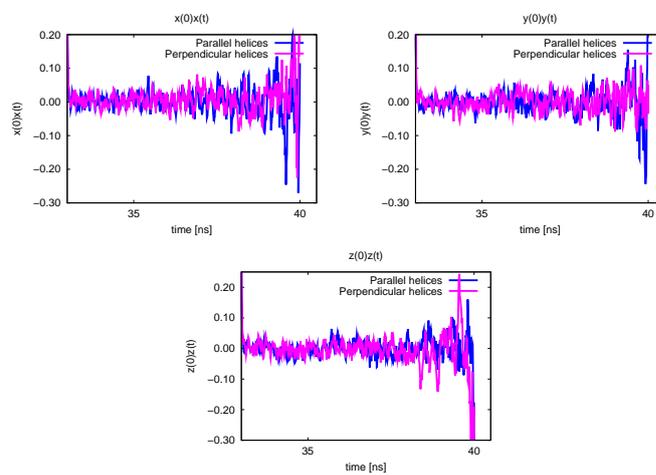
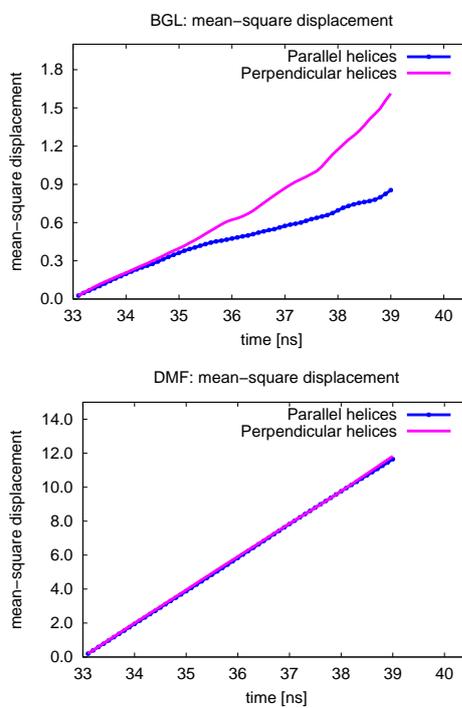


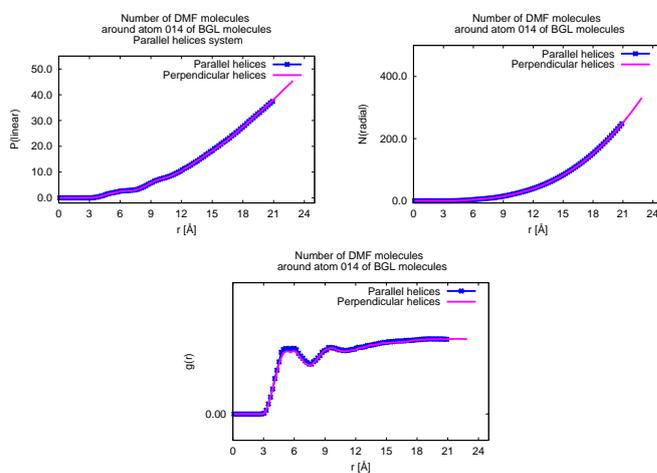
Figure 4.18: Time correlation function: $x(0)x(t)$, $y(0)y(t)$, and $z(0)z(t)$.

Displacement analysis

**Figure 4.19:** Displacement analysis.

As expected, the isotropic solvent presents a faster diffusion with respect to the BGL, independently to the laboratory frame orientation.

Solvation sphere

**Figure 4.20:** Solvation sphere.

The trends of both systems are similar. Probably the distance between helices centers of mass is very low and the solvation spheres consider only DMF molecules outside the interaction section of the helices.

Atom–atom distance trend

Atom–atom distance trend is broadened, even if in a similar way for both systems: helices do not remain stable. Indeed, the peak at the end of the simulation is shifted with respect to the one at the beginning.

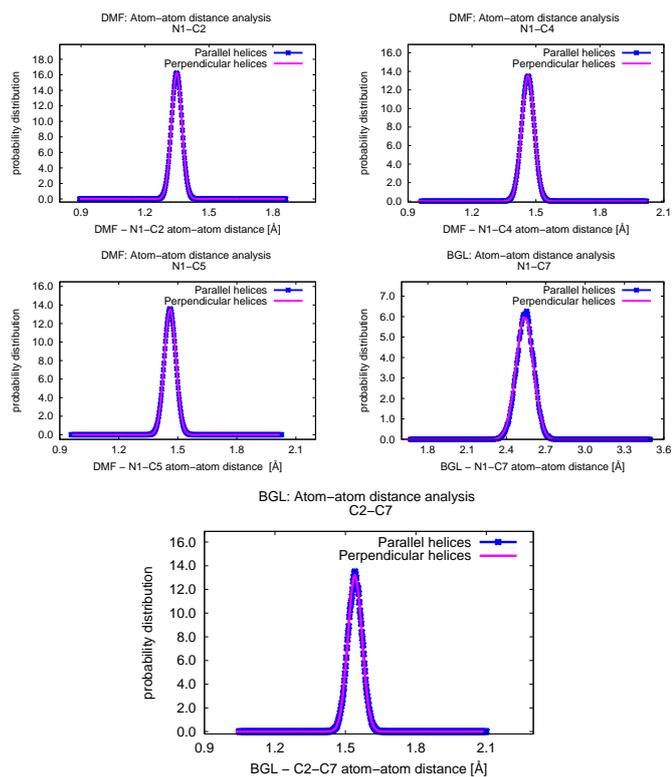


Figure 4.21: Atom–atom distance: comparison between the two systems.

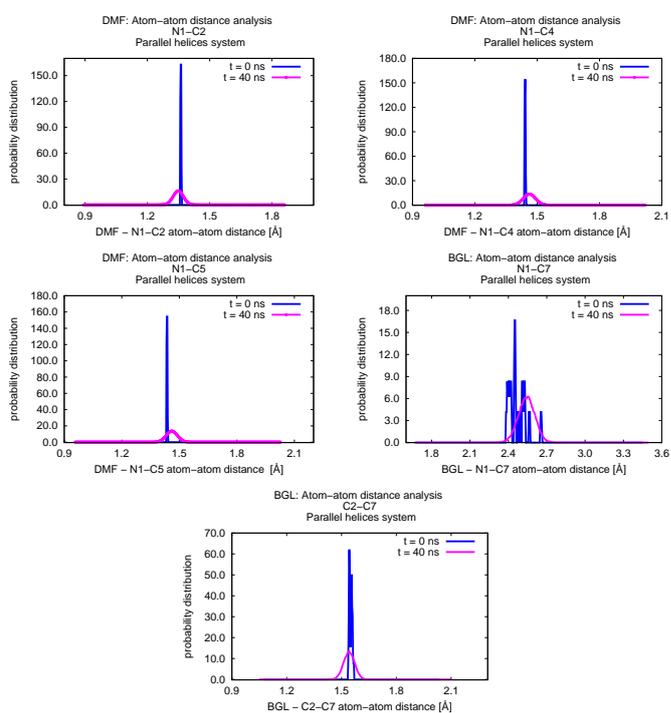


Figure 4.22: Atom-atom distance for the system with two parallel helices.

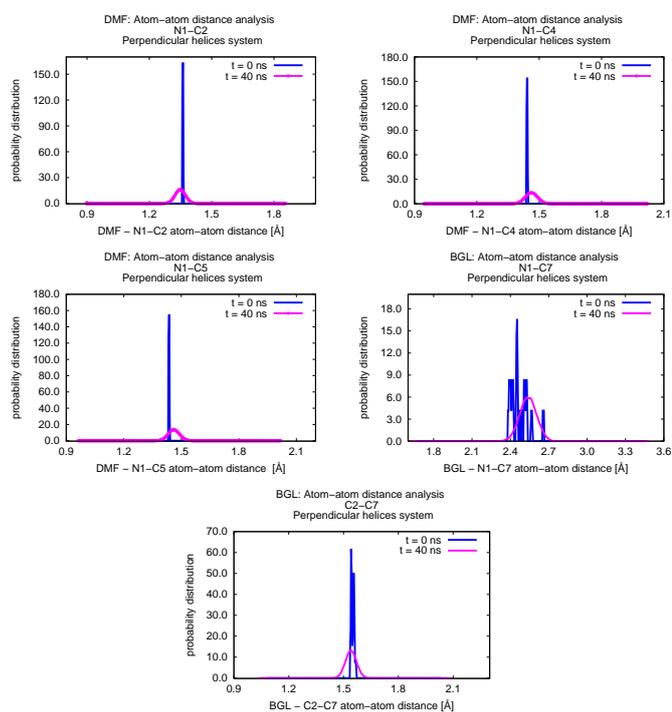


Figure 4.23: Atom-atom distance for the system with two perpendicular helices.

4.4.1.3 Comparison with Helfrich data [89]

For what regard $g(r)$, the trend we obtained is in good agreement with literature data [89], as shown in Figure 4.24. The first two solvation shells are almost placed at the same distances (first peak at $\sim 5.0 \text{ \AA}$ and the second one at $\sim 10.0 \text{ \AA}$).

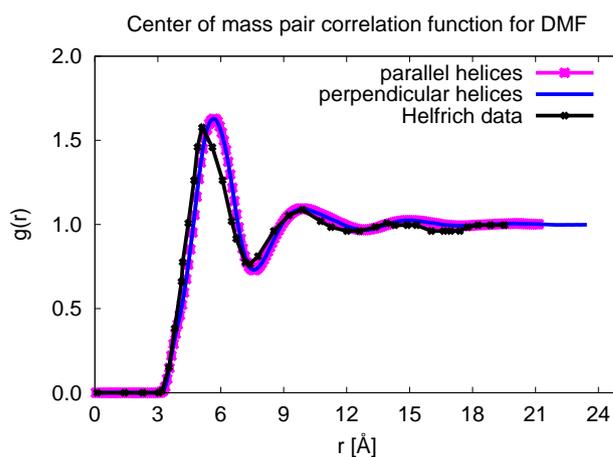


Figure 4.24: Comparison with literature: $g(r)$.

We also compare diffusion coefficients. Table 4.5 sums up values obtained from our simulations and those reported in literature.

DMF diffusion coefficient data we obtained are two order of magnitude greater (faster) than those obtained by Helfrich *et al.* [89]. The latter are in good agreement with the experimental values. BGL molecules we simulated move a lot, maybe because we just keep the first and the last atom of the backbone constrained to certain positions, while in the past work the entire backbone was fixed.

From diffusion coefficients computed for BGL, we could see that also such molecules show a fast motion. In particular, parallel helices separate more along x - and y -axis with respect to the z -axis, while perpendicular helices move away mainly along y -axis.

Diffusion coefficient for	Experimental data [89]	Literature data [89]	System with	
			parallel helices	perpendicular helices
DMF	2.0×10^{-9} [m ² /s]	1.98×10^{-9} [m ² /s]	1.68×10^{-7} [m ² /s]	1.67×10^{-7} [m ² /s]
THF	3.1×10^{-9} [m ² /s]	3.40×10^{-9} [m ² /s]	/	/
BGL	/	/	$D_x = 1.12 \times 10^{-8}$ [m ² /s] $D_y = 1.12 \times 10^{-8}$ [m ² /s] $D_z = 0.99 \times 10^{-9}$ [m ² /s]	$D_x = 1.81 \times 10^{-8}$ [m ² /s] $D_y = 2.00 \times 10^{-8}$ [m ² /s] $D_z = 1.91 \times 10^{-8}$ [m ² /s]

Table 4.5: Solvent diffusion coefficient: comparison between experimental data, data obtained by Helfrich *et al.* [89] and our results.

4.4.1.4 Comparison with the parameters of articles used as reference

Before starting our simulations, we compare our force field parameters and partial charges for DMF and BGL with those reported in the article of Helfrich *et al.*[88].

For what concern atom type, we could see some differences. C4 and C5 atom (see Figure 4.56) of DMF are defined as CT atom, i.e. as sp^3 aliphatic, while in reference article the atom type is C3, a carbon with 3 hydrogens. For what concern BGL, various atom are defined in distinct ways. The carbon atoms belonging to the chain (namely C5, C6 and C8) are defined as CT atoms in our simulation, while Helfrich *et al.* assign to them the C2 atom type, typical of a carbon with two hydrogens. The O15 oxygen is specified as a O2 oxygen (carboxyl and phosphate group oxygen), while in the cited article it is simply defined as O (carbonyl oxygen). Finally, a CA carbon atom type, typical of sp^2 pure aromatic (benzene) carbon, is assigned to C9, C10, C11, C12, C13 carbons, along with C4, while Helfrich *et al.* defined as CA only the C4 carbons and the other five as CD. In this case, however, CA stands for an aromatic carbon in a six-membered ring with one substitute, while CD for the same carbon belonging to a six-membered ring but with only an hydrogen bounded in this case. However, atom type are pretty the same for both simulations.

Bonds, angles and dihedral angles force field parameters specified in the reference article [88] are almost the same as the ones used for our simulations. Only Lennard–Jones parameters slightly differ.

Finally, for what concern charges, Table 4.6 and Table 4.7 sum up differences between atomic charges used in our simulations (see Table 4.13 and Table 4.12) and those of the article by Helfrich *et al.* [88]. This disagreement is due to the fact that we computed them with quantum chemistry calculations as previously mentioned, not with the charge equilibrium algorithm¹ [110] used by Helfrich group [88].

¹Charge equilibration method predicts charge distributions in molecules to be used in molecular dynamics simulations. The input data are experimental atomic ionization potentials, electron affinities and atomic radii. An atomic chemical potential is

Atom	Charges used		
	in the reference article [88]	our simulation	
		summing hydrogens	ignoring hydrogens
N1	-0.39	0.0570	0.0570
C2	0.51	0.3153	0.2790
O3	-0.48	-0.4496	-0.4496
C4	0.24	0.0386	-0.30875
C5	0.12	0.0386	-0.30875

Table 4.6: DMF: comparison between charges used in Helfrich *et al.* [89] and those of our simulations. Note that atom names reported here are the one used in our simulations. Charges written in the column called “summing hydrogens” are computed summing the charges of hydrogens linked to them.

constructed by using these quantities plus shielded electrostatic interactions between charges. Requiring equal chemical potentials, this method leads to equilibrium charges that depend upon geometry.

Atom	Charges used		
	in the reference article [88]	our simulation	
		summing hydrogens	ignoring hydrogens
N1	-0.498	-0.8887	-0.8887
H25	0.309	0.3535	0.3535
C2	0.177	0.2371	0.2573
C5	0.067	0.1681	0.1343
C6	0.007	-0.2848	-0.5712
C3	0.613	0.7977	0.7977
O15	-0.475	-0.5177	-0.5177
O14	-0.536	-0.3518	-0.3518
C8	0.193	0.1672	0.0061
C4	0.082	0.1596	0.1596
C10	0.019	-0.0604	-0.1953
C12	0.011	0.0232	-0.1013
C13	0.015	-0.0320	-0.1614
C11	-0.004	0.0429	-0.0816
C9	0.081	-0.1101	-0.2549
C7	0.541	0.5891	0.5891
O16	-0.575	-0.4821	-0.4821

Table 4.7: BGL: comparison between charges used in Helfrich *et al.* [89] and those for our simulations. Note that atom names reported here are the ones used in our simulations. Charges written in the column called “summing hydrogens” are computed summing the charges of hydrogens linked to them.

4.4.1.5 Conclusions

Results are summarize below:

- DMF molecules behave as an isotropic solvent;
- DMF arrange around BGL molecules in two solvation shells;
- $g(r)$ trends are more structured for the perpendicular system;
- $g(z)$ shows initial wave-trend typical for helicoidal systems;
- aspect ratio distribution is wider due to helix distortion;
- P_2 order parameter values are typical for isotropic phase, since the correct parameter would be an helicoidal one;
- R_{02}^2 introduces a molecular uniaxial behavior, despite high uncertainties;
- time correlation functions do not change, even if helices do distort;
- isotropic solvent presents a faster diffusion;
- solvation spheres for both systems are similar, since only DMF molecules placed outside the interaction section of helices are taken into account due to the small distance between centers of mass;
- atom-atom distance distribution is broadened and the peak shifts to underline helices distortion.

We also analyze snapshots (Figure 4.90) of last configurations of systems under study, where it is possible to see the effect of the periodic boundary conditions. From such images, it is clear that helices get more distant during the simulations.

In conclusion, since helices are too near, they undergo deformation. Another possible problem could be the shortness of such molecules which could make the system loose an additional stability. Hence, a bigger sample is needed. A possible better system to set up simulations with, could

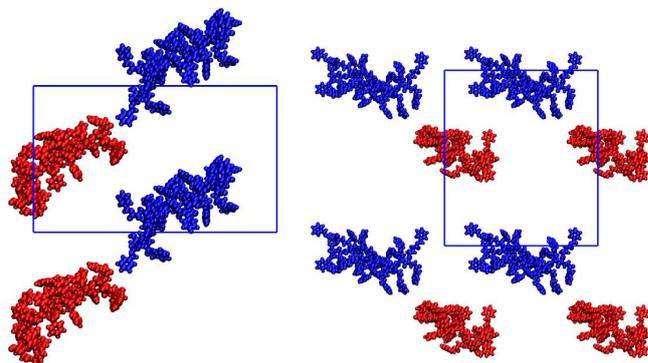


Figure 4.25: Snapshots of the effect of periodic boundary condition: on the left, an image of parallel system after a simulation of 40 ns, on the right a snapshot of the perpendicular system after 40 ns.

be composed by four helices of 36 monomers each, keeping them more distant than the ones of the samples just explained, or an unique helix of 144 residues. We analyze literature and decide to choose an average value of concentration: our new sample could be made up by 144 molecules of PBLG, 1456 DMF and 203 molecules of a suitable solute. In order not to let the helices deform and move too fast (high diffusion coefficient), the terminal atoms of the backbone should be fixed.

4.4.2 Parametrization and simulation of a system composed by a 144-residues long PBLG and heptyl butyrate in DMF

We analyze more than one hundred compounds with a chiral index (see Appendix: “Chiral index analysis for various solutes”) and choose the heptyl butyrate (quite high chiral index and low steric hindrance) as a pro-chiral solute that should interact differently with PBLG helices. We add thousands of this molecule to the PBLG solvated in DMF and study such system.

4.4.2.1 Setting up simulation

Molecules configuration file

For what concern solute (heptyl butyrate) and solvent (DMF), .pdb files were created simply drawing the chemical formula with Pymol [98] than optimizing it with an Universal Force Field (UFF) [100], using Avogadro [99] (see Figure 4.26).

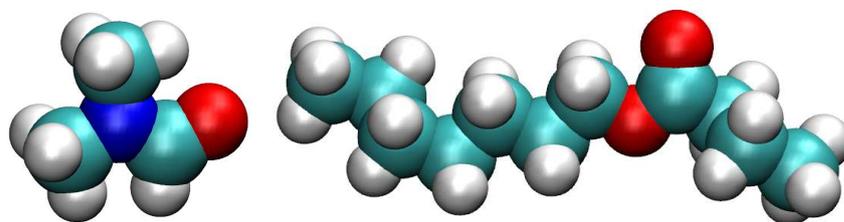


Figure 4.26: Solvent and solute snapshots. Structure of DMF (*right*) and of heptyl butyrate (*left*) are shown.

On the other hand, the first residue of benzyl L-glutamate was created and optimized (with the Universal Force Field, UFF [100]) using ISIS/Draw [103], Discovery Studio Visualizer [104] and Argus Lab [105]. Then, atom names were changed according to topology file. This last modified .pdb file is duplicated using “pdb_translation”, which performs a translation along z -axis of 1.5\AA and a rotation around the z -axis of the laboratory frame of

100°. Then, using Avogadro [99] all residues are put in a unique file, missing bonds between them are created and the resulting structure is optimized with UFF [100]. The resulting helix is shown in Figure 4.27.

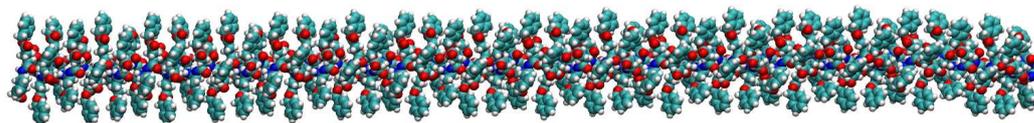


Figure 4.27: Helix composed by 144 residues of L-benzyl glutamate.

To verify that PBLG molecule built has a correct α helix structure, a Ramachandran plot has been computed. As shown in Figure 4.28, the molecule is clearly characterized by an α -helix structure.

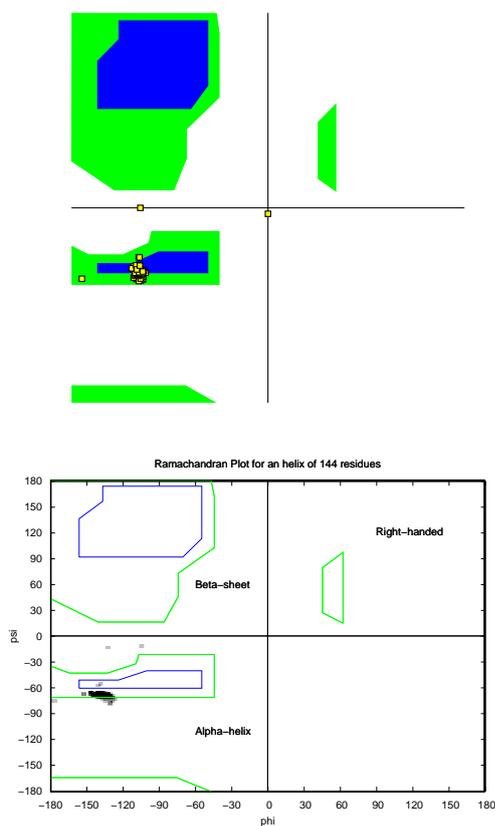


Figure 4.28: Ramachandran plot of the helix composed by 144 residues of L-benzyl glutamate: computed with VMD [106] (*left*) and using Outside (*right*).

Concentration

The choice of the most promising concentration is quite important in this study. To choose a good value, various articles are analyzed. Using densities, molecular weights, concentrations themselves and the Avogadro constant, number of molecules needed to achieve such concentrations are computed. Table 4.8 sums up molecules numbers for each articles to achieve those concentrations and averages of such values (on all results and excluding ibuprofen and *cis*-decalin concentrations, which are characterized by a huge amount of DMF molecules).

Solute	Number of molecules of			Reference
	solute	PBLG	DMF	
various chiral compounds	188	174	4120	[83]
ibuprofen	438	3300	56100	[111]
bza	278	275	1760	[87]
IPC	130	137		[86]
<i>cis</i> -decalin	871	2200	29500	[112]
solute 3	338	275	1760	[81]
solute 4	654	275	1760	[81]
solute 5	1310	297	2007	[81]
solute 6	182	275	2480	[81]
average	479	812	2314	
average without largest values	479	275	1760	

Table 4.8: Concentration of PBLG used in various articles. Note that in the article of Meddour *et al.* [83] various organic solvent are used.

The helix should be long enough not to undergo distortion, which is the reason why we decide to build a chain of 144 benzyl glutamate (BGL) molecules. Hence, 203 molecules of heptyl butyrate and 1456 molecules of DMF have to be added.

Initial sample

Once we obtained coordinates and parameters for these three molecules, we are to set up the system, using “Packmol” [113] and “Psfgen” [114]. Note that to create the psf file, topology file has been modified in order to contain description for each molecule and for last and final residue of PBLG chain: however, it has not been possible to discriminate these two peculiar residues for what concern charges, since they will not be recognized as part of the same chain. So, all atoms belonging to the chain will have the same charge in each residue, even if it should be a little different for the extreme ones. Initial sample is showed in Figure 4.29.

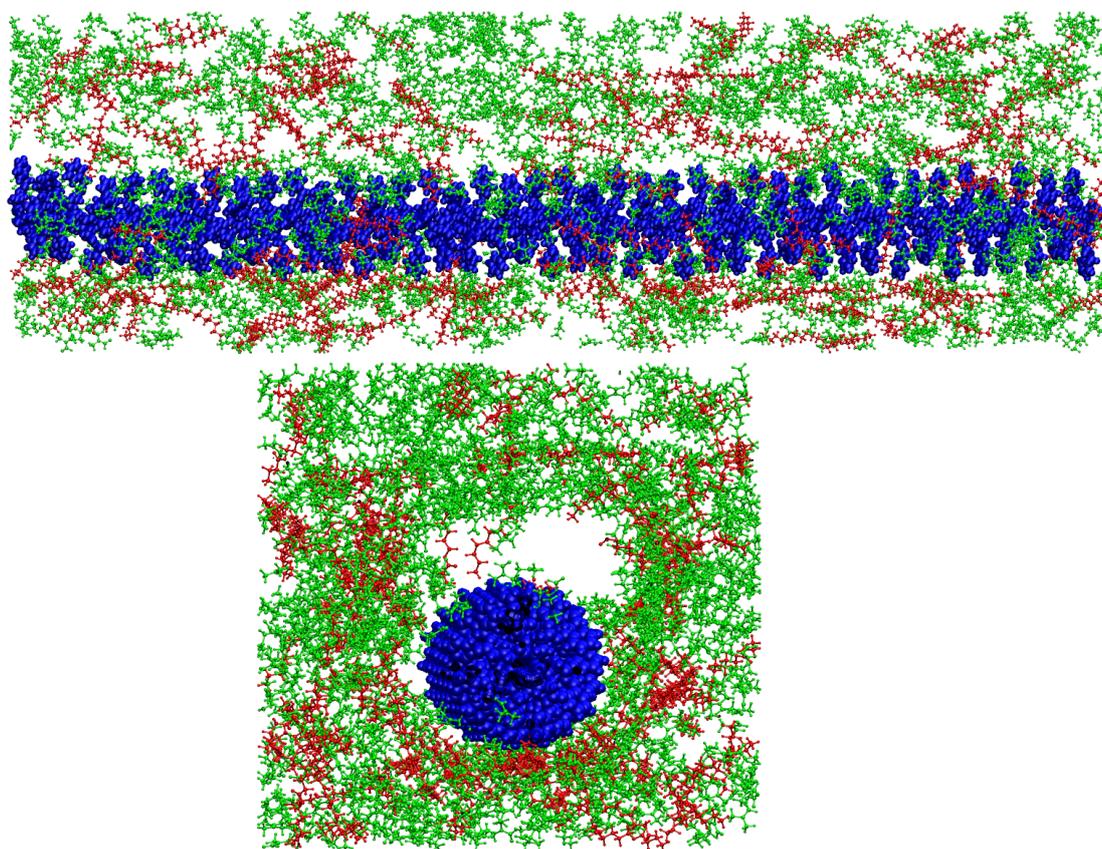


Figure 4.29: A snapshot of the initial sample, created with “Packmol” [113] and “Psfgen” [114]: green molecules are DMF, red ones HEP (heptyl butyrate) and blue ones are PBLG.

4.4.2.2 Running simulation

Minimization

After adding parameters for the last and the first residues, a simulation is run with NAMD [115], keeping atoms belonging to the backbone fixed. It will immediately stop due to “atoms moving too fast”. So a minimization is needed.

A snapshot of the resultant configuration is showed in Figure 4.30.

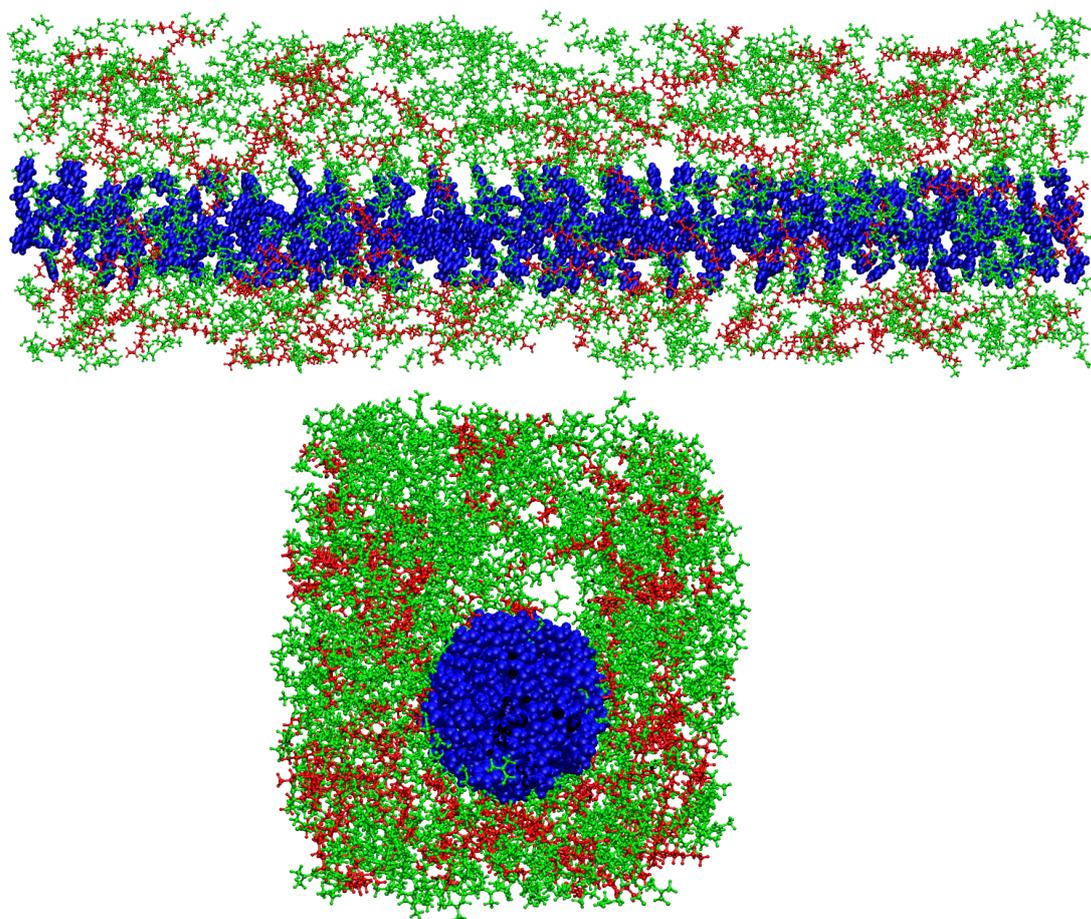


Figure 4.30: A snapshot of the sample, after a minimization done with NAMD [115]: green molecules are DMF, red ones HEP (heptyl butyrate) and blue ones are PBLG.

Increasing temperature

Next, temperature is risen step by step (first taking it to 5K, 10K, 25K, 50K, 100K, 120K, 140K, 160K, 180K, 200K, 220K, 250K) since otherwise the simulation stops due to the high motion of atoms. At $T=250\text{K}$, as showed in Figure 4.31, a cluster of molecules and empty spaces appears. To make them disappear, pressure has been increased.

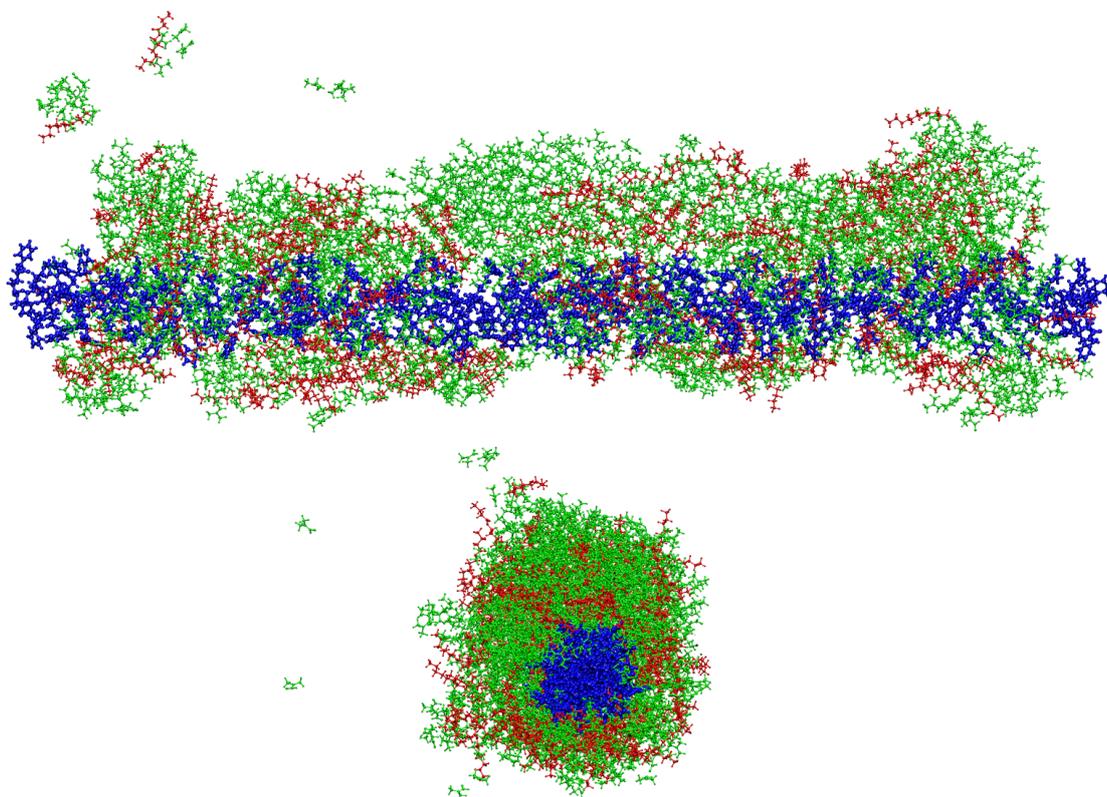


Figure 4.31: A snapshot of the sample, after having rescaled temperature up to $T=250\text{K}$ with NAMD [115]: green molecules are DMF, red ones HEP (heptyl butyrate) and blue ones are PBLG.

Compression

Pressure is risen in order to decrease the box sides and make the empty space disappears. The box actually becomes smaller, but the cluster still survives. Moreover, this high pressure (100.0 bar) makes the free-to-move molecules, i.e. DMF and heptyl butyrate, to gather towards to the center of the helix, which could not crush itself due to fixed backbone atoms (Figure 4.32).

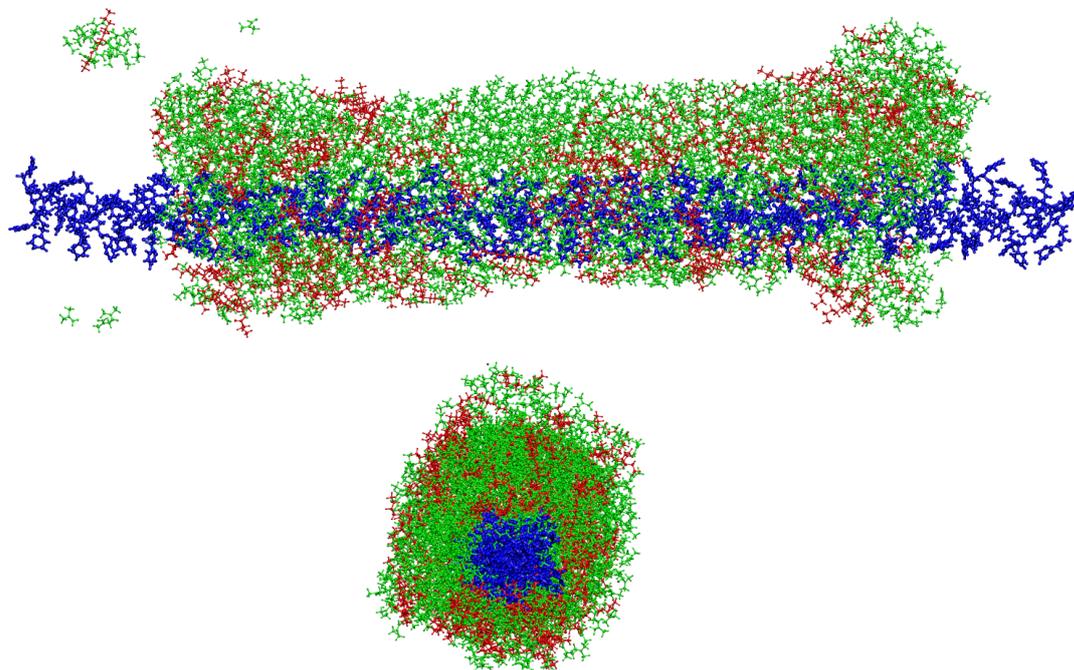


Figure 4.32: A snapshot of the sample, after a compression run: green molecules are DMF, red ones HEP (heptyl butyrate) and blue ones are PBLG.

NVT run

To let molecules occupy the whole sample, an NVT simulation is set up. Unfortunately, the resultant configuration shows that DMF and HEP are even more compressed towards the center of the helix than before (Figure 4.33).

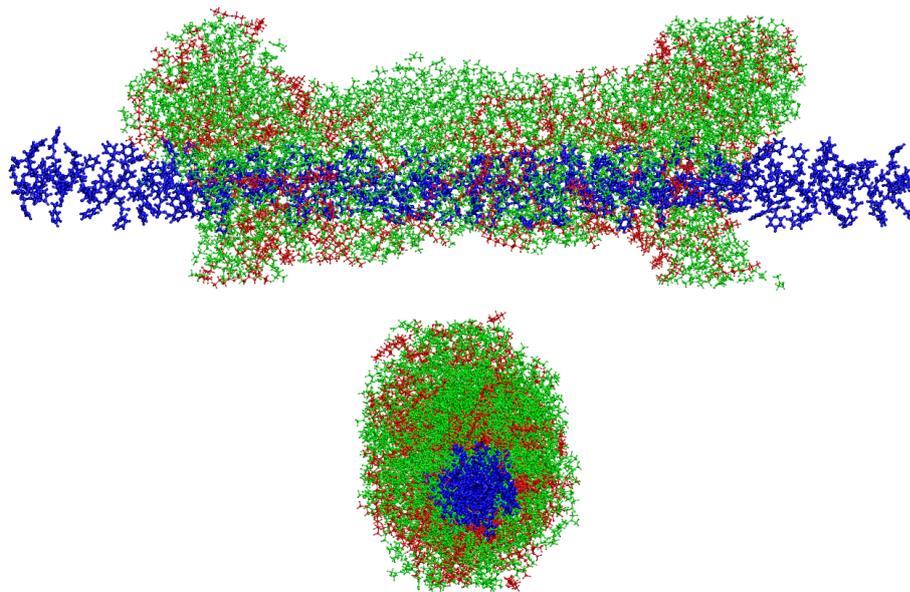


Figure 4.33: A snapshot of the sample, after a NVT run: green molecules are DMF, red ones HEP and blue ones are PBLG.

Since increasing pressure and relaxing the sample (using the command “useFlexibleCell yes” in order to let the box sides change freely) have not been useful for creating an uniform distribution of molecules, the sample has been enlarged, in order to avoid problems due to vacuum and small cell sides.

4.4.2.3 Increasing initial sample

The new sample is composed by the same α_{144} helix of PBLG but the number of DMF and HEP molecules is doubled (respectively 2912 and 406). Figure 4.34 shows a snapshot of this new system.

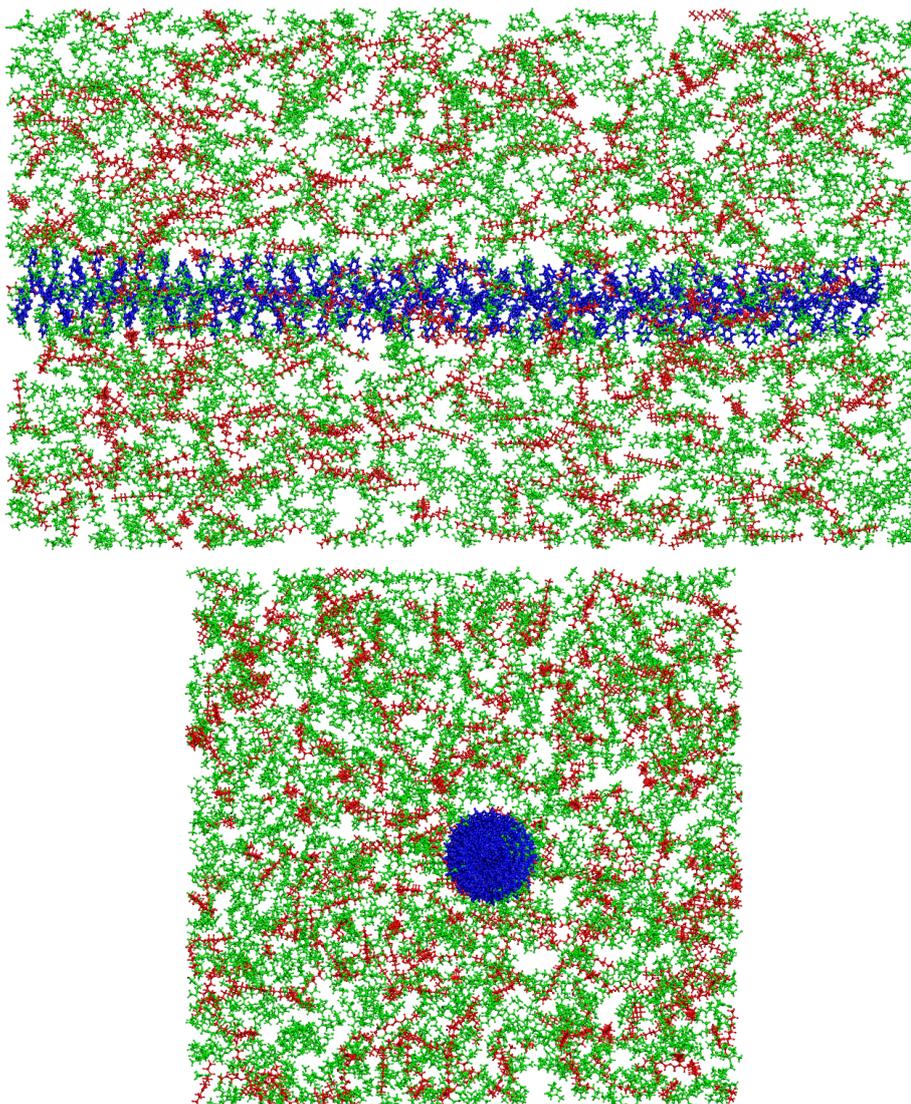


Figure 4.34: A snapshot of the new initial sample: green molecules are DMF, red ones HEP and blue ones are PBLG.

4.4.2.4 Results

Finally, we succeed in resizing the simulation box and the simulation run properly.

The following table Table 4.9 sums up average energy, temperature, volume and density values.

Also other parameters are analyzed:

- radial distribution function $g(r)$ and density one $g(z)$ for BGL-BGL, BGL-DMF, BGL-HEP interactions;
- aspect ratio histograms;
- orientational order parameters (first rank order parameter $\langle P_2 \rangle$ and molecular biaxiality $\langle R_{02}^2 \rangle$) for BGL;
- time correlation functions;
- displacement analysis;
- solvation sphere.

Parameter	(Units)	Value
Bond energy	(kcal/mol)	12320.362930
Angle energy	(kcal/mol)	18725.138506
Dihedral energy	(kcal/mol)	4594.978736
Electrostatic energy	(kcal/mol)	-29141.035627
Van der Waals energy	(kcal/mol)	-21096.462478
Total energy	(kcal/mol)	22265.724045
Temperature	(K)	249.709765
Volume	(\AA^3)	-39.674633
Density	(g/cm ³)	776764.799500

Table 4.9: Results of a simulation 9.54 ns long.

Radial distribution functions and density distribution functions

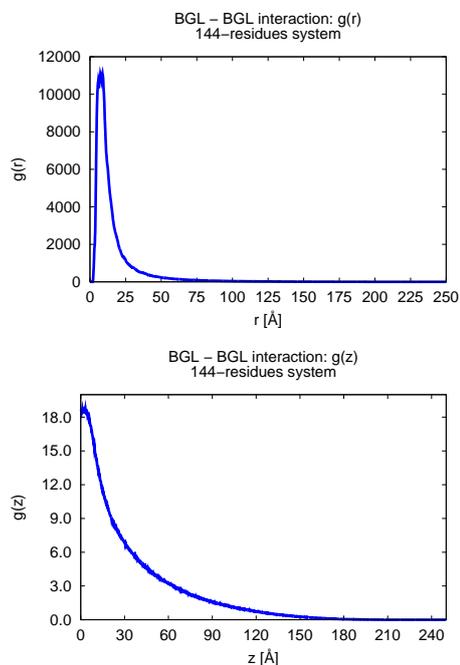


Figure 4.35: $g(r)$ and $g(z)$ computed with respect to two BGL molecules.

Since each molecule belonging to the helix has neighbor molecules placed at the same distance all along the polymer, the $g(r)$ shows a peak at about 7 Å, while the $g(z)$ shows a trend typical of such well-structured shape (Figure 4.35).

DMF molecules solvate uniformly the helix up to $\sim 40\text{Å}$ away from it (Figure 4.36).

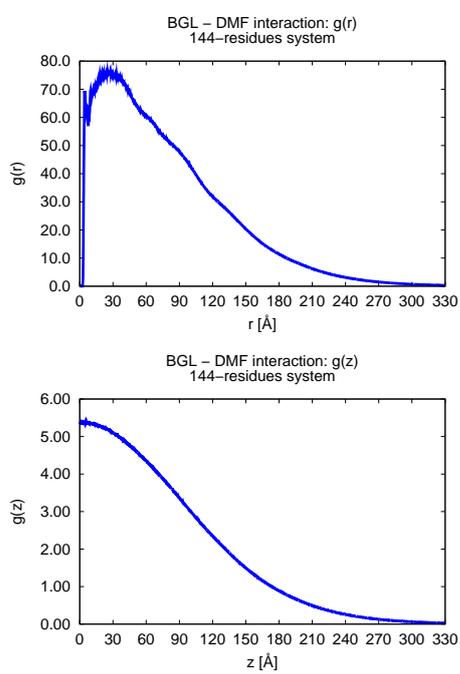


Figure 4.36: $g(r)$ and $g(z)$ for DMF molecules with respect to BGL helix.

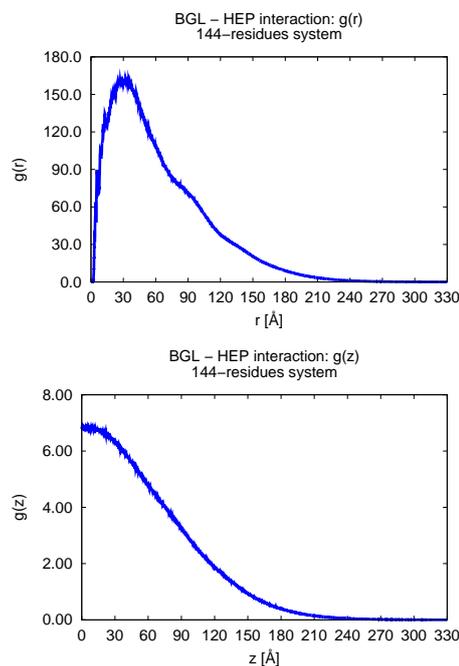


Figure 4.37: $g(r)$ and $g(z)$ for HEP molecules with respect to BGL helix.

Also HEP molecules solvate almost uniformly the helix (up to $\sim 35\text{\AA}$), with an increasing possibility of finding molecules within 5\AA from BGL molecules rather than DMF (Figure 4.37).

From trend of radial distribution of HEP molecules with respect to the helical axis along the x -direction (Figure 4.38), two peaks seems to characterized a bimodal distribution. Solute molecules could be oriented in two different ways which could be the reason of enantiomeric separation.

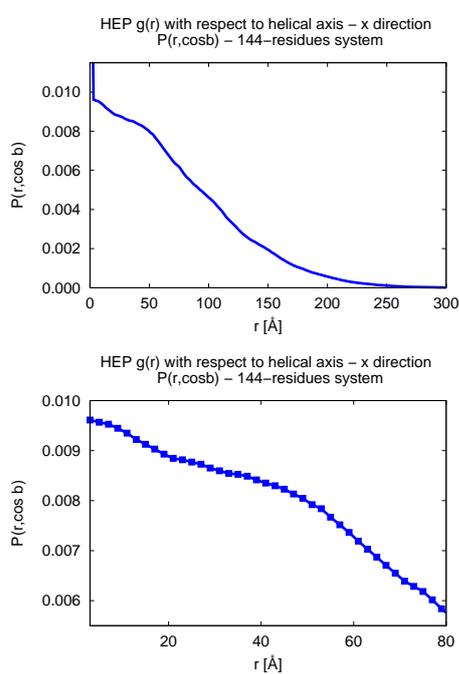


Figure 4.38: $g(r)$ for HEP molecules with respect to helical axis in the x -direction: overview (*left*) and zoom (*right*).

Aspect ratio

BGL molecules are only constrained by their backbone, not completely fixed so that lateral strands are now free to move. Also heptyl butyrate and dimethylformamide molecules change their conformation during our simulations, maybe to their interactions with helix. HEP molecules are those that undergo the biggest changing in shape, maybe due to a strongest interactions with helix.

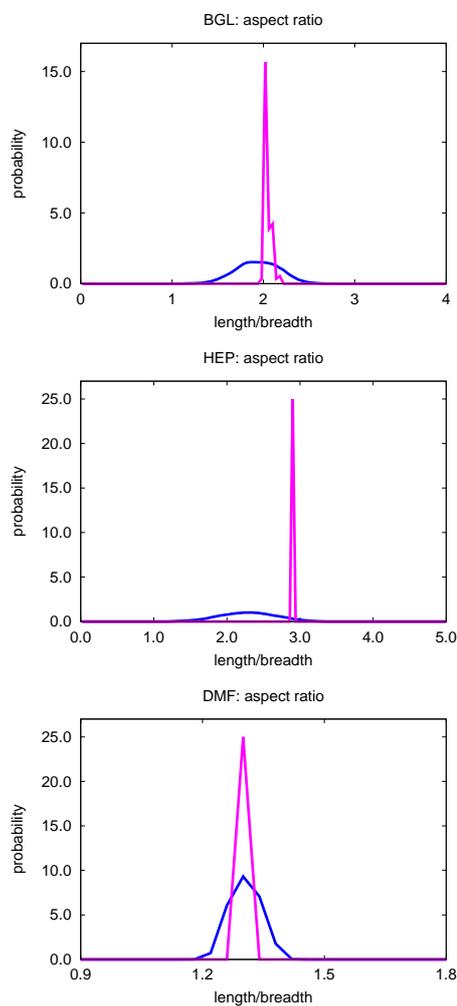


Figure 4.39: Aspect ratio for BGL and HEP.

Order parameters

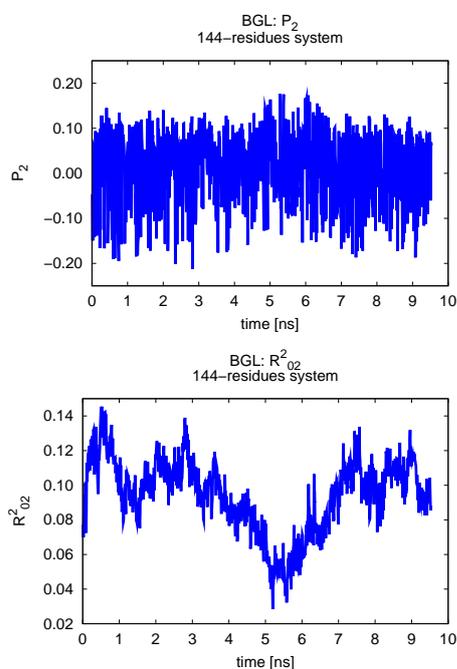


Figure 4.40: Order parameters computed with respect to BGL.

As previously discussed, $\langle P_2 \rangle$ values of the helix tends to zero Figure 4.40, since such order parameter does not take into account helicoidal ordering. Also HEP molecules are not aligned in an unique direction ($\langle P_2 \rangle = 0$). They do not show molecular biaxiality.

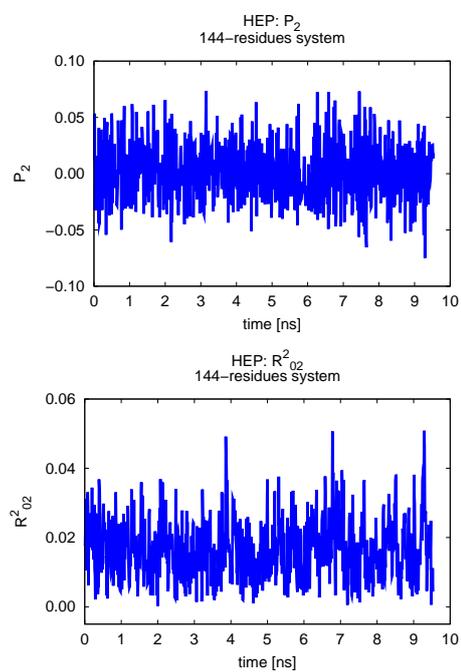


Figure 4.41: Order parameters computed with respect to HEP.

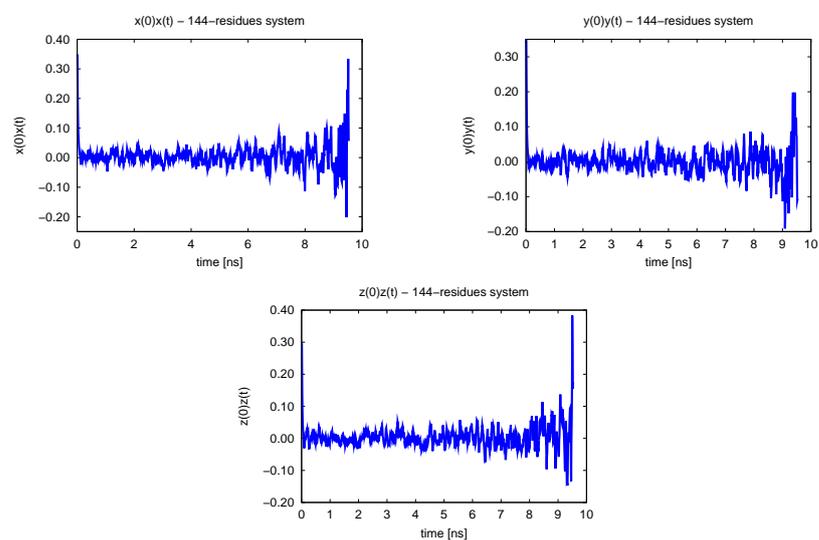
Time correlation functions

Figure 4.42: Time correlation function: $x(t)x(t)$, $y(t)y(t)$ and $z(0)(t)$.

Time correlation functions do not change actually: conformational changes do not break completely the structure of the helix and of the system.

Displacement analysis

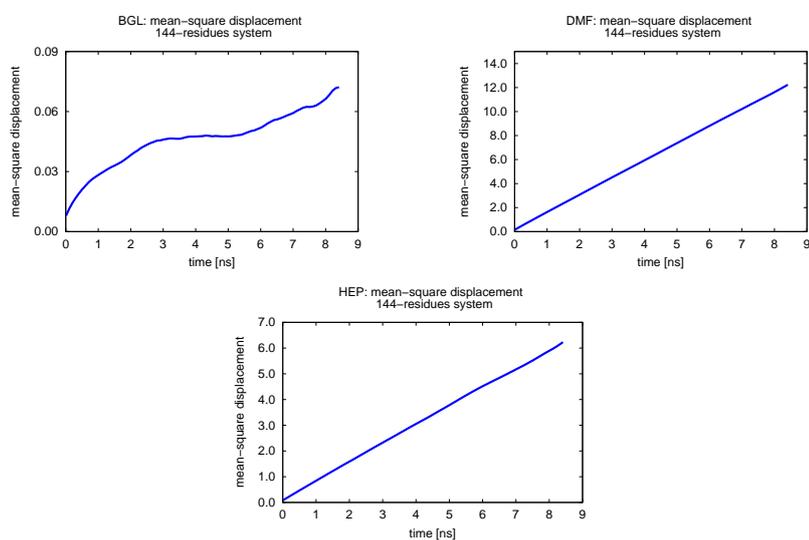


Figure 4.43: Displacement analysis.

The DMF solvent presents a faster isotropic diffusion with respect to both HEP molecules and the helix.

Solvation sphere

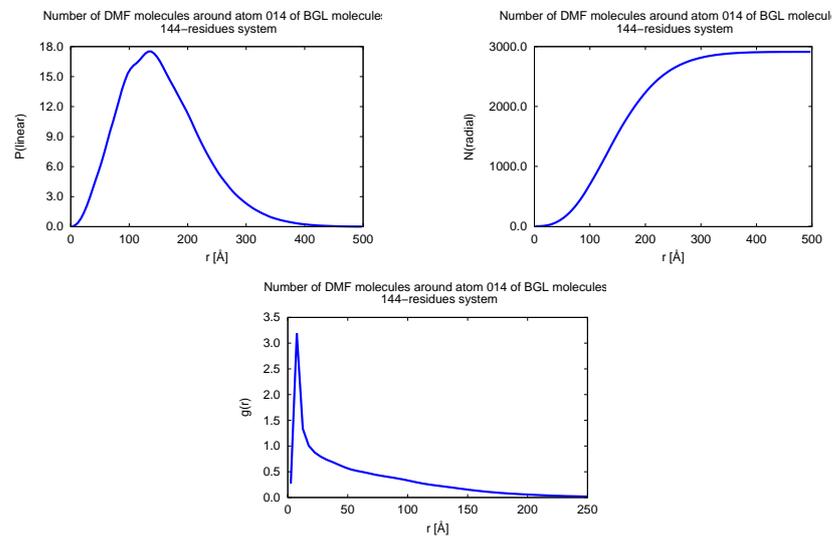


Figure 4.44: Solvation sphere for DMF.

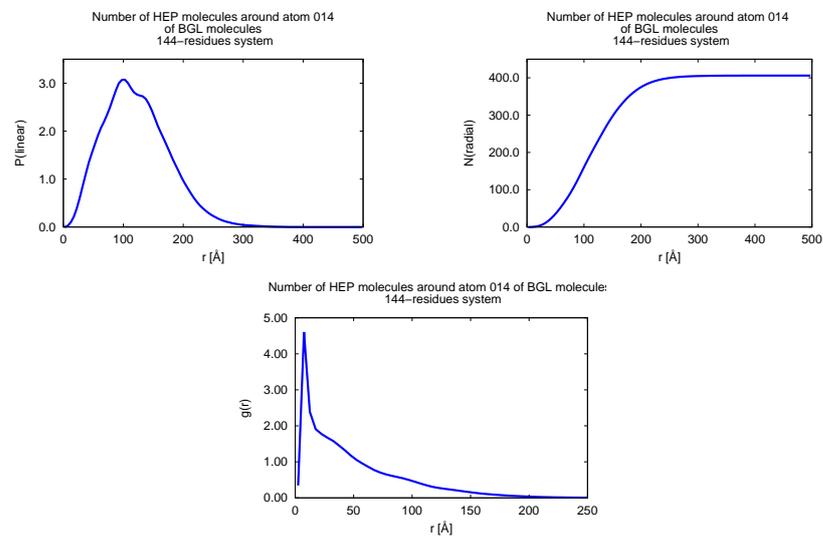


Figure 4.45: Solvation sphere for HEP.

Graphs show that the first peak of solvation, considering both HEP and DMF, is placed at about 7Å from PBLG helix.

Dipolar couplings

Dipolar coupling for hydrogen and for carbon atoms of HEP residue are computed. Data are not reliable due to the fact that there is no average orientation of solute molecule with respect to the z -axis ($\langle P_2 \rangle$ almost 0).

RMSD for H-bonds

From root mean square deviation is possible to see that, after a first equilibration region, a structure different from the initial one becomes stable.

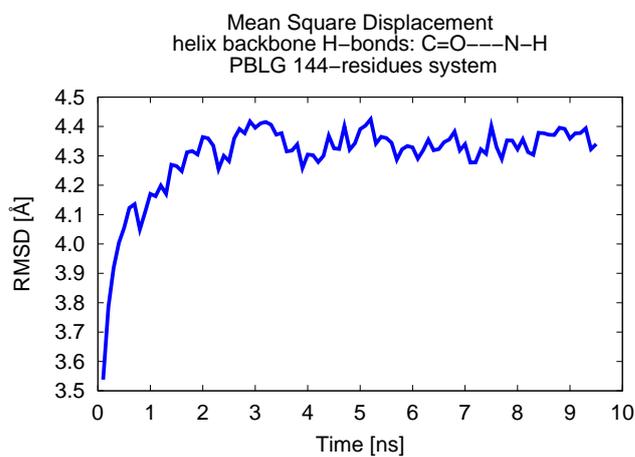


Figure 4.46: Root mean square displacements computed between H-bonds of the helix backbone.

Computation of order parameter in regions

In this section, $\langle P_2 \rangle$ is computed in regions, obtained by cutting slices of the sample along the z -axis. From the following graph Figure 4.47, we could see that molecules are found in interesting quantity only in the middle of the sample (considering the z direction).

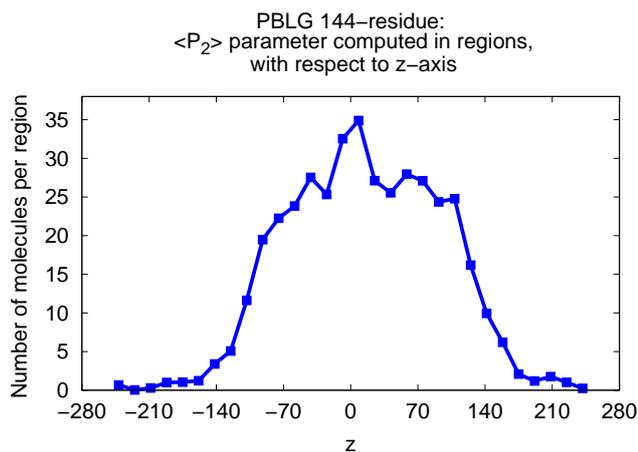


Figure 4.47: Number of molecules found along the z -axis.

In the populated region of the z -axis, the order parameter is constant Figure 4.48. Hence, HEP molecules in the center of the sample are placed randomly, not only with respect to the z -axis (along which the helix grown) but also with respect to solute molecules themselves.

The last figure Figure 4.49 shows that there is not a preferential order along none of the axes: all lines show average behavior.

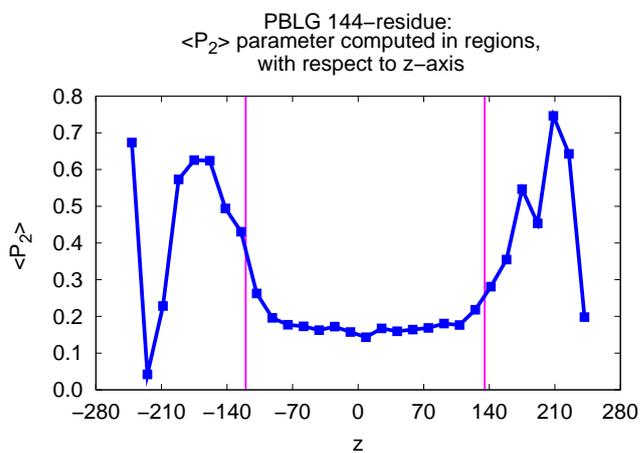


Figure 4.48: $\langle P_2 \rangle$ along the z -axis.

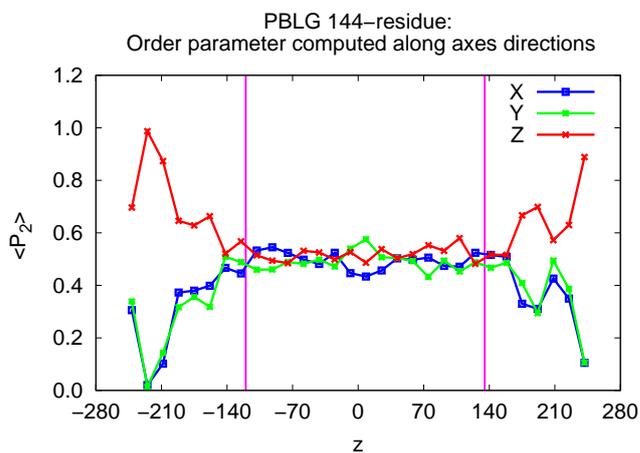


Figure 4.49: Order parameter computed along the axes.

Analysis with a chirality index

During the simulation, the chiral index [85] average value (explained in “Appendix: analysis with a chirality index for protein secondary structures investigation”) does not change basically. An average value of 0.04 is not one of those values characterizing various possible structures, but is almost the same as that of the initial configuration, suggesting that the first structure is almost maintained.

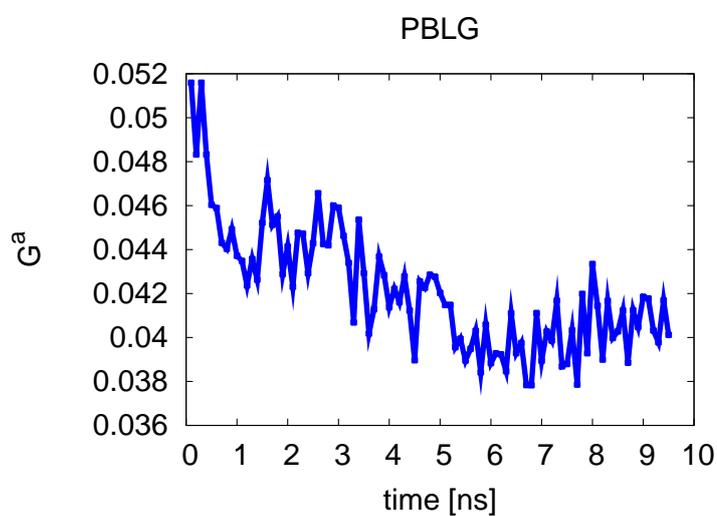


Figure 4.50: Trend of chiral index [85] during the simulation.

Ramachandran plot

Figure 4.51 shows that the structure is not well maintained during the simulation. Peaks are no more in the “ α helix” quadrant like in the initial sample (Figure 4.28). During the simulation the perfect helix is never regained, since points never go back in the middle of any defined area.

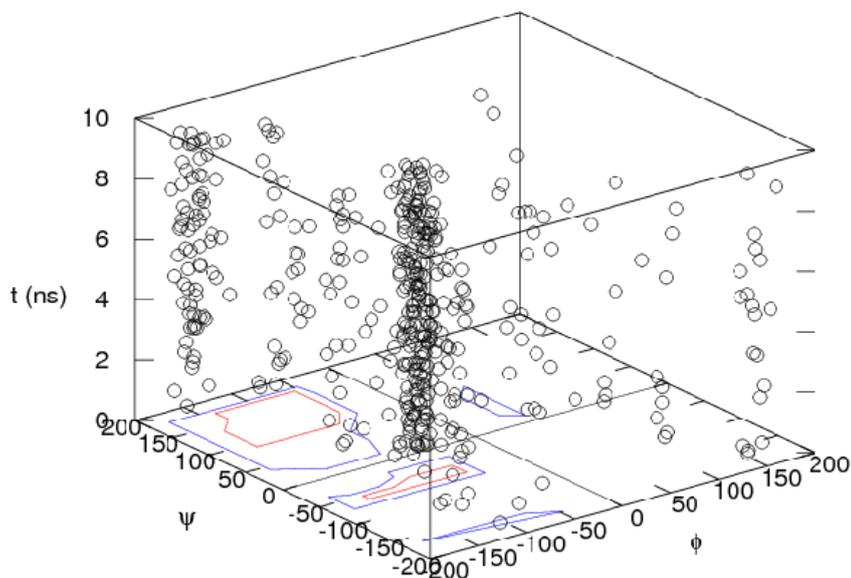


Figure 4.51: Ramachandran plot evolution during simulation time.

Analysis with helicoidal order parameters

A detailed description of various types of helicoidal order parameters, Appendix “Helicoidal order parameter analysis”. Following figures do not take into account the initial configuration.

PBLG: lateral chain helicoidal parameter

The lateral chain helicoidal parameter during the simulation tend to zero. This is due to the fact that PBLG is constrained only in the backbone structure, while lateral chains are left free to move.

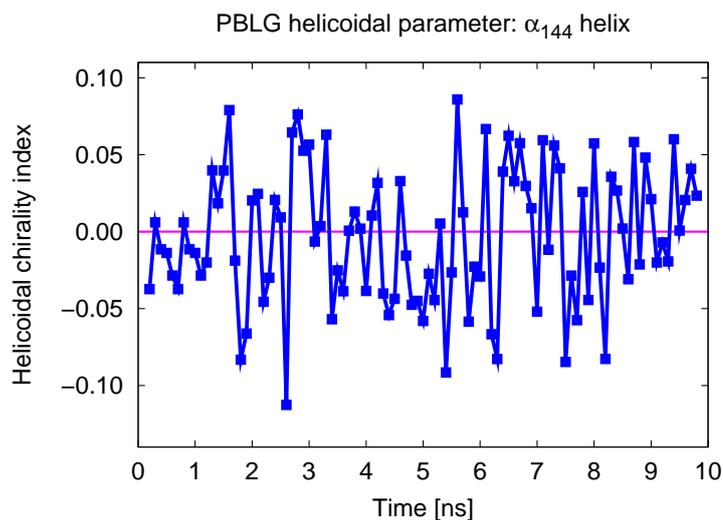


Figure 4.52: Trend of helicoidal parameter taking into account lateral chain.

PBLG: backbone helicoidal parameter

This backbone helicoidal parameter underline once again the loss of perfect α -helix structure of PBLG chain, in accordance with results shown in Figure 4.51 for what concern Ramachandran plot time evolution.

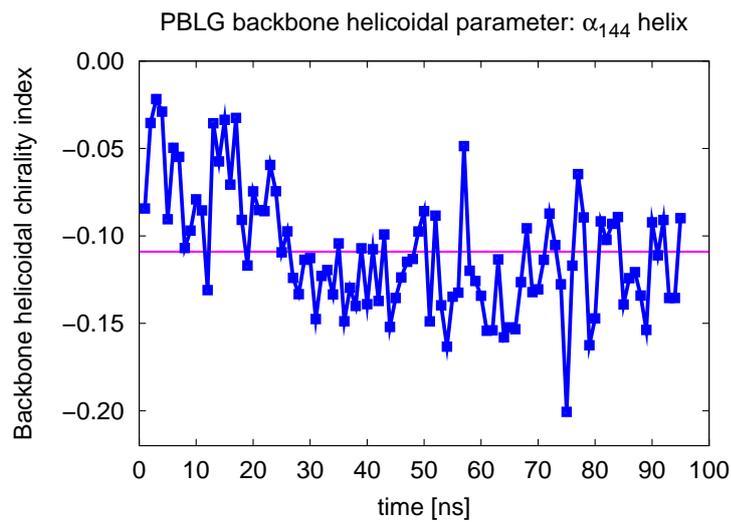


Figure 4.53: Trend of backbone helicoidal parameter.

HEP: solute-helix helicoidal parameter

This time, HEP helicoidal parameter average value similar to zero means that there are only few solute molecules which interacts with the helix and that they are oriented randomly (Figure 4.54).

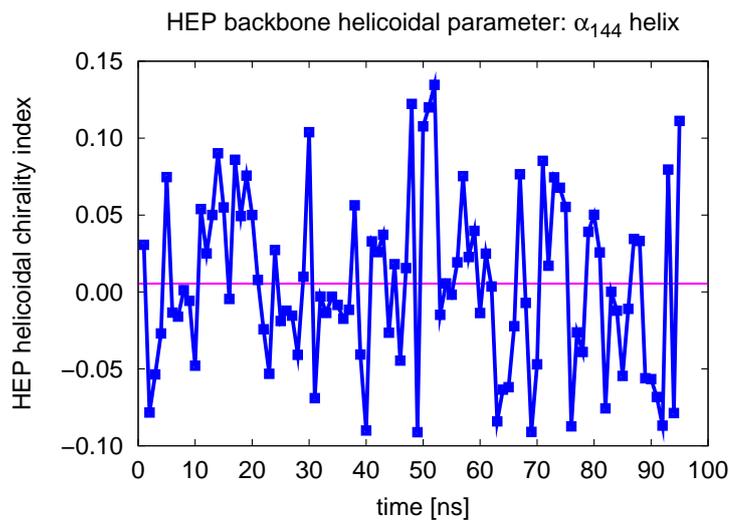


Figure 4.54: Trend of HEP–PBLG helicoidal parameter.

However, a deeper analysis of results of such parameter is attempted. First of all we sup up results obtained for the last configuration (Table 4.10). Next, we tried to find out if there were characteristic trends depending on HEP molecules orientations.

It seems that molecules enclosing the helix with the shorter part of the chain show an HEP helicoidal parameter average value in the interval $[-0.55-0.2]$, while those getting close to PBLG with the longer one shows positive values comprised between 0.2 and 0.55.

Docking analysis

Docking analysis were performed o both DMF and HEP molecules using ArgusLab [105]. Solvent molecule shows an minimum docking energy of

-4.15046 kcal/mol, while solute between -3.26642 kcal/mol (considering a flexible ligand) and -4.40980 kcal/mol (for a rigid one). For a more detailed treatment of this topic, see the relative section.

Residue number	HEP helicoidal parameter value	Type of orientation	Part of chain nearer to the helix	Distance
3057	0.24296	planar		far
3058	0.38984	planar		
3096	0.58845	planar		far
3098	-0.59003	zig-zag folded		far
3106	0.23700	folded	long	
3109	0.07385	planar		far
3169	0.35576	folded	long	far
3173	0.37277	folded	long	
3195	0.04133	zig-zag folded		almost inside
3215	0.54789	planar		far
3226	0.45395	folded	long	
3230	0.82117	folded	short	inside
3249	-0.91225	folded	short	far
3251	-0.82196	planar	long	
3282	-0.34246	folded	short	
3285	0.26106	folded	long	
3291	-0.46496	planar	short	
3309	-0.18256	folded	long	near
3343	0.13640	folded	short	
3361	0.97481	folded	middle	near
3362	-0.18273	planar	short	
3378	0.16283	folded	long	far
3384	0.49891	folded	long	far
3399	-0.72141	folded	long	far
3406	0.52033	planar	long	far
3416	-0.15325	planar	short	
3423	-0.91432	folded	long	far
3428	-0.46972	folded	short	
3445	0.16177	folded	long	far
3460	0.88503	planar	long	

Table 4.10: Values of HEP helicoidal parameter for each residue within 10 Å of PBLG. Also the type of orientation and distance with respect to the helix are shown.

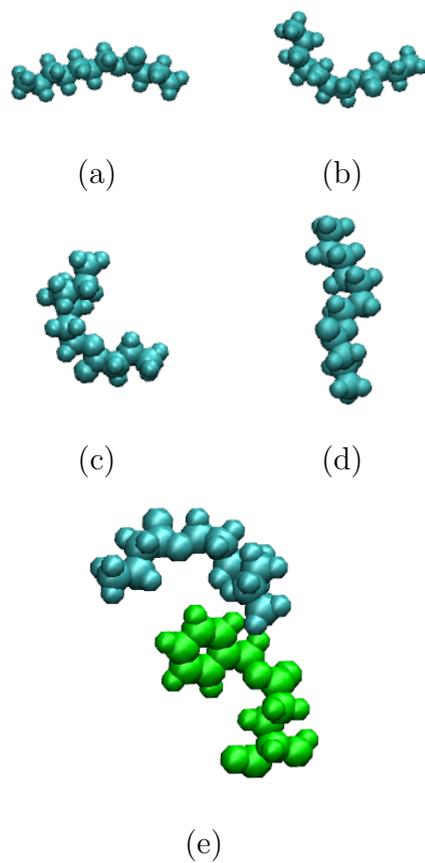


Table 4.11: Orientation of HEP molecules: (a) planar, (b) folded enclosing the helix with the long part of the C chain, (c) folded enclosing the helix with the short part of the C chain, (d) zig-zag and (e) middle. Note that the molecule showing a configuration named “middle” is forming H-bond with the aromatic ring.

4.4.2.5 Conclusions

Results are summarized below:

- DMF and HEP seem to solvate uniformly the PBLG helix, since there is only one solvation shell visible from the $g(r)$;
- HEP molecules may be found nearer the helix than DMF ones, which could mean depend on a higher affinity between these two residues;
- $g(r)$ and $g(z)$ of BGL molecules are typical of an helical structure;
- aspect ratio distributions are wide, but the HEP one is even wider;
- $\langle P_2 \rangle$ order parameter values are typical of an isotropic phase (an helicoidal parameter should be used);
- $\langle R_{02}^2 \rangle$ shows zero values so any biaxiality is introduced;
- time correlation functions are almost constant: conformational changes do not break completely the structure of the helix and of the system;
- DMF presents the faster isotropic diffusion, twice of HEP one;
- solvation sphere for both DMF and HEP are placed at about 7Å from the helix, but there are more probability to find an HEP than a DMF molecule in the nearby of the helix.

4.4.2.6 Future works

Starting from the initial sample, with the correct helicoidal structure, a simulation at room temperature, keeping fixed all helix atoms, is currently running. After the thermalization of this sample, we plan to keep fixed only backbone atoms and then to let them vibrate (using constrains), in order to see if the destroy of the perfect helix has influenced the interactions of HEP molecules with the helix itself.

Next, a sample made of two parallel helices of PBLG (36 residues long) in 3500 molecules of DMF is thermalized at room temperature. One molecule of solvent will be substituted with a solute (HEP, at first instance) and ABF runs would be carried on. We plan to change solutes and the distance between helices. The aim of using this method is that to figure out how solute molecules interact with PBLG and deepen the knowledge of the enantiomeric discrimination process.

Finally, docking analysis will be performed.

4.5 Conclusions

The simulation of short helices in DMF solvent shows that PBLG molecules undergo deformation. Hence, we set up another sample composed of a 144-residues long PBLG helix (18% w/w), 2912 molecules of the solvent DMF and 406 molecules of the chiral solute heptyl butyrate (HEP), corresponding 25% w/w. This molecule (HEP) has been chosen after having evaluated more than one hundred compounds with our chiral index. High chirality solutes should interact better with the helix itself, making the study of the mechanism of the discrimination easier to follow and understand. Results show that DMF and HEP solvate uniformly the PBLG helix, but molecules of the pro-chiral solute is found nearer to the helix with respect to the solvent molecules. It is worth noting that BGL molecules do not lose their helical structure during the simulation: conformational changes do not break completely the helical structure. The solvent presents the faster isotropic diffusion, twice of HEP, indicating a stronger interaction of the solute with the helix. The solvation spheres for both DMF and HEP are placed at about 7 Å from the helix, but there are more probability to find an HEP than a DMF molecule in the nearby of the helix.

We plan to measure residual dipolar (RDC) and quadrupolar couplings from MD simulation. This index could be used to derive an order parameter, giving information on the conformational distribution.

Finally, we also plan to do some Adaptive Biasing Force (ABF) runs [108], to follow the free energy profile of a solute approaching the helix, in order to detect and quantify differences for the different enantiomers. These runs will be carried using a sample composed of one helix of PBLG in DMF with only one molecule of solute. The profiles will be computed keeping fixed the position of the helix and PBLG will be set in the xy plane and the solute molecule will approach it along the z direction.

4.6 Appendices

4.6.1 CHARMM topology file

A force field topology file contains all information needed to convert a list of residue names into a complete .psf structure file. It also contains internal coordinates that allow the automatic assignment of coordinates to hydrogens and other atoms missing from a crystal .pdb file.

At the beginning of the file there is always an header which explains the version of CHARMM that generated the file, followed by a comment section.

The topology file must define the type, mass, and charge of every atom in every residue, so that a .psf file can be constructed. While the partial charges assigned to atoms of the same type vary between residues, their masses do not. Therefore, the mass of every atom type is declared once at the beginning of the file in a MASS statement. This statement also pairs an integer with each type name, which is used in CHARMM formatted .psf files. The type indexes are unique but not necessarily consecutive.

Atom types represent classes of chemical environments assigned to each atom in a force field calculation. The characteristics of an environment include hybridization, formal charge, and immediate bonded neighbors. For example, the AMBER atom type C represents an sp^2 carbonyl carbon, CT represents a tetrahedral carbon and CH is a united atom representation of an sp^3 carbon including one hydrogen. Each force field has a different set of atom types. In the case of benzyl glutamate (Figure 4.89), as described in Table 4.12, along with partial charges, atom types are: N = sp^2 nitrogen in amide groups, C = sp^2 C carbonyl group, CT = sp^3 aliphatic C, CA = sp^2 C pure aromatic (benzene), O = carbonyl group oxygen, O2 = carboxyl and phosphate group oxygen, OS = ether and ester oxygen, H = H bonded to nitrogen atom, H1 = H aliphatic bonded to C with 1 electron donor group, HC = H aliphatic bonded to C without electron donor groups, HA = H aromatic bonded to C without electron donor groups.

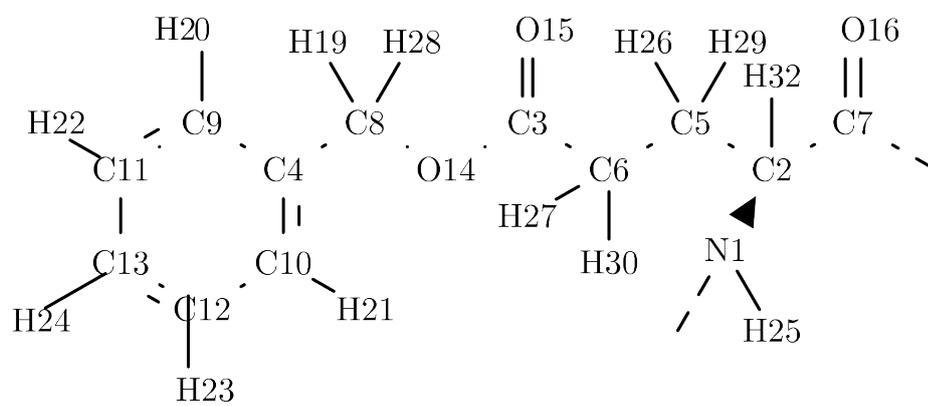


Figure 4.55: Benzyl glutamate: atom numbers.

Atom	Atom type	Partial charge
N1	N	-0.8887
C2	CT	0.2573
C3	C	0.7977
C4	CA	0.1596
C5	CT	0.1343
C6	CT	-0.5712
C7	C	0.5891
C8	CT	0.0061
C9	CA	-0.2549
C10	CA	-0.1953
C11	CA	-0.0816
C12	CA	-0.1013
C13	CA	-0.1614
O14	OS	-0.3518
O15	O2	-0.5177
O16	O	-0.4821
H19	HC	0.0838
H20	HA	0.1448
H21	HA	0.1349
H22	HA	0.1245
H23	HA	0.1216
H24	HA	0.1294
H25	H	0.3535
H26	HC	0.0298
H27	HC	0.1538
H28	HC	0.0773
H29	HC	0.0040
H30	HC	0.1326
H32	H1	-0.0202

Table 4.12: Atom types and partial atomic charges of benzyl glutamate. Note that charges for aromatic carbons and hydrogens are symmetrized (values are averaged in order to be equal for atoms with the same chemical environment). Dipole moment components, computed with Gaussian [101], are $M_x = 1.3476$, $M_y = -3.6094$, $M_z = 0.2243$.

Table 4.13 sums up atom types and partial charges for dimethylformamide (Figure 4.56). In this case, atom types are the following: N = sp^2 nitrogen in amide groups, C = sp^2 C carbonyl group, CT = sp^3 aliphatic C, O = carbonyl group oxygen, H1 = H aliphatic bonded to C with 1 electron donor group, HC = H aliphatic bonded to C without electron donor groups.

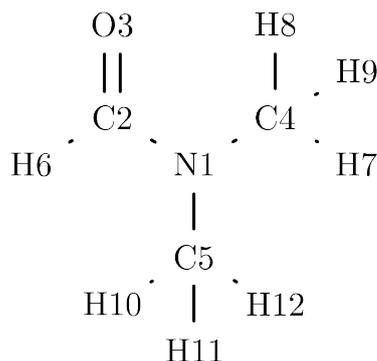


Figure 4.56: Dimethylformamide: atom numbers.

Topology file continues including a list of keywords. The first one is:

```
AUTOGENERATE ANGLES DIHEDRAL
```

which is the AUTOgenerate default options to be used when building a structure. ANGLEs specifies that all possible angles and DIHEdral specifies that all possible dihedral angles will be generated when building a structure. If these options are not included angles and/or dihedrals must be listed explicitly in the topology file. Next, there are the DECL keywords, which has to be added when specifying the connectivity of a chain of residues in a protein.

```
DECL -C2
DECL -C7
DECL -O16
DECL +N1
DECL +H25
DECL +C2
```

These declarations specify atoms covalently linked between previous and next residues. Following keyword is completely described by a single line:

```
DEFAULT FIRST NTER LAST CTER
```

Atom	Atom type	Partial charge
N1	N	0.0570
C2	C	0.2790
O3	O	-0.4496
C4	CT	-0.30875
C5	CT	-0.30875
H6	H1	0.0363
H7	HC	0.1158
H8	HC	0.1158
H9	HC	0.1158
H10	HC	0.1158
H11	HC	0.1158
H12	HC	0.1158

Table 4.13: Atom types and partial atomic charges of DMF. Note that charges for C4 and C5 carbons and for H7, H8, H9, H10, H11, H12 hydrogens are symmetrized. Dipole moment, computed with Gaussian [101], is $\mu = (0.0006, -0.4251, 3.7019)$.

Here the DEFAult patches used on FIRSt and LAST residues in a generated segment are specified. In this case, the NTERminus and CTERminus of a polypeptide chain are the last and first segment.

Finally, there is another keyword:

```
PATCH FIRST NBGL LAST CBGL
```

This line contains the PATChes to the FIRSt and LAST atoms of the residue. Note that all residues in the .pdb file have to be called simply BGL, even if in topology file there is a differentiation between the two terminal residues (NBGL and CBGL) and the central ones (BGL), otherwise they will not be recognized as belonging to the same molecule. The PATCH keyword is used to automatically find out which residue is the first one and which the last one and to modify them.

The actual residue definitions is now explained. A residue is indicated

by the RESI statement with the residue name (DMF in this case) and total charge (0.00). Next are listed all of the atoms in the residue in ATOM statements with the atom name (N1, C2, O3, O4), type (N, C, O, CT), and partial charge (0.0570, 0.2790, -0.4496, -0.30875). The GROUP statements, dividing the atoms into integer-charge groups, are not used by NAMD.

```
RESI DMF 0.00

*   O3   H8
*   ||   |
* H6--C2--N1--C4--H9
*       | |
*       | H7
*       |
*   H10--C5--H12
*       |
*       H11

group
atom N1  N   0.0570
atom C2  C   0.2790
atom O3  O  -0.4496
atom C4  CT -0.30875
atom C5  CT -0.30875
atom H6  H1  0.0363
atom H7  HC  0.1158
atom H8  HC  0.1158
atom H9  HC  0.1158
atom H10 HC  0.1158
atom H11 HC  0.1158
atom H12 HC  0.1158
```

The residue continues by defining connectivity, with each BOND statement followed by a list of atoms pairs to be connected with bonds. Observe that the atom C7 is bonded to +N1, the N of the following residue. A bond between N1 and -C7 will be provided by the preceding residue. The order of bonds, or of the atoms within a bond, is not significant.

```
bond C5 C2 C6 C5 C3 C6 014 C3
bond N1 H25 N1 C2 C7 C2 C6 H27
bond C2 H32 C5 H26 C5 H29
bond C6 H30 016 C7 C3 015 C13 C12
bond C8 014 C4 C8 C10 C4 C11 C9
bond C8 H19 C8 H28 C9 H20 C10 H21
bond C9 C4 C13 C11 C12 C10 C11 H22
bond C12 H23 C13 H24 N1 H31
bond C7 +N1
```

As noted above, the angle and dihedral terms will be autogenerated and are therefore not listed for this residue. The less common improper dihedrals

(normally just called impropers), however, must be listed explicitly. In this case there are two impropers, which maintain the planarity of the peptide bonds. As with dihedrals, the order of atoms within an improper may be reversed. As shown below, impropers are specified by the IMPR statement followed by sets of four atoms, with the central atom to which the other three are bonded typically listed first.

```

IMPROPER -C7 C2 N1 H25
IMPROPER C2 +N1 C7 O16
IMPROPER C7 C2 N1 H25
IMPROPER C2 N1 C7 O16
IMPROPER C6 O15 C3 O14
IMPROPER C4 C12 C10 H21
*IMPROPER C10 C13 C12 H23
IMPROPER C13 C10 C12 H23
IMPROPER C11 C12 C13 H24
IMPROPER C9 C13 C11 H22
IMPROPER C4 C11 C9 H20
IMPROPER C9 C10 C4 C8

```

Topology file ends with an END statement.

4.6.2 CHARMM parameter file

A force field parameter file contains all of the numerical constants needed to evaluate forces and energies, given a .psf structure file and atomic coordinates. The parameter file is closely tied to the topology file that was used to generate the .psf file, and the two are typically distributed together and given matching names.

This file starts explaining the version of the force field itself and with some comments.

The first set of entries in the parameter file are those for bonds, indicated by the BONDS keyword. Each entry consists of a pair of atom types, a spring constant and an equilibrium length. Entries are present for every type of bond present in the topology file.

The next section gives parameters for every type of angle present in the topology file, indicated by the THETAS keyword. Since angles are formed from combinations of bonds, there are many more types of angles than types of bonds. Each entry consists of three atom types, a spring constant, and an equilibrium angle.

The next section gives parameters for every type of dihedral (PHI) present in the topology file; there are even more dihedrals than angles. Since dihedrals represent the energy of rotation around a covalent bond, which is the source of most conformational flexibility in biomolecules, they must provide a smooth energy for 360 degrees. You may observe in the excerpts below that the dihedral spring constants are one to two orders of magnitude lower than for angles, with an order of magnitude of difference between flexible and inflexible dihedrals. Because of the large numbers of dihedral terms required to describe a complete protein, the wildcard atom type “x” is occasionally used. These parameters will be used in NAMD if a more specific match is not found elsewhere in the parameter file.

The final bond-like terms in the parameter file are improvers (IMPDI), which are used exclusively and explicitly in the molecular topology to maintain planarity. The harmonic form with a large spring constant and an equilibrium value typically zero is used to restrain deformations among an atom and three atoms bonded to it. As with dihedrals, the angle is the one between the plane containing the first three atoms and the plane containing the last three. Wildcard atom types (“x”) occur in the second and third positions, rather than the first and fourth as in dihedrals. The NONBONDED statement includes a list of parameters. Those shown below correspond to the NAMD settings exclude (1-4)-interactions, switching on, pairlistdist 14.0, cutoff 12.0, switchdist 10.0, dielectric 1.0, and 1-4scaling 0.833333:

```
NONBONDED
NBXMOD 5 GROUP SWITCH CDIEL - CUTNB 14.0 CTOFNB 12.0 CTONNB 10.0 EPS
1.0 E14FAC 0.83333333 WMIN 1.4
```

Recall that the partial charge of each atom is specified in the topology and .psf files and is independent from the atom type. Therefore the only type-based parameters are for the van der Waals interactions, which are represented by the classic Lennard-Jones potential:

$$U = \varepsilon \left[\left(r_{min}/r \right)^{12} - 2 \left(r_{min}/r \right)^6 \right]. \quad (4.3)$$

Observe that at $r = r_{min}$ the force is zero and the energy is $-\varepsilon$. Rather than providing a different value of epsilon for every possible combination of atom types, only one value is provided per type and inter-type interactions are calculated using the sum of the radii $r_{min}/2$ and the geometric mean of the well-depths (ε). By convention, the values are negative in the parameter file.

```
!           Emin      Rmin/2      Emin/2      Rmin (for 1-4's)
!           (kcal/mol) (A)
H  0.0 -0.0157  0.6000  0.0 -0.00785  0.6000 ! Ferguson base pair geom.
HO 0.0 -0.0000  0.6000  0.0 -0.0000  0.6000 ! OPLS, JACS,110,(1988),1657
```

As topology files, also this parameter file terminates with an END keyword.

4.6.3 How to create initial sample

In order to create initial samples, Packmol [113] has been used. The input file looks like:

```
tolerance 2.0
filetype pdb
output parallel_system.pdb
structure dmf.pdb
  number 1144
  inside box -70. -70. -70. 70. 70. 70.
end structure
structure nil.pdb
  number 1
  inside box -22.5 -22.5 -22.5 22.5 22.5 22.5
end structure
```

First of all, it is necessary to specify the distance tolerance required. Then the output filetype and name. There follows two *structures*: the first one generates DMF solution, while the second one adds the two helices. “dmf.pdb” contains only one DMF molecule, while “nil.pdb” the two helices. The latter file was created using a code called “pdb_translation” in order to translate and rotate original helix, which had been copied in this new modified .pdb along with the original one. Since “nil.pdb” already comprehends the two helices in the right position, there is just one copy of this file, while DMF has to be “multiplied” (*number 1144*). The command *inside box* specifies dimension of box in which we want molecules (for each structure) to be put in. The smaller box for PBLG is due to the fact that we wanted helices to be in the middle of the sample. Since code crashed using the correct box side, which turned out to be too small, these sides had to be enlarged (Figure 4.57): boxes bigger than the original ones need a compression, whose input file is described in Appendix 4.6.4. After this compression, samples are ready.

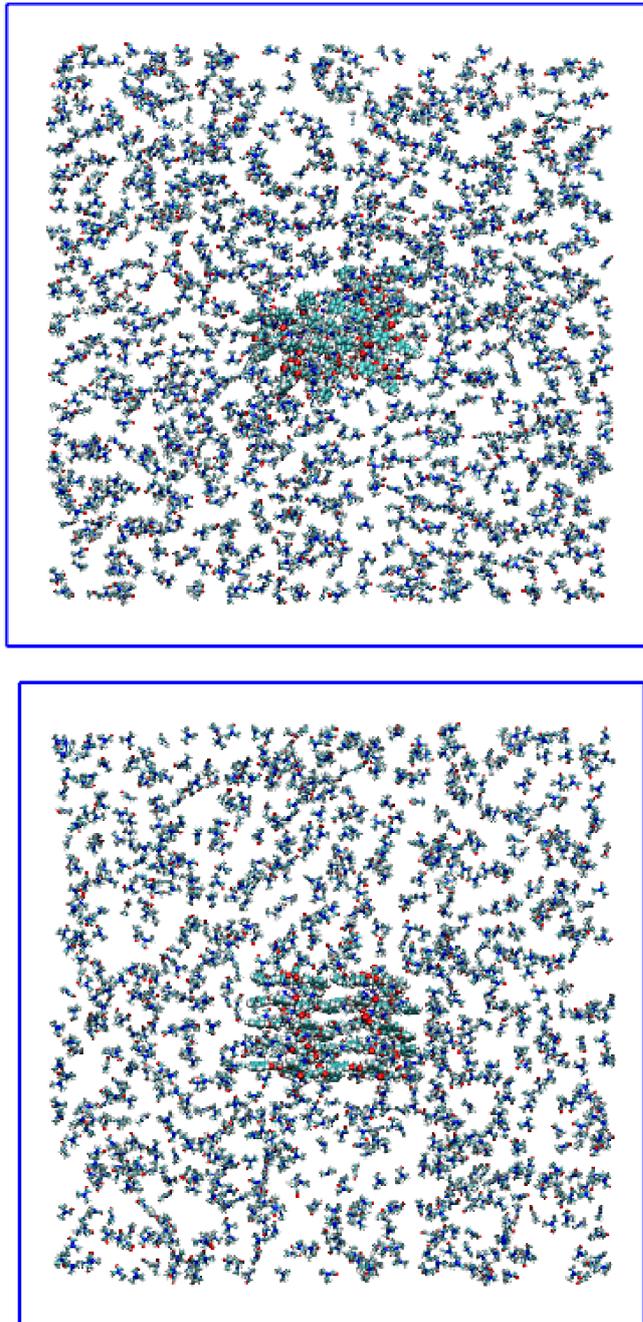


Figure 4.57: The upper figure is a snapshot of the perpendicular system, while the one below represents a parallel system.

4.6.4 About NAMD input file

The following section shows the input file used to compress samples, using NAMD [115].

```
#--- integrator

numsteps          20000000
timestep          1
nonbondedFreq    2
fullElectFrequency 8

#--- Parameter options (14 scaling is for electrostatics)

structure          parallel_2pblg.psf
paratypeCharmm    on
parameters        par_amber_cornell.inp
exclude           scaled1-4
1-4scaling        0.8333333

switching         on
switchdist        10.
cutoff            12.
pairlistdist      13.5
margin            0
stepspercycle     40

#--- Thermodynamic

coordinates        parallel_2pblg.pdb
#temperature       300
#seed              74269

rescaleTemp       300K
rescalefreq       100
COMmotion         no

BerendsenPressure          on
BerendsenPressureTarget    100.01325
BerendsenPressureCompressibility 0.000045
BerendsenPressureRelaxationTime 10000.
BerendsenPressureFreq      200

useFlexibleCell yes

#--- PBC

extendedSystem pblg300r.xsc

cellBasisVector1  77.0  0.0  0.0
cellBasisVector2  0.0  65.0  0.0
cellBasisVector3  0.0  0.0  77.0

#--- PME

PME              on
PMEGridSpacing   1.2
zeroMomentum     no

#--- Input coords
bincoordinates   pblg300r.coor
binvelocities    pblg300r.vel
```

```

#--- Output & Restart

binaryoutput    no
outputname      parallel_2pblg300

binaryrestart   yes
restartname     pblg300r
restartfreq     10000

DCDfile         parallel_2pblg300.dcd
DCDfreq         10000
XSTfreq         10000

#--- Standard Output

outputEnergies  10000
outputtiming     5000

#--- Constraints

constraints on
consexp 2

consref ref.pdb
conskfile kcol.pdb
conskcol 0

#--- Fixed atoms

fixedAtoms on
fixedAtomsForces off
fixedAtomsFile ref_fixed.pdb
fixedAtomsCol B
ExcludeFromPressure on
ExcludeFromPressureFile ref_fixed.pdb
ExcludeFromPressureCol B

```

The first part of this file (“Integrator”) declares the number of steps (*num-steps*), the timestep itself (*timestep*, fs), a parameter (*nonbonded Freq*) that specifies how often short-range nonbonded interactions should be calculated and the number of timesteps between each full electrostatic description (*fullElectFrequency*). The second section describes “Parameter Options”. First of all, the topology file (.psf) is declared (*structure*). Then, there is an explication about the parameter files: in this case, they are written in CHARMM format (*paratypeCharmm on*). Next, the name of the parameter file is specified (*parameters*). The *exclude* command define the exclusion of all 1–4 interactions from non-bonded interactions, while *1–4 scaling* specifies the constant factor by which electrostatic interactions of such pairs will be modified. The *switching on* command let smoothing functions to be applied to both electrostatic and van der Waals forces,

starting from the distance *switchdist*. *Cutoff* parameter (always bigger than *switchdist*) represents the local interaction distance, i.e. the distance within which electrostatic pairs will be directly computed step by step, while, outside of this distance, interactions will be calculated only periodic. The *pairlistdist* parameter (bigger than *cutoff*) is used when switching is turned on, in order to specify the allowable distance between atoms for inclusion in the pair list. The next parameter (*margin*) would not influence the physics of the system, since it is only an internal tuning parameter used to determine the size of the cubes of space with which NAMD partitions the system. Finally, *stepspercycle* specifies the number of timesteps in each cycle (a cycle represents the number of timesteps between atoms reassignment).

The “Thermodynamic” section starts with the specification of the *coordinates* file, the *temperature* (K) and the *seed* (number used to seed the random number generator). If you start a run from a saved starting point, the latter two parameters have to be commented, while the rescaling parameters would be used. *rescaleTemp* declares the temperature to which all velocities will be rescaled every *rescalefreq*. The *COMmotion* command does not allow the motion of the center of mass of the entire system. Next, Berendsen pressure bath coupling is parametrized: after having switch it on (*BerendsenPressure on*), the target pressure (bar) is specified (*BerendsenPressureTarget*). Then compressibility (*BerendsenPressureCompressibility*) and relaxation time (*BerendsenPressureRelaxationTime*), along with the number of timesteps between position rescaling (*BerendsenPressureFreq*), are described. Flexible cell are used in order to obtain a non-squared box with a non-isotropic barostat.

The first command in “PBC” section (*extended System*) allows NAMD to read the *.xsc* file generated in a previous run and use these periodic cell parameters. If this command is not commented, the various *cellBasisVector*, which specify a basis vector for periodic boundary condition, will not be taken into account.

Next, the “PME” section declare that Particle Mesh Ewald method is used for electrostatic calculation (*PME on*). The *PMEGridSpacing* value is used

to determine accuracy and efficiency of PME. Finally, *zeroMomentum no* specifies that any drift in the net momentum of simulation is not removed before every full electrostatic step.

The following section (“Input coords”) specifies the name of file containing coordinates (*bincoordinates*) and velocities (*binvelocities*) of a previous run.

The “Output & Restart” section begins underlining that output file will not be in binary format (*binaryoutput no*). Then, the output filename is specified (*output*). The name of restart files (*restartname*), which now will actually be binary file (*binaryrestart yes*), along with the frequency of saving (*restartfreq*), is declared. The name of .dcd file (*DCDfile*) and frequency of saving both .dcd (*DCDfreq*, trajectory) and .xst (*XSTfreq*, eXtended System Trajectory, contains a record of periodic cell parameters and extended system variables during the simulation) files are specified.

The “Standard Output” section specifies the number of timesteps between each energy output of NAMD (*outputEnergies*) and the number of timesteps between each timing output of NAMD (*outputtiming*).

The section “Constraints” is used in the first part of the simulation in order to keep atoms fixed, without destroy the PBLG helices. First of all, the harmonic constraints were switched on (*constraints on*) and an exponent to be used in this harmonic constraint energy function (*consexp*) is specified. *consref* command defines a .pdb file to be used for reference position for harmonic constraints, which specifies atoms which will be constrained to their positions. *conskfile* command specifies the .pdb file to use for force constants for harmonic constants. Finally, *conskcol* defines the column of .pdb file in which read the constrained atoms (a value of ”0” indicates that the atom will not be constrained). Note that columns in .pdb files are called X, Y, Z, O, B.

After the sample had almost been equilibrated and compressed, the “constrained” were switched off and the atom are “fixed”. Note that the “atom fixed” were applied to all atoms belonging to the helices, while the “constrained” to just the first and the last C_α . After switching on the fixed atoms method (*fixedAtoms on*), it should be specified that forces

between fixed atoms have not to be computed (*fixedAtomsProces off*). The *fixedAtomsFile* specified the name of the file containing indications about atoms fixed, while *fixedAtomsCol* specifies the column of the latter .pdb files containing the flag for fixing (or not) atoms. Then there is another .pdb file (*ExcludeFromPressureFile*, column specified by *ExcludeFromPressureCol*) which specifies atoms to be excluded from pressure rescaling (*ExcludeFromPressure on*): in this case, all atoms belonging to the PBLG helices will be excluded.

4.6.5 Chiral index analysis for various solutes

The choice of the compound to be used as solute in our PBLG solution is difficult. Various molecules (chiral compounds already studied with such mixture and some odorants) are analyzed and a chiral index for every solute has been assigned. A compound with a high chiral index and low steric effects would be a greater choice since it should be easy to analyze in simulations.

4.6.5.1 Molecular chirality index

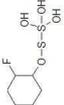
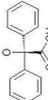
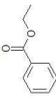
The molecular chirality index used is defined as followed [116, 117]:

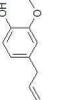
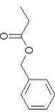
$$G_{0S} = \frac{4!}{3N^4} \sum_{ijkl} \hat{P}_{ijkl} m_i m_j m_k m_l \frac{[(\mathbf{r}_{ij} \times \mathbf{r}_{kl}) \cdot \mathbf{r}_{il}](\mathbf{r}_{ij} \cdot \mathbf{r}_{jk})(\mathbf{r}_{jk} \cdot \mathbf{r}_{kl})}{(r_{ij} r_{jk} r_{kl})^2 r_{il}} \quad (4.4)$$

where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is an interatomic vector and \mathbf{r}_i and m_i are the position and mass in atomic units of the i -th atom. The summations is performed over all sets of four atoms i, j, k, l out of the N atoms of the molecule, while $\hat{P}_{ijkl} A_{ijkl}$ is an operator generating a sum over all the permutations of i, j, k, l . The index G_{0S} assigns to each observed molecular conformation a measure of chirality obtained as a mass-weighted degree of asymmetry of the atomic coordinates, with a sign that can be used for a right/left classification.

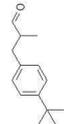
4.6.5.2 Compounds

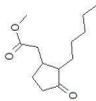
Molecules analyzed are taken from two type of articles. The first one treats compounds which have already been used experimentally to be discriminated with a system composed by an helix of PBLG and a co-solvent. For this chiral discrimination, ^{13}C - and ^1H -NMR techniques are used. Along with these molecules, other compounds known to be odorant stimuli are investigated [118].

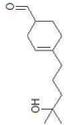
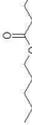
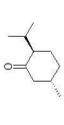
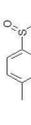
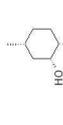
Number [ref.]	Common compound name	IUPAC compound name	Chemical formula	Functional group name	G_{05} value
109 [83]	Fluoro-oxytiosulfate- cyclohexane			Chiral fluorinated compounds	-78.8093
73 [83]	(1R,2R)-1,2-Bis-(2- methoxy-phenyl)-ethane- 1,2-diol	(1R,2R)-1,2-Bis-(2-methoxy- phenyl)-ethane-1,2-diol		Diol	-39.4067
114 [83]	2,2,2-Trifluoro-1-phenyl- ethanol	2,2,2-Trifluoro-1-phenyl-ethanol		Chiral fluorinated compounds	-30.3150
70 [87]	Benzylic acid (-)	Hydroxy-diphenyl-acid acetic		Carboxylic acid	-23.7087
48 [118]	Ethyl benzoate	Benzoic acid ethyl ester		Ester	-22.9395

95 [83]	2,3,3-Trichloro THF	2,3,3-Trichloro-tetrahydrofuran		Heterocycle	-19.0195
45 [118]	Valeraldehyde	Pentanal		Aldehyde	-12.7526
9 [118]	Eugenol	4-Allyl-2-methoxy-phenol		Alcohol	-12.3368
94 [83]	TCPO	2-Trichloromethyl-oxirane		Heterocycle	-9.00588
24 [83]	Oct-1-yn-3-ol	Oct-1-yn-3-ol		Alcohol	-7.30105
42 [118]	Citral	(E)-3,7-Dimethyl-octa-2,6-dienal		Aldehyde	-6.07588
46 [118]	Benzyl propionate	Propionic acid benzyl ester		Ester	-5.81916
113 [83]	2-Fluoro-2-pentyl-ethanol	2-Fluoro-2-pentyl-ethanol		Chiral fluorinated compounds	-5.13705
27 [83]	α -terpineol	2-(4-Methyl-cyclohex-3-enyl)-propan-2-ol		Alcohol	-5.03800

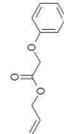
32 [118]	Citronella-(+)	(S)-4,8-Dimethyl-non-7-en-2-one		Aldehyde	-4.76917
78 [118]	Menthone(3)	(2R,5R)-2-Isopropyl-5-methylcyclohexanone		Ketone	-4.43287
97	(R)-Limonene	(R)-4-Isopropenyl-1-methylcyclohexene		Hydrocarbon	-4.34756
100 [83]	Limonene	4-Isopropenyl-1-methylcyclohexene		Hydrocarbon	-4.19734
16 [118]	Menthol(6)	(1R,2R,5S)-2-Isopropyl-5-methylcyclohexanol		Alcohol	-4.14967
66 [83]	2-Methylbutyric acid	2-Methylbutyric acid		Carboxylic acid	-3.85035
19 [83]	1-Isopropyl-4-methylcyclohex-3-enol	1-Isopropyl-4-methylcyclohex-3-enol		Alcohol	-3.59768

30 [81]	Absolute alcohol	Ethanol		Alcohol	-3.43546
63 [83]	Methoxy-phenyl-acetic acid	Methoxy-phenyl-acetic acid		Carboxylic acid	-3.18345
3 [118]	Butanol	Butanol		Alcohol	-3.16012
37 [118]	Lilial	3-(4-tert-Butyl-phenyl)-2-methyl-propionaldehyde		Aldehyde	-3.07894
101 [83]	Methyl-pentene	3-Methyl-pent-1-ene		Hydrocarbon	-2.81892
29 [86]	(-)-Isopinocampheol (-IPC)	(1R,2R,3R,5S)-2,6,6-Trimethyl-bicyclo-[3.1.1]heptan-3-ol		Alcohol	-2.74123
116 [83]	2-Methyl bromobutane	1-Bromo-2-methyl-butane		Other compounds	-2.42807
26 [83]	Isopinocampheol	(1R,3R,5S)-2,6,6-Trimethyl-bicyclo[3.1.1]-heptan-3-ol		Alcohol	-2.39029
102 [112]	<i>cis</i> -Decalin	1,2,3,4,5,6,7,8,9,10-decahydronaphthalene		Hydrocarbon	-1.62606

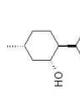
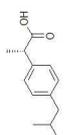
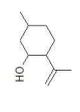
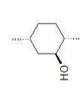
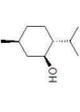
88 [118]	Hedione	(3-Oxo-2-pentyl-cyclopentyl)- acetic acid methyl ester		Ketone	-1.48391
5 [118]	Hexanol	Hexanol		Alcohol	-1.42877
98	(S)-Limonene	(S)-4-Isopropenyl-1-methyl- cyclohexene		Hydrocarbon	-1.31482
58 [118]	Butyric acid	Butyric acid		Carboxylic acid	-0.974937
99 [83]	(4S,6R)-1,4,5,5,6- Pentamethyl-cyclohexene	(4S,6R)-1,4,5,5,6-Pentamethyl- cyclohexene		Hydrocarbon	-0.780563
86 [118]	β -ionone	(E)-4-(2,6,6-Trimethyl- cyclohex-1-enyl)-pent-3-en-2- one		Ketone	-0.629614

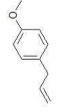
38 [118]	Lyral	4-(4-Hydroxy-4-methyl-pentyl)- cyclohex-3-enecarbaldehyde		Aldehyde	-0.624416
4 [118]	Geraniol	(E)-3,7-Dimethyl-octa-2,6-dien- 1-ol		Alcohol	-0.562556
51 [118]	Amyl butyrate	Butyric acid pentyl ester		Ester	-0.503050
77 [118]	Menthone(2)	(2R,5S)-2-Isopropyl-5-methyl- cyclohexanone		Ketone	-0.496209
106 [83]	1-methyl-4- (methylsulphanyl)benzene	1-Methanesulfinyl-4-methyl- benzene		Chiral sulfur compounds	-0.495053
6 [118]	Linalool(S)	(S)-3,7-Dimethyl-octa-1,6- dien3-ol		Alcohol	-0.489926
25 [83]	Capryl alcohol	Octan-2-ol		Alcohol	-0.443572
17 [118]	Menthol(7)	(1R,2R,5R)-2-Isopropyl-5- methyl-cyclohexanol		Alcohol	-0.437589
8 [118]	Octanol	Octanol		Alcohol	-0.254708

67 [83]	4-Methylcaprylic acid	4-Methyl-octanoic acid		Carboxylic acid	-0.249275
2 [118]	2-Phenylethyl alcohol	2-Phenylethyl alcohol		Alcohol	-0.104049
85 [118]	Acetophenone	1-Phenyl-ethanone		Ketone	-0.0190704
92 [118]	Propiophenone	1-Phenyl-propan-1-one		Ketone	-0.0144947
62 [118]	Propionic acid	Propionic acid		Carboxylic acid	-0.00886745
105 [81]	Bicycloheptadiene	Bicyclo[2.2.1]hepta-2,5-diene		Hydrocarbon	-0.00452401
34 [118]	Butanal	Butyraldehyde		Aldehyde	-0.00182548
72 [118]	Eucalyptol	1,3,3-Trimethyl-2-oxa-bicyclo[2.2.2]octane		Ether	0.000456677
47 [118]	Benzyl acetate	Acid acetic benzyl ester		Ester	0.000807496
43 [118]	Heptanal	Heptanal		Aldehyde	0.00140304
83 [118]	2-Hexanone	2-Hexanone		Ketone	0.00825162

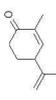
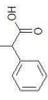
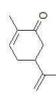
89 [118]	2,4-Dimethylacetophenone	1-(2,4-Dimethyl-phenyl)-ethanone		Ketone	0.0103357
55 [118]	Butyl acetate	Acid acetic butyl ester		Ester	0.0116577
41 [118]	Ethyl vanillin	3-Ethoxy-4-hydroxy-benzaldehyde		Aldehyde	0.0165949
104 [81]	Acenaphthene	1,2-Dihydroacenaphthylene		Hydrocarbon	0.0186572
91 [118]	Butyrophenone	1-Phenyl-1-butan-1-one		Ketone	0.0209042
93 [118]	Valerophenone	1-Phenyl-pentan-1-one		Ketone	0.0250282
33 [118]	Benzaldehyde	Benzaldehyde		Aldehyde	0.0446760
44 [118]	Nonanal	Nonanal		Aldehyde	0.0479744
40 [118]	Cinnamaldehyde	(E)-3-Phenyl-propenal		Aldehyde	0.0528689
50 [118]	Allyl phenoxy acetate	Phenoxy-acetic acid allyl ester		Ester	0.0625678
87 [118]	Ethyl isoamyl ketone	6-Methyl-heptan-3-one		Ketone	0.0902273

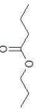
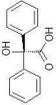
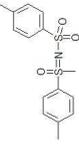
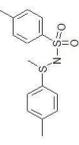
1 [118]	Decanol	Decanol		Alcohol	0.106919
7 [118]	Linalool(R)	(R)-3,7-Dimethyl-octa-1,6-dien3-ol		Alcohol	0.477855
90 [118]	2-Aminoacetone	2-Amino-1-phenyl-ethanone		Ketone	0.659987
21 [83]	Benzenethanol	1-Phenyl-ethanol		Alcohol	0.688619
71 [118]	2-Phenylethyl amine	Phenethylamine		Amine	0.708790
15 [118]	Menthol(5)	(1S,2S,5R)-2-Isopropyl-5-methyl-cyclohexanol		Alcohol	0.792153
75 [118]	<i>o</i> - <i>sec</i> -butyl cyclohexanone	2- <i>sec</i> -Butyl-cyclohexanone		Ketone	0.807936
35 [118]	Decanal	Decanal		Aldehyde	1.06357
20 [83]	Indan-1-ol	Indan-1-ol		Alcohol	1.09752
118 [81]	Propanedinitrile	Malonitrile		Other compounds	1.10567

12 [118]	Menthol(2)	(1R,2S,5R)-2-Isopropyl-5-methyl-cyclohexanol		Alcohol	1.22293
68 [111]	Ibuprofen	(S)-2-(4-Isobutyl-phenyl)-propionic acid		Carboxylic acid	1.59153
10 [118]	Isopulegol	2-Isopropenyl-5-methyl-cyclohexanol		Alcohol	1.69926
13 [118]	Menthol(3)	(1S,2R,5R)-2-Isopropyl-5-methyl-cyclohexanol		Alcohol	2.14045
36 [118]	Hexanal	Hexanal		Aldehyde	2.19886
103 [112]	<i>trans</i> -Decalin	or bicyclo[4.4.0]decane		Hydrocarbon	2.33024
11 [118]	Menthol(1)	(1S,2R,5S)-2-Isopropyl-5-methyl-cyclohexanol		Alcohol	2.33148

28 [86]	(+)-Isopinocampheol (+IPC)	(1S,2S,3S,5R)-2,6,6-Trimethylbicyclo-[3.1.1]heptan-3-ol		Alcohol	2.73720
23 [83]	4-Trimethylsilylbut-3-yn-2-ol	4-Trimethylsilylbut-3-yn-2-ol		Alcohol	3.07164
22 [83]	But-3-yn-2-ol	But-3-yn-2-ol		Alcohol	3.58548
56 [118]	Octyl acetate	Acid acetic octyl ester		Ester	3.82058
112 [83]	2-Fluoro-cyclohexanol	2-Fluoro-cyclohexanol		Chiral fluorinated compounds	3.89028
49 [118]	Estragole	1-Allyl-4-methoxy-benzene		Ester	4.08500
39 [118]	Octanal	Octanal		Aldehyde	4.11847
79 [118]	Menthone(4)	(2S,5S)-2-Isopropyl-5-methylcyclohexanone		Ketone	4.24311

111 [83]	3-Fluoro-butanol	3-Fluoro-butan-2-ol		Chiral fluorinated compounds	4.36227
76 [118]	Menthone(1)	(2S,5R)-2-Isopropyl-5-methyl-cyclohexanone		Ketone	4.56458
59 [118]	Decanoic acid	Decanoic acid		Carboxylic acid	4.65544
18 [118]	Menthol(8)	(1S,2S,5S)-2-Isopropyl-5-methyl-cyclohexanol		Alcohol	5.32573
31 [118]	Citronella(-)	(R)-4,8-Dimethyl-non-7-en-2-one		Aldehyde	5.33569
74 [83]	1,1-Dihydroxynaphthalene	[1,1']Binaphthalenyl-2,2'-diol		Diol	5.39607
84 [118]	2-Octanone	2-Octanone		Ketone	5.46623
60 [118]	Hexanoic acid	Hexanoic acid		Carboxylic acid	5.62799

115 [83]	2-Phenyl-oxirane		Other compounds	5.67646
61 [118]	Octanoic acid		Carboxylic acid	5.86449
14 [118]	Menthol(4)		Alcohol	5.92888
54 [118]	Hexyl acetate		Ester	6.70519
81 [118]	(-)-Carvone		Ketone	7.09319
82 [118]	2-Butanone		Ketone	7.26670
64 [83]	Hydratropic acid		Carboxylic acid	7.69169
53 [118]	Heptyl butyrate		Ester	8.27699
80 [118]	(+)-Carvone		Ketone	9.48229

52 [118]	Ethyl acetate	Acid acetic ethyl ester		Ester	14.8230
117 [81]	(1S,7R)-4-Oxa-tricyclo-[5.2.1.0*2,6*]dec-8-ene-3,5-dione	(1S,7R)-4-Oxa-tricyclo-[5.2.1.0*2,6*]dec-8-ene-3,5-dione		Other compounds	16.2414
57 [118]	Propyl butyrate	Butyric acid propyl ester		Ester	16.7403
69 [87]	Benzyllic acid (+)	Hydroxy-diphenyl-acid acetic		Carboxylic acid	17.1231
96 [83]	(R)-4-Methyl-oxetan-2-one	(R)-4-Methyl-oxetan-2-one		Heterocycle	21.6505
110 [83]	1-Chloro-3-fluoro-propan-2-ol	1-Chloro-3-fluoro-propan-2-ol		Chiral fluorinated compounds	39.3049
108 [83]	Dibenzyl-tio-sulfonamide	Dibenzyl-tio-sulfonamide		Chiral sulfur compounds	68.2121
107 [83]	Dibenzyl-tiomethyl-sulfonamide	Dibenzyl-tiomethyl-sulfonamide		Chiral sulfur compounds	71.5130

65 [83]	2-Bromo-propionic acid		2-Bromo-propionic acid	Carboxylic acid	87.8498
---------	------------------------	---	------------------------	-----------------	---------

Table 4.14: Compounds analyzed (organized by increasing G^a values).

4.6.5.3 Histograms for classes

Next figures show distributions of chiral index values for classes of compounds. Bins are defined as follows:

- A: [-80.0;-64.54546]
- B: [-65.54545;-49.09091]
- C: [-49.09090;-33.63636]
- D: [-33.63635;-18.18181]
- E: [-18.18180;-2.72726]
- F: [-2.72725;12.72728]
- G: [12.72729;28.18184]
- H: [28.18183;43.63637]
- I: [43.63638;59.09092]
- J: [59.09093;74.54547]
- K: [74.54548;90.00000]

Even if we have analyzed only few compounds of this class, sulfurated compounds are characterized by zero or positive chiral index values. Zero or negative G^a values are typical for diols, while carboxylic acid sometimes show highly positive chiral index values. Also fluorinated chiral compounds can show big negative G^a values.

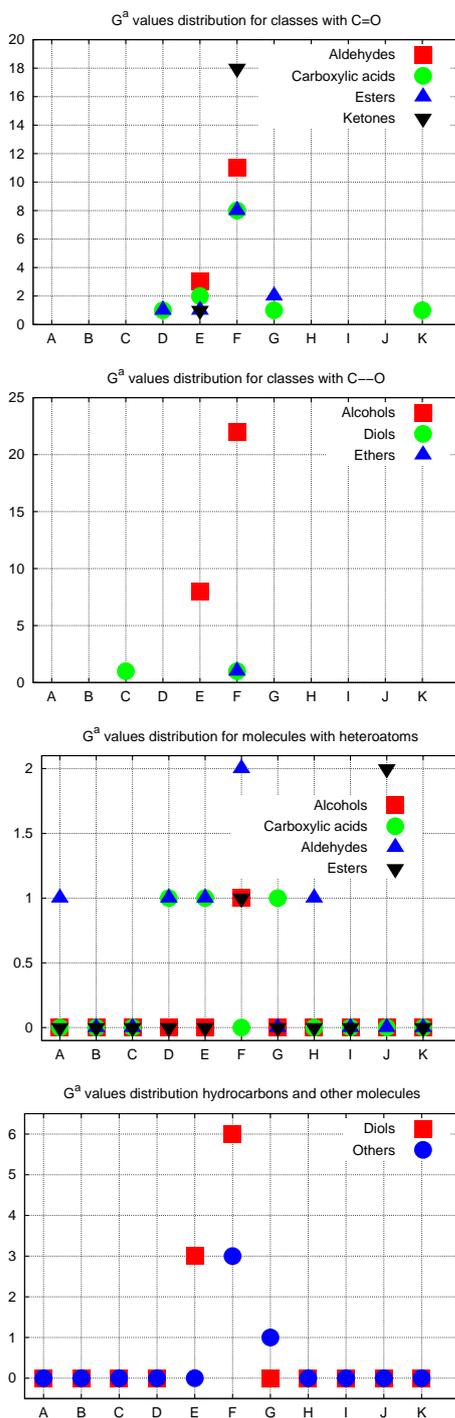


Figure 4.58: Histograms of G^a values distribution for classes of organic compounds. For the definition of bins see text.

4.6.5.4 Other properties

To discover if there are features that increases (for what concern the absolute values) or turn to zero the G^a parameter, some graphs are built to monitor the behavior of chiral index depending on various characteristics, like the presence of heteroatoms, the number of double bonds, the quantity of benzylic group and so on.

Adding benzylic group to compounds does not seem to influence the range of chiral index values (see Figure 4.59): however, the presence of two benzylic groups can furnish quite high positive G^a values.

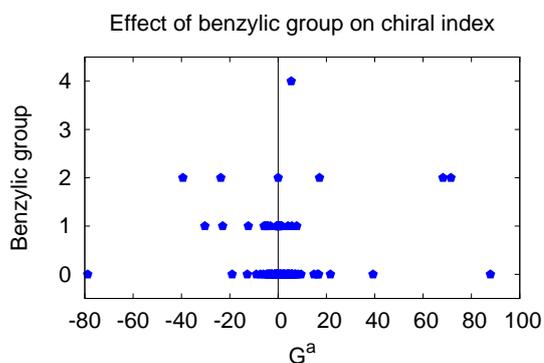


Figure 4.59: Effect of the number benzylic group on chiral index.

Compounds with no cyclohexyl group, on the other hand, show chiral index values ranging from quite highly negative values to highly positive one (Figure 4.60): in one case, the addition of one cyclohexyl group decreases the G^a value to one of the most negative index.

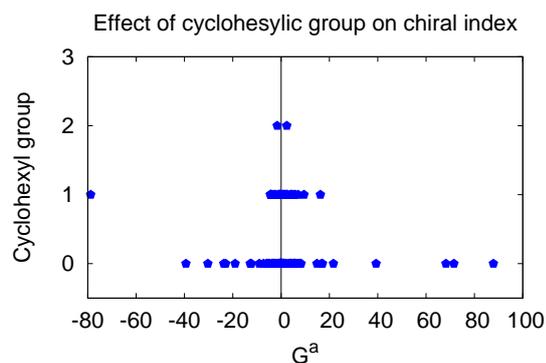


Figure 4.60: Effect of the number of cyclohexyl group on chiral index.

From Figure 4.61, it is clear that the presence of double bonds in cyclic group reduce the absolute chiral index value, which tends to zero.

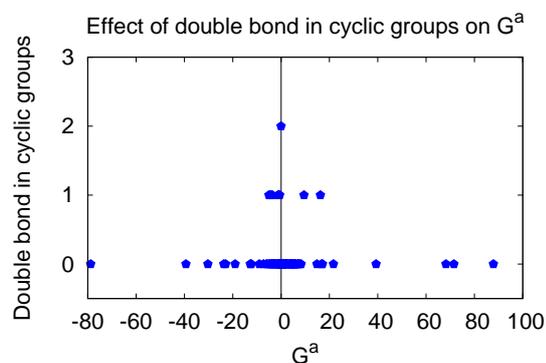


Figure 4.61: Effect of double bonds in cyclic group on chiral index.

Compounds studied have often linear chain of carbon atoms linked, that lower the chiral index values. However, their length is not responsible of any typical effects (Figure 4.63).

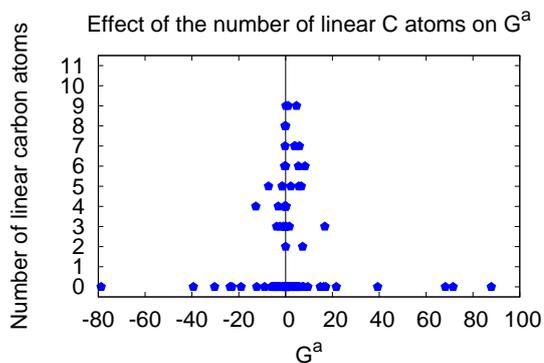


Figure 4.62: Effect of the number of carbons linked in a linear chain on chiral index.

Figure 4.63 shows that the presence of a linear chain of C atoms linked to cyclic groups has no effect on G^a values, even if it seems that this parameter tends to zero more frequently when such chain is present.

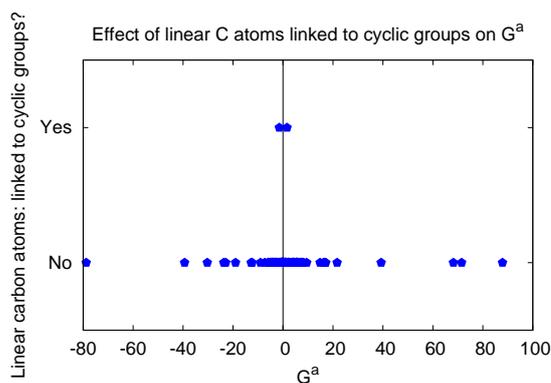


Figure 4.63: Effect of number of carbons belonging to a linear chain linked to cyclic group on chiral index.

Next figure (Figure 4.64) shows that adding heteroatoms to molecules have different effects. When sulfur, fluorine, nitrogen, oxygen and bromide are linked to solutes, the range of possible G^a values is enlarged to reach also high absolute values of G^a . On the other hand, chlorine seems to narrow the range of possible chiral index values.

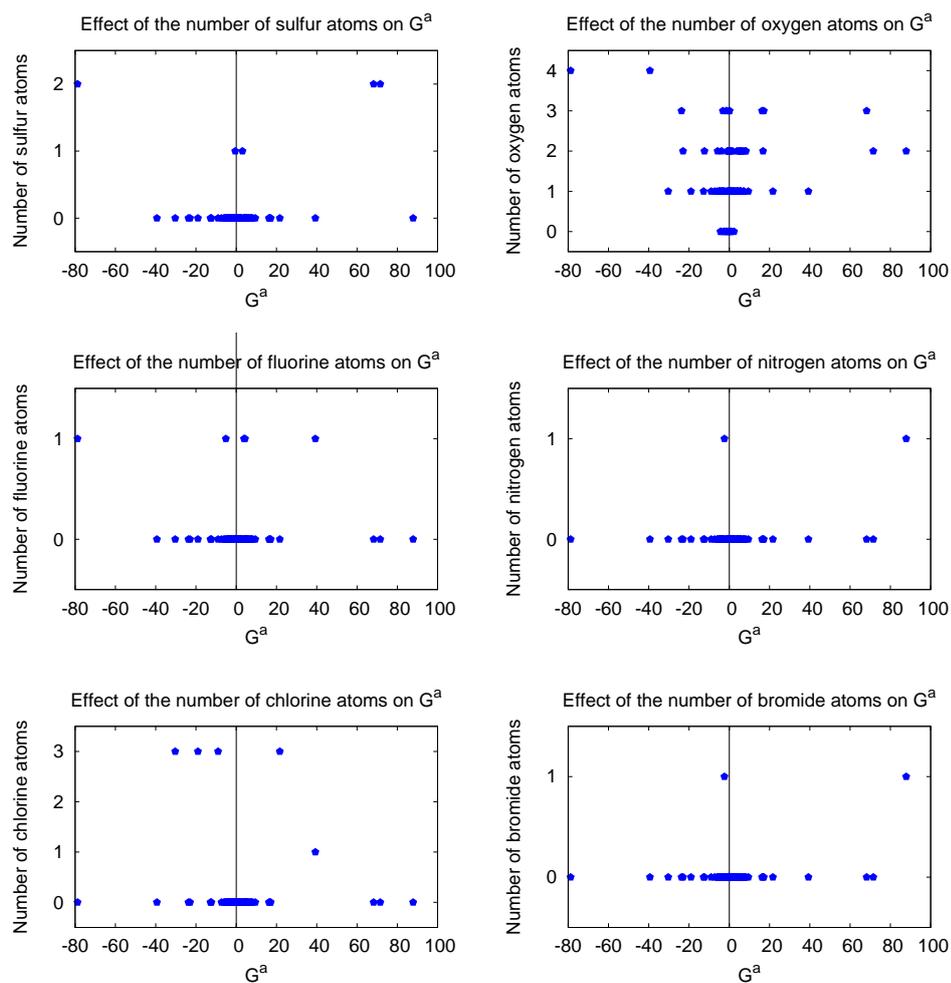


Figure 4.64: Effect of the number of heteroatoms on chiral index.

4.6.5.5 Conclusions

From the analysis of a total of 118 compounds, we decide to choose molecule number 53, heptyl butyrate, which shows a value of $G^a \sim 8.27699$.

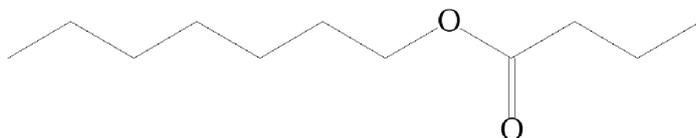


Figure 4.65: Formula of heptyl butyrate.

This is not one of the highest value of such parameter but such molecule is a good compromise between a sufficient high value of chiral index and a low steric effect (since there are no cyclic groups).

Moreover, some graphs were built to verify if there are classes of compounds or peculiar atoms/groups which show typical high values. We found out that sulfur compounds show high values (modulus) and that adding benzylic group, sulfur and oxygen atoms to molecules enlarges the range of possible G^a values.

4.6.6 Behavior of a poly-glutamate helix in dimethylformamide

The aim of this study is to discover the behavior of a poly-glutamate (poly-GLU) helix in dimethylformamide (DMF) in order to compare it with that of a poly-(γ -benzyl L-glutamate) in DMF and find out if the set of parameters for this last system has been incorrectly chosen or if extending the length of the chain is the key for making simulations work correctly.

Glutamic acid (Glu or E) is one of the 20 proteinogenic amino acids. This molecule is responsible for the so called “umami” flavor) of many food as seaweed, cheese and soy sauce. Glutamate is a key molecule in metabolism [53] and the most abundant excitatory neurotransmitter [119]. Poly-glutamate is now being used as a drug carrier releasing molecule [120,

121] in response to a change in the pH value of the media. This technology does not damage healthy tissues since polymer inactivates drug until it reaches the tumor cells, which are more porous to poly-glutamate than healthy blood vessels. Moreover, poly-glutamate is used to increase the bioavailability of folate. Folate is an important nutrient in the daily diet, since its deficiency can cause megaloblastic anemia [122, 123], neural tube defects [124, 125], colon cancer [126], plasma homocystein (a potential risk factor for cardiovascular disease [127, 128] and depression [129]).

4.6.6.1 Initial sample

The initial sample is composed by an α -helix containing 36 residues of glutamate, 36 sodium atoms to obtain a neutral solution and 1320 DMF molecules. This quantity has been chosen to make simulation comparable to PBLG solution simulations, currently under study.

Poly-glutamate helix

Poly-glutamate shows an α -helix backbone which gives rigidity to the structure, thanks to the N-H \cdots O=C hydrogen bonds [130, 131].

First of all we create an helix of 36 glutamate residues using the “Build residue” tool of Pymol [98]. We decide to use the original parm94 (AMBER Cornell) force field[90]. A residue of glutamate and its atom labels are shown in Figure 4.66. ²

²Parameter and topology files for a single glutamate molecule (ready to be bound to another amino acid, hence without the water molecule responsible of the peptide bond) are called “par_amber_cornell.inp” and “top_amber_cornell.inp”. In order to obtain an initial sample with explicit coordinate in pdb format suitable for the topology of the residue called “GLU” in such force field, some atom names have been modified: the atom that Pymol [98] labels H corresponds to a HN atom of the residue described in the topology files, 2HB has become HB2, 3HB has been renamed HB1, 2HG has its name changed in HG1 while 3HG has been modified in HG2 (Figure 4.66). The topology file contains patches for linking various residues together, keeping trace of the first and the last one. Since these parameter and topology files had been written for nucleic acids and amino acids, no further parametrization is needed.

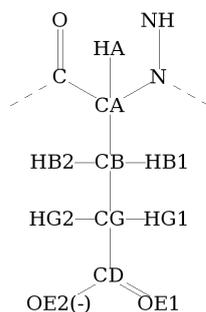


Figure 4.66: A residue of glutamate and atom labels.

Sodium and DMF

DMF has previously been optimized and parametrized (assigning to each atom the appropriate charge) with Gaussian: so the description of such residue has been embedded to the original topology file. Sodium has already been defined in the original topology.

Sample generating

Now, we setup an input file for generating a .pdb file for such system containing the helix of 36 residues, 36 Na^+ atoms and 1320 DMF molecules with “packmol” [113]. Such .pdb file has been used for generating a new .pdb file with an appropriate topology file. This file has to be built in pdf format for using the molecular dynamics (MD) engine NAMD [115] with the executable “psfgen” that is able to perform the atom types present in a topology file corresponding to determined force field parameters.

Figure 4.67 shows a snapshot of the initial sample.

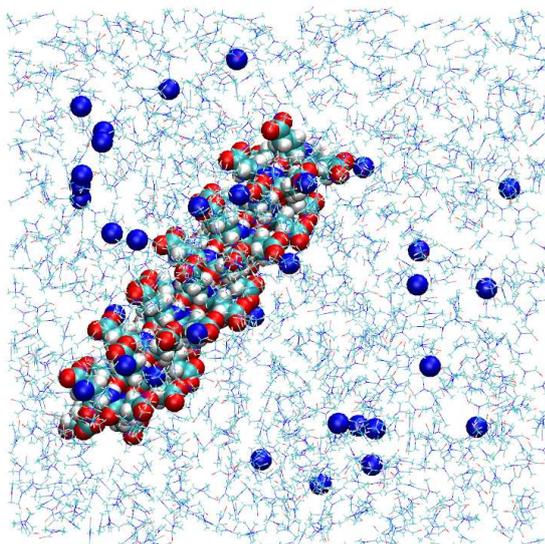


Figure 4.67: Snapshot of initial sample: 36 residues long poly–glutamate, 36 Na^+ ions and 1320 DMF molecules.

Setting up simulation

Simulations are run using “fixed atoms” for keeping all atoms of the glutamate helix fixed, pressure is set constant to 1.01325 and the *zeroMomentum* is not set, because the flag “fixed atoms” does not allow a control of the of atoms momenta. Simulations are characterized by the following features:

- timestep = 0.5 fs;
- total simulation time = 20 ns;
- T = 300 K (with a rescale velocities algorithm as thermostat);
- P = 1.01325 bar (Berendsen barostat);
- isotropic isobaric ensemble.

Following figure (Figure 4.68) shows how the system changes during the simulation.

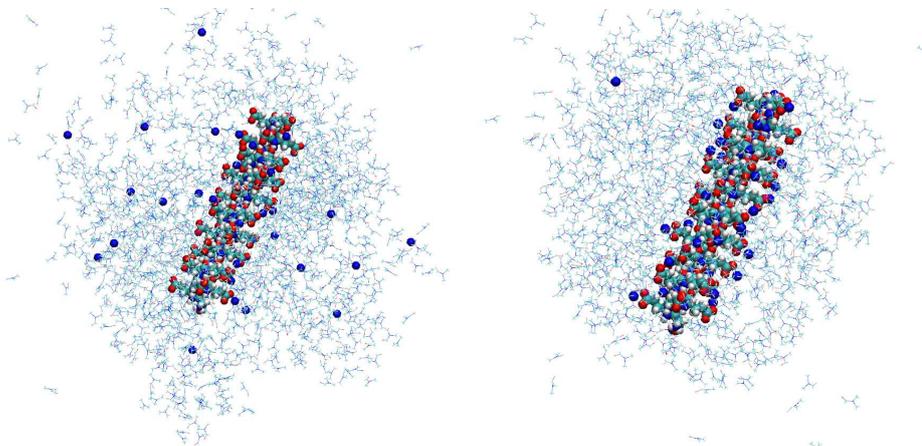


Figure 4.68: Snapshots of the sample at the beginning (*right*) and at the end of the simulation (20 ns, *left*).

4.6.6.2 Results

The following table (Table 4.15) sums up average energy, temperature, volume and density values for the system studied. Note that energy values are not normalized, since they refer to the whole sample.

Parameter	(Units)	Value
Bond energy	(kcal/mol)	4583.652080
Angle energy	(kcal/mol)	6728.316644
Dihedral energy	(kcal/mol)	1498.76209
Electrostatic energy	(kcal/mol)	-5913.712391
Van der Waals energy	(kcal/mol)	-4902.442294
Total energy	(kcal/mol)	16217.493045
Temperature	(K)	300.00282
Volume	(Å ³)	763284.511637
Density	(g/cm ³)	0.0102

Table 4.15: Parameters for a 20 ns long simulation of poly-glutamate in DMF.

Also other parameters are analyzed:

- radial distribution function $g(r)$ and the density one $g(z)$ for DMF-DMF, DMF-GLU and GLU-GLU interactions;
- aspect ratio histograms;
- orientational order parameters (first rank order parameter $\langle P_2 \rangle$ and molecular biaxiality $\langle R_{02}^2 \rangle$) for GLU;
- time correlation functions;
- displacement analysis;
- solvation sphere.

Diffusion coefficient is also computed: the result obtained is $0.72 \times 10^{-9} \text{m}^2/\text{s}$, while the literature value is $1.98 \times 10^{-9} \text{m}^2/\text{s}$ [89]. These values are almost in good agreement (on the contrary of what we had seen when treating the two helices of poly-benzyl glutamate).

Aspect ratio

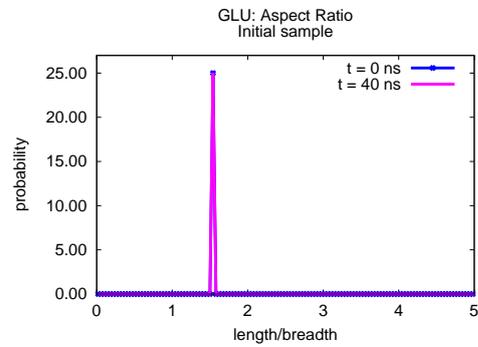


Figure 4.69: GLU aspect ratio: initial sample and sample after a simulation time of 20 ns.

The distribution of the aspect ratio does not become wider, since we have constrained all atoms belonging to the helix to maintain their positions fixed.

Order parameters

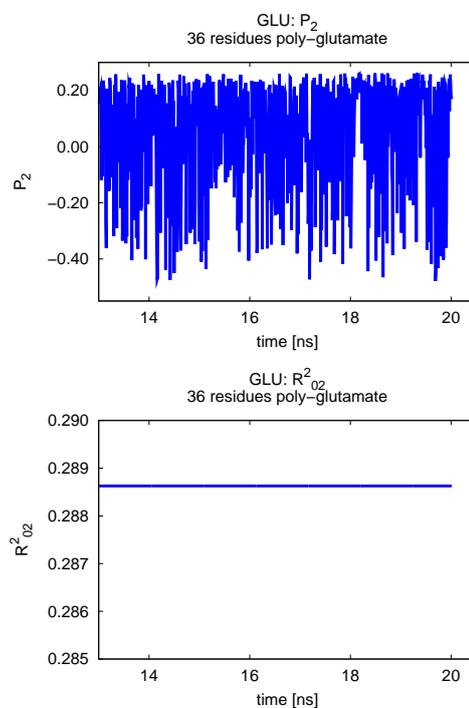


Figure 4.70: Order parameters computed with respect to GLU.

The orientational order parameter, P_2 , shows typical isotropic zero values. Even if helices are truly ordered, residues are arranged in kind a random way. Further analysis, in particular using an helicoidal order parameter, are needed (see further below). Molecular biaxiality R_{02}^2 shows high values typical of a biaxial molecule.

Time correlation functions

Since time correlation functions do not change, helix does not become spatially modified, which is not surprising since all its atoms are fixed.

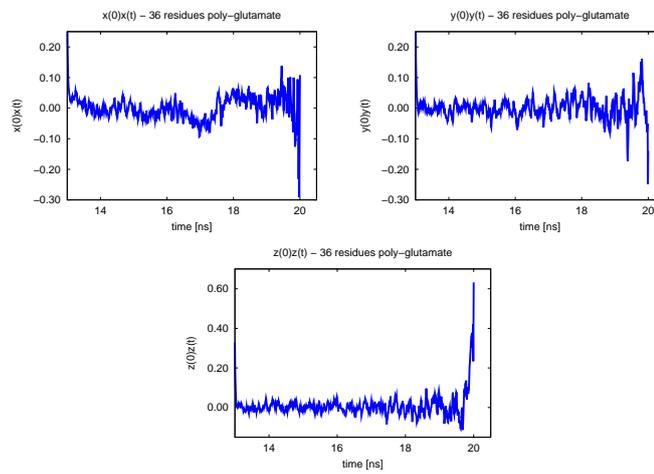


Figure 4.71: Time correlation functions: $x(0)x(t)$, $y(0)y(t)$, and $z(0)z(t)$.

Displacement analysis

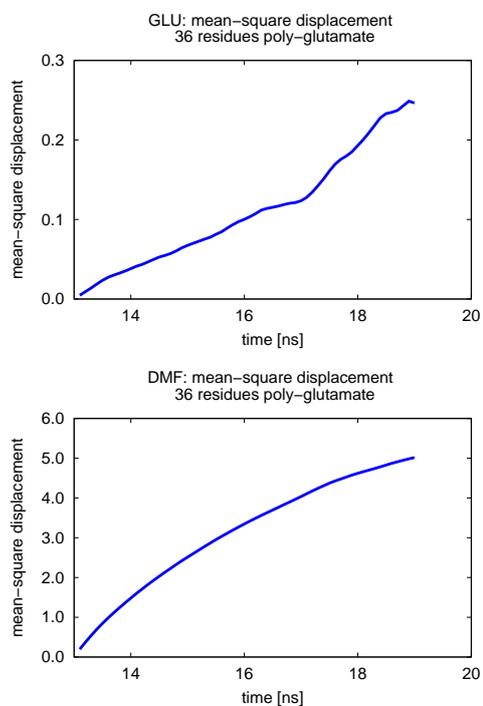


Figure 4.72: Displacement analysis.

As expected, the isotropic solvent presents a faster diffusion with respect to the poly-glutamate helix, independently to the laboratory frame orientation.

Solvation sphere

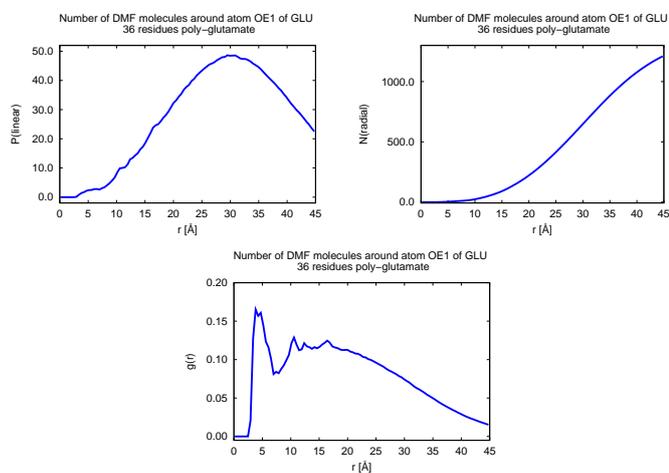


Figure 4.73: Solvation sphere.

Graphs show a first peak of solvation at about 5 Å. This parameter is computed with respect to one of the two more external oxygen atom (OE1) of glutamate residues.

4.6.6.3 Comparison with simulation of two PBLG helices in DMF

First of all, we should notice that samples are different for dimensions and solvent molecules number so energy values are not really comparable. However, energy terms values are of the same order of magnitude a part from the electrostatic energy which shows positive values for PBLG samples³ (1759–2268 kcal/mol) against the negative one typical of the poly-glutamate sample. The latter system expands a lot since its volume is almost four times that of the two helices PBLG systems (763284 Å³ for the poly-glutamate sample vs 192605 Å³ of the perpendicular PBLG helices one): hence, also densities are much lower in the case of the functionalized helices. For what concern solvation shells, they are placed almost at the same distance ($\sim 5\text{Å}$ and $\sim 10\text{Å}$) in all of the three samples. The $g(r)$ typical of the helix structure is much more structured for the poly-glutamate helix than for PBLG systems as show from $g(r)$ computed with respect to the interactions between residues of the helix itself. The distortion of the helix is removed by fixing the positions of atoms belonging to the helix, as underlined by the aspect ratio analysis. P_2 order parameter is about zero for all samples while R_{02}^2 (0.2885) is almost 50 times higher than those of PBLG systems (0.05–0.07). Displacement analysis shows a slower motion for the system with poly-glutamate (GLU = 0.01 Å/ns vs BGL = 0.05 Å/ns, DMF (GLU)= 0.26 Å/ns vs DMF(BGL) = 0.31 Å/ns). For what concern solvation sphere, the trend characteristic of the poly-glutamate sample is more structured with respect to those of PBLG samples.

³Two samples of PBLG in DMF have been studied: the first one is characterized by two parallel helices and the second one by two perpendicular helices.

4.6.6.4 Ramachandran plot

Next, we decide to analyze such proteic structure with Ramachandran Plot (Figure 4.74): results show that poly-glutamate structure is a perfect helix except for the first and the last residue.

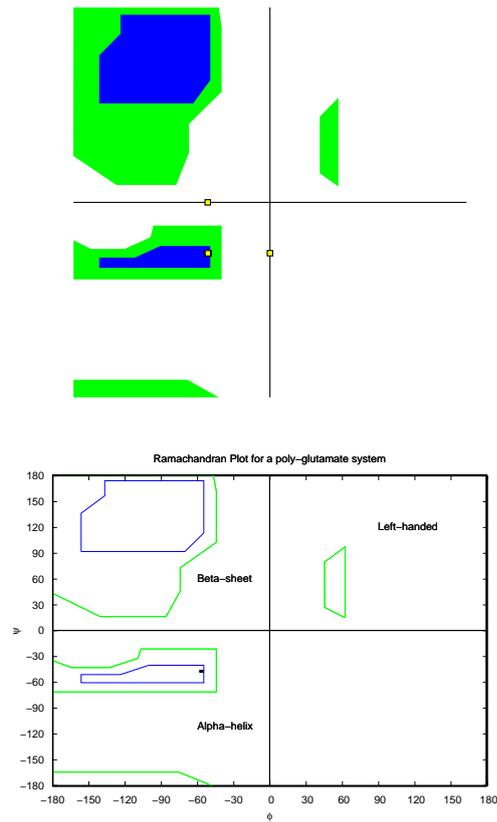


Figure 4.74: Ramachandran plot: computed with VMD (*left*) and with Outside (*right*).

4.6.6.5 Conclusions

Results are summarize below:

- DMF molecules do not solvate uniformly the helix, since various solvation shells are present;
- DMF arranges around GLU molecules in three solvation shells;
- $g(z)$ shows wave-like trend typical for helicoidal systems;
- aspect ratio distribution does not change during the simulation, since helix atoms are kept fixed;
- P_2 order parameter values are typical for isotropic phase, since the correct parameter would be an helicoidal one;
- R_{02}^2 introduces a biaxial behavior of the molecules;
- time correlation functions do not change, since helix does not distort;
- isotropic solvent presents a diffusion coefficient similar to the one reported in literature: they are of the same order of magnitude, even if tone is the double than the other, maybe due to the fact that the helix is not free to move;
- the first solvation sphere is at about 5 Å from the OE2 oxygen.

Moreover, from the trajectory it is clear that sodium ions get close as soon as possible to the negative charge of the oxygen atom.

4.6.7 PBLG systems: a chirality index investigation

4.6.7.1 Method

The assignment of secondary structure of proteins is a field of great interest since its importance in determining whether protein's function is undamaged and in new applications like drug design and design of novel enzymes. There are several studies based on both experimental and computational approach about predicting these structures starting from the sequence of amino acid, like I-TASSER [132]. A newly tool is the previously explained methodology [85] based on a chirality index. We applied this procedure to samples of one and two (parallel or perpendicular) helices of poly(γ -benzyl L-glutamate) (PBLG) in dimethylformamide (DMF) solution. The chiral index is described by the following equation:

$$G^{a,N_a} = \frac{4!}{3N_a^4} \times \sum_{\substack{\text{all permutation} \\ \text{of } i,j,k,l=1,\dots,N}} \left\{ \begin{array}{ll} \frac{[(\mathbf{r}_{ij} \times \mathbf{r}_{kl}) \cdot \mathbf{r}_{il}](\mathbf{r}_{ij} \cdot \mathbf{r}_{jk})(\mathbf{r}_{jk} \cdot \mathbf{r}_{kl})}{(r_{ij}r_{jk}r_{kl})^2 r_{il}} & \text{if } r_{ij}, r_{kl}, r_{il}, r_{jk} < r_c \\ & \text{and } a \leq i, j, k, \\ & l \leq N_a + a - 1 \\ 0 & \text{otherwise} \end{array} \right. \quad (4.5)$$

where i, j, k, l are four of the N_a atoms belonging to the sequence of connected atoms (backbone), r_{ab} are interatomic distance vectors, r_c is a cut-off radius to avoid the computation of unnecessary long-range terms, that give a negligible contribution to the overall chirality. The latter parameter should be greater than 10 Å in order to achieve the stability of G^{N_a} values [85]: the r_c value chosen is 12 Å. The N_a value that allows the best differentiation [85] of the secondary structure is 15, corresponding to five consecutive residues. Note that, since structural motifs represent a local property of a small group of amino acids, this index will be computed only for backbone atoms (N, C $_{\alpha}$ and C).

From published data [85], it is possible to describe structures like the right handed α -helix ($G^a = -0.05 / -0.04$), type I β turn ($G^a = -0.10 / -0.06$) and 3_{10} helix ($G^a = -0.09 / -0.07$) with negative chiral index values, the

left handed helix of poly-L-proline II with a positive sign of chiral index ($G^a = 0.10/0.12$), while the β sheet structures ($G^a = 0.00$), having a flat shape and symmetric ϕ and ψ dihedrals, with a chiral index close to zero, as well as the π helix ($G^a = -0.01/0.00$).

4.6.7.2 Results

Poly(γ -benzyl L-glutamate)

Since the helix composed by 18 residues (Figure 4.76, *left*) has been created in order to be an α -helix (as confirmed in Figure 4.75), the expected G^a value should be around $-0.05/ -0.04$. But this supposition has not been confirmed (Figure 4.76, *right*): the G^a value ranges from 0.01 at one end and 0.09 at the other one. Moreover, all residues should have almost the same value of G^a , without any drift contrarily to results obtained from our calculations.

α_{144} **PBLG**

Results (Figure 4.77) show that the average value of G^a typical of such a long helix is ~ 0.045 , which is in good agreement with values previously obtained for systems with two parallel or perpendicular helices (see section: "PBLG systems: a chirality index investigation"), even if it is typical for a PPII structure (an helix with an opposite handedness with respect to the α helix).

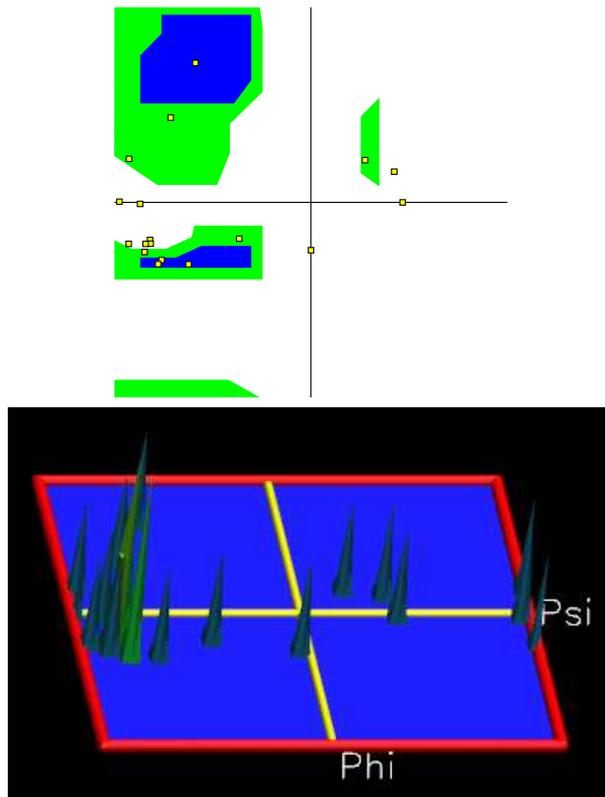


Figure 4.75: Ramachandran plot for single PBLG helix: on the left, there is the plot itself, while, on the right, a 3D histogram is shown. Both images are created using VMD [106]

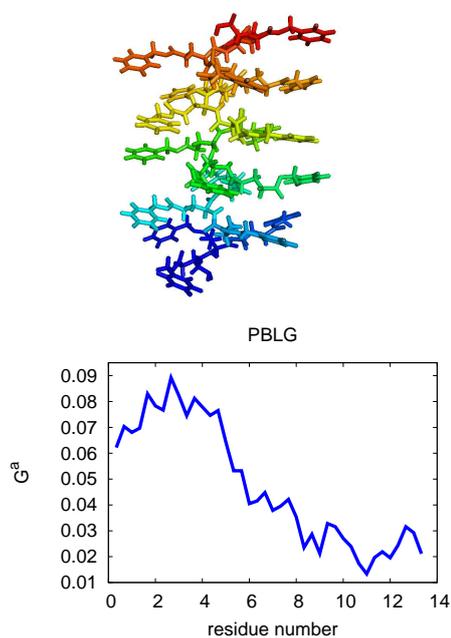


Figure 4.76: A single PBLG helix: on the left, a snapshot of the molecule and, on the right, the chiral index (G^a) values, along the backbone, for this PBLG α -helix.

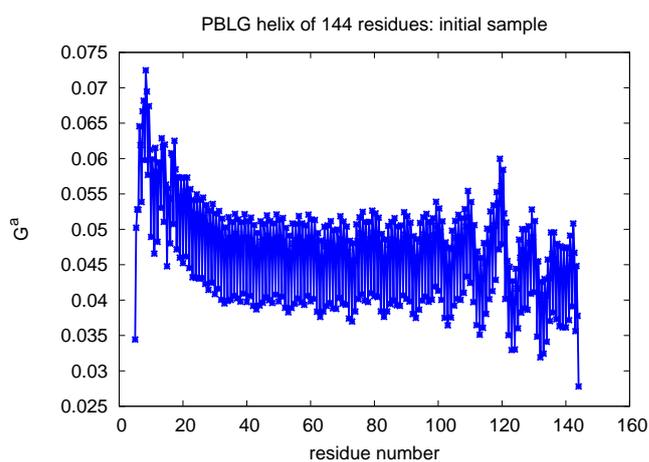


Figure 4.77: Analysis of a 144 residues long PBLG helix with a chiral index used for protein secondary structures investigation [85].

Sample with two parallel PBLG helices

In Figure 4.78 there is a snapshot of the system with two parallel PBLG helices, without DMF molecules. The chirality index trend is almost symmetric, presenting two peaks at the extremes of the plot, while two smaller ones are placed in the middle of this.

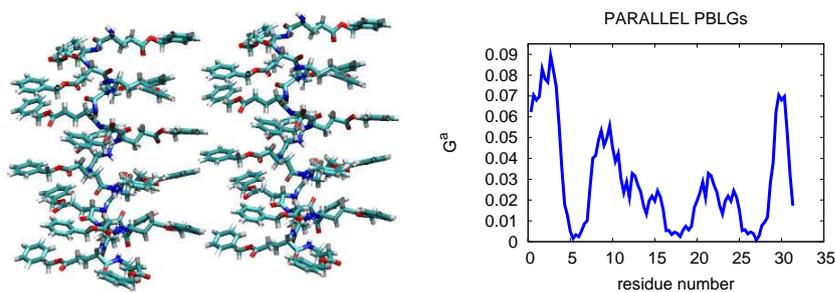


Figure 4.78: A sample with two parallel PBLG helices: (on the left) a snapshot of this system and (on the right) the chiral index (G^{α}) trend, along the backbone, for these PBLG α -helices.

Sample with two perpendicular PBLG helices

Figure 4.79 reports a snapshot of the system with two perpendicular PBLG helices (*left*) and the chirality index behavior (*right*) that shows a sort of drift which is repeated twice, a time for each helix.

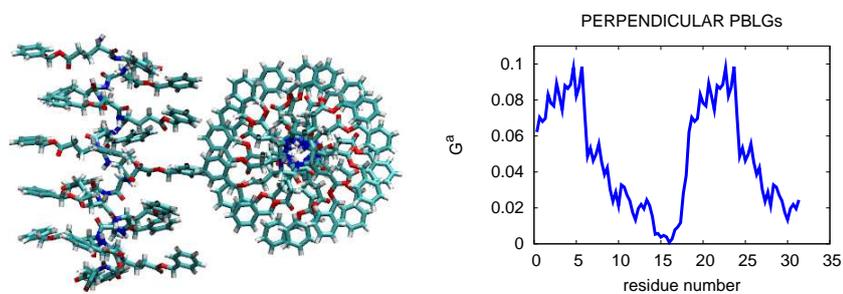


Figure 4.79: A sample with two perpendicular PBLG helices: (on the left) an snapshot of this system and (on the right) the chiral index (G^{α}), along the backbone, for these PBLG α -helices.

Parallel system: time analysis

The following figure (Figure 4.80) shows how chirality index varies along the backbones of the two helices at different time.

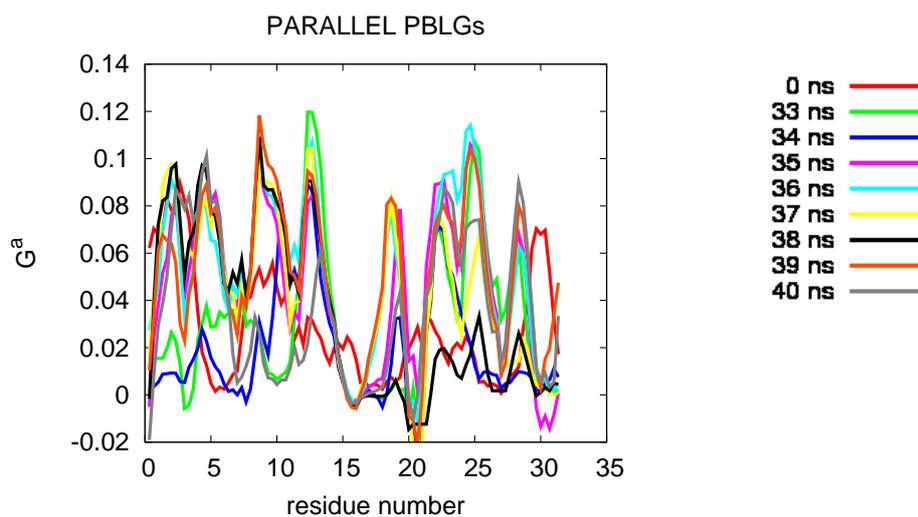


Figure 4.80: Parallel helices system: chiral index trend.

All chirality indexes show positive values, ranging from the values typical of a poly-proline II (PPII) structure and of a π -helix.

It is also possible to observe how chirality index changes for each residue, as a function of simulation time. The code used (“chirality.f90”) analyzes the relationship between five backbone atoms ($N_a = 5$), so the number of residue is decreased of such value. Thus residue will go from 1 to 18 for what concern the first helix and from 19 to 31 for the second one, even if values characteristic of residues between the first and the second one have not to be considered (since they are not really spatially related).

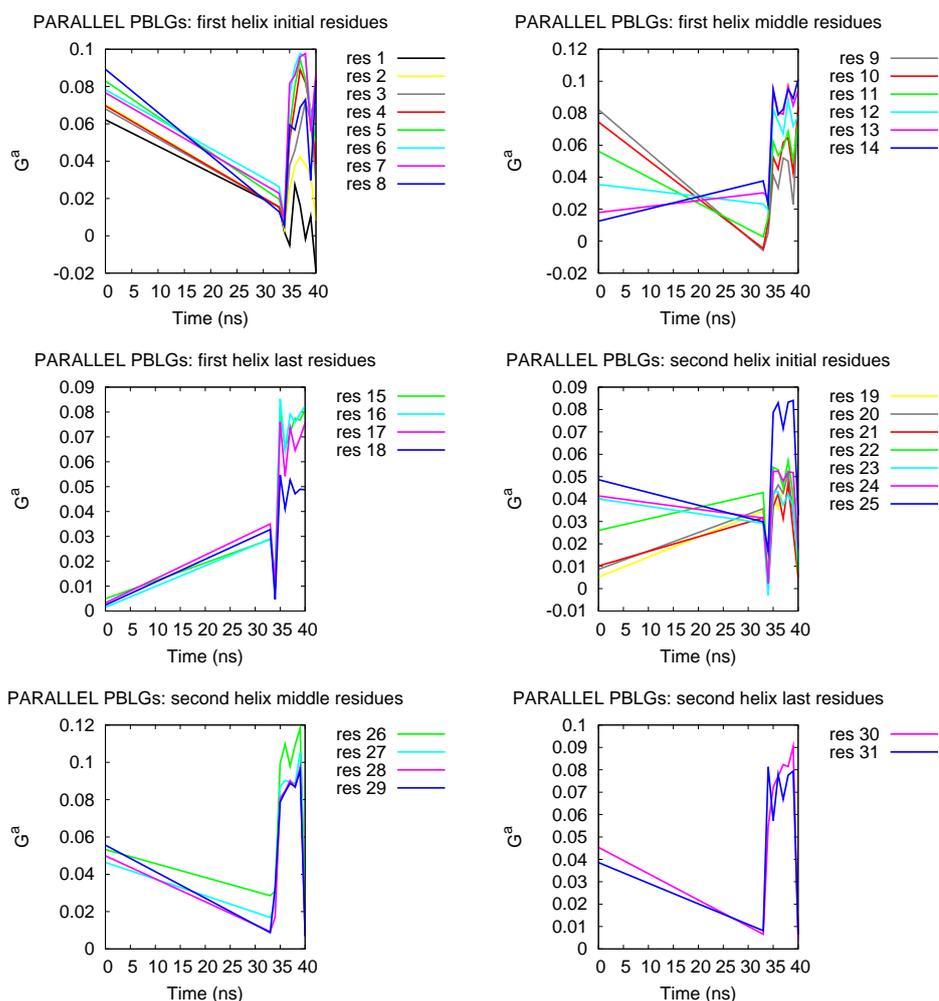


Figure 4.81: Parallel helices system: chiral index variation for each residue, depending on simulation time.

Perpendicular system: time analysis

The following figure (Figure 4.82) shows how chirality index varies along the backbones of the two helices at different analysis time.

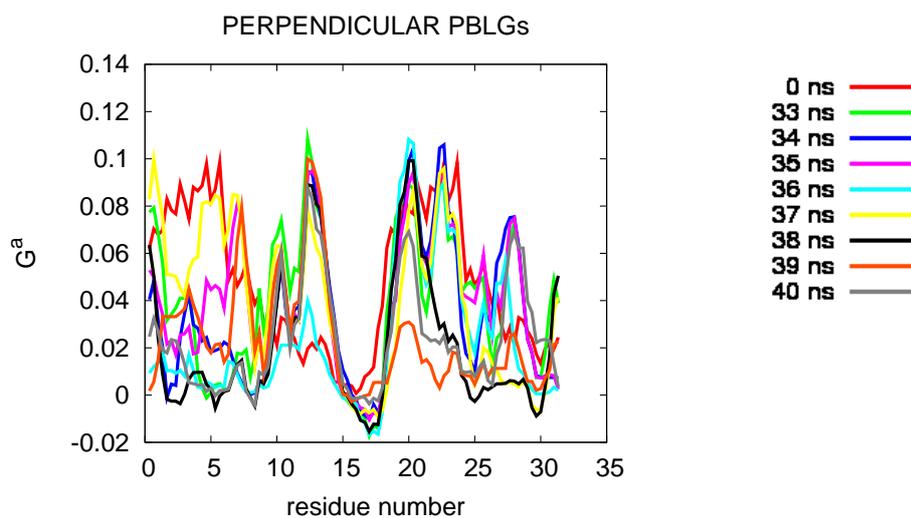


Figure 4.82: Perpendicular helices system: chiral index trend.

This system show positive chiral index values, a part for residue at the end of the first helix (residues 16, 17 and 18) and at the beginning of the second one. However, these residues have not to be considered since they belong to different molecules, not spatially related.

It is also possible to observe how chirality index changes for each residue, as a function of simulation time.

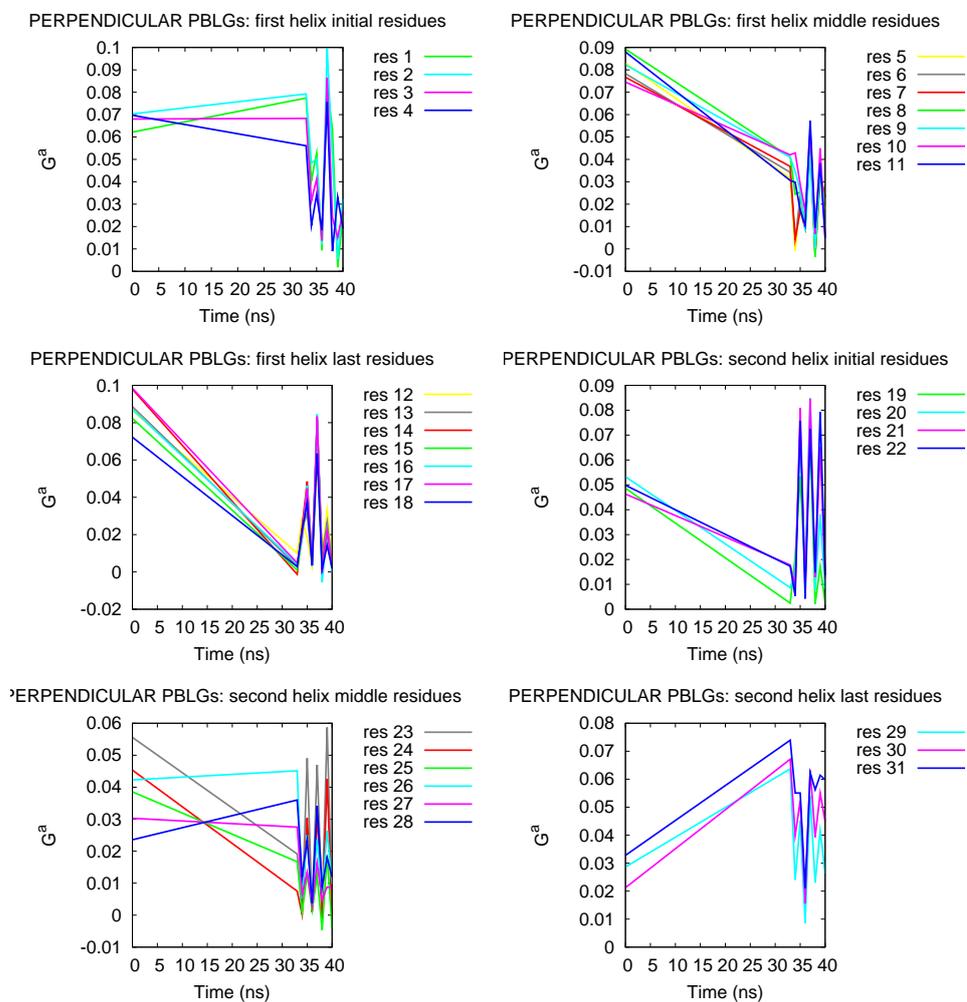


Figure 4.83: Perpendicular helices system: chiral index variation for each residue, depending on simulation time.

Molecular shape analysis

Moreover, we run a molecular shape analysis in order to define if helices are deformed during the simulation. We analyze the aspect ratio (the ratio between the length and the breadth of a molecule) distribution.

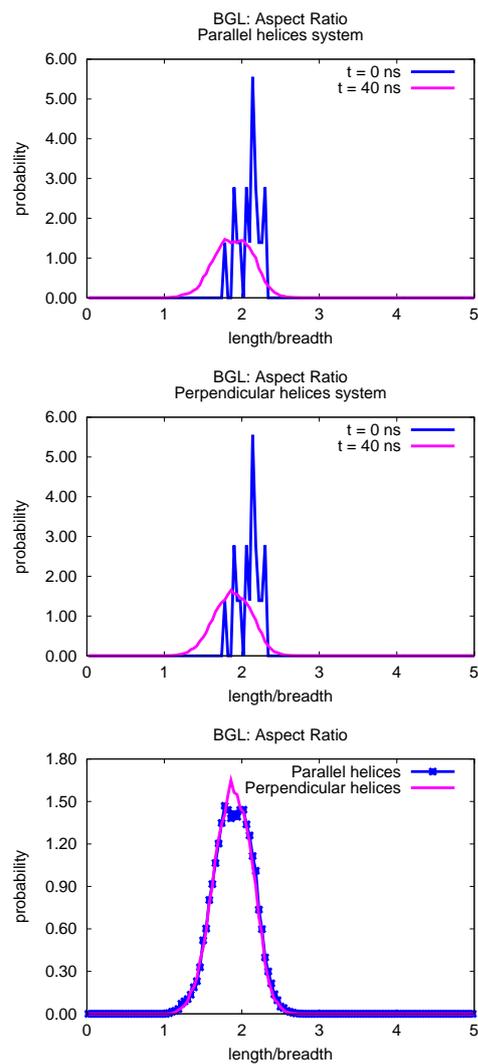


Figure 4.84: Analysis of helices deformation. Comparison between aspect ratio computed at $t = 0$ ns and $t = 40$ ns for each system. The last graph shows the difference of aspect ratio between the system with parallel helices and the one with perpendicular molecules at $t = 40$ ns.

Both systems, after a simulation time corresponding to 40 ns, show helices deformation. The peak is not narrow enough to affirm helices are stable. What is more, comparing initial and last samples, the central peak of aspect ratio distribution is shifted toward smaller values of such parameter: helices do undergo deformation.

4.6.7.3 Validation

Other α -helix

In order to discover if the way of building our molecule was correct, we built other three α -helices: poly-alanine, -glutamate and -proline.

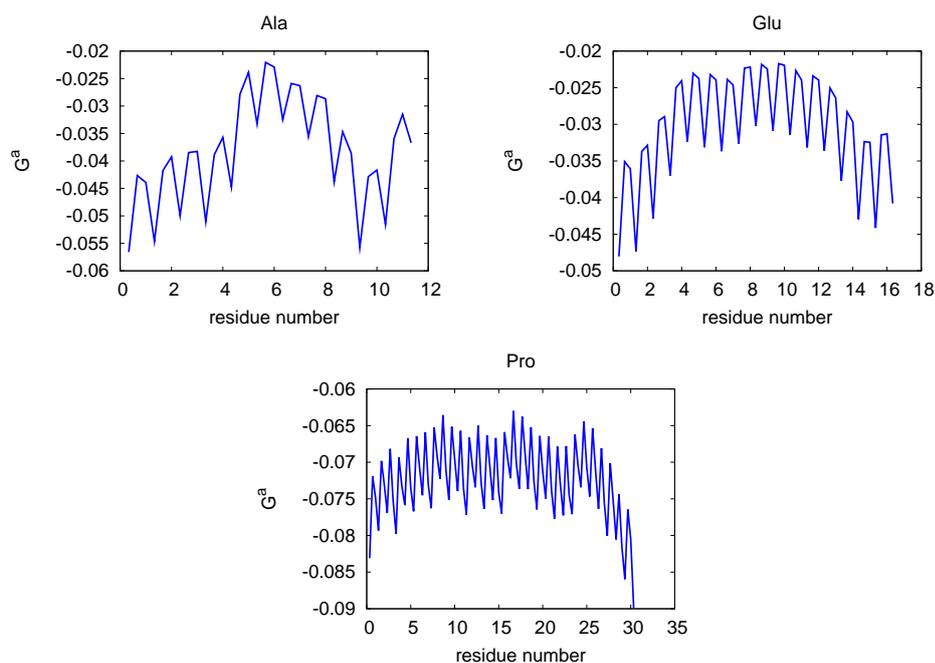


Figure 4.85: G^a values for poly-alanine, -glutamate and -proline.

Alanine and glutamate helices show chiral index values corresponding to those typical of an α -helix ([85]), while poly-proline shows more negative data, maybe due to the fact that proline should acquire an opposite handedness (PPII) but we forced it to be an α -helix. From these data, we could affirm that the method we built our molecules with was consistent.

Parametrization of N_a and r_c

As previously explained, N_a represents the number of atoms belonging to the sequence of connected atoms (backbone) and r_c is a cutoff radius chosen to avoid the computation of unnecessary long-range terms. We also tried to modify these parameters in order to see if chirality index values will change. Figure 4.86 shows results of such attempts.

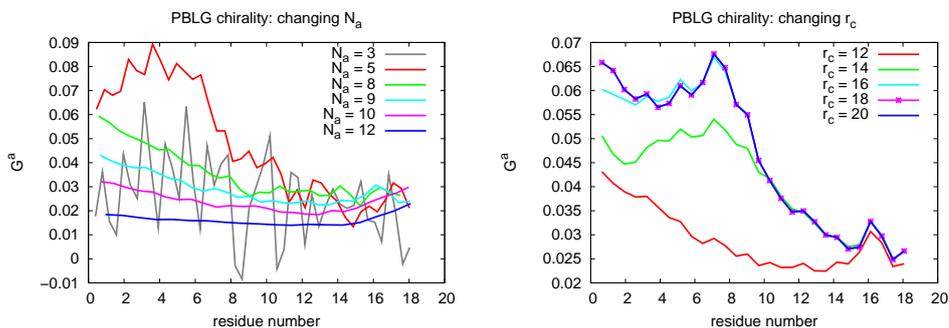


Figure 4.86: G^a values modification depending on N_a and r_c .

No modifications of such parameters force chiral index values to be negative. Moreover, the values of N_a and r_c suggested in the article [85] are those that keep the G_a values lower. Thus, we decide to maintain the suggested ones.

Ramachandran plot

To confirm the structure of BGL molecules in our systems, we also analyze Ramachandran plot for systems at different time, using “Outside”, a code developed by our group. Figure 4.87 shows Ramachandran plots for initial samples of both systems: points distribution is almost the same, but there is not a large majority of angles typical of α -helix, whose population is comparable to that of the β -sheet region. When simulations start, all points tend to have ψ equal to 0, +180 or -180. This is probably due to the fact that helices distort. Since such disruption of the initial molecular geometry occurs almost immediately, we only reported Ramachandran plots for the last frame of each simulations (see Figure 4.88).

In conclusion, a single BGL helix, as explained before, is defined as an α -helix using Ramachandran plot, while systems composed by two helices are not well-described using such method: results shows that initial structures could be both α -helix or β -sheet. Suddenly, systems distort and the ψ angle tends to 0.0, +180.0 or -180.0.

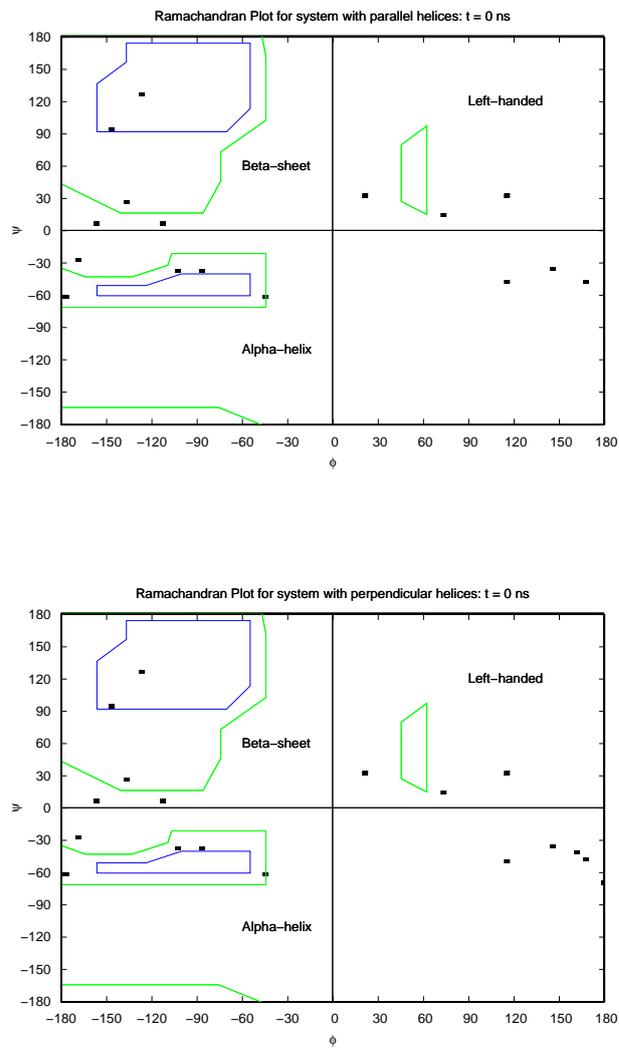


Figure 4.87: Ramachandran plot for initial samples of parallel and perpendicular systems.

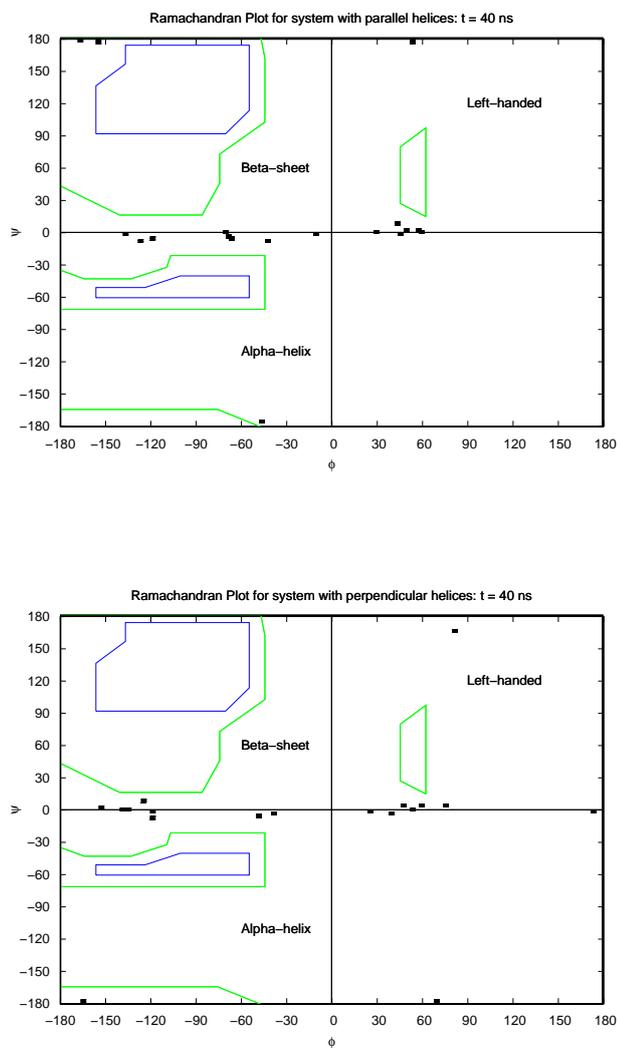


Figure 4.88: Ramachandran plot for parallel and perpendicular systems at $t = 40$ ns.

4.6.7.4 Conclusions

The first important feature is the positive value of G^a of a single PBLG helix and the drift that this parameter shows when moving along the backbone (Figure 4.76). A positive value is characteristic of a PPII helix (or a π -helix, when values tend to zero, but still remaining positive): so we tested, using VMD [106], the handedness of our helix, which turned out to be an α -helix (Figure 4.75) as it should be. In order to explain this failure of the code, we looked up closely to all files, codes and methodologies used: the format of the input files (parameters and configuration .pdb files), the order in which backbone atoms appear (which is the same as the .pdb file found on line at Protein Data Bank site, www.pdb.org, and always respected: N, C_α , C), the uselessness of comment preceding lines beginning with “ATOM” (the only lines considered by the code), the justness of calling N the atom type previously defined as N1, C the C7 atom type and CA (C_α) the C2 atom type (Figure 4.89).

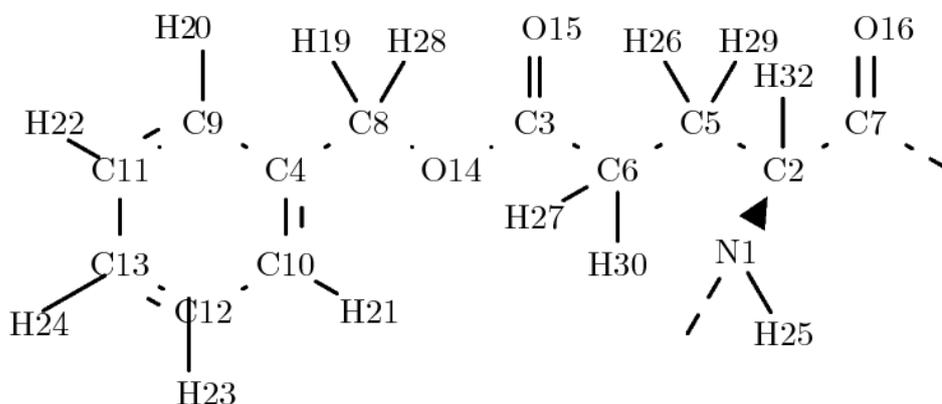


Figure 4.89: Atom label of benzyl glutamate. C2 is the carbon atom attached at the first (α) position, C7 and N1 are the two atoms which connect other residues.

At a first glance, a possible solution for the positive G^a values for an α -helix structure could reside in the fact that periodic boundary conditions were used during the simulation: it may be hard for the code to recognize an helix which is split up in more segments (Figure 4.90). But, actually,

we find that this could not be the cause of the incorrect sign of G^a values, since starting configurations are not afflicted by PBC but chirality index values are still positive.

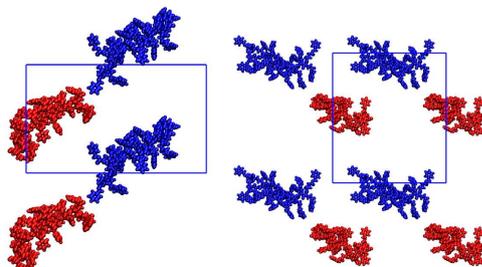


Figure 4.90: Snapshot of the effect of periodic boundary condition: on the left, an image of parallel system after a simulation of 40 ns, on the right a snapshot of the perpendicular system after 40 ns.

Chirality index analysis of starting configurations shows that perpendicular helices seem not to be afflicted by the near molecule: the plot simply shows two subsequential drifts, one for each helix. On the contrary, the chirality of the other system is modified by the nearness of two parallel helices.

For what concern the time evolution analysis for a parallel system, the main feature is the rise of peaks, which become higher and more well-defined. Highest peaks shift on the left side of the plot (first helix region). Looking at residues' time evolution, it is clear that all residues at the beginning of each helix have a chirality index which changes a lot, while at the end of each helix this value tends to reach a constant value (peaks are smaller).

Perpendicular system, on the contrary, starts with a drift (from highest to lowest values) repeated twice, one for each molecule, as previously described. After 10 ns, all highest peaks are in the first part of the plot. Some positive maxima will appear also in the second helix section. Residues at the beginning of helices show the highest peaks, while the ones at the end of these helices the lower ones (G^a index has almost become stable). For the second helix the plateau value is reached earlier (the first helix has

stable values for residue placed at the middle of the molecule). Note that when index becomes stable, it also reaches smaller values: the last three residues of both helices show (small) negative values.

Finally, we have analyzed helices' deformations: from aspect ratio of BGL helices: these structures change during simulations, since distribution is not narrow enough and the peak shifts toward smaller aspect ratio values. In conclusion, helices are too near, they undergo deformation. Another possible problem could be the shortness of our molecules (even if other shorter helices gives appropriated chiral index values, as seen in one the "in-depth analyses"). In conclusion, a bigger sample is needed. A possible solution is to set up a simulation with four helices composed by 36 monomers each, keeping them more distant than the ones of the samples just explained. We analyze literature and decide to choose an average value of concentration: our new sample will be made up by 144 molecules of PBLG (4 helices of 36 monomers each), 1456 DMF and 203 menthol. In order not to the helices deform (as confirmed by Ramachandran plot, see "Appendix"), the terminal atoms of the backbone should be fixed, not only constrained.

4.6.8 Helicoidal parameter

Some structures could not be studied with the classic order parameters. For example, helices show an average value of the first rank order parameter ($\langle P_2 \rangle$) equal to zero, since the average distribution of residues composing such structure seems “casual” to such parameter. So, we decide to use another index suitable for helical disposition of particles, previously used in the analysis of the chirality of a nanodroplet. This chiral index is an orientational parameter, on the contrary of the G^a index which is positional.

4.6.8.1 Helicoidal order parameter

The helicoidal order parameter is an average chirality index $\chi = \langle \hat{\mathbf{z}} \cdot (\hat{\mathbf{r}} \times \hat{\mathbf{u}}) \rangle$, where the three unit vectors are: $\hat{\mathbf{z}}$ the direction along which the helix is built $\hat{\mathbf{r}}$ the orientation of the a radial position vector in cylindrical coordinates for each residue and $\hat{\mathbf{u}}$ the orientation of each residue. Figure 4.91 explains how this order parameter is computed.

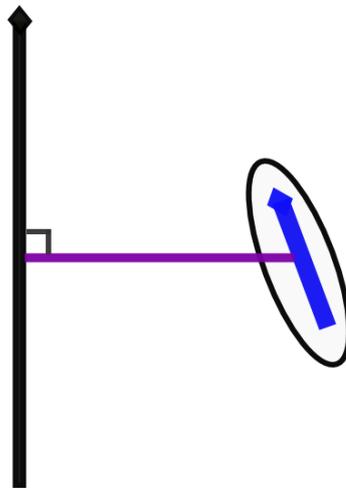


Figure 4.91: How the helicoidal order parameter is computed: $\hat{\mathbf{z}}$ is the black versor, $\hat{\mathbf{u}}$ is the blue one, while the purple one is $\hat{\mathbf{r}}$.

4.6.8.2 Computational details

Setting up an helicoidal chiral index needs to pay a particular attention. First of all, the helix should be translated to its center of mass, whose coordinates are subtracted to those of the atoms.

The $\hat{\mathbf{z}}$ -axis is computed simply by evaluating the inertia tensor of all atoms belonging to the molecule, diagonalizing such matrix and choosing the eigenvector characterized by the biggest eigenvalue.

$\hat{\mathbf{u}}$ is needed to be chosen in such a way that the versor itself and $\hat{\mathbf{z}}$ have the same direction (i.e. $\hat{\mathbf{u}} \cdot \hat{\mathbf{z}} > 0$).

Moreover, computing $\hat{\mathbf{r}}$ in Cartesian coordinates and then transpose them into cylindrical ones is a big source of error. So, we decide to find it as the difference between other two vectors ($\mathbf{r} = \mathbf{d} - (\mathbf{d} \cdot \mathbf{z})\mathbf{z} = \mathbf{d} - \mathbf{c}$) and then normalize it, as shown in Figure 4.92.

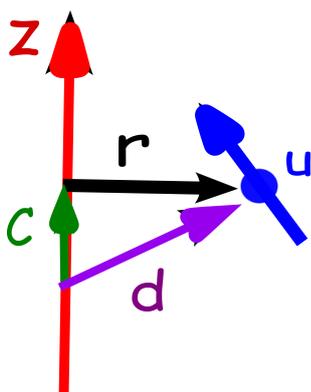


Figure 4.92: The easiest way to compute $\hat{\mathbf{r}}$.

4.6.8.3 PBLG: lateral chain helicoidal parameter

This index is set up for poly-(γ -benzyl L-glutamate), PBLG, finding first of all the \hat{z} axis, which is the eigenvector corresponding to the biggest eigenvalue of the inertia matrix (i.e. the axis along which the helix evolves). The \hat{r} versor is computed as the direction between the C9 and the C10 atoms, while the \hat{u} versor is computed between the center of mass of the carbon atoms belonging to the benzene ring and the \hat{z} -axis (Figure 4.93).

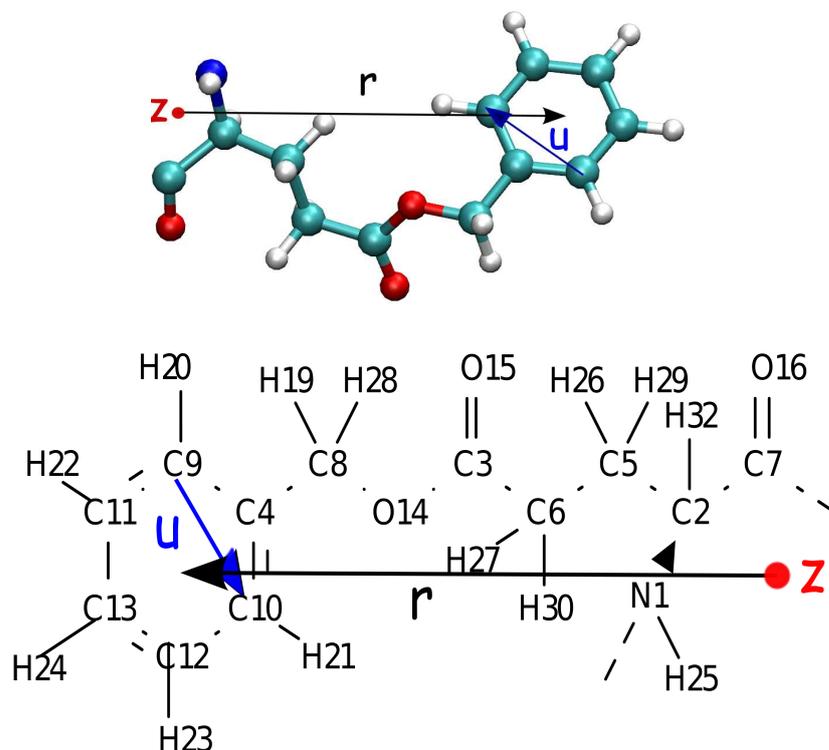


Figure 4.93: How to compute the helicoidal parameter for a BGL residue.

α_{18} PBLG helix

We analyzed an helix 18-residues long. The average value for such a molecule is $\chi = -0.571141 \pm 0.346811$: Figure 4.94 (*left*) shows the trend of the helicoidal chiral index along the helix.

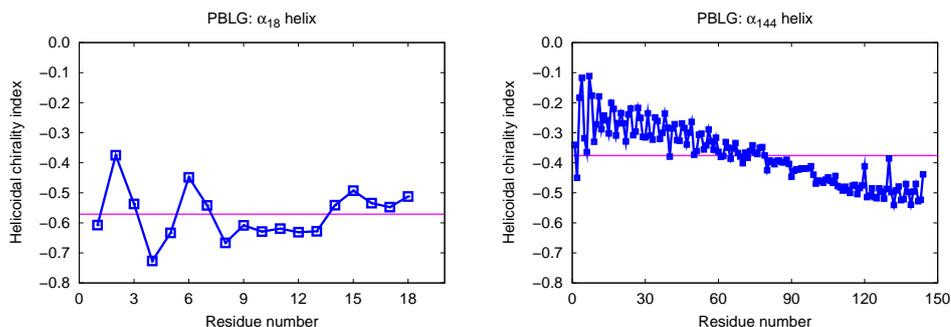


Figure 4.94: The trend of helicoidal chiral order parameter along the α_{18} helix (*left*) and along the α_{144} helix (*right*).

 α_{144} PBLG helix

Also an helix composed by 144 residues is analyzed (Figure 4.95). The average value for this molecule is -0.376152 ± 1.15133 : Figure 4.94 (*right*) shows the trend of the helicoidal chiral index along the helix.

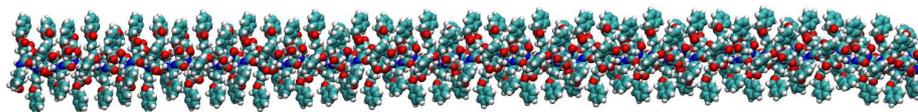


Figure 4.95: An image of a α_{144} helix of PBLG: the structure seems almost perfect.

System with two α_{18} PBLG helices (parallel and perpendicular)

The time evolution of the average of this helicoidal chiral index for the system with two PBLG parallel helices and for that with two perpendicular helices is showed in Figure 4.96.

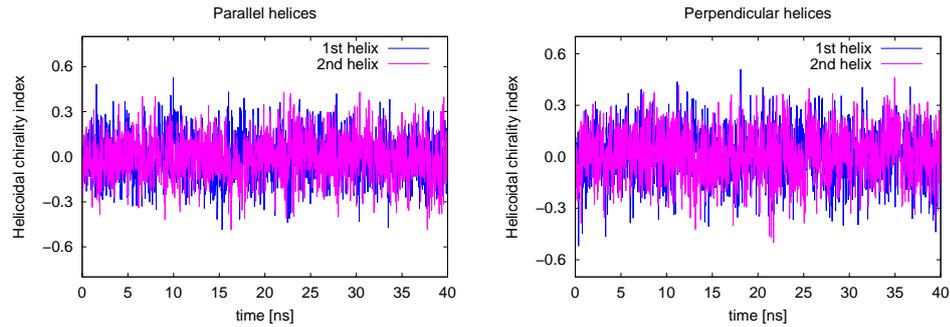


Figure 4.96: The time evolution of the average of the helicoidal chiral order parameter for system with two (perpendicular or parallel) helices.

It is clear that systems loose their helicoidal structure, typical of a perfect α_{18} helix ($\chi = -0.571141 \pm 0.346811$), since the index values immediately tends to zero.

4.6.8.4 PBLG: backbone helicoidal parameter

The previous index is also modified (χ_B) in order to take into account the backbone structure (Figure 4.97). The \hat{z} axis is computed as previously described while the \hat{r} versor is the direction between the C2 and the C7 atoms and the \hat{u} versor that between the center of mass of these two carbon atoms and the \hat{z} -axis (Figure 4.97). This kind of order parameter does not consider lateral chain of molecule, but only the backbone.

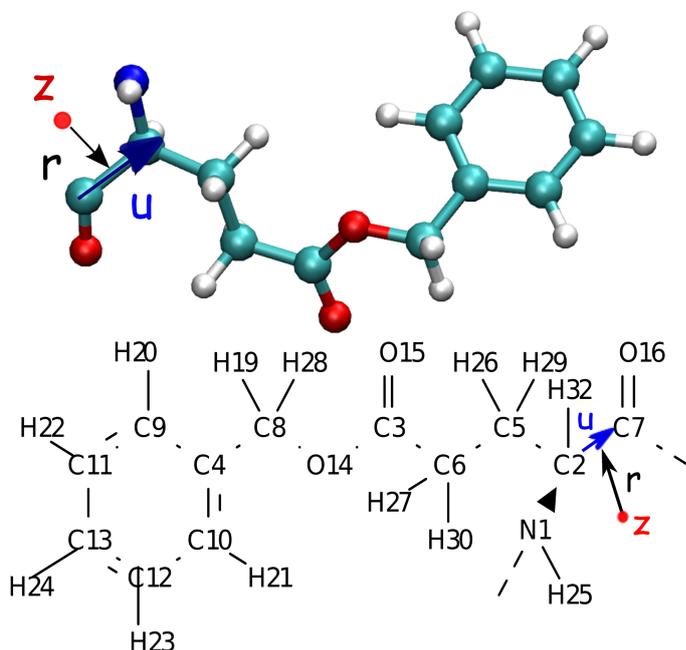


Figure 4.97: How to compute the helicoidal parameter for a BGL residue.

α_{18} and α_{144} helix

In this case, the χ_B value is -0.574001 ± 1.95921 for the shortest helix and -0.934066 ± 1.11017 for the longer one. Figure 4.98 sums up the backbone chiral index trend along the helix structure for both molecules.

As seen for the helicoidal index related to lateral chains, the parameter tends to decrease and stabilize increasing the length of the helix. In this case, the value for a α_{144} helix is very low (tends to -1.0), so the structure is almost perfect.

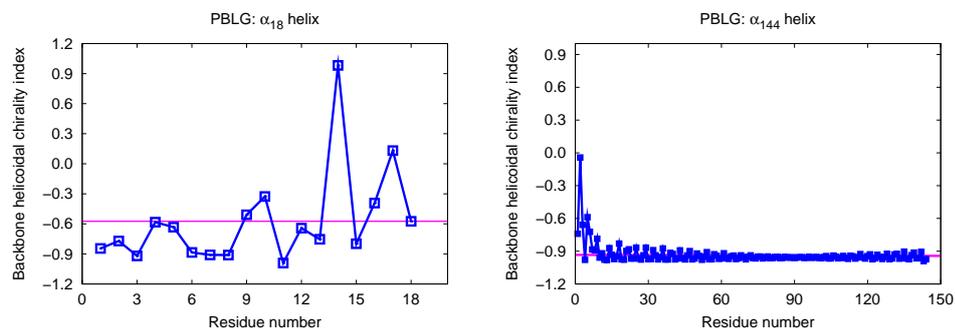


Figure 4.98: The trend of backbone helicoidal chiral order parameter along the α_{18} helix (*left*) and along the α_{144} helix (*left*).

System with two α_{18} PBLG helices (parallel and perpendicular)

The time evolution of the average of this new backbone helicoidal chiral index for the system with two PBLG parallel helices and for that with two perpendicular helices is showed in Figure 4.99.

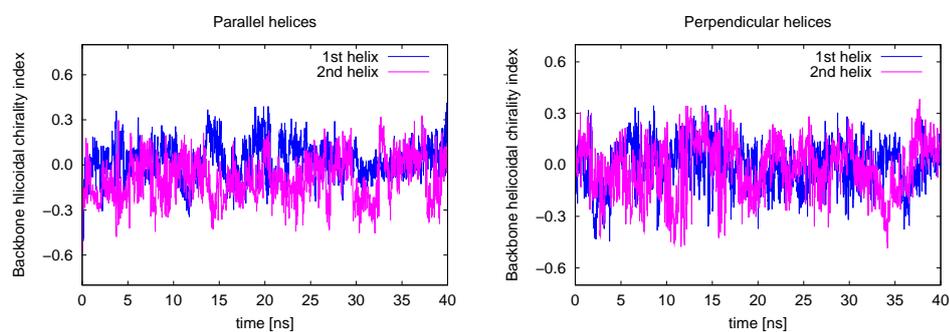


Figure 4.99: The time evolution of the average of the helicoidal chiral order parameter for system with two (perpendicular or parallel) helices.

Also this parameter tends to zero, since structures distort a lot.

4.6.8.5 Poly-glutamate: lateral chain helicoidal parameter

This index is modified also for analyzing a poly-glutamate helix ($\chi_{polyglu}$). The \hat{z} axis is always the eigenvector corresponding to the biggest eigenvalue of the inertia matrix (i.e. the axis along which the helix evolves). The \hat{r} versor is computed as the direction between the OE1 and the OE2 atoms, while the \hat{u} versor is computed between the center of mass of these two oxygen atoms and the z-axis (Figure 4.100).

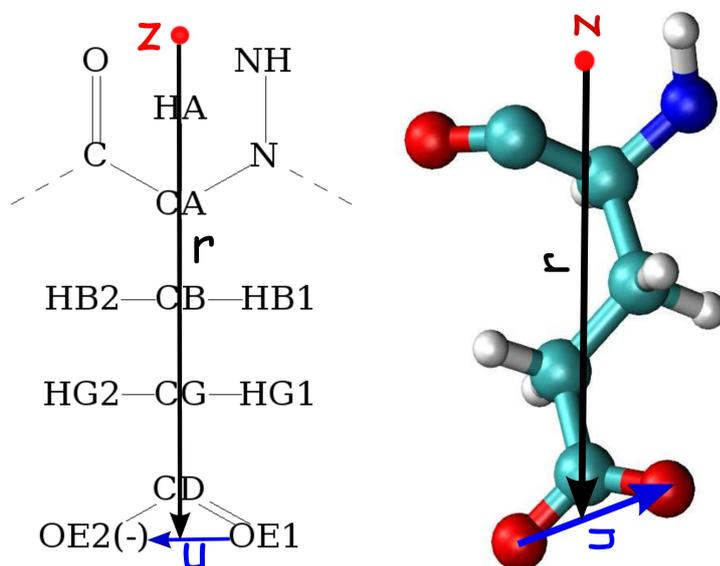


Figure 4.100: How to compute the helicoidal parameter for a GLU residue.

A single α_{36} helix of poly-glutamate shows an average chiral index value of $\chi_{polyglu} = 0.0564239 \pm 0.0738306$, which is slightly positive on the contrary of χ value typical for PBLG molecules, maybe due to the fact that it is computed in a different way (since lateral chains differ). Figure 4.101 shows the value of $\chi_{polyglu}$ for each residue of the helix at $t = 0$ ns and at $t = 35$ ns, that have exactly the same trend depending on the fixation of atomic positions.

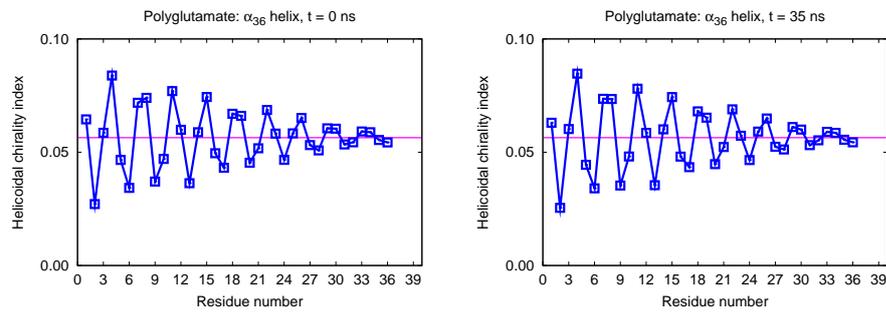


Figure 4.101: Values of chiral index along the helix, a $t = 0.0\text{ns}$ and $t = 35.0\text{ns}$.

4.6.8.6 Poly–glutamate: backbone helicoidal parameter

This index is modified also for analyzing the backbone of the poly–glutamate helix ($\chi_{B,polyglu}$). It is computed in the same way it has been done for the backbone of PBLG helix, since all protein have the backbone structure in the common; so this parameter could actually be applied to all proteins structures. The $\hat{\mathbf{z}}$ axis is still the eigenvector corresponding to the biggest eigenvalue of the inertia matrix (i.e. the axis along which the helix evolves). The $\hat{\mathbf{r}}$ versor is computed as the direction between the C and the CA atoms, while the $\hat{\mathbf{u}}$ versor is computed between the center of mass of these two carbon atoms and the $\hat{\mathbf{z}}$ –axis (Figure 4.102).

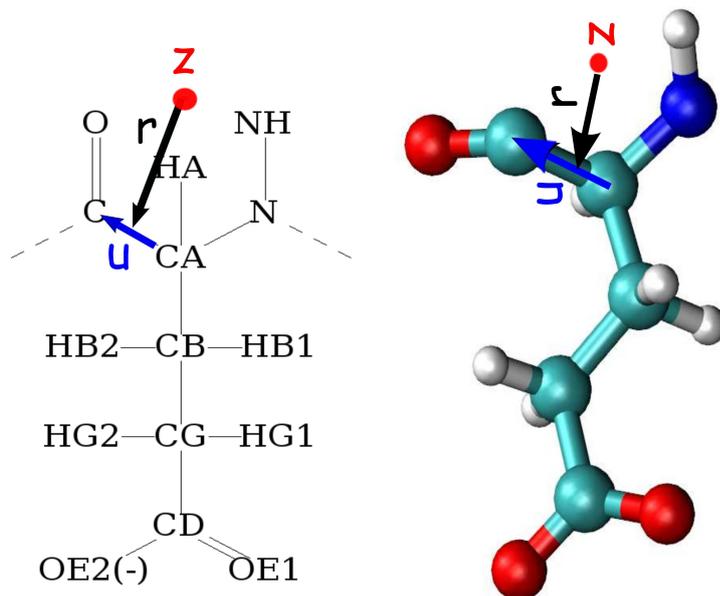


Figure 4.102: How to compute the backbone helicoidal parameter for a GLU residue.

The average value of the backbone chiral index for a poly-glutamate molecule is $\chi_{B,polyglu} = -0.630456 \pm 0.202224$. This value is constant during the simulation, as shown in Figure 4.103, since all atoms of such helix are kept fixed at their positions. Such average value is in good agreement with that of a perfect α_{18} helix of PBLG.

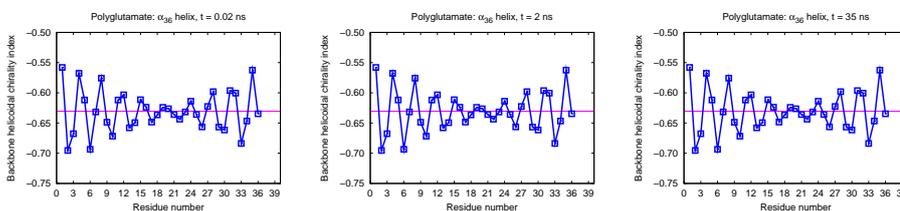


Figure 4.103: Values of backbone chiral index value along the helix, a $t = 0.02$ ns, $t = 2.0$ ns and $t = 35.0$ ns.

4.6.8.7 HEP: solute helicoidal parameter

In order to try and understand better how solute molecules interact with the helix, another version of the previous helicoidal parameter is set up. The \hat{z} axis is computed as previously described (the orientation of the helix itself) while the \hat{r} versor is computed as the direction between the center of the PBLG helix (namely the \hat{z} -axis) and the C7 atom of the HEP molecule. The \hat{u} versor is computed between the C7 atom of HEP, considered as the center of mass of the solute molecule, and the last atom of the carbon chain of HEP, C11 (Figure 4.161). Only HEP molecules nearer than 10 Å from the PBLG helix are considered.

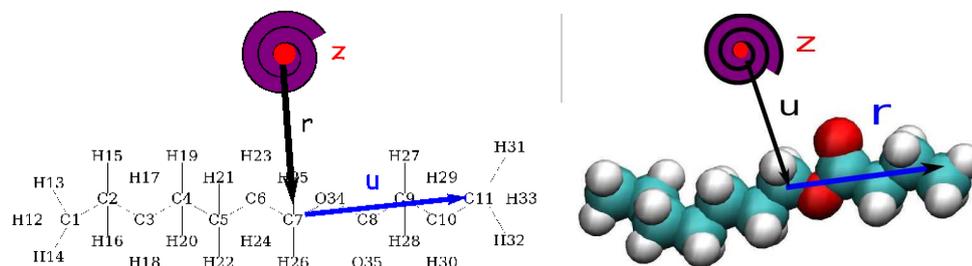


Figure 4.104: How to compute the helicoidal parameter for a HEP residue.

In this case, the χ_H value is -0.025848 ± 6.14469 . Figure 4.105 sums up the HEP chiral index values for all solute molecules. Note that the average value is almost zero, since in the initial configuration HEP molecules are randomly distributed.

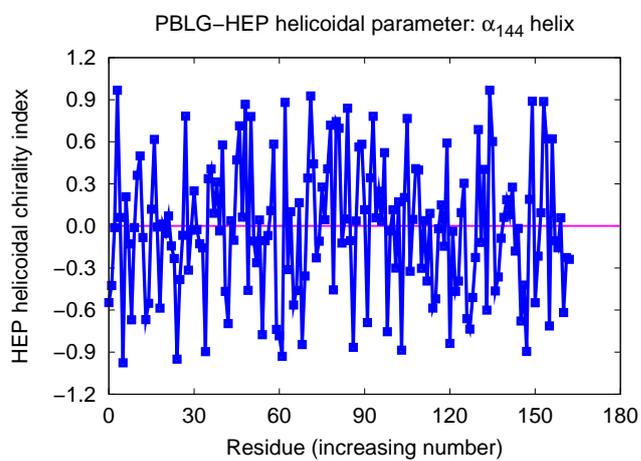


Figure 4.105: The trend of HEP helicoidal chiral order parameter computed for all random solute molecules.

4.6.8.8 Conclusions

Some structures could not be studied with the classic order parameters. For example, helices show an average value of the first rank order parameter ($\langle P_2 \rangle$) equal to zero, since the average distribution of residues composing such structure seems “casual” to such parameter. So, we decide to use another index suitable for helical orientational disposition of particles. This helicoidal order parameter is an average chirality index $\chi = \langle \hat{\mathbf{z}} \cdot (\hat{\mathbf{r}} \times \hat{\mathbf{u}}) \rangle$, where the three unit vectors are: $\hat{\mathbf{z}}$ the direction along which the helix is built $\hat{\mathbf{r}}$ the orientation of the a radial position vector in cylindrical coordinates for each residue and $\hat{\mathbf{u}}$ the orientation of each residue.

First of all, this index has been set up for PBLG and poly–glutamate helices in different way, since they have distinct lateral chains. A negative chiral parameter value ($\chi \sim -0.5$) is associated to the PBLG helices while the average value is positive ($\chi_{polyglu} \sim 0.05$) for the poly–glutamate molecule. When the helix distorts, the chiral index tends to zero, as expected.

This helicoidal parameter is also modified for taking into account only the backbone structure: this type of index can be applied for studying all protein structures. The α_{144} PBLG helix shows a very negative value ($\chi_B \sim -0.9$) and for the α_{36} poly–glutamate helix the χ_B is ~ -0.63 . Also in this case, the index tends to zero when the molecule loses its helicoidal structure.

The parameters explained are a good tool for obtaining a quantitative measure of the helicity of a structure.

4.6.9 Docking

We also decided to study the docking of HEP molecules in PBLG chains, to get a more detailed overview of the mechanism of this process.

4.6.9.1 Docking parameters

We use a code called *Hex* [133] to dock a chain of 19 residue of BGL with different solute.

Correlation Type:	Shape+Electrostatics
FFT Mode:	3D
Post Processing:	Bumps + Volumes / MM Energies / MM Minimisation
Grid Dimension:	0.6
Solutions:	2000
Receptor Range:	180
Step Size:	7.5
Ligand Range:	180
Step Size:	7.5
Twist Range:	360
Step Size:	5.5
Distance Range:	40
Scan Step:	0.8
SubSteps:	0
Steric Scan:	18
Final Search	31

4.6.9.2 Various post processing

We apply docking studies to various type of post processing. The “Bumps + Volumes” take into account the steric effect and the effect of molecules clashing. The second one, “MM energies”, computes the energies of the complexes, using Molecular Mechanics. The last approach, “MM minimization”, performs a minimization of the complexes.

Heptyl butyrate

First of all we use the heptyl butyrate (HEP) Figure 4.194. Results of docking are reported in Table 4.16

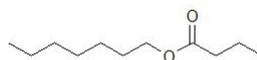


Figure 4.106: Formula of heptyl butyrate.

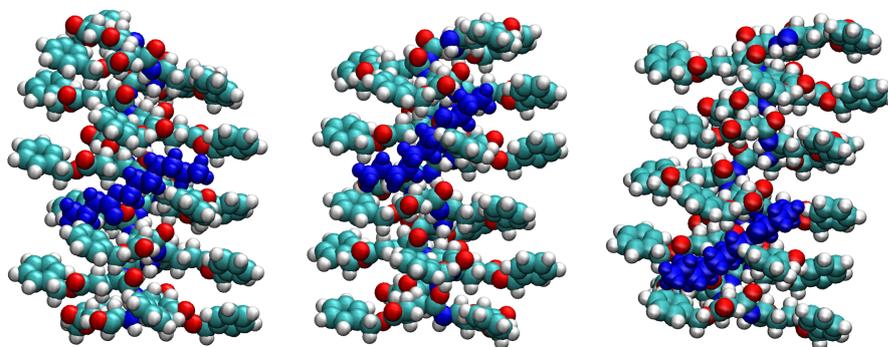


Figure 4.107: Best configuration obtained from different types of docking: bumps+volumes, MM energies and MM minimization (*from left to right*).

Post processing	E_{total}	E_{shape}	E_{force}	% E_{shape}	% E_{force}
Bumps+Volumes	-164.5	-164.5	0.0	100	0
MM energies	-211.8	-155.0	-56.3	73	27
MM minimization	-215.0	-155.2	-59.9	72	28

Table 4.16: Results of docking of PBLG with HEP, using different techniques [134].

Ethyl acetate

First of all we use the ethyl acetate (EAC) Figure 4.109. Results of docking are reported in Table 4.17



Figure 4.108: Formula of heptyl butyrate.

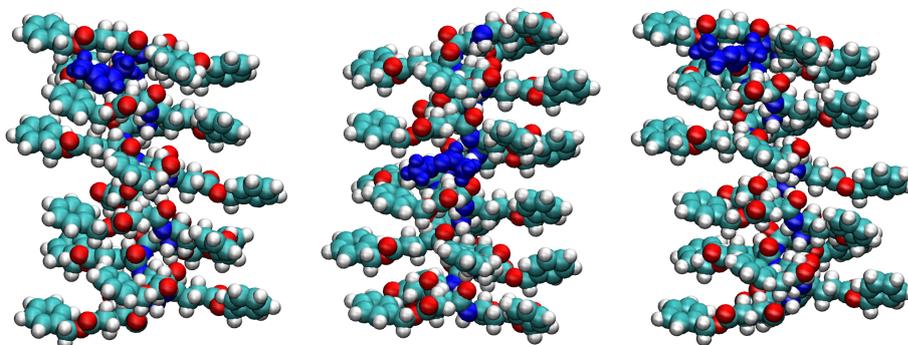


Figure 4.109: Best configuration obtained from different types of docking: bumps+volumes, MM energies and MM minimization (*from left to right*).

Post processing	E_{total}	E_{shape}	E_{force}	% E_{shape}	% E_{force}
Bumps+Volumes	-104.2	-104.2	0.0	100	0
MM energies	-129.8	-95.7	-34.1	74	26
MM minimization	-132.9	-102.2	-30.7	77	23

Table 4.17: Results of docking of PBLG with EAC, using different techniques [134].

4.6.9.3 Comparing different molecules

This section summarizes results of docking of different molecules, obtained using MM minimization. We decided to use only this post-processing because it takes into account not only volume but also energy; moreover, it minimizes the structure to find the better solution. We use different molecules: heptyl butyrate, ethyl acetate and different enantiomers of limonene and menthol. We also dock a very big diol in order to understand if a bigger volume would not fit the free binding volume of the protein. Lastly, also dimethyl formamide used as solvent is docked.

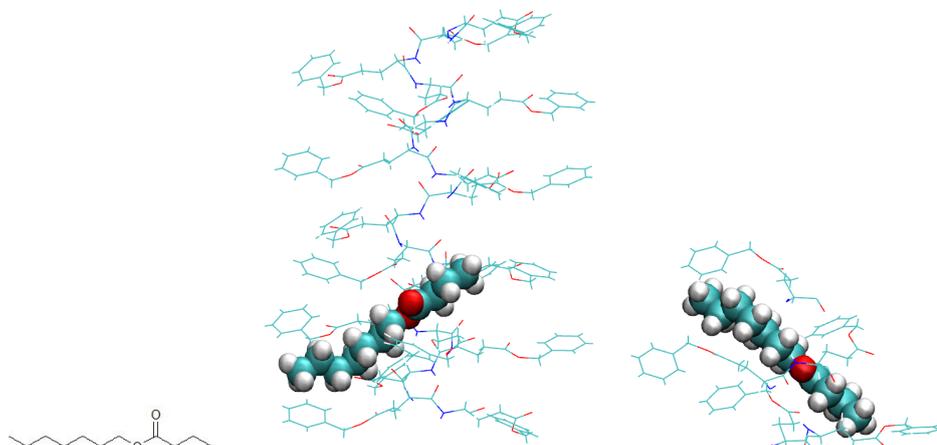


Figure 4.110: Best configuration for docking a PBLG helix of 19 residues and a heptyl butyrate molecule (IUPAC name: butyric acid heptyl ester).

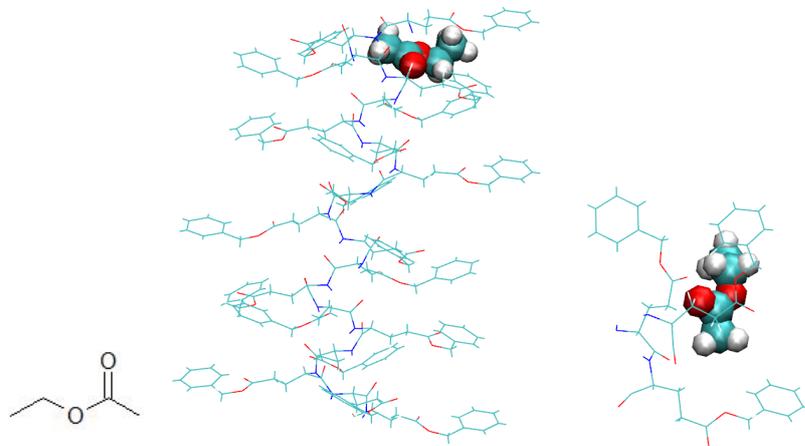


Figure 4.111: Best configuration for docking a PBLG helix of 19 residues and an ethyl acetate molecule (IUPAC name: acid acetic ethyl ester).

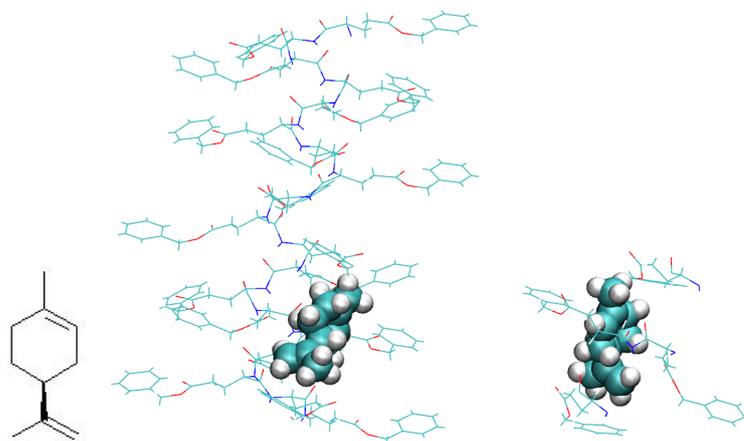


Figure 4.112: Best configuration for docking a PBLG helix of 19 residues and an (R) enantiomer of limonene (IUPAC name: (R)-4-isopropenyl-1-methylcyclohexene).

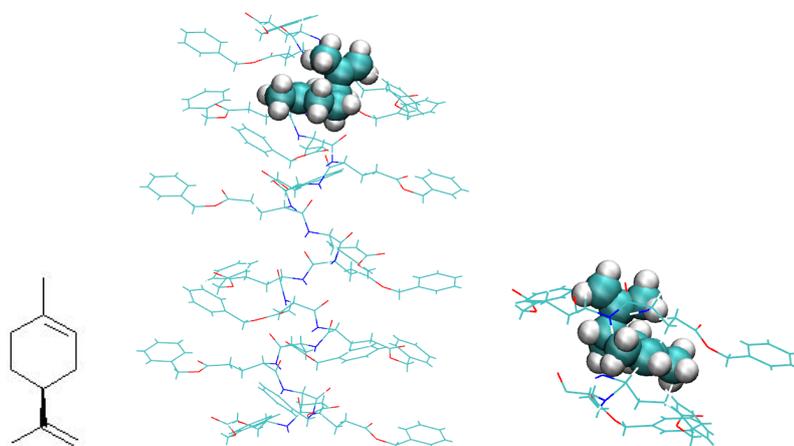


Figure 4.113: Best configuration for docking a PBLG helix of 19 residues and an (*S*) enantiomer of limonene (IUPAC name: (*S*)-4-isopropenyl-1-methylcyclohexene).

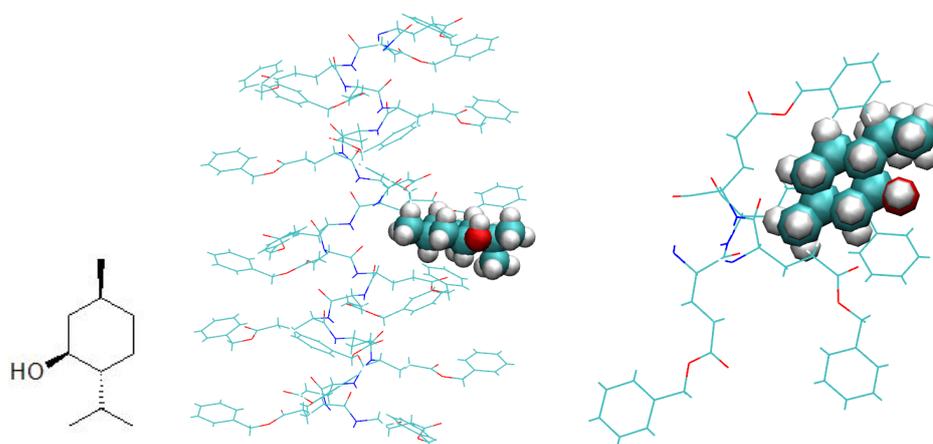


Figure 4.114: Best configuration for docking a PBLG helix of 19 residues and a (+)-menthol molecule (IUPAC name: (1*S*,2*R*,5*S*)-2-Isopropyl-5-methylcyclohexanol).

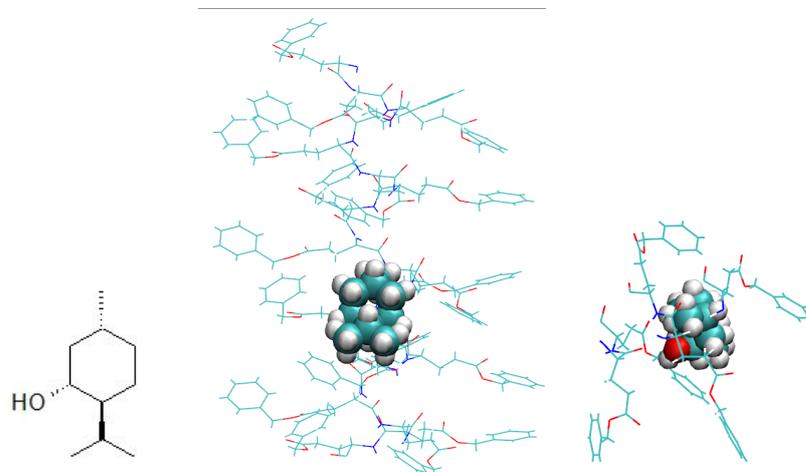


Figure 4.115: Best configuration for docking a PBLG helix of 19 residues and a (-)-menthol molecule (IUPAC name: (1R,2S,5R)-2-Isopropyl-5-methylcyclohexanol).

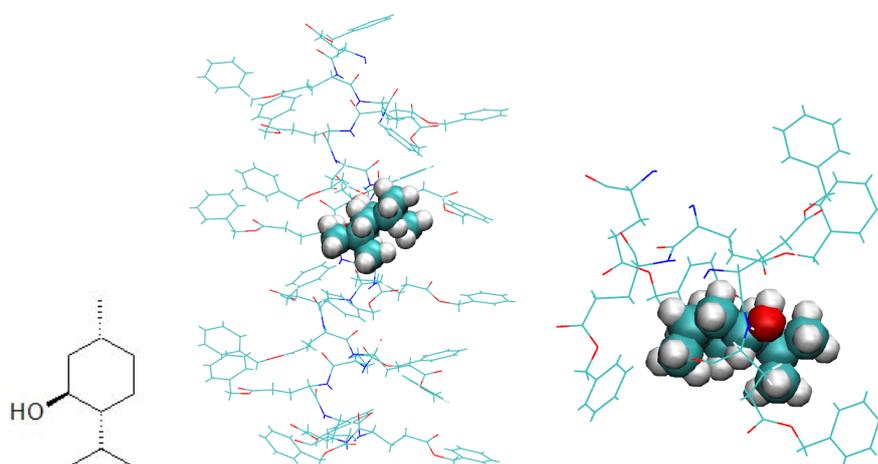


Figure 4.116: Best configuration for docking a PBLG helix of 19 residues and a (+)-isomenthol molecule (IUPAC name: (1S,2R,5R)-2-Isopropyl-5-methylcyclohexanol).

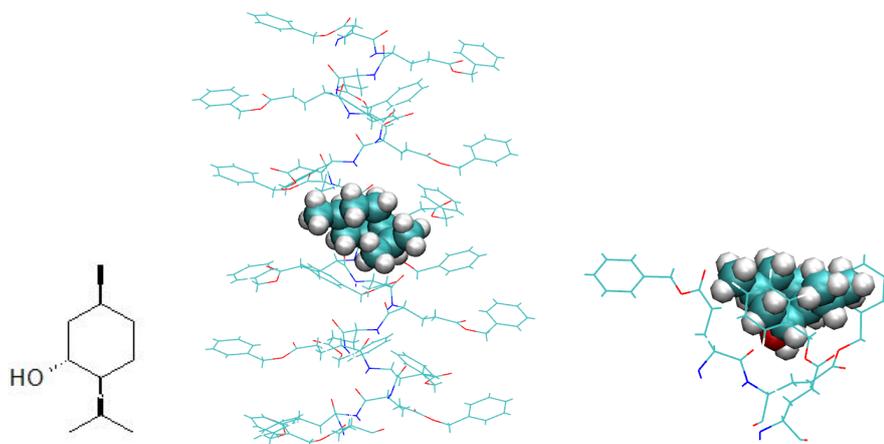


Figure 4.117: Best configuration for docking a PBLG helix of 19 residues and a (-)-isomenthol molecule (IUPAC name: (1R,2S,5S)-2-Isopropyl-5-methyl-cyclohexanol).

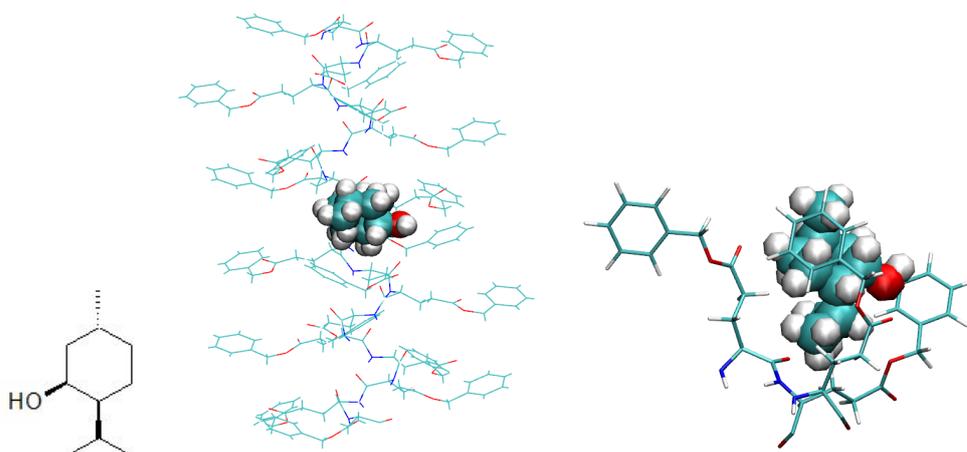


Figure 4.118: Best configuration for docking a PBLG helix of 19 residues and a (+)-neomenthol molecule (IUPAC name: (1S,2S,5R)-2-Isopropyl-5-methyl-cyclohexanol).

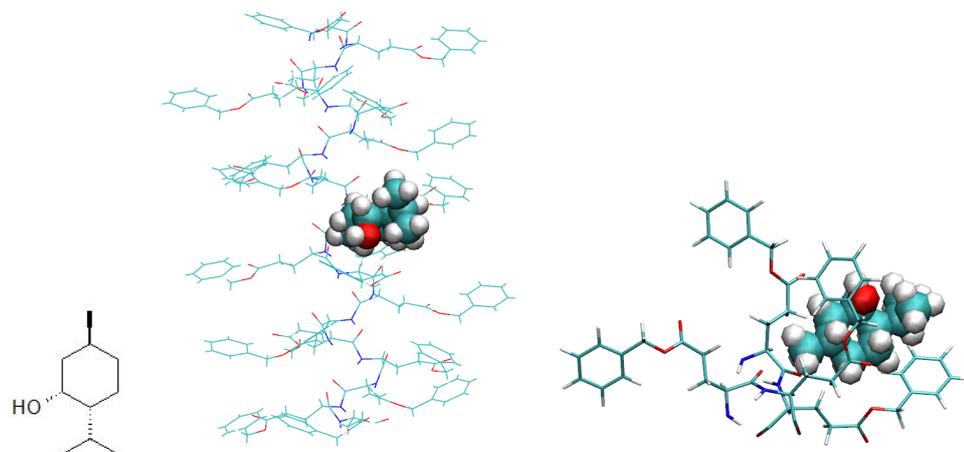


Figure 4.119: Best configuration for docking a PBLG helix of 19 residues and a (-)-neomenthol molecule (IUPAC name: (1R,2R,5S)-2-Isopropyl-5-methyl-cyclohexanol).

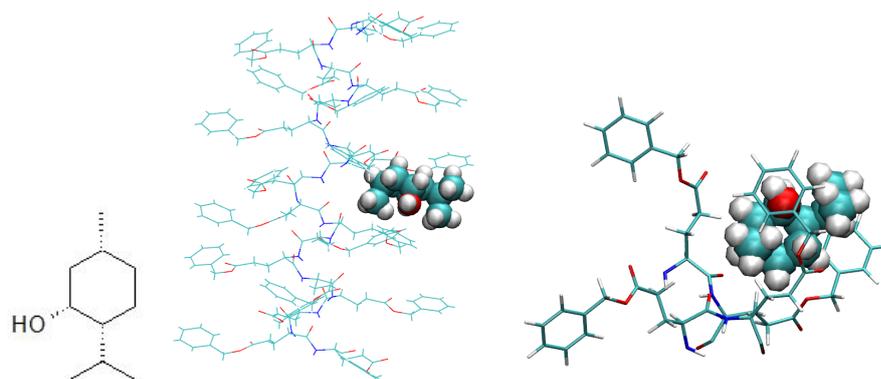


Figure 4.120: Best configuration for docking a PBLG helix of 19 residues and a (+)-neoisomenthol molecule (IUPAC name: (1R,2R,5R)-2-Isopropyl-5-methyl-cyclohexanol).

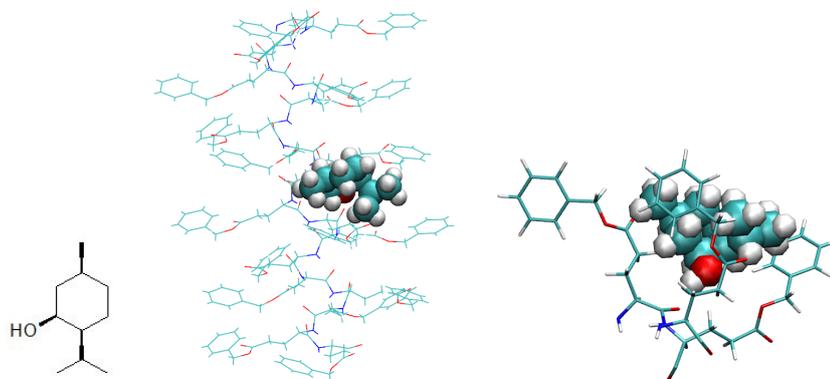


Figure 4.121: Best configuration for docking a PBLG helix of 19 residues and a (-)-neoisomenthol molecule (IUPAC name: (1S,2S,5S)-2-Isopropyl-5-methylcyclohexanol).

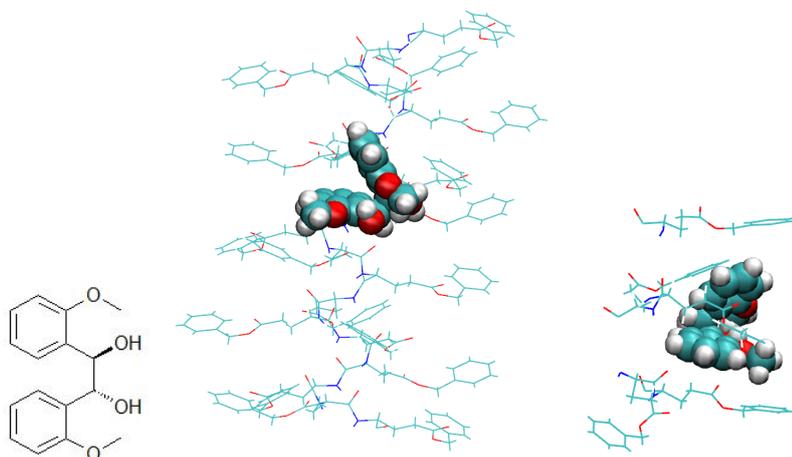


Figure 4.122: Best configuration for docking a PBLG helix of 19 residues and a (1R,2R)-1,2-bis-(2-methoxyphenyl)ethane-1,2-diol molecule.

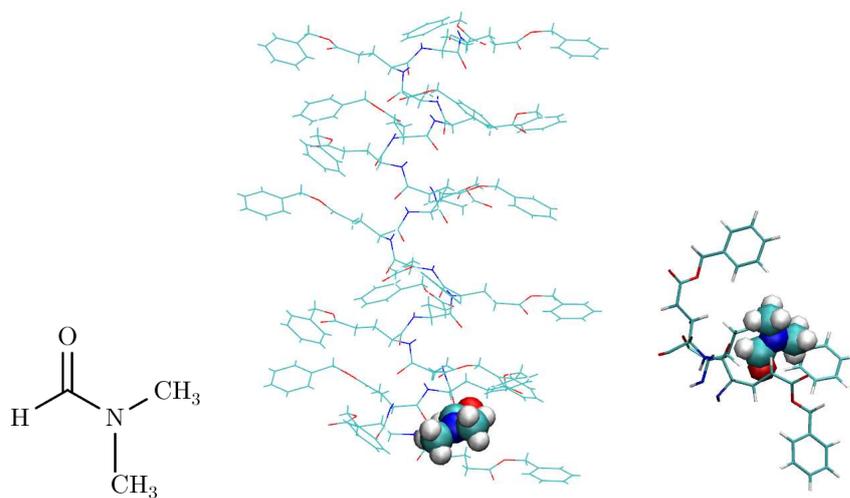


Figure 4.123: Best configuration for docking a PBLG helix of 19 residues and a dimethylformamide molecule.

Results

Following graphs and tables show the trend of affinities for molecules above described.

Molecule	Functional group	G_{0S}	E_{total}	E_{shape}	E_{force}	% E_{shape}	% E_{force}
Dimethylformamide	Amide	$0.290495 \cdot 10^{-2}$	-121.6	-88.0	-33.6	72.4	27.6
Heptyl butyrate	Ester	8.27699	-215.0	-155.2	-59.9	72.2	27.8
Ethyl acetate	Ester	14.8230	-132.9	-102.2	-30.7	76.9	23.1
(R)-Limonene	Hydrocarbon	-4.34756	-180.2	-134.3	-45.9	74.5	25.5
(S)-Limonene	Hydrocarbon	-1.31482	-183.2	-132.9	-50.4	72.5	27.5
(+)-menthol	Alcohol	2.33148	-179.7	-132.3	-47.4	73.6	26.4
(-)-menthol	Alcohol	1.22293	-165.9	-138.9	-27.0	83.7	16.3
(+)-isomenthol	Alcohol	2.14045	-174.8	-133.6	-41.2	76.4	26.6
(-)-isomenthol	Alcohol	5.92888	-171.5	-134.0	-37.4	78.1	21.9
(+)-neomenthol	Alcohol	0.792153	-178.9	-141.3	-37.6	79.0	21.0
(-)-neomenthol	Alcohol	-4.14967	-180.3	-137.3	-43.0	76.2	23.8
(+)-neoisomenthol	Alcohol	-0.437589	-172.9	-128.5	-44.4	74.3	26.7
(-)-neoisomenthol	Alcohol	5.32573	-186.4	-144.7	-41.7	77.6	22.4
(1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol	Diol	-39.4067	-255.6	-191.6	-64.0	75.0	25.0

Table 4.18: Results of docking of PBLG with different molecules.

Molecule	Mass [a.m.u.]	Volume [$\frac{4}{3}\pi R_{g,x}R_{g,y}R_{g,z}$]	Mass affinity			Volume affinity		
			E_{total}	E_{shape}	E_{force}	E_{total}	E_{shape}	E_{force}
Dimethylformamide	73.0947	3.45262	-1.6636	-1.20392	-0.459678	-0.631205	-0.456793	-0.174412
Heptyl butyrate	186.294	13.5482	-1.15409	-0.833092	-0.321535	-0.379921	-0.27425	-0.105848
Ethyl acetate	88.1063	2.031	-1.50841	-1.15996	-0.348443	-1.20316	-0.925228	-0.27793
(R)-Limonene	136.237	5.3472	-1.3227	-0.985782	-0.336913	-0.541723	-0.403737	-0.137986
(S)-Limonene	136.237	6.36286	-1.34472	-0.975506	-0.369944	-0.380409	-0.275963	-0.104654
(+)-menthol	156.268	6.28753	-1.14995	-0.846622	-0.303325	-0.382247	-0.281421	-0.100826
(-)-menthol	156.268	6.99255	-1.06164	-0.888858	-0.17278	-0.283911	-0.237705	-0.0462061
(+)-isomenthol	156.268	7.29654	-1.11859	-0.854942	-0.26365	-0.292811	-0.223796	-0.069015
(-)-isomenthol	156.268	8.0063	-1.09747	-0.857501	-0.239332	-0.249479	-0.194928	-0.0544053
(+)-neomenthol	156.268	7.38978	-1.14483	-0.904216	-0.240612	-0.278857	-0.220249	-0.0586083
(-)-neomenthol	156.268	6.62796	-1.15379	-0.878619	-0.275168	-0.360343	-0.274404	-0.0859388
(+)-neoisomenthol	156.268	6.92407	-1.10643	-0.822305	-0.284127	-0.303611	-0.225645	-0.077966
(-)-neoisomenthol	156.268	7.97562	-1.19282	-0.925973	-0.266849	-0.269205	-0.208981	-0.0602246
(1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol	274.317	18.2669	-0.931769	-0.698462	-0.233307	-0.11528	-0.0864146	-0.028865

Table 4.19: Mass and volume affinities of docking of PBLG with different molecules.

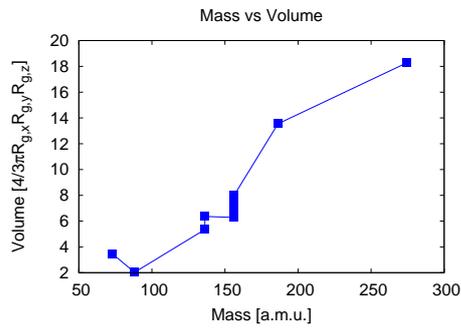


Figure 4.124: Plot of mass versus volume.

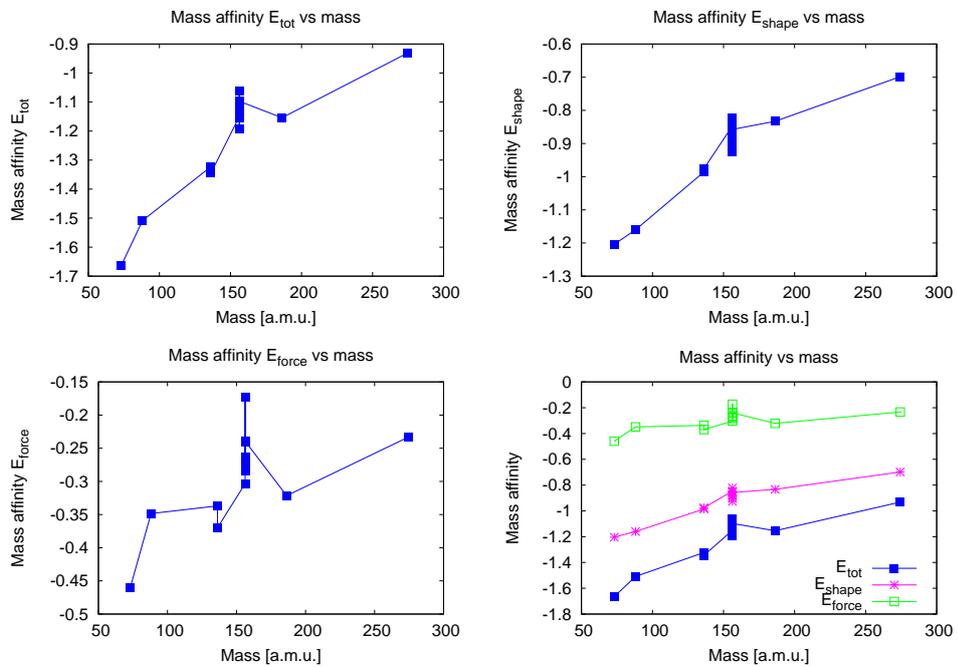


Figure 4.125: Plot of mass affinity for E_{tot} , E_{shape} and E_{force} versus masses.

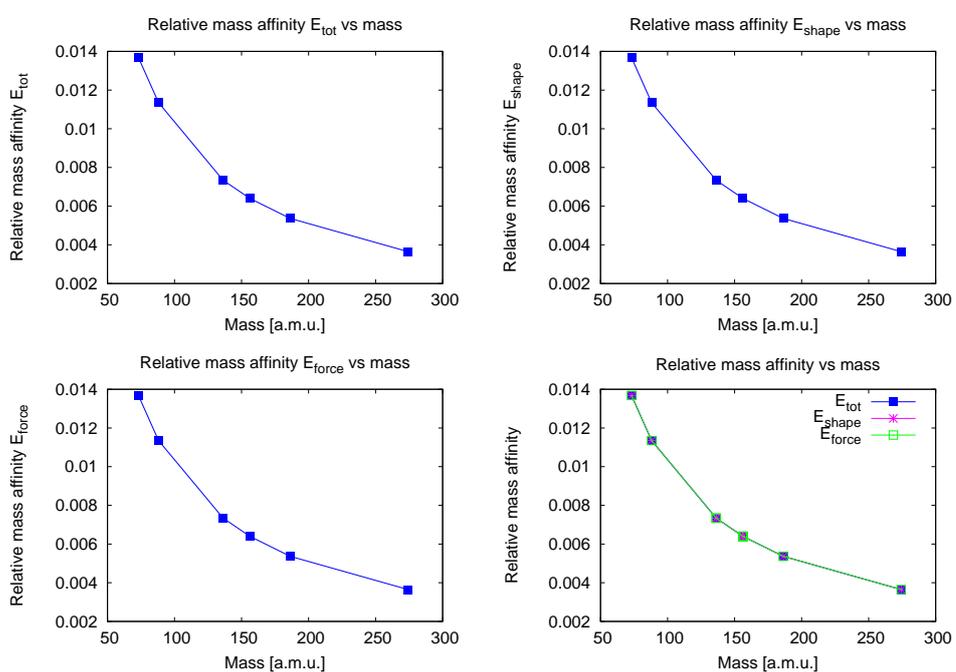


Figure 4.126: Plot of relative mass affinity for E_{tot} , E_{shape} and E_{force} versus masses.

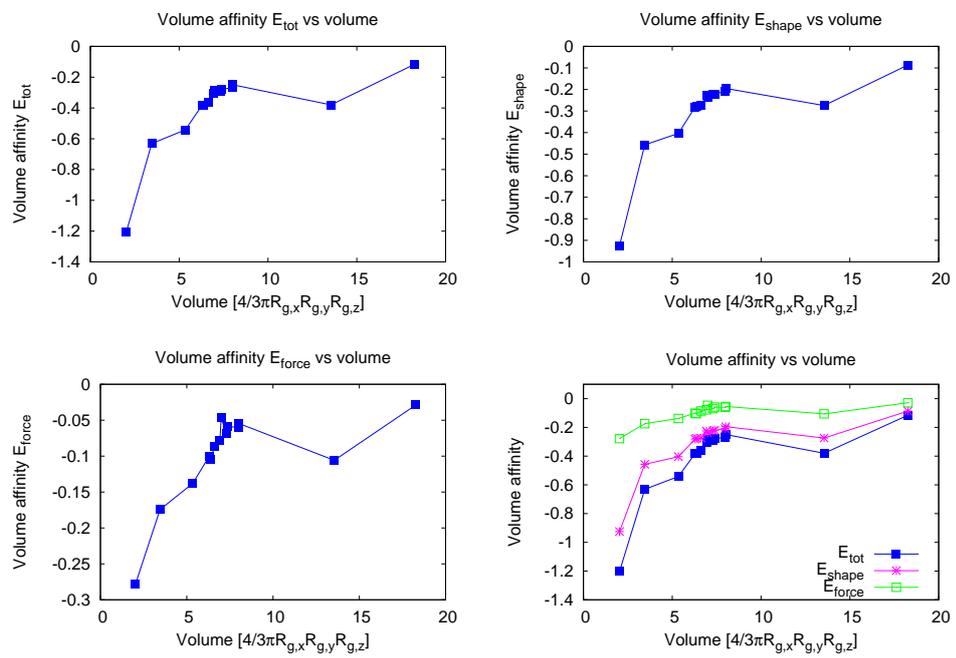


Figure 4.127: Plot of volume affinities for E_{tot} , E_{shape} and E_{force} versus volumes.

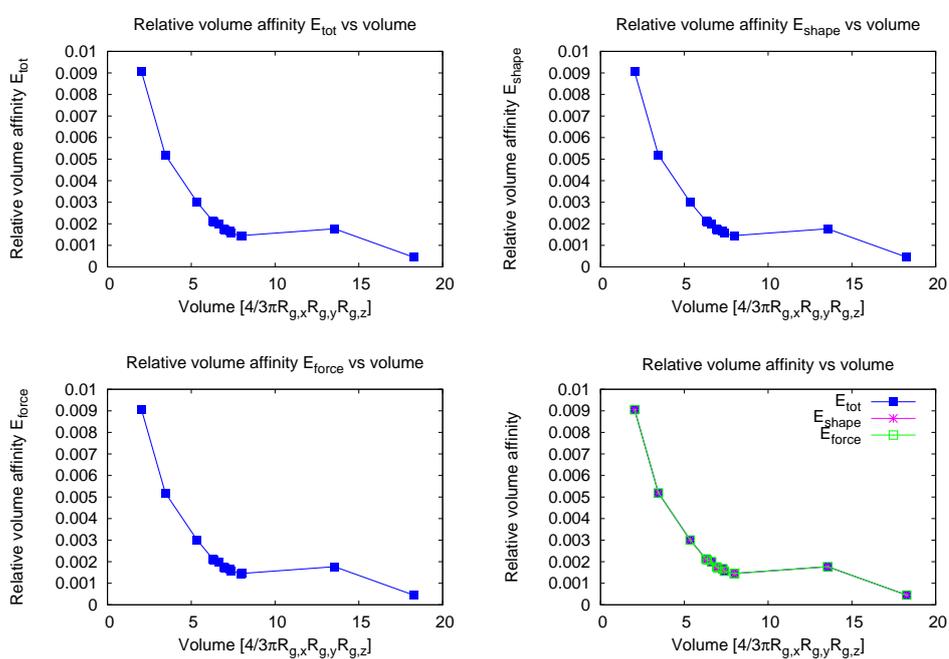


Figure 4.128: Plot of relative volume affinity for E_{tot} , E_{shape} and E_{force} versus volumes.

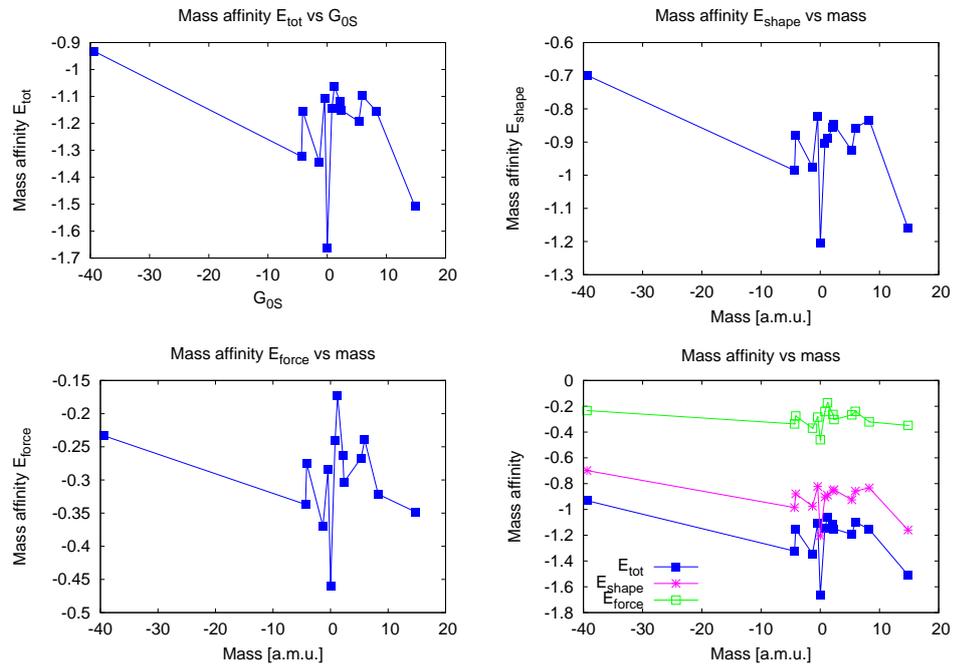


Figure 4.129: Plot of mass affinity for E_{tot} , E_{shape} and E_{force} versus G_{0S} .

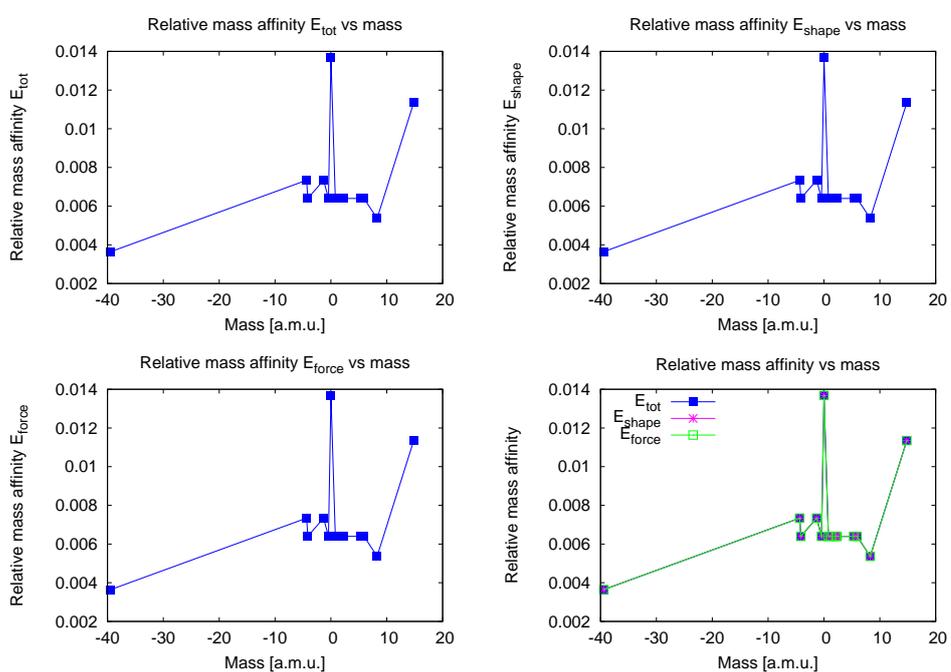


Figure 4.130: Plot of relative mass affinity for E_{tot} , E_{shape} and E_{force} versus G_{0S} .

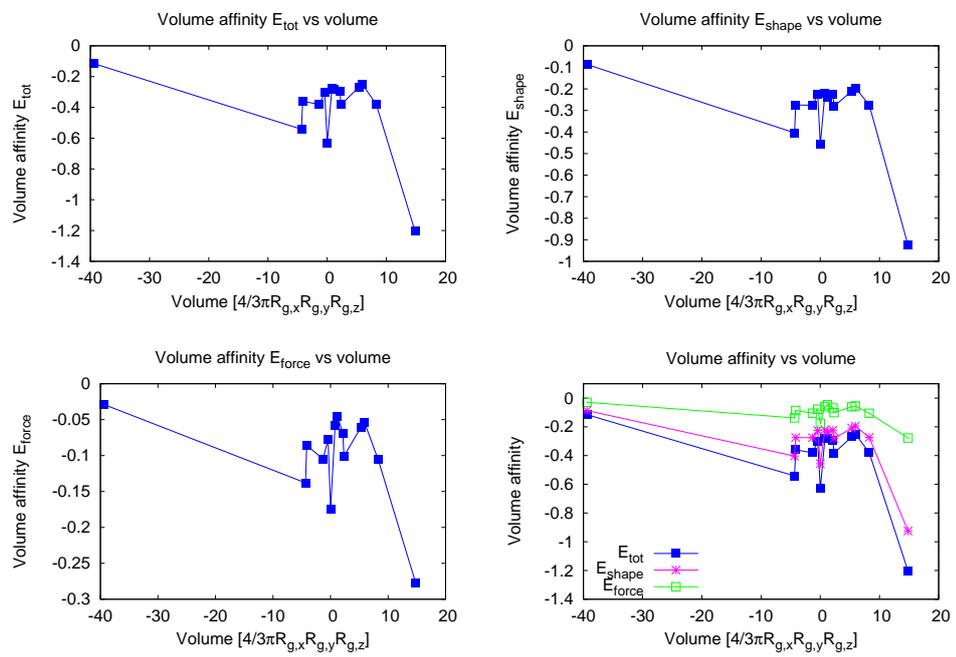


Figure 4.131: Plot of volume affinities for E_{tot} , E_{shape} and E_{force} versus G_{0S} .

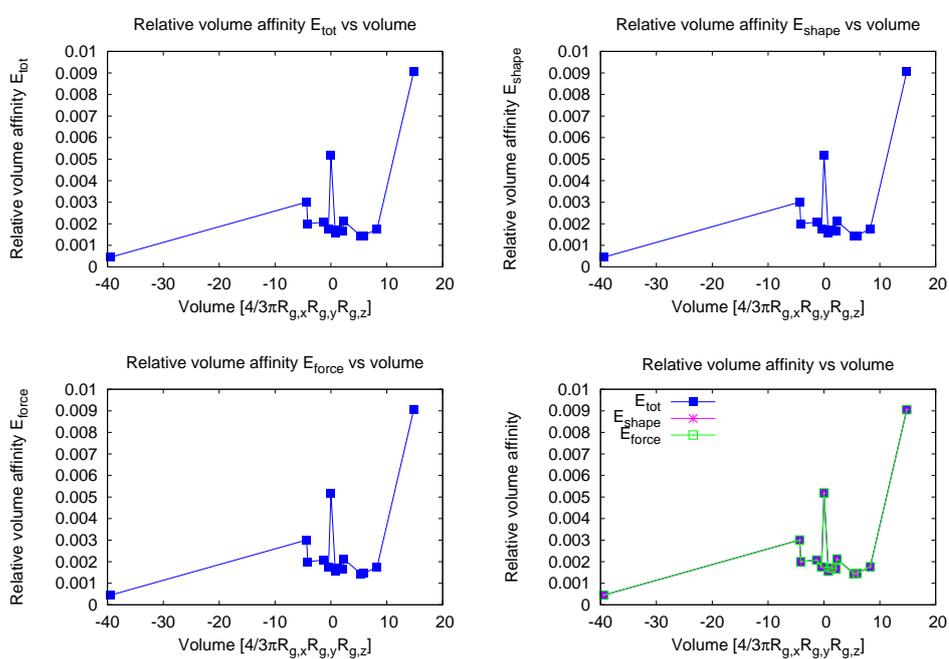


Figure 4.132: Plot of relative volume affinity for E_{tot} , E_{shape} and E_{force} versus G_{0S} .

4.6.9.4 Deprotonated molecules

Since the solvent is polar, we also dock the anion of the molecules described above.

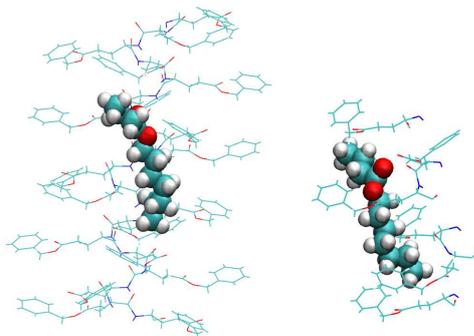


Figure 4.133: Best configuration for docking a PBLG helix of 19 residues and a deprotonated heptyl butyrate molecule (IUPAC name: butyric acid heptyl ester, =O deprotonated in O^-).

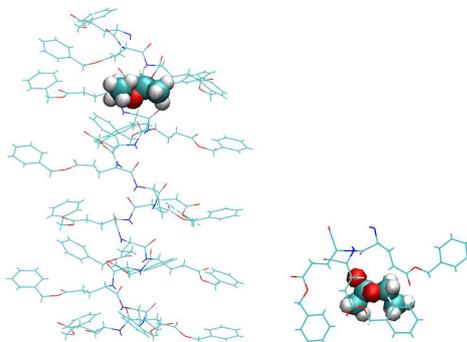


Figure 4.134: Best configuration for docking a PBLG helix of 19 residues and a deprotonated ethyl acetate molecule (IUPAC name: acid acetic ethyl ester, =O deprotonated in O^-).

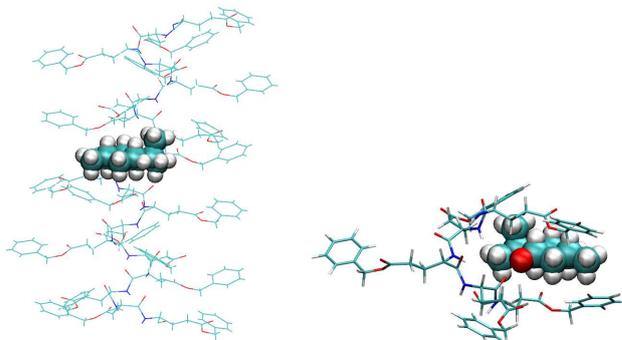


Figure 4.135: Best configuration for docking a PBLG helix of 19 residues and a (+)-menthol molecule (IUPAC name: (1S,2R,5S)-2-Isopropyl-5-methylcyclohexanol, -OH deprotonated in $-O^-$).

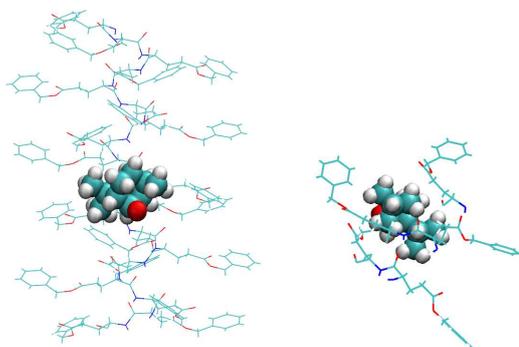


Figure 4.136: Best configuration for docking a PBLG helix of 19 residues and a (-)-menthol molecule (IUPAC name: (1R,2S,5R)-2-Isopropyl-5-methylcyclohexanol, -OH deprotonated in $-O^-$).

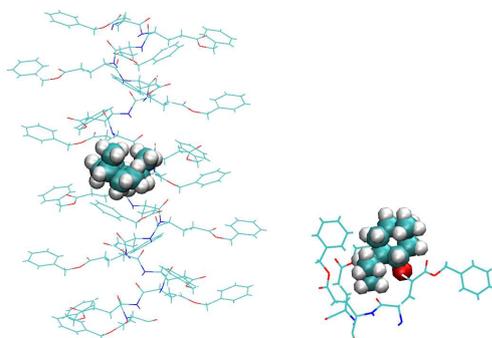


Figure 4.137: Best configuration for docking a PBLG helix of 19 residues and a (+)-isomenthol molecule (IUPAC name: (1S,2R,5R)-2-Isopropyl-5-methylcyclohexanol, -OH deprotonated in $-O^-$).

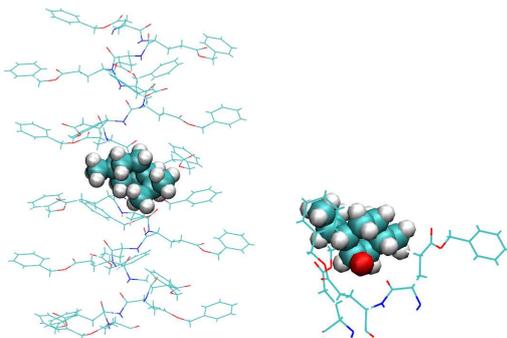


Figure 4.138: Best configuration for docking a PBLG helix of 19 residues and a (-)-isomenthol molecule (IUPAC name: (1R,2S,5S)-2-Isopropyl-5-methylcyclohexanol, -OH deprotonated in $-O^-$).

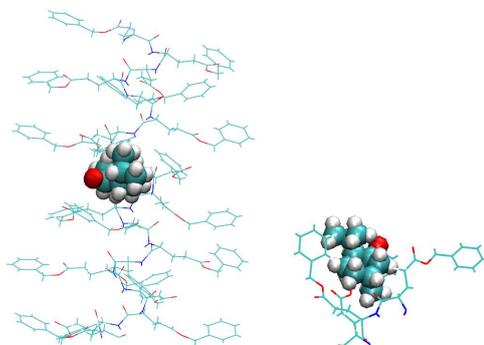


Figure 4.139: Best configuration for docking a PBLG helix of 19 residues and a (+)-neomenthol molecule (IUPAC name: (1S,2S,5R)-2-Isopropyl-5-methylcyclohexanol, -OH deprotonated in -O^-).

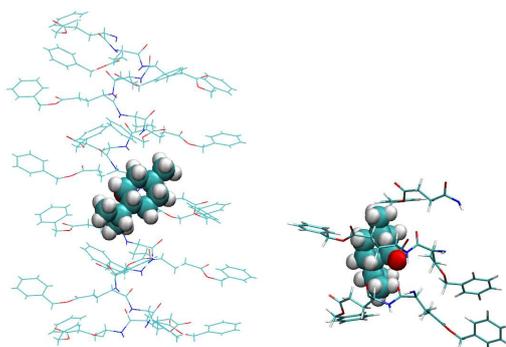


Figure 4.140: Best configuration for docking a PBLG helix of 19 residues and a (-)-neomenthol molecule (IUPAC name: (1R,2R,5S)-2-Isopropyl-5-methylcyclohexanol, -OH deprotonated in -O^-).

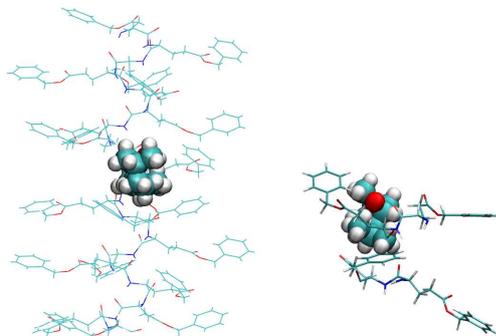


Figure 4.141: Best configuration for docking a PBLG helix of 19 residues and a (+)-neoisomenthol molecule (IUPAC name: (1R,2R,5R)-2-Isopropyl-5-methyl-cyclohexanol, -OH deprotonated in -O^-).

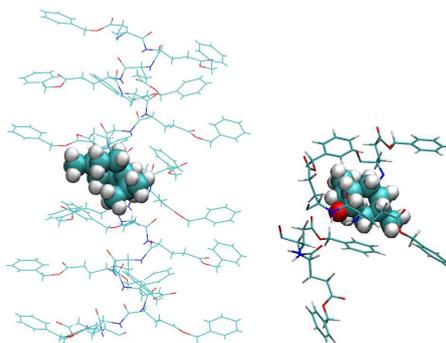


Figure 4.142: Best configuration for docking a PBLG helix of 19 residues and a (-)-neoisomenthol molecule (IUPAC name: (1S,2S,5S)-2-Isopropyl-5-methyl-cyclohexanol, -OH deprotonated in -O^-).

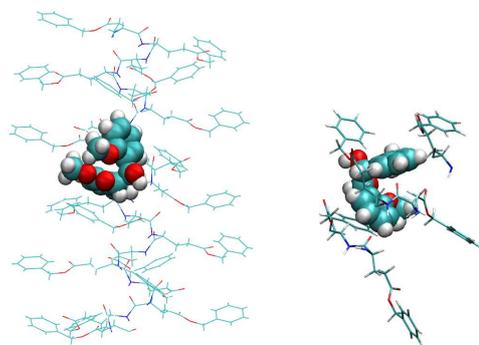


Figure 4.143: Best configuration for docking a PBLG helix of 19 residues and a (1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol with one -OH deprotonated in $-O^-$.

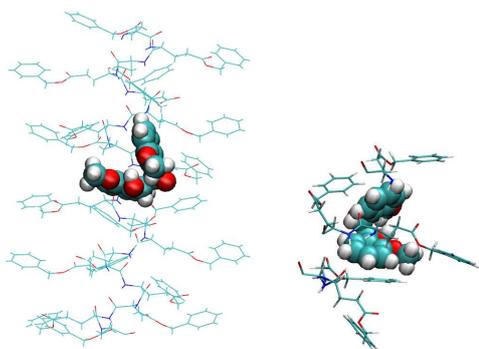


Figure 4.144: Best configuration for docking a PBLG helix of 19 residues and a (1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol molecule, with the other one -OH deprotonated in $-O^-$.

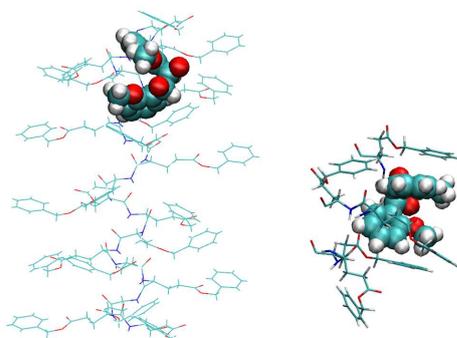


Figure 4.145: Best configuration for docking a PBLG helix of 19 residues and a (1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol molecule, with both -OH deprotonated in $-O^-$.

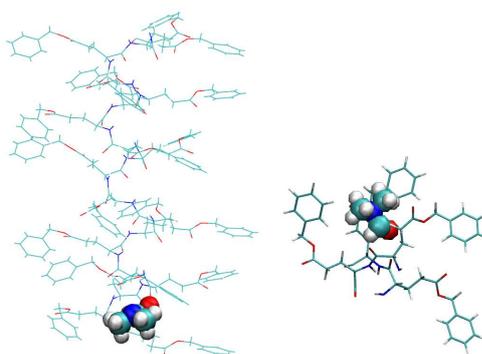


Figure 4.146: Best configuration for docking a PBLG helix of 19 residues and an anion molecule of dimethylformamide, =O deprotonated in $-O^-$.

Results

Molecule	Deprotonated group	E_{total}	E_{shape}	E_{force}	% E_{shape}	% E_{force}
Dimethylformamide	=O	-119.1	-88.1	-30.9	74.0	26.0
Heptyl butyrate	=O	-211.4	-165.8	-45.6	78.4	21.6
Ethyl acetate	=O	-125.7	-93.1	-32.5	74.1	25.9
(+)-menthol	-OH	-182.0	-132.6	-49.4	72.9	27.1
(-)-menthol	-OH	-174.3	-145.2	-29.1	83.3	16.7
(+)-isomenthol	-OH	-181.7	-140.8	-40.9	77.5	22.5
(-)-isomenthol	-OH	-181.9	-140.2	-41.7	77.1	22.9
(+)-neomenthol	-OH	-176.6	-137.1	-39.5	77.6	22.4
(-)-neomenthol	-OH	-176.0	-142.0	-34.0	80.7	19.3
(+)-neoisomenthol	-OH	-179.9	-141.0	-38.9	78.4	21.6
(-)-neoisomenthol	-OH	-170.6	-132.5	-38.0	77.7	22.3
(1R,2R)-1,2-Bis-	1 st -OH	-256.4	-192.2	-64.2	75.0	25.0
(2-methoxy-phenyl)-	2 nd -OH	-264.1	-202.2	-61.9	76.6	23.4
-ethane-1,2-diol	both -OH	-256.7	-190.7	-66.0	74.3	25.7

Table 4.20: Results of docking of PBLG with different deprotonated molecules.

Molecule	Mass [a.m.u.]	Volume $[\frac{4}{3}\pi R_{g,x}R_{g,y}R_{g,z}]$	Deprotonated	Mass affinity		Volume affinity	
				E_{total}	E_{shape}	E_{total}	E_{shape}
Dimethylformamide	73.0947	3.52237	=O	-1.62939	-1.20529	-0.422739	-0.441388
Heptyl butyrate	186.294	16.8398	=O	-1.13477	-0.889991	-0.244774	-0.2262
Ethyl acetate	88.1063	2.22814	=O	-1.42669	-1.05668	-0.368873	-0.70662
(+)-menthol	155.26	6.17723	-OH	-1.17223	-0.854051	-0.318176	-0.288959
(-)-menthol	155.26	6.87953	-OH	-1.12263	-0.935205	-0.187428	-0.255568
(+)-isomenthol	155.26	7.14213	-OH	-1.17029	-0.906866	-0.263429	-0.242521
(-)-isomenthol	155.26	7.91385	-OH	-1.17158	-0.903001	-0.268582	-0.20881
(+)-neomenthol	155.26	7.2466	-OH	-1.13745	-0.883035	-0.254412	-0.220558
(-)-neomenthol	155.26	6.5075	-OH	-1.13358	-0.914595	-0.218988	-0.292333
(+)-neoisomenthol	155.26	6.8213	-OH	-1.1587	-0.908154	-0.250547	-0.254288
(-)-neoisomenthol	155.26	7.86755	-OH	-1.0988	-0.853407	-0.244751	-0.197419
(1R,2R)-1,2-Bis-	273.309	17.7739	1 st -OH	-0.938132	-0.703233	-0.234899	-0.0822964
(2-methoxy-phenyl)-	273.309	18.3445	2 nd -OH	-0.966306	-0.739822	-0.226484	-0.0903378
-ethane-1,2-diol	272.301	17.5125	both -OH	-0.942707	-0.700328	-0.242379	-0.0831796

Table 4.21: Mass and volume affinities of docking of PBLG with different deprotonated molecules.

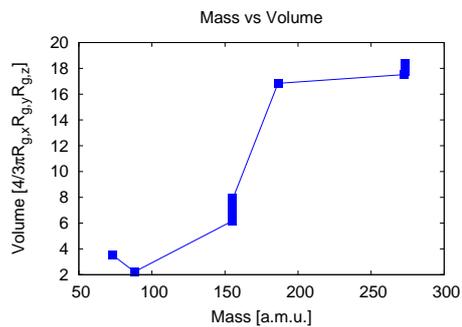


Figure 4.147: Plot of mass versus volume.

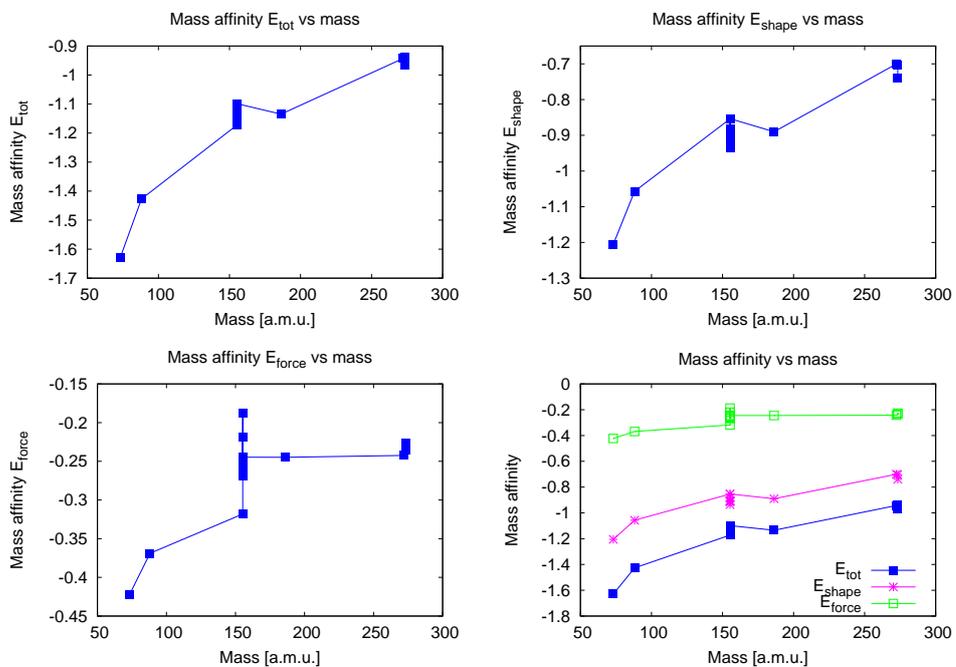


Figure 4.148: Plot of mass affinity for E_{tot} , E_{shape} and E_{force} versus masses.

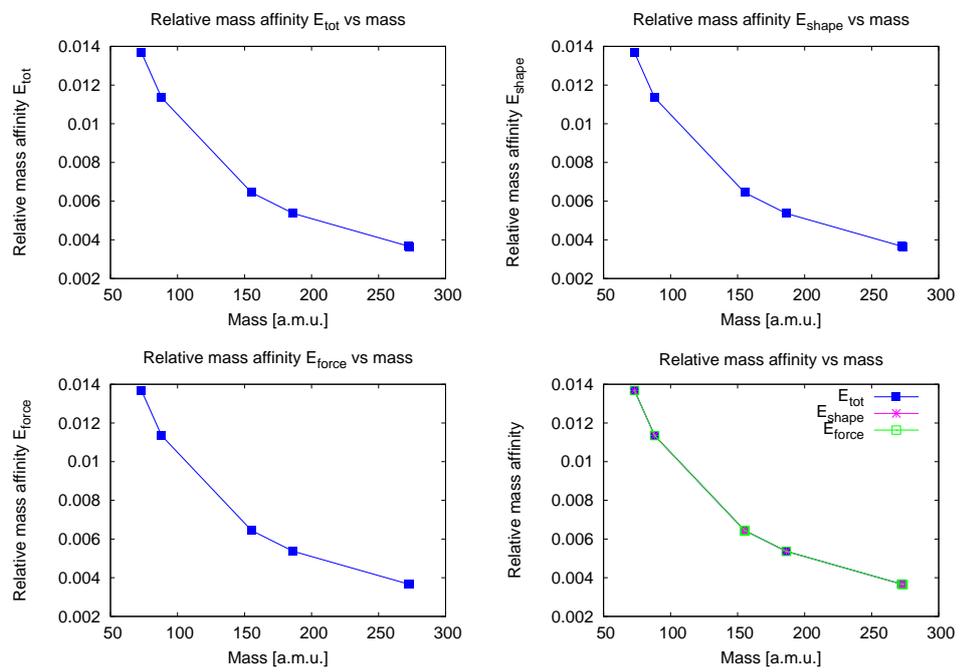


Figure 4.149: Plot of relative mass affinity for E_{tot} , E_{shape} and E_{force} versus masses.

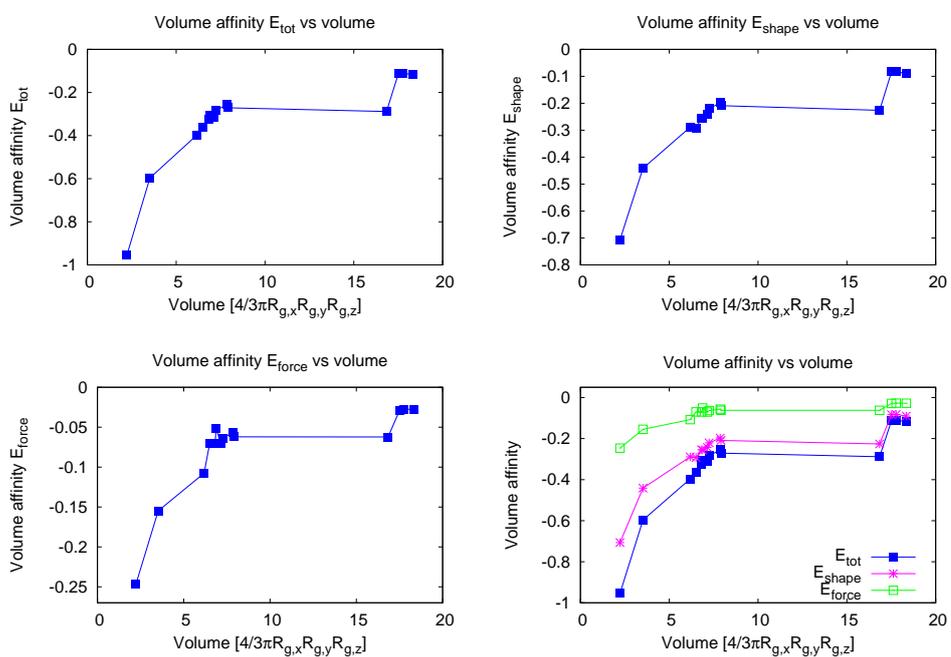


Figure 4.150: Plot of volume affinities for E_{tot} , E_{shape} and E_{force} versus volumes.

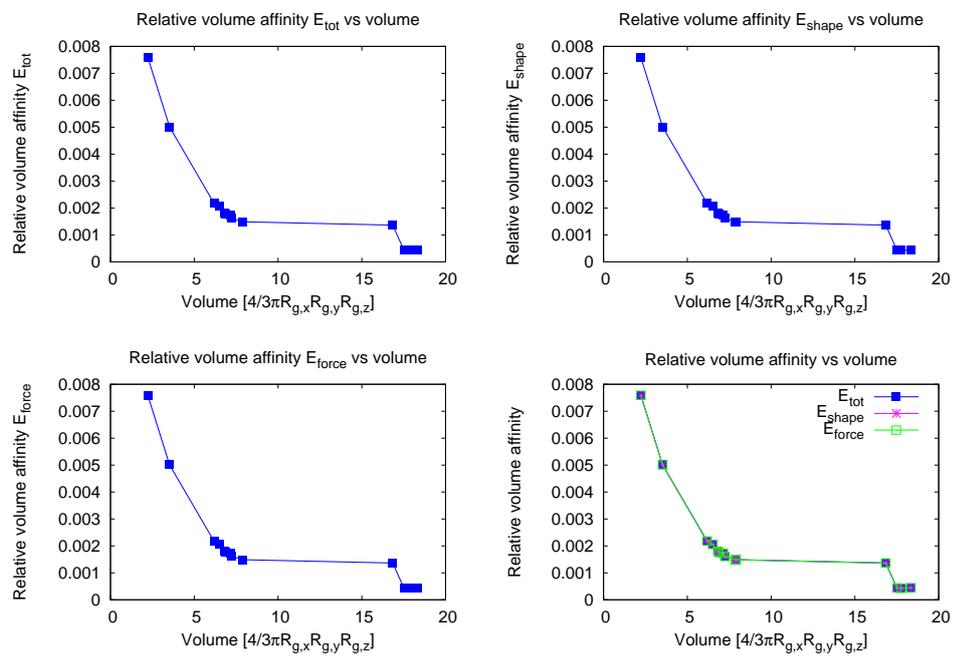


Figure 4.151: Plot of relative volume affinity for E_{tot} , E_{shape} and E_{force} versus volumes.

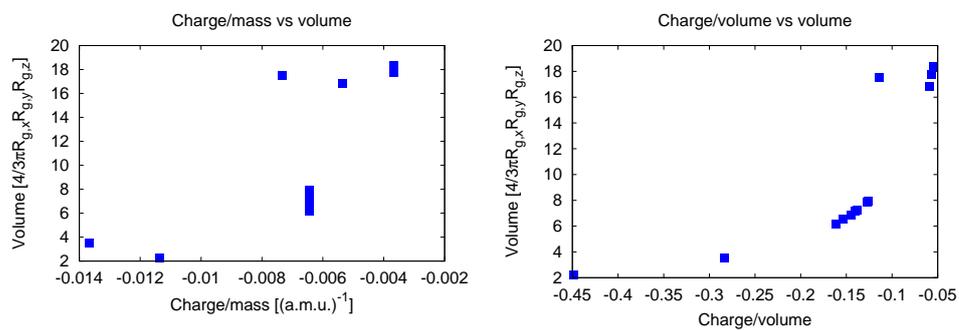


Figure 4.152: Plot of $\frac{charge}{mass}$ and of $\frac{charge}{volume}$ versus volume.

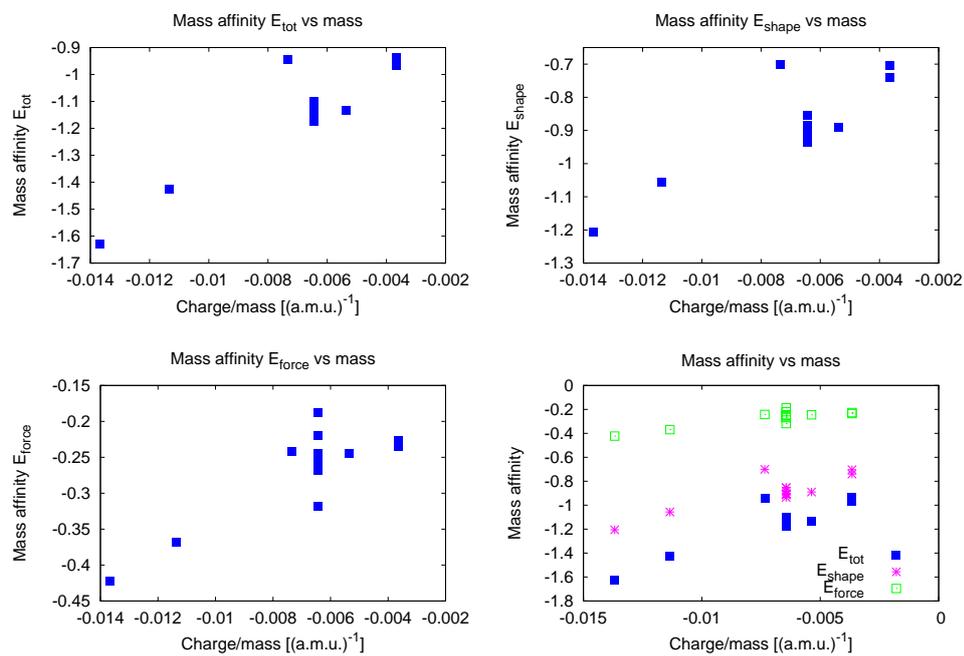


Figure 4.153: Plot of mass affinity for E_{tot} , E_{shape} and E_{force} versus $\frac{charge}{mass}$.

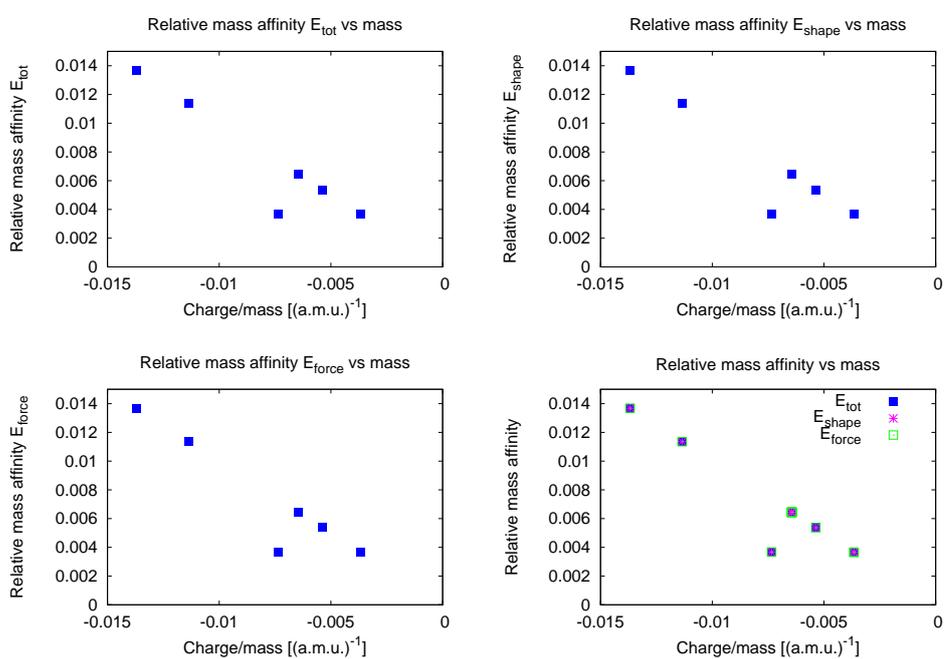


Figure 4.154: Plot of relative mass affinity for E_{tot} , E_{shape} and E_{force} versus $\frac{charge}{mass}$.

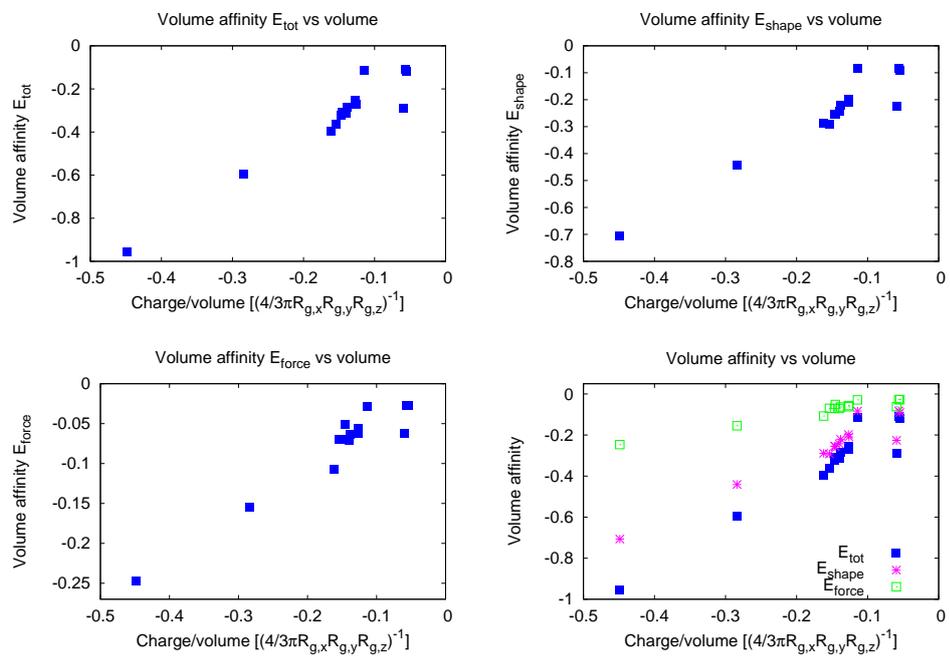


Figure 4.155: Plot of volume affinities for E_{tot} , E_{shape} and E_{force} versus $\frac{charge}{volume}$.

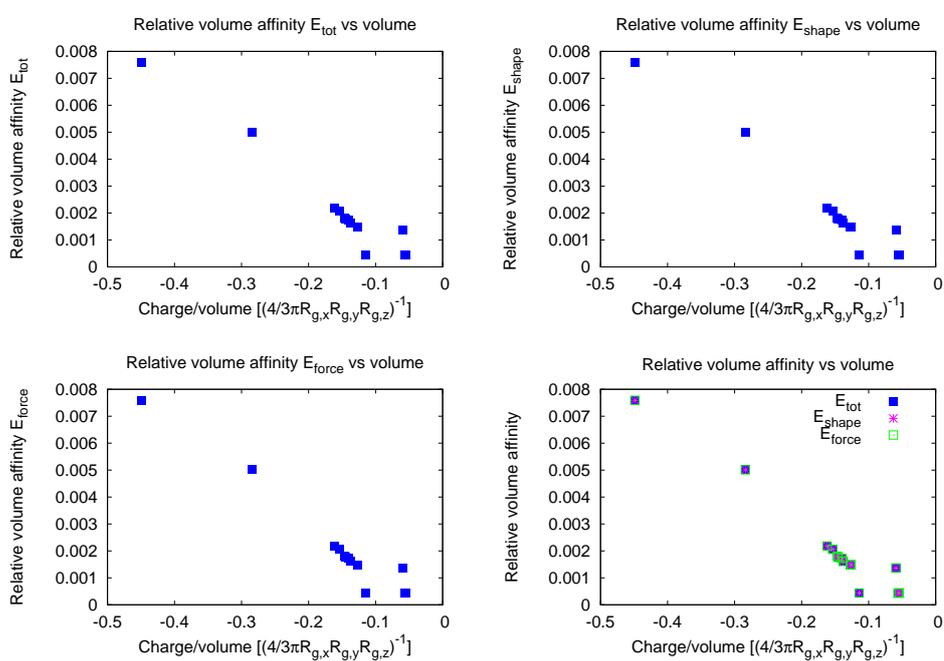


Figure 4.156: Plot of relative volume affinity for E_{tot} , E_{shape} and E_{force} versus $\frac{charge}{volume}$.

4.6.9.5 Comparison between neutral and deprotonated molecules

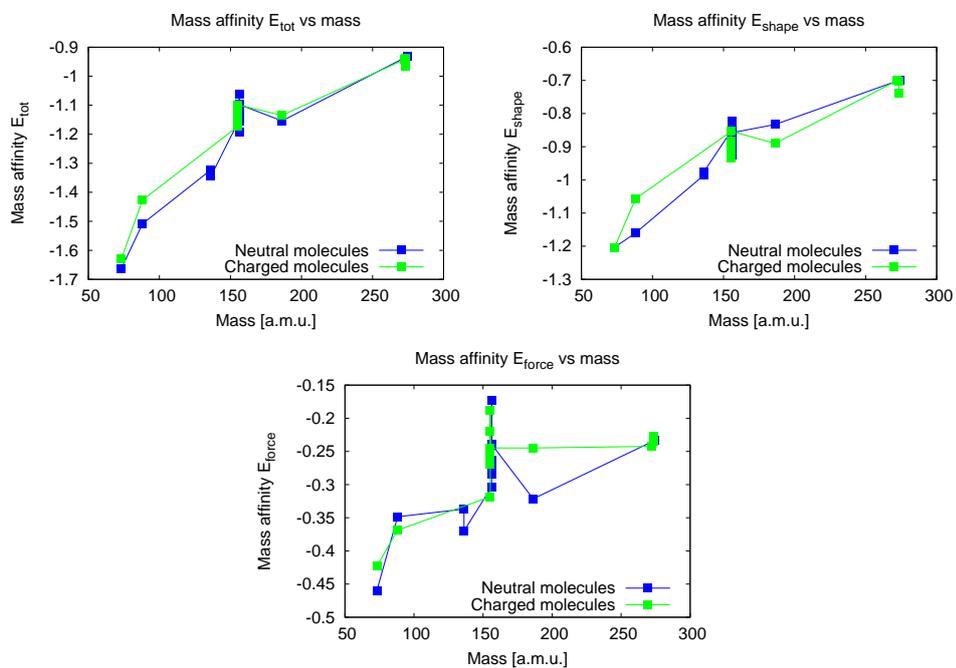


Figure 4.157: Plot of mass affinity for E_{tot} , E_{shape} and E_{force} versus masses.

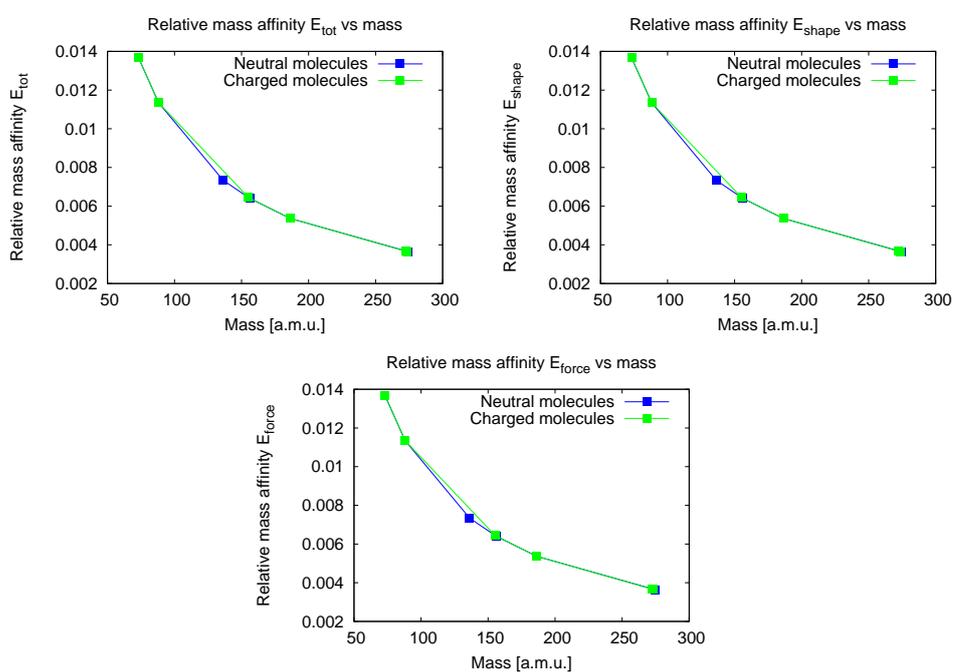


Figure 4.158: Plot of relative mass affinity for E_{tot} , E_{shape} and E_{force} versus masses.

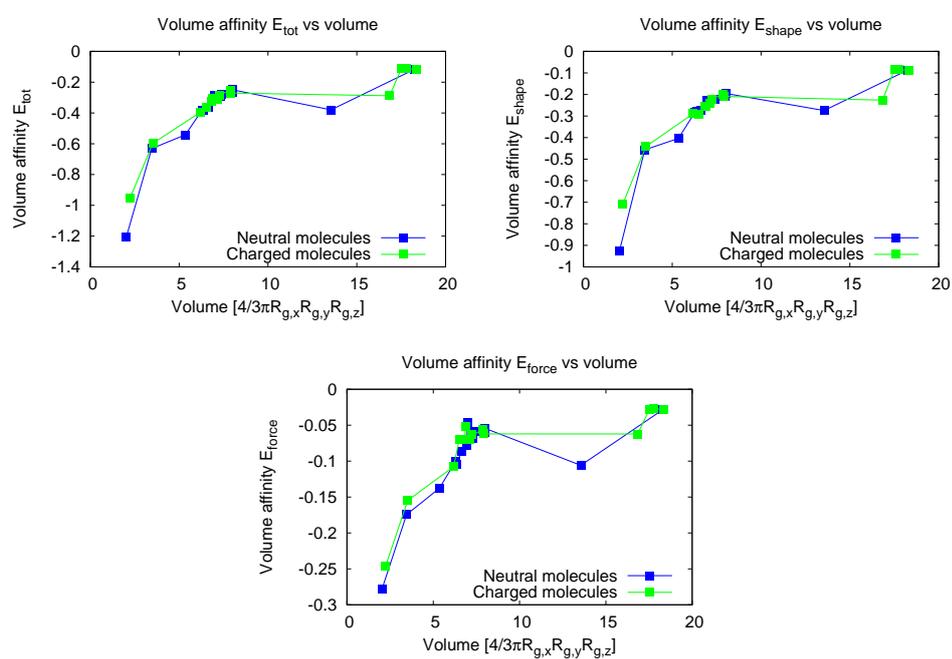


Figure 4.159: Plot of volume affinities for E_{tot} , E_{shape} and E_{force} versus volumes.

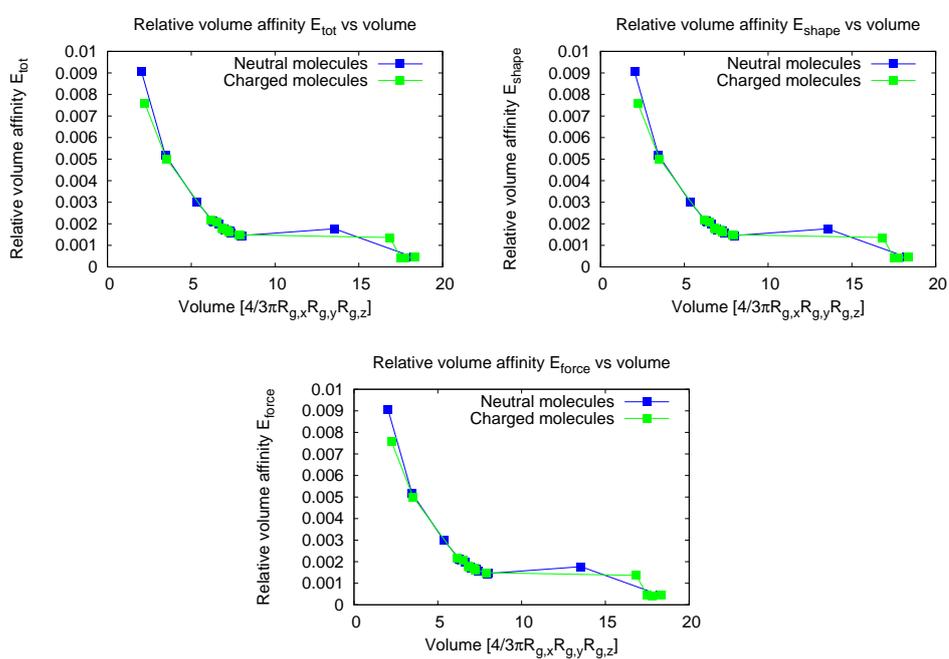


Figure 4.160: Plot of relative volume affinity for E_{tot} , E_{shape} and E_{force} versus volumes.

4.6.9.6 Helicoidal order parameter

In order to try and understand better how solute molecules interact with the helix, a more general version of the helicoidal order parameter is set up: $\chi = \langle \hat{\mathbf{z}} \cdot (\hat{\mathbf{r}} \times \hat{\mathbf{u}}) \rangle$. The $\hat{\mathbf{z}}$ axis is computed as the orientation of the helix, while the $\hat{\mathbf{r}}$ versor is computed as the direction between the center of the PBLG helix (namely the $\hat{\mathbf{z}}$ -axis) and the center of mass of the solute molecule. The $\hat{\mathbf{u}}$ versor is computed between the center of mass of the solute and an atom of the farthest part of the solute molecule. Figure 4.161 shows how this parameter is computed for a molecule of heptyl butyrate: the last atom of the carbon chain of HEP, C11 has been chosen.

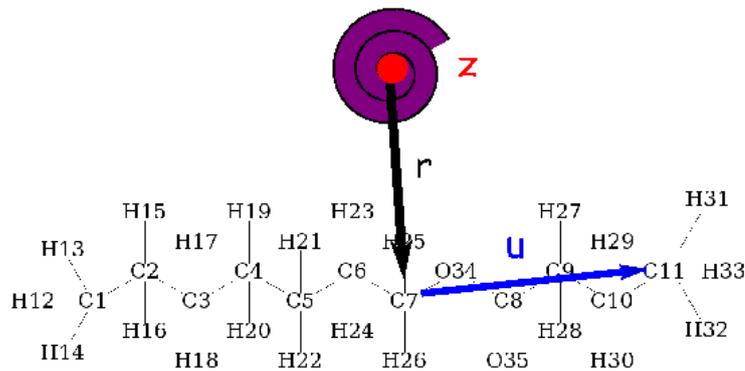


Figure 4.161: How to compute the helicoidal parameter for a HEP residue.

Molecule	Atom chosen	Helicoidal parameter	
		Neutral	Anion
Dimethylformamide	C4	0.658128	0.738936
Heptyl butyrate	C11	-0.242118	-0.0416994
Ethyl acetate	C05	-0.757072	-0.0676533
(R)-Limonene	C07	-0.304415	
(S)-Limonene		-0.169283	
(+)-menthol	C07	-0.204495	0.155806
(-)-menthol		-0.439276	0.570589
(+)-isomenthol		0.491796	-0.589456
(-)-isomenthol		-0.276025	-0.302937
(+)-neomenthol		-0.64615	0.227983
(-)-neomenthol		0.0531706	-0.0200674
(+)-neoisomenthol		0.0316853	-0.446866
(-)-neoisomenthol		-0.460454	-0.476245
(1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol	H02	-0.338039	-0.362316 0.334234 -0.182923

Table 4.22: Helicoidal order parameters.

4.6.9.7 Computing distance between backbone chiral atoms and solute

We also compute distances between the chiral atoms of the PBLG helix (namely N1, C2 and C7) and the center of mass of the solute molecule or the (pro-)chiral atom/atoms of the solute itself.

Table 4.23: Distance between the center of mass and chiral atom of solute from the PBLG backbone chiral atoms.

Molecule	Distance of the center of mass of solute from the nearest			Chiral atom	Distance of the chiral atom of solute from the nearest		
	N1	C2	C7		N1	C2	C7
Dimethylformamide	5.15178	4.93674	4.54671	N1	5.03637	4.82881	4.44523
Heptyl butyrate	4.15569	4.60501	3.77723	C8	4.86073	4.45556	4.54311
Ethyl acetate	5.85718	5.24887	5.10777	C04	1.90334	2.12201	1.52376
(R)-Limonene	4.42175	4.95636	4.16193	C05	4.90155	5.40795	4.7158
(S)-Limonene	4.73821	5.3846	4.76561	C05	4.02469	4.42715	4.37251
(+)-menthol	4.33627	4.77756	4.35213	C02	1.90334	2.12201	1.52376
				C04	1.52376	1.90334	2.12201
				C05	1.90334	2.12201	1.52376
(-)-menthol	6.11736	5.63906	6.21017	C02	4.64722	5.06579	4.85597
				C04	5.81149	5.31416	5.48635
				C05	6.11046	5.25568	6.08725
(+)-isomenthol	5.41445	4.48968	5.35455	C02	4.06591	4.66522	4.10539
				C04	4.33868	3.53543	4.28163
				C05	5.8571	5.05224	5.73356
(-)-isomenthol	5.87883	5.88762	5.43252	C02	5.98769	5.27878	5.27014
				C04	4.99065	5.48298	4.79137
				C05	5.9441	6.91284	5.59878
(+)-neomenthol	5.60157	5.53777	5.06161	C02	7.17873	7.96254	8.68434
				C04	6.71355	6.92525	6.28147
				C05	5.51371	5.87861	5.24081
(-)-neomenthol	5.03956	5.19383	4.67522	C02	4.43111	5.00937	4.38723
				C04	6.32587	6.12377	5.77543
				C05	6.43352	6.58278	5.92416
(+)-neoisomenthol	5.50188	5.78248	5.16804	C02	3.7396	4.14584	3.46166
				C04	6.23312	6.05809	5.66695
				C05	6.42783	6.72558	5.9973
(-)-neoisomenthol	6.01619	6.09408	5.56738	C02	5.90999	5.16577	5.18732
				C04	4.40043	4.82865	4.10077
				C05	5.59771	5.90374	5.19474
(1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol	5.28587	5.91821	5.37225	C07	4.80325	5.16586	5.11904
				C08	1.90334	2.12201	1.52376

4.6.9.8 Charge distribution

Molecules

Next, we have computed the charge distribution, in order to see how can solutes interact from the electrostatic point of view with the helix. We used the AM1 method for calculating atomistic charges with GAMESS [135].

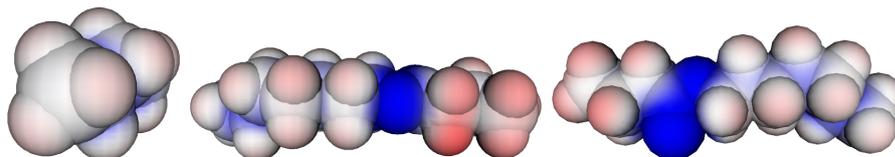


Figure 4.162: Charge distribution of heptyl butyrate.

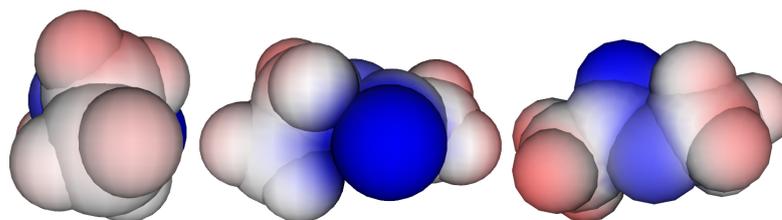


Figure 4.163: Charge distribution of ethyl acetate.

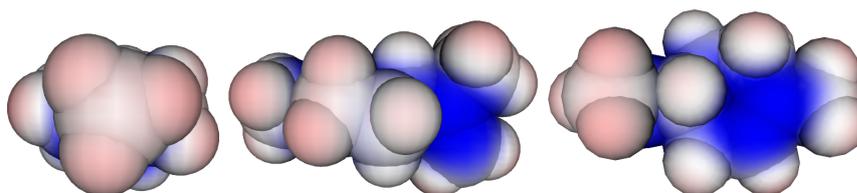


Figure 4.164: Charge distribution of (R)-limonene.

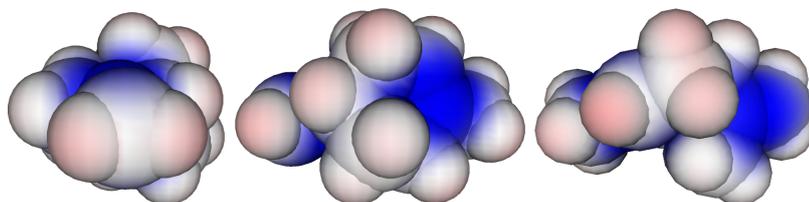


Figure 4.165: Charge distribution of (S)-limonene.

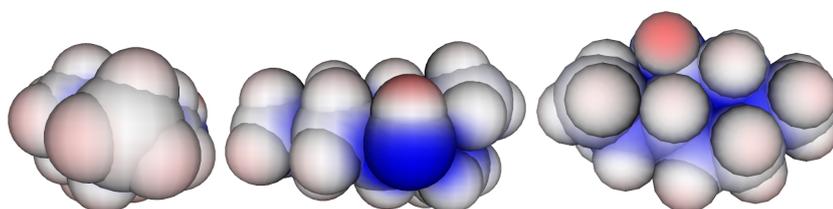


Figure 4.166: Charge distribution of (+)-menthol.

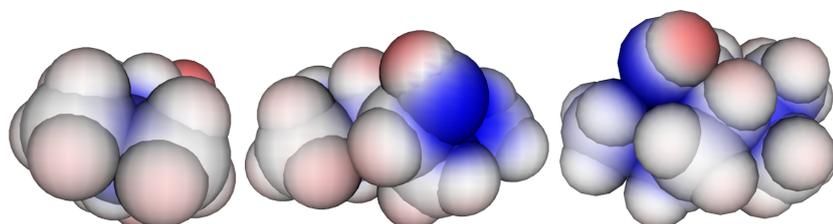


Figure 4.167: Charge distribution of (-)-menthol.

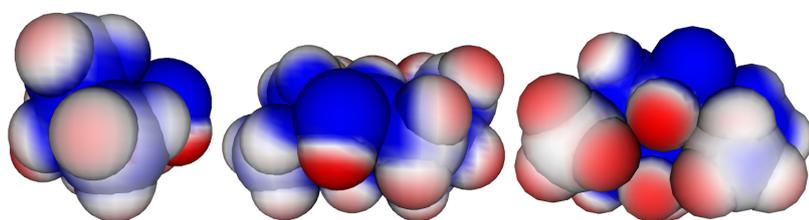


Figure 4.168: Charge distribution of (+)-isomenthol.

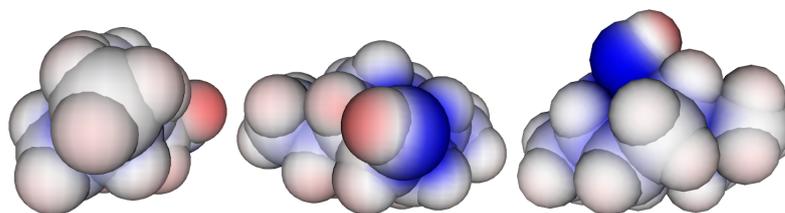


Figure 4.169: Charge distribution of (-)-isomenthol.

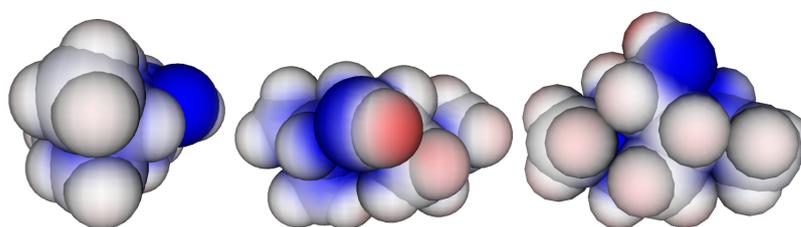


Figure 4.170: Charge distribution of (+)-neomenthol.

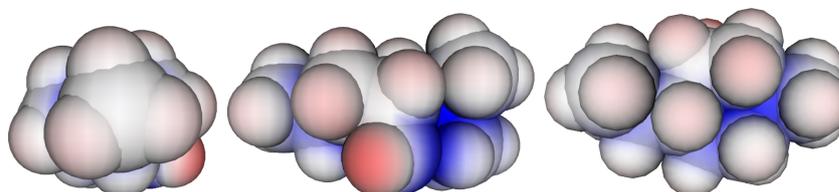


Figure 4.171: Charge distribution of (-)-neomenthol.

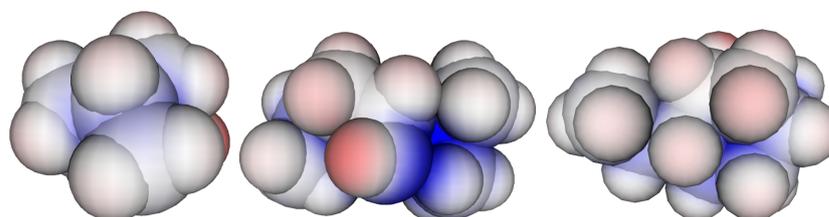


Figure 4.172: Charge distribution of (+)-neoisomenthol.

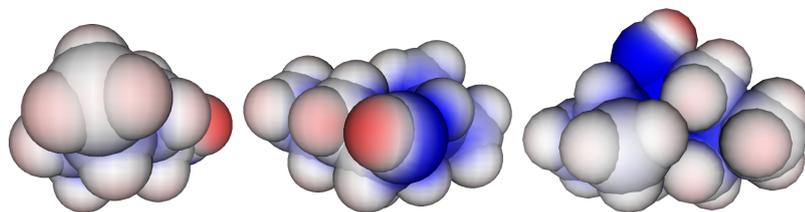


Figure 4.173: Charge distribution of (-)-neoisomenthol.

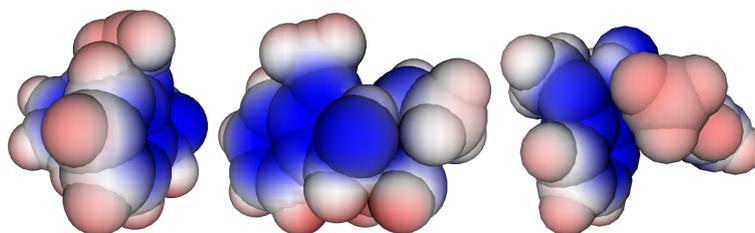


Figure 4.174: Charge distribution of (1R,2R)-1,2-bis-(2-methoxy-phenyl)-ethane-1,2-diol.

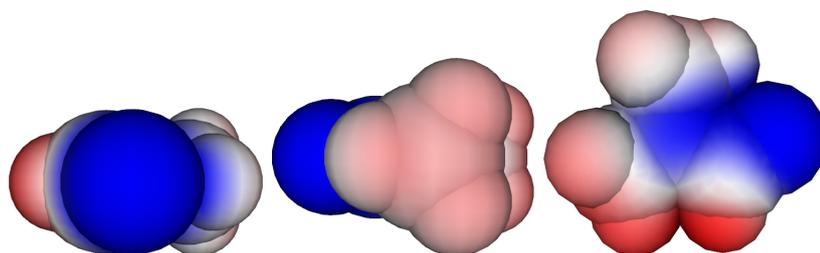


Figure 4.175: Charge distribution of dimethylformamide.

Deprotonated

Also for anions we computed the atomistic charge distribution.

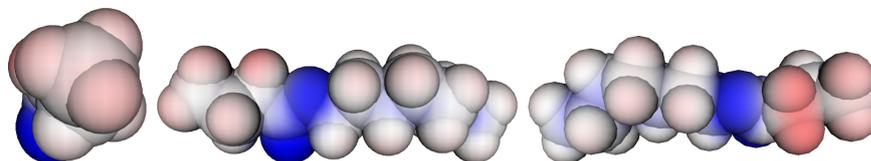


Figure 4.176: Charge distribution of an anion of heptyl butyrate.

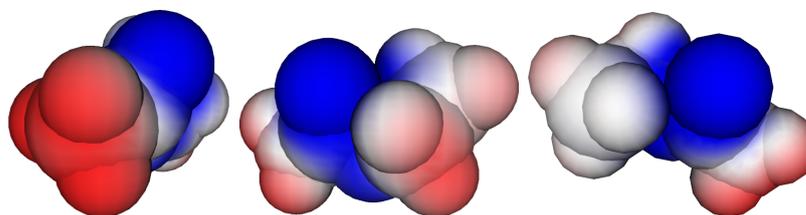


Figure 4.177: Charge distribution of an anion of ethyl acetate.

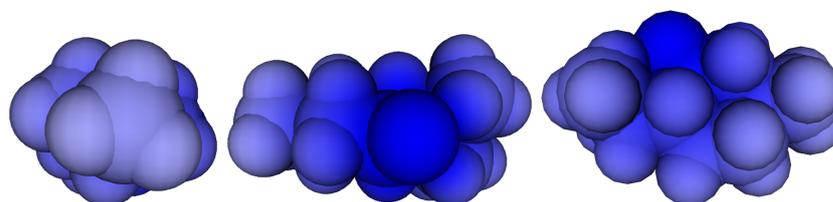


Figure 4.178: Charge distribution of an anion of (+)-menthol.

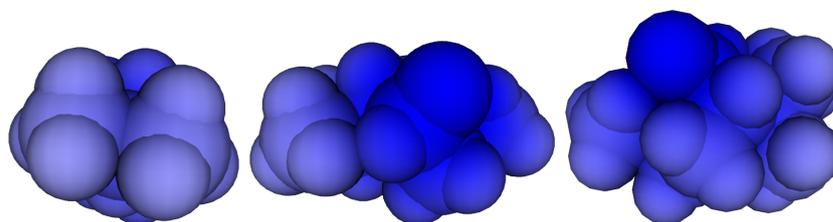


Figure 4.179: Charge distribution of an anion of (-)-menthol.

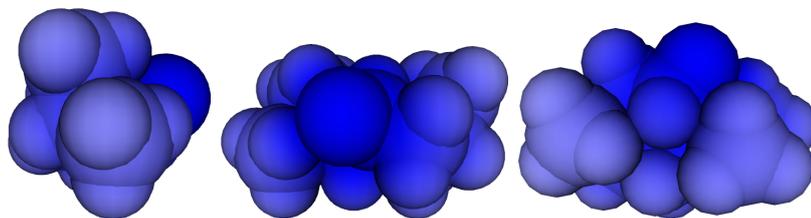


Figure 4.180: Charge distribution of an anion of (+)-isomenthol.

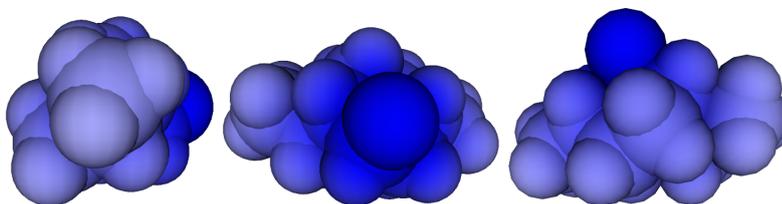


Figure 4.181: Charge distribution of an anion of (-)-isomenthol.

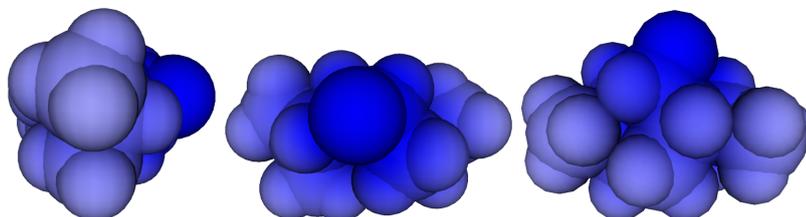


Figure 4.182: Charge distribution of an anion of (+)-neomenthol.

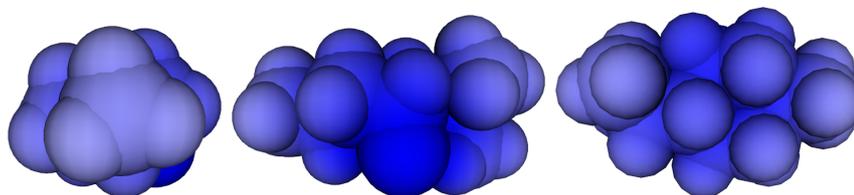


Figure 4.183: Charge distribution of an anion of (-)-neomenthol.

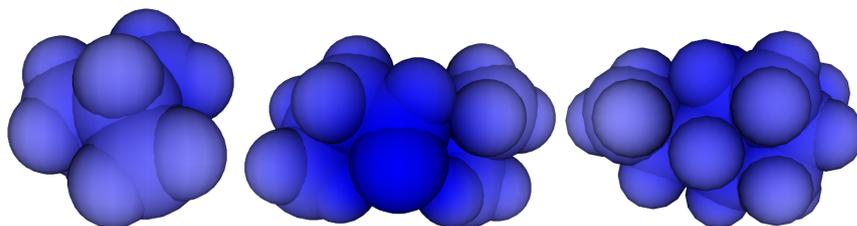


Figure 4.184: Charge distribution of an anion of (+)-neoisomenthol.

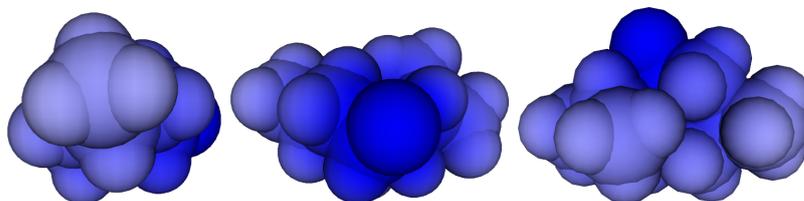


Figure 4.185: Charge distribution of an anion of (-)-neoisomenthol.

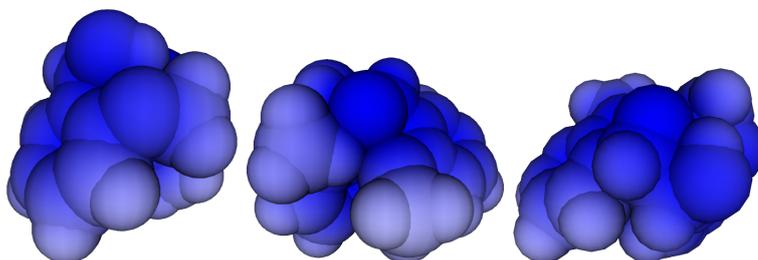


Figure 4.186: Charge distribution of an anion of (1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol.

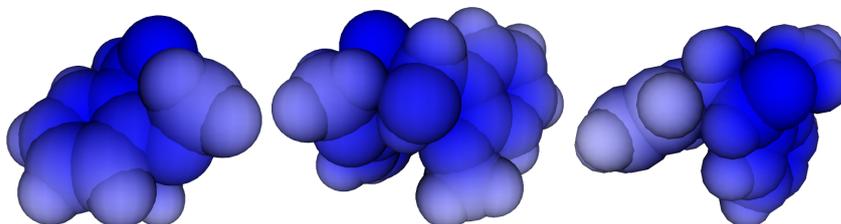


Figure 4.187: Charge distribution of another anion of (1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol.

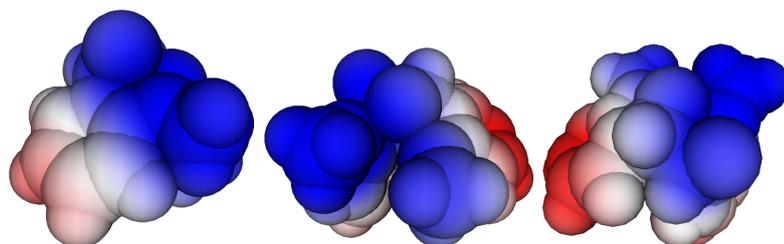


Figure 4.188: Charge distribution of a dianion of (1R,2R)-1,2-Bis-(2-methoxy-phenyl)-ethane-1,2-diol.

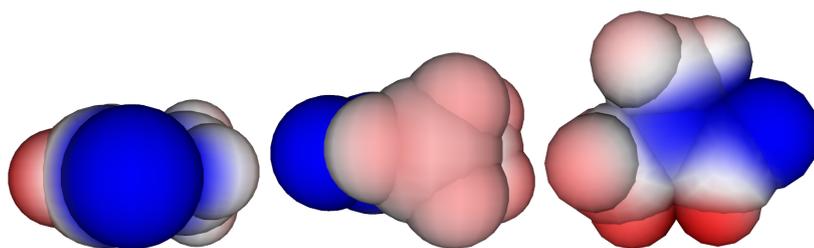


Figure 4.189: Charge distribution of an anion of dimethylformamide.

PBLG

It follows charge distribution of the helix of PBLG (19 residues long) that has been used for docking runs.

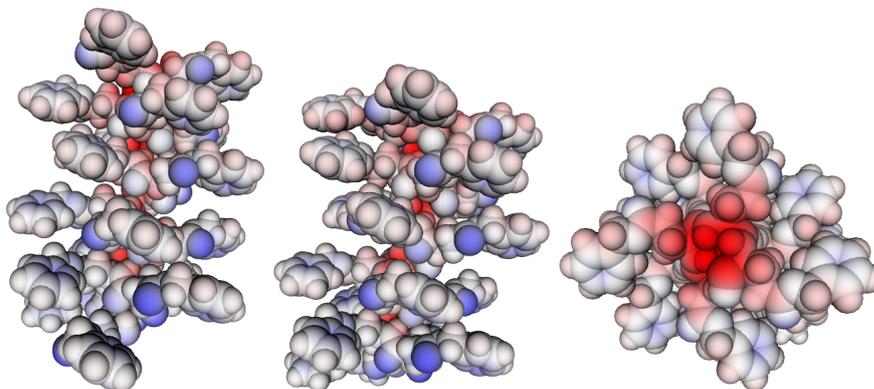


Figure 4.190: Charge distribution of an helix of PBLG.

4.6.9.9 Pocket volume

Finally, using the software “Pocket Finder” [136], we have computed the volume of the two predicted binding site found.

First pocket

Min. Coords: (-2, -4, 7)

Max Coords: (5, 4, 15)

Predicted site 1

Site Volume: 11 A³

Protein Volume: 1326 A³

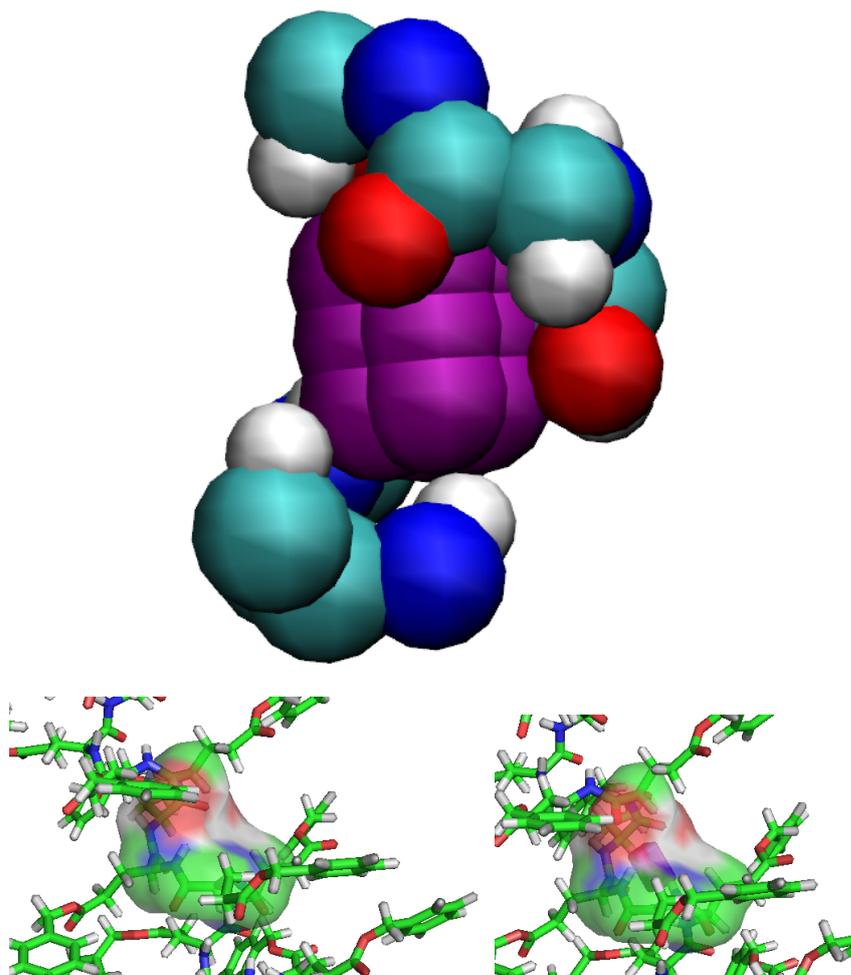


Figure 4.191: First binding site found for a molecule, with the volume showed in purple.

Second pocket

Min. Coords: (-2, -7, 4)

Max Coords: (9, 4, 15)

Predicted site 2

Site Volume: 7 Å³

Protein Volume: 1326 Å³

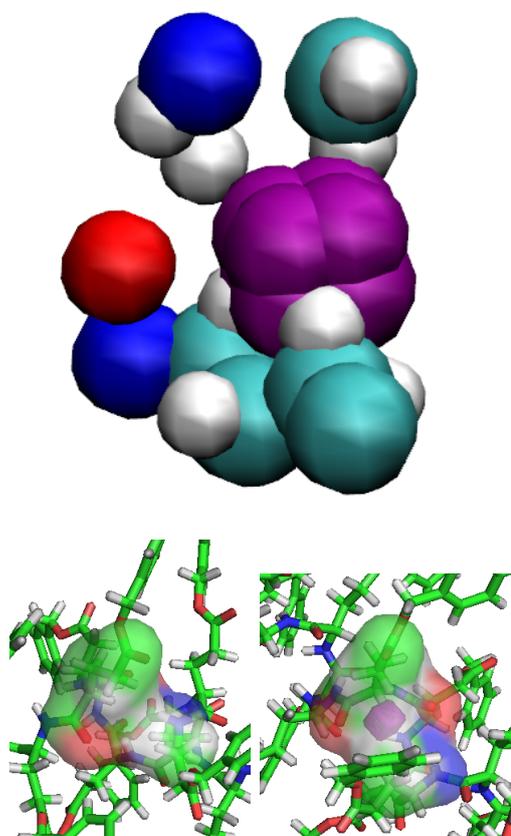


Figure 4.192: Second binding site found for a molecule with the volume showed in purple.

4.6.10 Docking of HEP starting from various rotated configuration

4.6.10.1 Docking parameters

We use a code called *Hex* [133] to dock a chain of 19 residue of BGL with heptyl butyrate.

Correlation Type:	Shape+Electrostatics
FFT Mode:	3D
Post Processing:	MM Minimisation
Grid Dimension:	0.6
Solutions:	2000
Receptor Range:	180
Step Size:	7.5
Ligand Range:	180
Step Size:	7.5
Twist Range:	360
Step Size:	5.5
Distance Range:	40
Scan Step:	0.8
SubSteps:	0
Steric Scan:	18
Final Search	31

4.6.10.2 Initial sample

PBLG

The receptor is an α_{18} helix of PBLG.

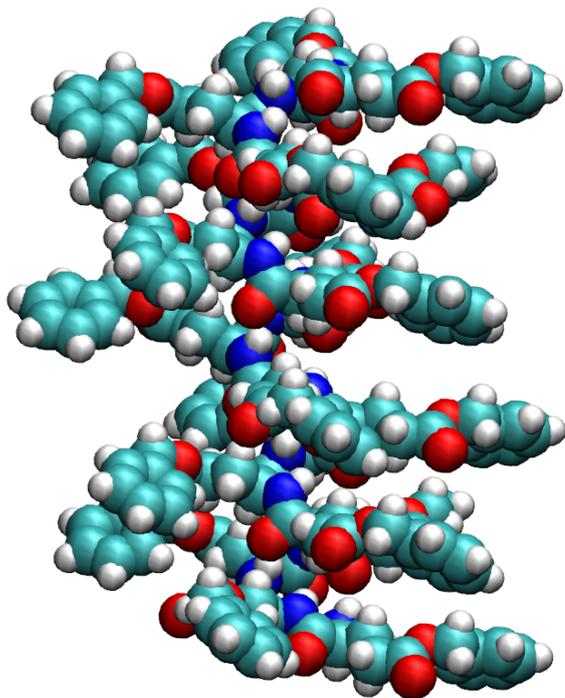


Figure 4.193: Snapshot of the helix of PBLG used.

HEP

The ligand is heptyl butyrate.

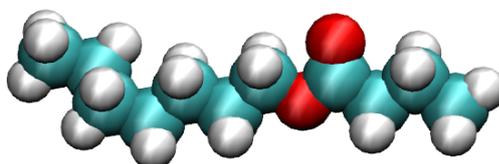


Figure 4.194: Snapshot of the heptyl butyrate molecule used.

This is the first orientation attempted. Next, we rotate this molecule using a set of 15 Euler angles (Table 4.24) and dock the resulting configuration with the PBLG helix.

Orientation	α	β	γ
1	0	15	30
2	0	70	90
3	15	0	50
4	90	0	30
5	80	5	0
6	15	70	0
7	10	50	80
8	50	10	80
9	80	10	50
10	80	50	10
11	70	5	35
12	65	15	55
13	25	30	60
14	70	55	45
15	85	5	25

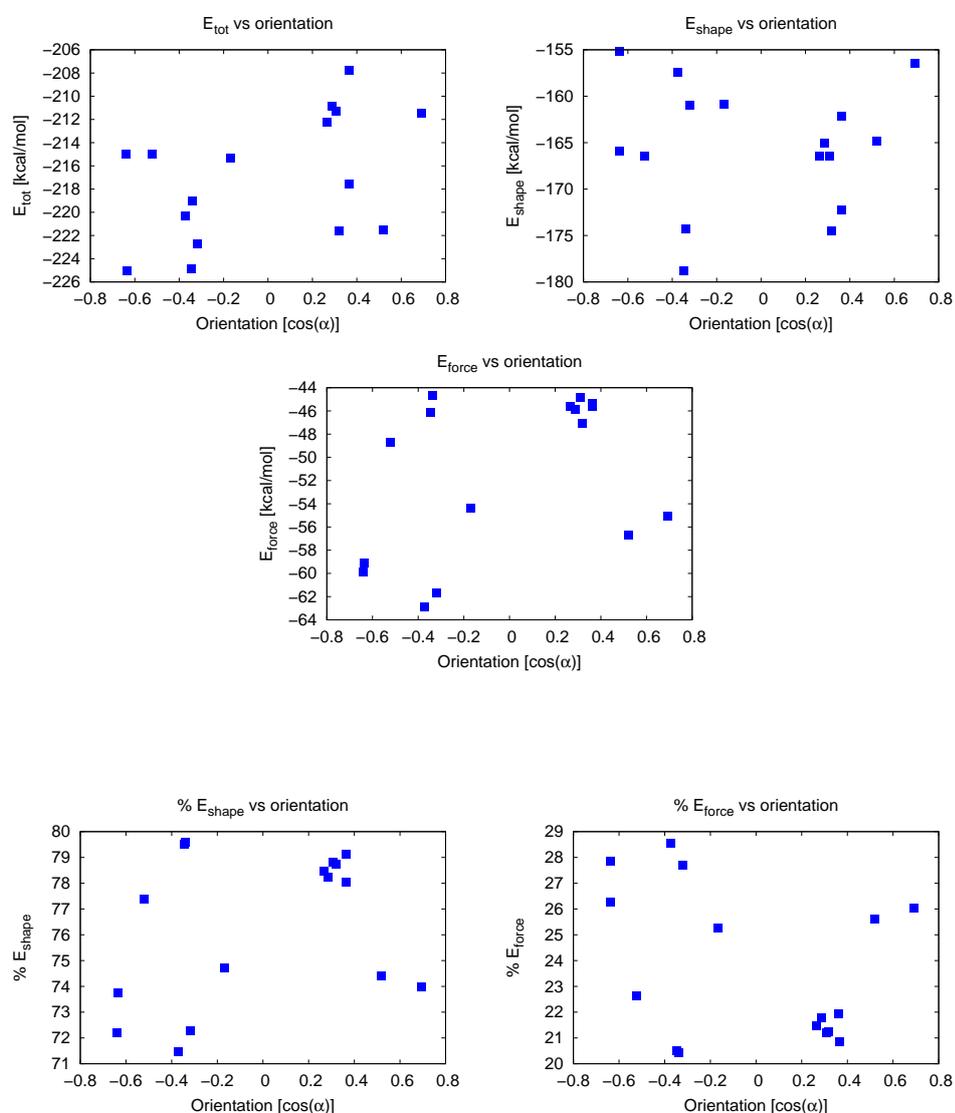
Table 4.24: Orientations: Euler angles.

4.6.10.3 Parameters analyzed

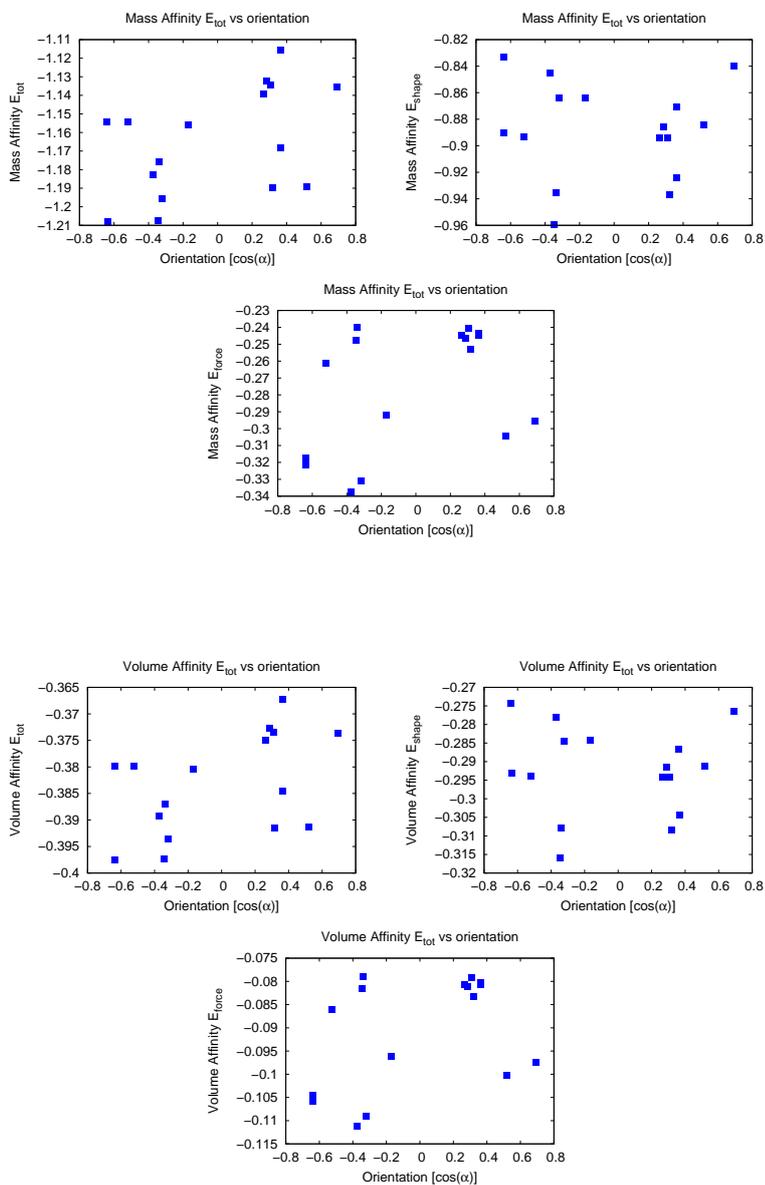
From *HEX* [133] code, we have found collected data of energy: E_{tot} , E_{shape} and E_{force} (kcal/mol). Dividing this parameters by the molecular mass and the volume, respectively molecular and volume affinity are computed. Next, we computed the distance from the three atoms of PBLG backbone (N1, C2 and C7) of the geometric center and the prochiral atom (C8) of the solute. In order to understand in which pocket the heptyl butyrate bonds, also distances from the center of mass to atom C3 and C6 of the PBLG helix (typical of the second, smaller pocket) were computed. Finally, the helicoidal order parameter has been computed: $\chi = \langle \hat{\mathbf{z}} \cdot (\hat{\mathbf{r}} \times \hat{\mathbf{u}}) \rangle$. The $\hat{\mathbf{z}}$ axis is computed as the orientation of the helix, while the $\hat{\mathbf{r}}$ versor as the direction between the center of the PBLG helix (namely the $\hat{\mathbf{z}}$ -axis) and the geometrical center of the solute molecule. The $\hat{\mathbf{u}}$ versor is computed as the three inertial axes of the solute (x is the shortest and z the longest). The orientation of the solute is computed as the cosine of the angle between the axis of the helix and the longest inertial axis of the solute.

4.6.10.4 Results

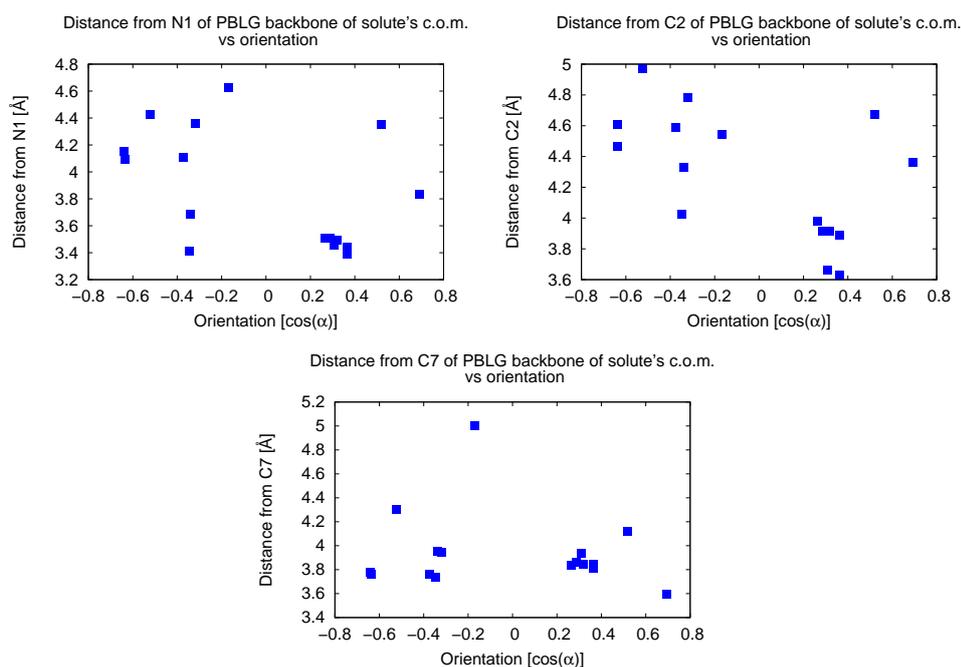
From the figures below, showing the energy terms versus the orientation, we can see that “negative orientations” of the solute show a little more stable complex with the helix (more negative energy values). For what concern the percentage of energy due to electrostatic interactions and to shape interaction, it seems that the electrostatic forces are stronger for “negative orientations” of HEP, while the shape term is not really determined by the orientation of the solute.



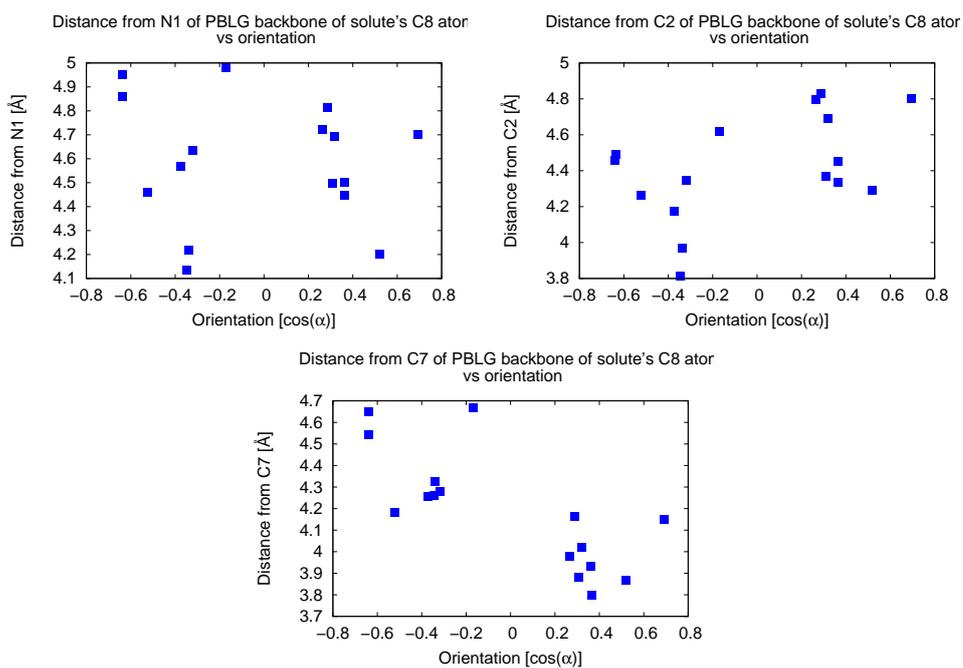
The lowest (negative) mass affinities are shown for “negative orientations” of solute. The volume ones show the same trend even if more slightly, according with what was seen in the previous page.



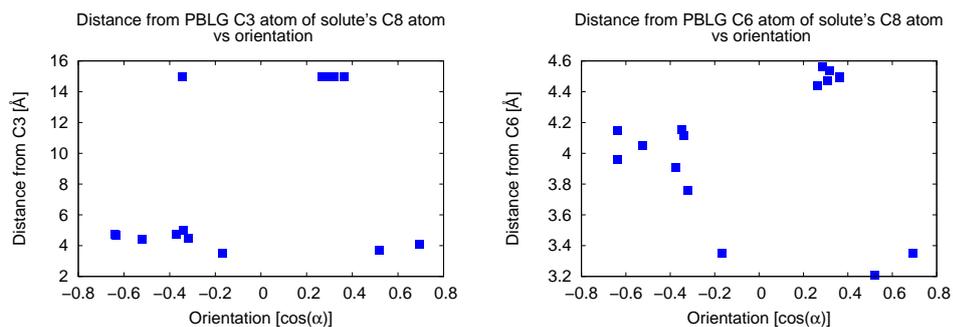
On the contrary, the distance of the center of mass of the solute and the backbone atoms of the helix are almost independently of the orientations. However, the distance from the C2 atom is shorter for “positive orientations” of the solute. So it seems that “positively oriented” solutes tend to interact better with the backbone atoms of the PBLG, even if their complex is less stable.



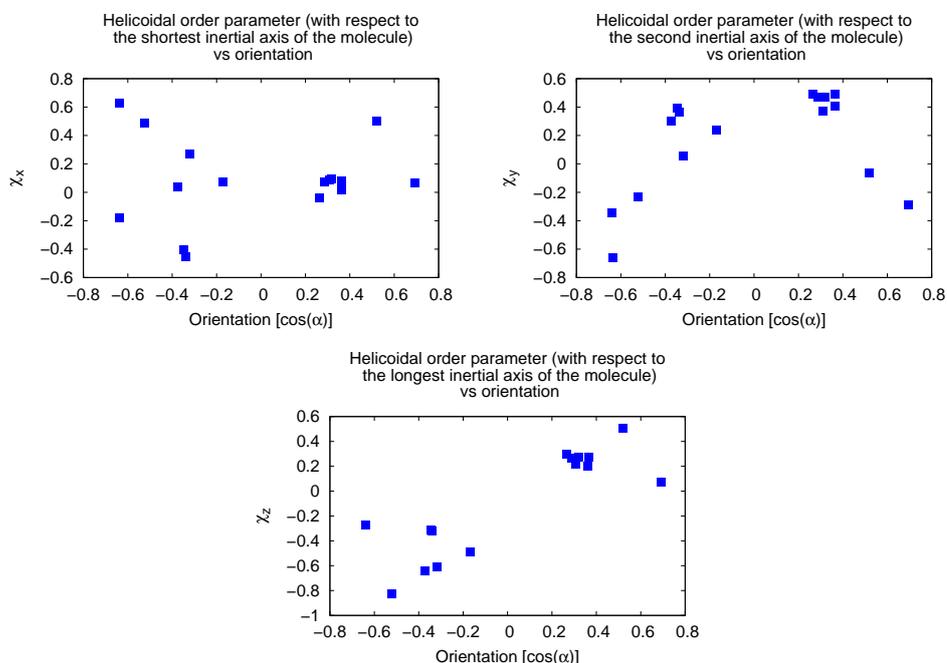
On the other hand, if we consider the distances of the C8 atom of the solute with respect to the three atoms of the backbone of the PBLG helix, we can see that this prochiral atom of the solute is nearer to the C2 atom of the backbone if the molecule approaches the helix with a “negative orientation” and nearer to the C7 one if approaching it with a “positive orientation”.



From the below graphs, we can suggest that the solute is near enough to the atoms typical only of the smallest pocket to consider this pocket as the one in which the solute interacts. However, it seems that when the solute is in a orientation between 0.0 and 0.4 the preferred pocket of docking is the biggest one (the distances are greater than for the other orientations).



In these graphs, we have computed the helicoidal order parameters taking into account the three inertial axes of the solute molecules. From the last one, showing the interactions with the longest inertial axes of the solute, we can see that if it approaches the PBLG helix with a “negative orientation” (so similar to the helicity of the PBLG, i.e. right-handed), the solute will have negative values of chiral index, meaning that the interaction will be in the sense of the helix; while if it approaches the helix with a “positive orientation” it will interact with the helix taking a left-handed orientation, contrary to the helicity of the PBLG.



Bibliography

- [1] T. Bellini, R. Cerbino, and G. Zanchetta. Dna-based soft phases. *Top. Curr. Chem.*, 318:225–280, 2012.
- [2] L.A. Madsen, T.J. Dingemans, M. Nakata, and E.T. Samulski. Thermotropic biaxial nematic liquid crystals. *Phys. Rev. Lett.*, 92:145505, 2004.
- [3] R. Berardi, L. Muccioli, S. Orlandi, M. Ricci, and Zannoni C. Computer simulation of biaxial nematics. *J. Phys.: Condens. Matter*, 20:464101, 2008.
- [4] P.G. de Gennes and J. Prost, editors. *The Physics of Liquid Crystals*. Clarendon Press, Oxford, 1993.
- [5] S. Chandrasekhar, editor. *Liquid Crystals*. Cambridge University Press, 2nd ed. edition, 1993.
- [6] Dierking, editor. *Textures of Liquid Crystals*. Wiley-VCH., Weinheim, 2003.
- [7] L. J. Yu and A. Saupe. Observation of a Biaxial Nematic Phase in Potassium Laurate-1-Decanol-Water Mixtures. *Phys. Rev. Lett.*, 45:1000–1003, 1980.
- [8] C. Tschierske and D. J. Photinos. Biaxial nematic phases. *J. Mater. Chem.*, 20:4263–4294, 2010.
- [9] G. R. Luckhurst. Liquid crystals: A missing phase found at last? *Nature*, 430:413–414, 2004.

- [10] R. Berardi, L. Muccioli, S. Orlandi, M. Ricci, and C. Zannoni. Computer simulation of biaxial nematics. *J. Phys. Condens. Matter*, 20:464101, 2008.
- [11] R. Berardi, L. Muccioli, and C. Zannoni. Field response and switching times in biaxial nematics. *J. Chem. Phys.*, 128:024905.1–12, 2008.
- [12] M. J. Freiser. Ordered states of a nematic liquid. *Phys. Rev. Lett.*, 24:1041–1043, 1970.
- [13] F. Biscarini, C. Chiccoli, P. Pasini, F. Semeria, and C. Zannoni. Phase Diagram and Orientational Order in a Biaxial Lattice Model: A Monte Carlo Study. *Phys. Rev. Lett.*, 75:1803–1806, 1995.
- [14] M. P. Allen. Computer simulation of a biaxial liquid crystal. *Liq. Cryst.*, 8:499–511, 1990.
- [15] R. Berardi and C. Zannoni. Do thermotropic biaxial nematics exist? A Monte Carlo study of biaxial GayBerne particles. *J. Chem. Phys.*, 113:5971–5979, 2000.
- [16] G. R. Luckhurst. Biaxial nematic liquid crystals: fact or fiction? *Thin Solid Films*, 393:40–52, 2001.
- [17] F. Hessel and H. Finkelmann. Optical biaxiality of nematic LC-side chain polymers with laterally attached mesogenic groups. *Polymer Bulletin*, 15:349–352, 1986.
- [18] K. Severing and K. Saalwacher. Biaxial nematic phase in a thermotropic liquid–crystalline side–chain polymer. *Phys. Rev. Lett.*, 92:125501, 2004.
- [19] K. Severing, E. Stibal-Fischer, A. Hasenhindl, H. Finkelmann, and K. Saalwacher. Phase biaxiality in nematic liquid crystallin side–chain polymers of various chemical consistutions. *J. Phys. Chem. B*, 110:15680, 2006.

- [20] L. A. Madsen, T. J. Dingemans, M. Nakata, and E. T. Samulski. Thermotropic biaxial nematic liquid crystals. *Phys. Rev. Lett.*, 92:145505, 2004.
- [21] B. R. Acharya, A. Primak, and S. Kumar. Biaxial nematic phase in bent-core thermotropic mesogens. *Phys. Rev. Lett.*, 92:145506, 2004.
- [22] K. Merkel, A. Kocot, J. K. Vij, R. Korlacki, G. H. Mehl, and T. Meyer. Thermotropic biaxial nematic phase in liquid crystalline organo-siloxane tetrapodes. *Phys. Rev. Lett.*, 93:237801, 2004.
- [23] R. Berardi, S. Orlandi, and C. Zannoni. Antiphase structures in polar smectic liquid crystals and their molecular origin. *Chem. Phys. Lett.*, 261:357, 1996.
- [24] M. Bruno, T. Pierre, T. Hideo, and W. Junji. Theory of polar biaxial nematic phases. *Phys. Rev. E*, 66:031701, 2002.
- [25] N.V. Madhusudana. Role of molecular dipoles in liquid crystals. *Mol. Cryst. Liq. Cryst.*, 409:371–387, 2004.
- [26] H. Murthy and B.K. Sadashiva. A polar biaxial smectic a phase in new unsymmetrical compounds composed of bent-core molecules. *Liq. Cryst.*, 31:567–578, 2004.
- [27] R. Berardi, S. Orlandi, and C. Zannoni. Molecular dipoles and tilted smectic formation. A Monte Carlo study. *Phys. Rev.E.*, 124:041708.1–14, 2003.
- [28] S. Orlandi, R. Berardi, J. Stelzer, and C. Zannoni. A Monte Carlo study of the mesophases formed by polar bent-shaped molecules. *J. Chem. Phys.*, 124:124907.1–9, 2006.
- [29] A. Januszko, K.L. Glab, P. Kaszynski, K. Patel, R.A. Lewis, G.H. Mehl, and M.D. Wand. The effect of carborane, bicyclo[2.2.2] octane and benzene on mesogenic and dielectric properties of laterally fluorinated three-ring mesogens. *J. Mat. Chem.*, 16:3183–3192, 2006.

- [30] R. Berardi, C. Fava, and C. Zannoni. A generalized Gay–Berne intermolecular potential for biaxial particles. *Chem. Phys. Lett.*, 236:462–468, 1995.
- [31] R. Berardi, C. Fava, and C. Zannoni. A Gay-Berne potential for dissimilar biaxial particles. *Chem. Phys. Lett.*, 297:8–14, 1998.
- [32] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.*, 102:5451–5459, 1995.
- [33] R. Berardi, J.S. Lintuvuori, M.R. Wilson, and C. Zannoni. Phase diagram of the uniaxial and biaxial soft-core gay-berne model. *J. Chem. Phys.*, 135:134119.1–134119.10, 2011.
- [34] R. Berardi and C. Zannoni. Low-temperature biaxial nematic from rod and disc mesogen mixture. *Soft Matter*, 8:2017–2025, 2012.
- [35] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola, and J. R. Haak. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, 81:3684–3690, 1984.
- [36] R. Berardi, L. Muccioli, S. Orlandi, M. Ricci, and C. Zannoni. Computer simulation of biaxial nematics. *J. Phys. Condens. Matter*, 20:464101, 2008.
- [37] R. Berardi, S. Orlandi, and C. Zannoni. Columnar phases and field induced biaxiality of a Gay-Berne discotic liquid crystals. *Phys. Chem. Chem. Phys.*, 2:2933–2942, 2000.
- [38] F. Biscarini, C. Chiccoli, P. Pasini, F. Semeria, and C. Zannoni. Phase Diagram and Orientational Order in a Biaxial Lattice Model: A Monte Carlo Study. *Phys. Rev. Lett.*, 75:1803–1806, 1995.
- [39] R. Berardi and C. Zannoni. Do thermotropic biaxial nematics exist? A Monte Carlo study of biaxial GayBerne particles. *J. Chem. Phys.*, 113:5971–5979, 2000.

- [40] M. P. Allen. Computer simulation of a biaxial liquid crystal. *Liq. Cryst.*, 8:499–511, 1990.
- [41] R. Berardi and C. Zannoni. Biaxial discotic gay - berne mesogens and biaxial nematics. *Mol. Cryst. Liq. Cryst.*, 396 PN Part 1:177–186, 2003.
- [42] M. Ricci, R. Berardi, and C. Zannoni. Columnar liquid crystals formed by bowl-shaped mesogens. a monte carlo study. *Soft Matter*, 4:2030–2038, 2008.
- [43] F.H. Stillinger. A topographic view of supercooled liquids and glass formation. *Science*, 267:1935–1939, 1995.
- [44] L. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.*, 102:5451–5459, 2009.
- [45] D. Wolf, S. R. Koblinski, and J. Eggebrecht. Exact method for the simulation of coulombic systems by spherically truncated, pairwise r^*-1 summation. *J. Chem. Phys.*, 110:8254–8282, 1999.
- [46] Sponer and Lankas, editors. *Computational Studies of RNA and DNA*. Sponer and Lankas.
- [47] M. Mandelkern, J. Elias, D. Eden, and D. Crothers. The dimensions of dna in solution. *J. Mol. Biol.*, 152:153–161, 1981.
- [48] Chieffi, Dolfini, Malcovati, Pierantoni, and Tenchini, editors. *Genetica e Biologia*. EdiSES.
- [49] J.D. Watson and F.H.C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
- [50] Sinden and Richard, editors. *DNA structure and function*. Academic Press., 1994.

- [51] A.H.J. Wang, G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. van Boom, G. Van der Marel, and A. Rich. Molecular structure of a left-handed double helical dna fragment at atomic resolution. *Nature*, 282:680–686, 1979.
- [52] R.H. Garrett and Grisham C.M., editors. *Biochimica*. Zanichelli, 2002.
- [53] D. Voet, G.J. Voet, and C.W. Pratt, editors. *Fondamenti di Biochimica*. Zanichelli.
- [54] S. Neidle, B. Schneider, and H.M. Berman, editors. *Fundamentals of DNA and RNA structure*. Wiley–Liss, 2003.
- [55] B. Basham, G.P. Schroth, and P.S. Ho. An a-dna triplet code: thermodynamic rules for predicting a- and b-dna. *Proc. Natl. Acad. Sci.*, 92:6464–6468, 1995.
- [56] A. Rich, A. Norheim, and A.H.J. Wang. The chemistry and biology of left-handed z-dna. *Ann. Rev. Biochem.*, 53:791–846, 1984.
- [57] Y.-W. Kwon, C.H. Lee, D.-H. Choi, and Jin J.-I. Materials science of dna. *J. Mat. Chem.*, 19:1353–1380, 2009.
- [58] I. Willner and E. Katz, editors. *Bioelectronics An Introduction*.
- [59] T.X. Hoang and M. Cieplak. *J. Chem. Phys.*, 112:6851, 2000.
- [60] P. Poulain, A. Saladin, B. Hartmann, and C. Prévost. Insights on protein–dna recognition by coarse grain modelling. *J. Comput. Chem.*, 29:2582, 2008.
- [61] R. Phillips. *The Bridge*, 34:22–27, 2004.
- [62] A.D. Jr. MacKerell, N. Banavali, and N. Foloppe. Development and current status of the charmm force field for nucleic acids. *Biopol.*, 56:257–265, 2001.

- [63] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Jr. Merz, 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, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.
- [64] G. D. Fasman, S. Arnott, P. J. C. Smith, and R. Chandrasekaran, editors. *Handbook of Biochemistry and Molecular Biology – Nucleic Acids – volume II*.
- [65] A. Blondel and M. Karplus. New formulation for derivatives of torsion angles and improper torsion angles in molecular mechanics: Elimination of singularities. *J. Comput. Chem.*, 17:1132–1141, 1996.
- [66] M. Nakata, G. Zanchetta, B.D. Chapman, C.D. Jones, J.O. Cross, R. Pindak, T. Bellini, and Clark N.A. End-to-end stacking and liquid crystal condensation of 6-to 20-base pair dna duplexes. *Science*, 318:1276–1279, 2007.
- [67] G. Zanchetta. Spontaneous self-assembly of nucleic acids: liquid crystal condensation of complementary sequences in mixtures of dna and rna oligomers. *Liquid Crystal Today*, 18:40–49, 2009.
- [68] A. Pietropaolo, L. Muccioli, R. Berardi, and C. Zannoni. A chirality index for investigatin protein secondary structures and their time evolution. *Proteins*, 70:667–677, 2008.
- [69] J. Bertram. The molecular biology of cancer. *Mol. Aspects. Med.*, 21 (6):167–223, 2000.
- [70] Y.T. Aminetzach, J.M. Macpherson, and D.A. Petrov. Pesticide resistance via transposition-mediated adaptive gene truncation in drosophila. *Science*, 309 (5735):764–7, 2005.
- [71] V Burrus and M. Waldor. Shaping bacterial genomes with interactive and conjugative elements. *Res. Microbiol.*, 155 (5):376–386, 2004.

- [72] J.W. Drake and J.J. Holland. Mutation rates among rna viruses. *Proc. Natl. Acad. Sci. U.S.A.*, 96 (24):13910–13913, 1999.
- [73] J. Holland, K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol. Rapid evolution of rna genomes. *Science*, 215 (4540):1577–1585, 1982.
- [74] users.rcn.com/jkimball.ma.ultranet/BiologyPages/M/Mutations.
- [75] V.A. Raker, A.A. Mironov, M.S. Gelfand, and D.D. Pervouchine. Modulation of alternative splicing by long-range rna structures in *drosophila*. *Nucleic Acid Research*, 37 (14):4533–4544, 2009.
- [76] G. Matlashewski, P. Lamb, D. Pim, J. Peacock, L. Crawford, and S. Benchimol. Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. *EMBO J.*, 3 (13):3257–3262, 1984.
- [77] M. Isobe, B. S. Emanuel, D. Givol, M. Oren, and C. M. Croce. Localization of gene for human p53 tumour antigen to band 17p13. *Nature*, 320 (6057):54–85.
- [78] S. E. Kern, K. W. Kinzler, A. Bruskin, D. Jarosz, P. Friedman, C. Prives, and B. Vogelstein. Identification of p53 as a sequence-specific DNA-binding protein. *Science*, 252 (5013):1708–1711, 1991.
- [79] A. P. Read and T. Strachan, editors. *Chapter 18: Cancer Genetics - Human molecular genetics 2*. New York: Wiley, 1999.
- [80] K.C. Timberlak, editor. *Chemistry*. Haper–Collins Publisher Inc., New York, 5th edition, 1992.
- [81] C. Aroulanda, D. Merlet, J. Courtieu, and P. Lesot. Nmr experimental evidence of the differentiation of enantiotopic directions c_s and c_{2v} molecules using partially oriented, chiral media. *J. Am. Chem. Soc.*, 123:12059–12066, 2001.

- [82] B. de Campos Vidal and M. L. Silveira Mello. Chirality and helicity of poly-benzyl-l-glutamate in liquid crystal and a wave structure that mimics collagen helicity in crimp. *Materials Research*, 4, n 3:169–173, 2001.
- [83] A. Meddour, P. Berdague, A. Hedli, J. Courtieu, and P. Lesot. Proton-decoupled carbon-13 nmr spectroscopy in a lyotropic chiral nematic solvent as an analytical tool for the measurement of the enantiomeric excess. *J. Am. Chem. Soc.*, 119:4502–4508, 1997.
- [84] M. P. Neal and M. Solymosi. Helical twisting power and scaled chiral indices. *J. Chem. Phys.*, 119, 8:3567–3573, 2003.
- [85] A. Pietropaolo, L. Muccioli, R. Berardi, and C. Zannoni. A chirality index for investigatin protein secondary structures and their time evolution. *Proteins*, 70:667–677, 2008.
- [86] A. Marx, V. Schmidts, and C. M. Thiele. How different are diastereomorphous orientations of enantiomers in the liquid crystalline phase of pblg and pbdg: a case study. *Magn. Reson. Chem.*, 47:734–740, 2009.
- [87] P. Lesot, O. Lafon, C. Aroulanda, and R.Y. Dong. 2h nmr studies on two-homopolypeptide lyotropic enantiodiscriminating mesophases: Experimental quatification of solute-fiber affinities. *Chem. Eur. J.*, 14:4082–4092, 2008.
- [88] J. Helfrich, R. Hentschke, and U.M. Apel. Molecular dynamics simulation study of poly(γ -benzyl l-glutamate) in dimethylformamide. *Macromolecules*, 27:472–482, 1994.
- [89] J. Helfrich and R. Hentschke. Molecular dynamics simulation of macromolecular interactions in solution: Poly(γ -benzyl glutamate) in dimethylformamide and tetrahydrofuran. *Macromolecules*, 28:3831–3841, 1995.

- [90] 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, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179, 1995.
- [91] K. Akagi, G. Piao, S. Kaneko, K. Sakamaki, H. Shirakawa, and M. Kyotani. *Science*, 282:1683–1686, 1998.
- [92] H. Athenstaedt. *Ann. N.Y. Acad. Sci.*, 238:7–25, 1974.
- [93] J.J.L.M. Cornelissen, M. Fisher, N.A.J.M. Sommerdick, and R.J.M. Nolte. *Science*, 280:1427–1430, 1998.
- [94] V.K. Gupta, J.J. Skaife, T.B. Dubrovsky, and N.L. Abbot. *Science*, 279:2077–2080, 1998.
- [95] I. Haller. *J. Am. Chem. Soc.*, 100:8050–8055, 1978.
- [96] S.A. Jenekhe and X.L. Chen. *Science*, 279:1903–1906, 1998.
- [97] J.K. Whitesell and H.K. Chang. *Science*, 261:73–76, 1993.
- [98] Schrödinger. *The PyMOL Molecular Graphics System, Version 1.2r1*, LLC.
- [99] Avogadro: an open-source molecular builder and visualization tool. <http://avogadro.openmolecules.net/>, Version 0.9.7.
- [100] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and Skiff W. M. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.*, 114:10024–10035, 1992.
- [101] Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [102] B. H. Besler, K. M. Jr. Merz, and P. A. Kollman. Atomic charges derived from semiempirical methods. *J. Comput. Chem.*, 11:431, 1990.

- [103] <http://www.symyx.com>.
- [104] accelrys.com.
- [105] M. A. Thompson. Arguslab 4.0.1. *Planaria Software LLC, Seattle, WA*.
- [106] W. Humphrey, A. Dalke, and K. Schulten. Vmd – visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996.
- [107] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Jr. Merz, 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, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.
- [108] E. Darve, D. Rodriguez-Gomez, and A. Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.*, 128:144120, 2008.
- [109] J. Hnin, G. Forin, C. Chipot, and M.L. Klein. Exploring multi-dimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theor. Comput.*, 6:35–47, 2010.
- [110] A. K. Rapp and W.A. Goddard. *J. Phys. Chem.*, 95:3358, 1991.
- [111] M. V. Marathias, G. J. Tawa, I. Goljer, and A. C. Bach II. Stereochemical identification of (*r*)- and (*s*)-ibuprofen using residual dipolar couplings, nmr and modeling. *Chirality*, 19:741–750, 2007.
- [112] M. Sarfati, C. Aroulanda, J. Courtier, and P. Lesot. Enantiomeric recognition of chiral invertomers through nmr in chiral oriented phases: a study of *cis*-decalin. *Asymmetry*, 12:737–744, 2001.
- [113] L. Martinez, R. Andrade, E.G. Birgin, and J.M. Martinez. Packmol: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.*, 30:2157–2164, 2009.

- [114] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. Namd2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.*, 151:283–312, 1999.
- [115] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with namd. *J. Comput. Chem.*, 26:1781–1802, 2005.
- [116] R. Berardi, G. Cainelli, P. Galletti, D. Giacomini, A. Gualandi, L. Muccioli, and C. Zannoni. Can the π -facial selectivity of solvation be predicted by atomistic simulations? *J. Am. Chem. Soc.*, 127:10699–10706, 2005.
- [117] M. Solymosi, R. J. Low, M. Grayson, and M. P. Neal. A generalized scaling of chiral index for molecules. *J. Chem. Phys.*, 116, 22:9875–9881, 2002.
- [118] S.E. Hall, W. B. Floriano, N. Vaidehi, and W. A. Goddard III. Predicted 3-d structures for mouse 17 and rat 17 olfactor receptors and comparison of predicted odor recognition profiles with experiments. *Chemical Senses*, 29 (8):595–616, 2004.
- [119] Y. Shigeri, R.P. Seal, and K. Shimamoto. Molecular pharmacology of glutamate transporters eac1 and vglut1. *Brain Res. Rev.*, 45, 3:250–265, 2004.
- [120] B. A. Chabner, C. J. Allegra, G. A. Curt, N. J. Clendeninn, J. Baram, S. Koizumi, Drake J. C., and J. Jolivet. Polyglutamation of methotrexate. is methotrexate a prodrug? *J. Clin. Invest.*, 76(3):907–912, 1985.
- [121] R.L. Schilsky. Methotrexate: an effective agent for treating cancer and building careers. the polyglutamate era. *The Oncologist*, 1 (4):244–247, 1996.

- [122] V. Herbert. Biochemical and hematologic lesions in folic acid deficiency. *Am. J. Clin. Nutr.*, 20:562–569, 1967.
- [123] J. D. Hines, C. H. Halsted, R. C. Griggs, and J. W. Harris. Megaloblastic anemia secondary to folate deficiency associated with hypothyroidism. *Ann. Intern. Med.*, 68:792–805, 1968.
- [124] MRC Vitamin Study Research Group. Prevention of neural tube defects: results of the medical research council vitamin study. *Lancet*, 338:131–137, 1991.
- [125] A. E. Czeizel and I. Duds. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N. Engl. J. Med.*, 327:1832–1835, 1992.
- [126] J. Ma, M. J. Stampfer, E. L. Giovannucci, C. Artigas, D. J. Hunter, C. Fuchs, W. C. Willett, J. Selhub, C. H. Hennekens, and R. Rozen. Methylene tetrahydrofolate reductase polymorphism, dietary interactions, and risk of colorectal cancer. *Cancer Res.*, 57:1098–1102, 1997.
- [127] J. Danesh and S. Lewington. Plasma homocysteine and coronary heart disease: systematic review of published epidemiological studies. *J. Cardiovasc. Risk*, 5:229–232, 1998.
- [128] C. J. Boushey, S. A. Beresford, G. S. Omenn, and A. G. Motulsky. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. probable benefits of increasing folic acid intakes. *J. Am. Med. Assoc.*, 274:1049–1057, 1995.
- [129] A. Coppen and C. Bolander-Gouaille. Treatment of depression: time to consider folic acid and vitamin b12. *Journal of Psychopharmacology*, 19 (1):59–65.
- [130] L. J. Watanabe and Y. Takahina. *Macromolecules*, 24:3423, 1991.
- [131] M. M. Mabuchi, S. Ito, M. Yamamoto, T. Miyamoto, A. Schmidt, and W. Knoll. *Macromolecules*, 31:8802, 1998.

- [132] Y. Zhang. I-tasser server for protein 3d structure prediction. *BMC Bioinformatics*, 9:40, 2008.
- [133] D. W. Ritchie and V. Venkatraman. Ultra-fast fft protein docking on graphics processors. *Bioinformatics*, 26:2398–2405, 2010.
- [134] D. W. Ritchie and G. J. Kemp. Protein docking using spherical polar fourier correlations. *Proteins*, 39:178–194, 2000.
- [135] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery. General atomic and molecular electronic structure system. *J. Comput. Chem.*, 14:1347–1363, 1993.
- [136] <http://www.modelling.leeds.ac.uk/pocketfinder/>.