Department of Mathematics and Applied Mathematics,
University of Crete

DIPLOMA THESIS

Mathematical Modeling of Diphenylalanine Peptides in Tetrahydrofuran (THF) Solution using Atomistic Simulations: Evaluation of Force Field

Author: Avramidou Aikaterini

Thesis committee: E. Harmandaris
                  A. Rissanou
                  A. Mitraki

Heraklion

July 2016

Thesis Committee

Evangelos Harmandaris
Department of Applied Mathematics, University of Crete, GR-71409, Heraklion, Crete, Greece.

Anastassia Rissanou
Crete Center for Quantum Complexity and Nanotechnology (CQCN), Department of Physics, University of Crete, Heraklion, Crete, Greece.
Department of Applied Mathematics, University of Crete, GR-71409, Heraklion, Crete, Greece.

Anna Mitraki
Department of Materials Science and Technology, University of Crete, Greece.
Institute for Electronic Structure and Laser, FORTH, Heraklion, Crete, Greece.

Contents

Abstract

The choice of a proper force field for the description of molecules is always a big issue in simulations. One of the more demanding case is that of biomolecules. In this case the development of a new force field appropriate for the specific molecule is the most accurate method.

The focus of the current study is on diphenylalanine (FF) peptides in tetrahydrofuran (THF) solvent. Our work highlights the effect of the force field on the structural and conformational characteristics of diphenylalanine peptides in THF. More specifically small changes of the charge distribution seem to play a crucial role on the behavior of FF peptides in THF.

The structures that are formed by the molecules studied are very sensitive to the parameters of the force field and this is what the current study tries to address from the point of view of the simulations.

I. Theoretical background

Biomolecular systems & self-assembly
The study of biomolecular systems is a very important topic. It gives information about interactions of vital biomolecules, such as proteins and DNA, their structure and how they behave under different conditions.

Self-assembly is a basic property of many biomolecules, which is the focus of the current work. Self-organized protein structures have either physiological or pathological activities. Many human disorders are associated with the protein self-organization, including Alzheimer is disease and Type II diabetes.[1] More specifically in the self-assembly observed in peptides, the most significant case is that of amyloid fibrils. Amino acid composition plays a crucial role in the conformation and stability of proteins in solution.

Research of the physical and chemical properties of peptides/proteins constitutes an important area for experimental and computational studies.

Diphenylalanine (FF, Phe-Phe)
FF is a very common and well-known peptide. It is the core recognition motif of Alzheimer's a-amyloid peptide. Its chemical formula is presented in Figure 1.
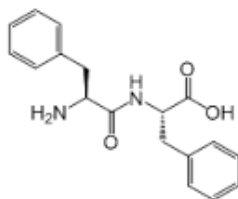
Figure 1. FF molecule.[2]

Particular interest is focused on the FF peptide by both sciences of medicine and nanotechnology[3]. It forms a variety of different structures, which have been observed experimentally and computationally. A model of diphenylalanine peptide is presented in Figure 2. The same building block can self-assemble into various structures depending on the conditions such as temperature, pressure and solvent.[4,5] FF can strongly interact with other molecules through hydrogen bond networks and π-π stacking because of the phenyl groups (two aromatic phenylalanine rings), which seem to be responsible for the self-assembly.[6]
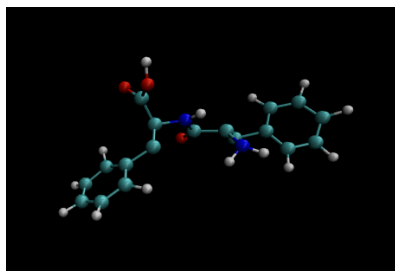
Figure 2. Snapshot of FF

An experimental work of Yan et al.[7] reports three morphologies of FF until now; nanotubes, nanowires and fibrils in gels, detected by scanning electron microscopy technique (SEM). All of them depend on the kind of solvent where they are diluted, and the environmental conditions (i.e. temperature etc.). More specifically, FF in aqueous solvent forms from long stiff nanotubes to nanowires. In tetrahydrofuran, FF forms structures like flakes called peony-flower-like[7], as they are presented and described in Figures 3a, b, c.
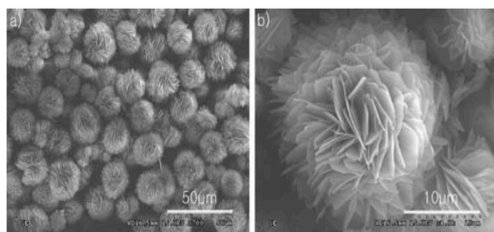
Figure 3a. Low (left) and high (right) magnified SEM images of peony-like flowers assembled by adding THF to a high concentration of HFIP solvent to FF.[7]
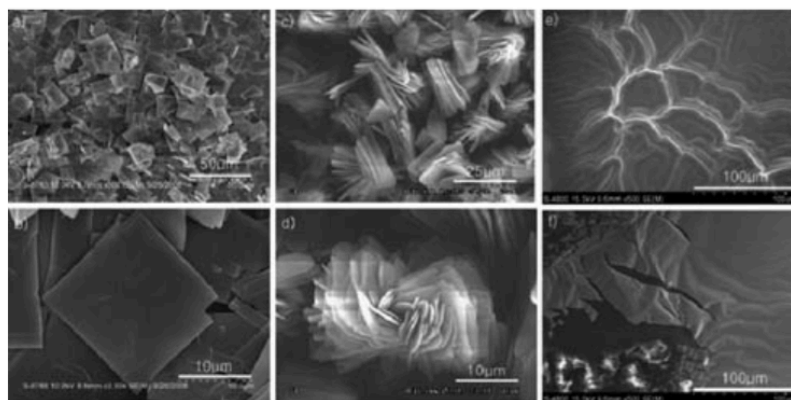
Figure 3b. Low and high magnification SEM images show the variation of FF morphologies in the appearance of different amount of HFIP. (a,b) Without HFIP before THF was added. (c,d) Double the amount of HFIP before THF was added. (e,f) Triple the amount of HFIP.[7]
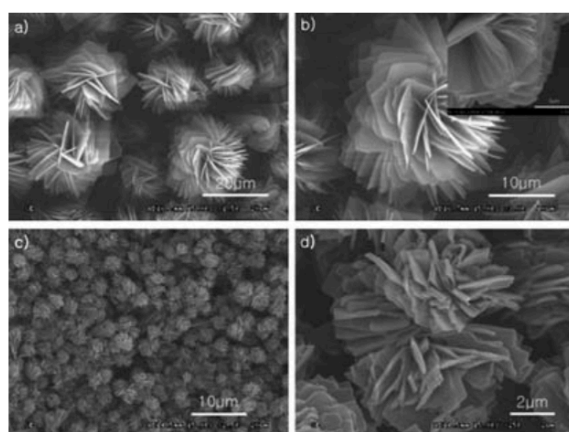
Figure 3c. Low and high magnification SEM images show the variation of FF morphologies in different temperature with the control experiment. (a,b) 40° C. (c, d) Around the melting point of THF (-108° C).[7]

8

Tetrahydrofuran (THF)

THF is a small cyclic ether frequently adopted as solvent due to its remarkable physicochemical properties. It is a five-membered ring with an oxygen (O), four carbons (C) and eight hydrogens (H) bounded to the carbons. The THF molecules are very interesting for studies related to structure of organic rings, torsional barriers and vibrational modes. This molecule presents a pseudo-rotation as it presented by a series of reported results. The chemical and formula of THF is presented in Figure 4.
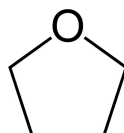
Figure 4. THF

THF is a precursor of many biologically active molecules. For instance, marine macrolides are secondary metabolites with very important biological activities as a defense mechanism to the aggression of the environment.[8-14] Because of that they are good candidates for finding new bioactive molecules with pharmacological potential. Some tetrahydrofuran lignans have showed a wide range of biological activities such as anticancer, anti-inflammatory, analgesic, trypanocidal, schistosomicidal, antileishmanial and antimalarial. THF is also used as analogous of the DNA skeleton, since DNA can be viewed as a chain consisting of various THF molecules linked to each other by phosphate ions. This model has been used in DNA mutations induced by ionizing radiation. A model THF peptide is presented in Figure 5.
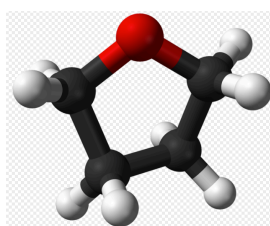
Figure 5. THF molecule. [15]

THF as a liquid solvent allows the creation of larger volumes in which free electrons are accommodated. For this reason, it is used for investigations of solvated electron of specific importance, such as electron-transfer reactions and radiation chemistry.

Most properties of the above have derived exclusively from computer simulations. All these investigations have been parameterized so that the results produce reasonable macroscopic physical properties, i.e. heat capacity.

Moreover, THF as a well-known clathrate former and a promoter for gas hydrate formation is particularly interesting due to its ability to form hydrogen bonds with water. Water and THF form weak hydrogen-bonded complex. Hydrogen bonds are important interactions because they determine the macromolecular structures of biologically active molecules such as proteins and DNA.

THF acts only as an H-bond acceptor, as it is shown in the following picture, and the oxygen of the ring interacts with other molecules in order to form hydrogen bonds.[16]
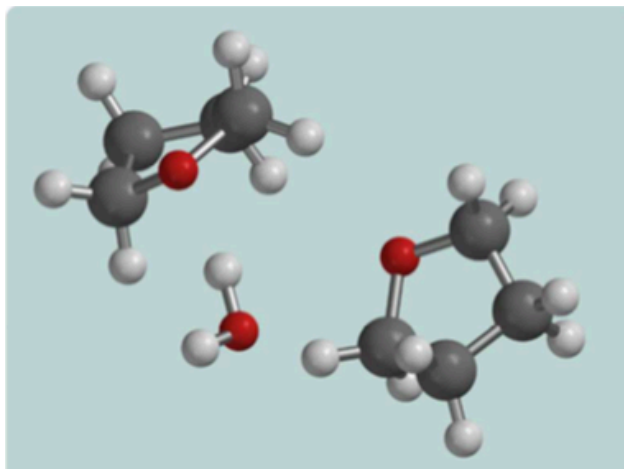
Figure 6. THF only acts as an H-bond acceptor. [16]
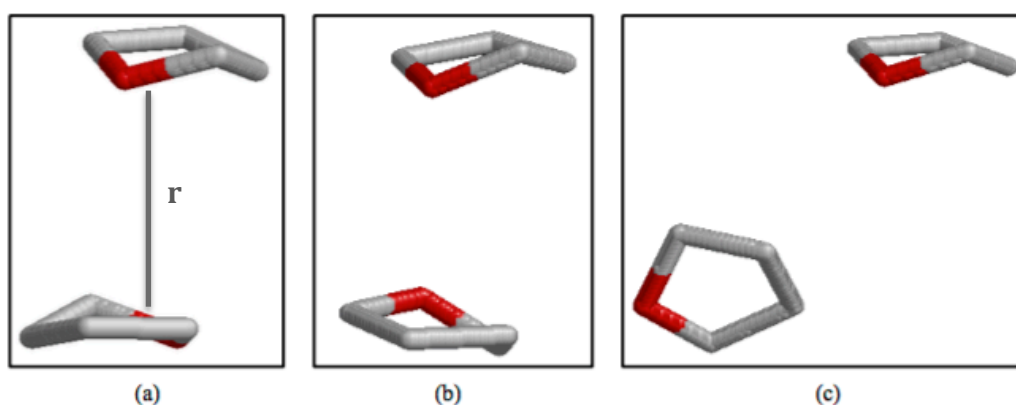
Figure 7. Parallel pairs of THF as a function of the intermolecular distance. (a) r=3-4A; (b) r=4-6A; (c) r=6-7.5A. [17]

A pair of THF molecules has a standard orientation, depending on the distance between the two molecules. Some preferable possible orientations are from antiparallel to T-shape. It is suggested that this may be related to the absence of packing and the formation of voids in the liquid.

It is clear from the figure above that the THF molecules in pairs are antiparallel to molecular distances between 3 and 4 Å, with the angle between the dipole moment vectors about 130° and that angles tend to values about 90° for larger molecular distances. [18]

II. Molecular Simulations

Simulations

Simulation is the imitation of the operation of a real-world process or system in time. For this purpose the development of a proper model is required. Mathematical models can be used in order to reproduce the behavior of a real system under specific conditions. Calculations are executed on a computer and output data are analyzed properly.

Doing simulations is very useful. First, it is way to do an experiment rather than perfoming it in a laboratory, especially when parameters such as temperature, pressure etc. are very high, when the experiment is risky in practice, or the materials studied have not been constructed yet. Simulations provide a satisfactory level of detail in a lot of properties that cannot be measured in a different way. This depends on the programming code, which calculates the given amounts.

On the other hand, simulations may lack accuracy in calculations. Considering that a mathematical model is used, errors are included in the results. Statistics are also a characteristic, which affects the properties and the results of a simulation. If statistics are not good, then simulation might be uncertain at least in a quantitative degree.

Computer simulations constitute a bridge between theory and experiment, as it is presented schematically in Figure 8.
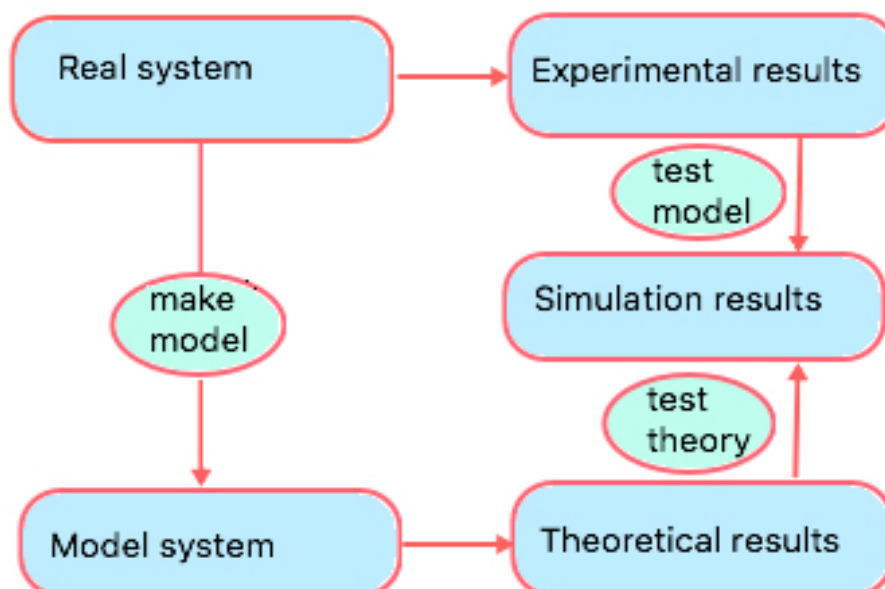
Figure 8. Computer simulations constitute a bridge between theory and experiment.

Molecular Dynamics (MD)

Molecular dynamics is a technique for studying equilibrium and dynamical properties of classical many-body systems. MD uses a numerical algorithm to compute some basic characteristics of the system. Systems consist of N molecules (ex. simple liquids, colloids, macromolecules, polymers etc.). MD can provide information with atomistic detail for a variety of properties, such as interactions between molecules, their structure and assembly. Historically, the first proper MD simulations were reported in 1956 by Alder and Wainwrght at Livermore.[19] They studied the dynamics of an assembly of hard spheres. Vineyard's group at Brookhaven, who simulated radiation damage of crystalline Cu, reported the first MD simulation of a real material in 1959. Also the first MD simulation of a real liquid (argon) was done in 1964 by Rahman at Argonne. [20] The development of science and technology of computers contributes to the development of molecular dynamics simulations and their widely use.

MD calculates many properties of classical many-body systems by solving the Newton's classical equations, described extensively in section II.a and Figure 9 is also a typical flow chart.

The steps of MD algorithm are the following [21, 22]:

1. Prepare the **initial configuration**. This means a "snapshot" of the corresponding physical system is constructed and initialized (positions and velocities for each particle within the system).

2. **Compute the forces** within the system; intermolecular forces act between atoms of different chains or molecules, and intramolecular between the same chain.

3. In order to estimate potential, next step is the **integration of equations of motion** with an appropriate method.

4. After system is well equilibrated, in **every step** of the algorithm positions, velocities, forces, energies etc. are measured and stored periodically.

5. Last step of the MD algorithm is to **average** all the measurements done and print out the final values.

MD Flow Diagram

Starting

1. Initial configuration

2. Compute forces of each atom within the system

3. Integrate equations of motion

N steps

4. Store positions, energies etc. every N steps

5. Print out final averages

END
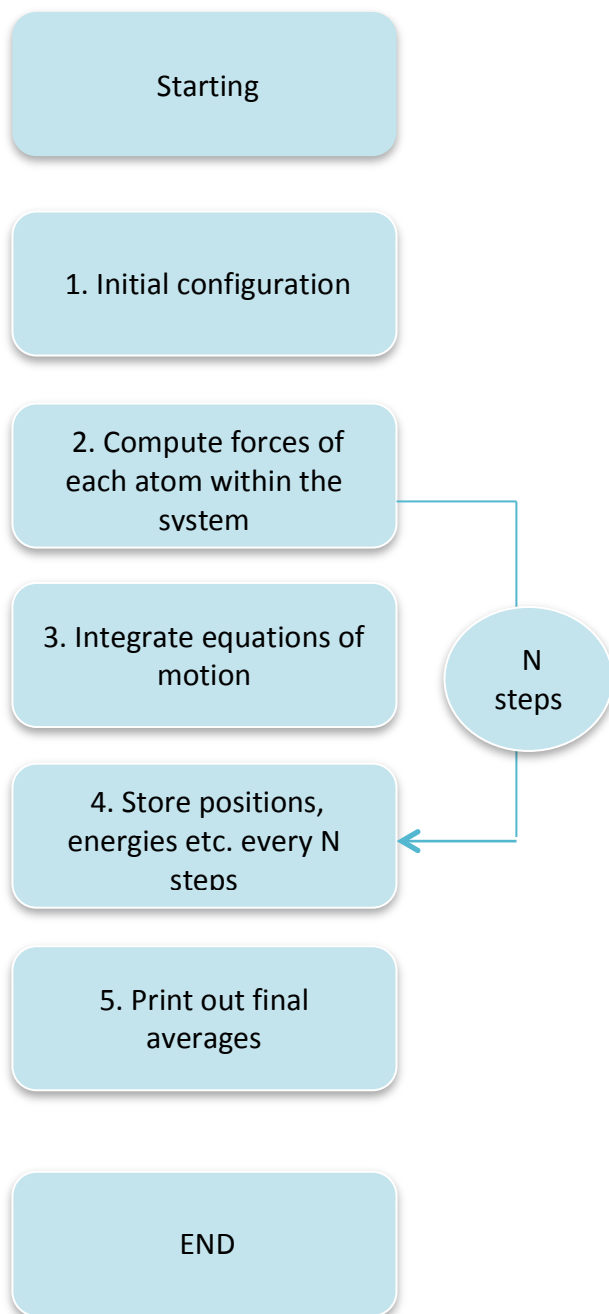
Figure 9. MD flow diagram. [21]

a. Force Calculation
Newton's equations of motion
Equations of motion are integrated numerically to give information for the positions and velocities of each atom of the system. [21-23]
Consider a system of N particles.
A simple atomic system can be written

$$m_i \ddot{r}_\iota = f_i$$

$$f_i = -\frac{\partial}{\partial r_i} \mathcal{U}$$

where

f_i the forces acting on the atoms
r_i the coordinates of every particle
$\mathcal{U}(r^N)$ the potential energy, where $r^N = (r_1, r_2, \ldots, r_N)$ represents the complete set of 3N atomic coordinates.

And we introduce the atomic momenta $p^N = (p_1, p_2, \ldots, p_N)$, in terms of which the kinetic energy is written

$$\mathcal{K}(p^N) = \sum_{i=1}^{N} |p_i|^2 / 2m_i$$

Then the Hamiltonian (energy) may be written as the sum of kinetic and potential energy

$$\mathcal{H} = \mathcal{K} + \mathcal{U}$$

Writing down the equations of motion as

$$\dot{r}_\iota = p_i / m_i$$

$$\dot{p}_i = f_i$$

lead us to a system of ordinary differential equations.

These equations present some interesting properties;

- Hamiltonian is a constant of motion, $\dot{\mathcal{H}} = d\mathcal{H}/dt$ is zero, if we assume $\mathcal{K}$ and $\mathcal{U}$ do not depend explicitly on time. This is the conservation law.
- Hamilton's equations of motion are reversible in time. That means all trajectories can retrace backwards with changing the signs of velocities.

b. Numerical Algorithms

Algorithms solve the ordinary differential equations of the form

$$m_i \, \ddot{\boldsymbol{r}}_i \equiv m_i \, \dot{\boldsymbol{q}}_\iota = \boldsymbol{F}_i$$

Some terms of choosing the most ideal algorithm are the following: [21-23]

- Algorithms should be accurate, fast and require not a lot of memory.
- The satisfaction of the energy conservation law.
- Large time step dt should be permitted by the algorithm
- Algorithm must not require an expensively large number of force evaluations per integration time step, i.e. Runge-Kutta is an inappropriate method.

It is clear that algorithms do not provide the exact solution of the integrations. But this is not a serious problem because MD simulations focus on thermodynamic properties and dynamics and not on the exact configuration.

b.1. Verlet Methods

A fast, accurate, simple and time reversible algorithm. It is used for integrating the classical equations of motion. Verlet equations are derived from the Taylor expansion at times t-dt and t+dt:

$$\mathbf{r}(t + dt) = \mathbf{r}(t) + \, dt\mathbf{v}(t) + \frac{dt^2}{2}\ddot{\mathbf{r}}(t) + \frac{dt^3}{6}\dddot{\mathbf{r}}(t) + \, \mathcal{O}(dt^4)$$

$$\mathbf{r}(t - dt) = \mathbf{r}(t) - \, dt\mathbf{v}(t) + \frac{dt^2}{2}\ddot{\mathbf{r}}(t) - \frac{dt^3}{6}\dddot{\mathbf{r}}(t) + \, \mathcal{O}(dt^4)$$

By summing the two equations, we get

$$\mathbf{r}(t + dt) = 2\mathbf{r}(t) - \, \mathbf{r}(t - dt) + \, dt^2\ddot{\mathbf{r}}(t) + \, \mathcal{O}(dt^4)$$

and $\ddot{\mathbf{r}}(t)$ is known from the forces, velocities and positions.

There are two modifications of the Verlet scheme. The first is the leap-frog algorithm, where positions and velocities are not calculated at the same time; velocities are evaluated at half-integer time steps:

$$\mathbf{r}(t + dt) = \mathbf{r}(t) + dt\mathbf{v}\left(t + \frac{dt}{2}\right)$$

$$\mathbf{v}\left(t + \frac{dt}{2}\right) = \mathbf{v}\left(t + \frac{dt}{2}\right) + \, dt\ddot{\mathbf{r}}(t)$$

Then we calculate velocities at time t as averages of the values at times $t + \frac{dt}{2}$ and $t - \frac{dt}{2}$, in order to find the Hamiltonian $\mathcal{H}$:

$$v(t) = \frac{1}{2}\left(v\left(t + \frac{dt}{2}\right) + v\left(t - \frac{dt}{2}\right) \right)$$

From the formula above we cannot find the positions and velocities at the same time, so there is another way to calculate them, with a new algorithm, the velocity-Verlet algorithm:

$$\mathbf{r}(t + dt) = \mathbf{r}(t) + dt\mathbf{v}(t) + \frac{dt^2}{2}\ddot{\mathbf{r}}(t) \qquad\qquad (1)$$

$$\mathbf{v}(t + dt) = \mathbf{v}(t) + \frac{dt}{2}\left[\ddot{\mathbf{r}}(t) + \ddot{\mathbf{r}}(t + dt)\right]$$

If we write equation (1) as:
$$\mathbf{r}(t) = \mathbf{r}(t + dt) - dt\mathbf{v}(t) - \frac{dt^2}{2}\ddot{\mathbf{r}}(t)$$
and add a dt :

$$\mathbf{r}(t + 2dt) = \mathbf{r}(t + dt) + dt\mathbf{v}(t + dt) + \frac{dt^2}{2}\ddot{\mathbf{r}}(t + dt)$$

we get:
$$\mathbf{r}(t) + \mathbf{r}(t + 2dt) = 2\mathbf{r}(t + dt) + \frac{dt^2}{2}\left[\ddot{\mathbf{r}}(t + dt) - \ddot{\mathbf{r}}(t)\right]$$

In conclusion, the accuracy of an algorithm depends on its order. A higher order method is much better for running a simulation for long times but it is not time reversible, so it doesn't satisfy the energy conservation law. On the other hand Verlet methods are reversible in time and ideal for running long MD simulations.

c. Statistical ensembles
Most quantities we wish to calculate depend on the temperature and the pressure. There are several reasons why it is necessary to control them. MD simulations are conducted under specified conditions of temperature and pressure, in order to give the desirable results.
There is a variety of methodologies for performing MD simulations under isothermal (Constant temperature) or isochoric (Constant pressure) conditions. Figure 18 refers to some statistical ensembles.

NPT ensemble: Isothermal-isobaric ensemble; N particles (N) of the system are kept constant; pressure (P) is constant and temperature (T) also.

NVT ensemble: Canonical ensemble; N particles (N) of the system are kept constant; volume (V) and temperature (T) are constant too.

Thermodynamics Kinetic Theory

NVE Microcanonical
NVT Canonical
NPT Isothermal-isobaric
NPH Isoenthalpic-isobaric
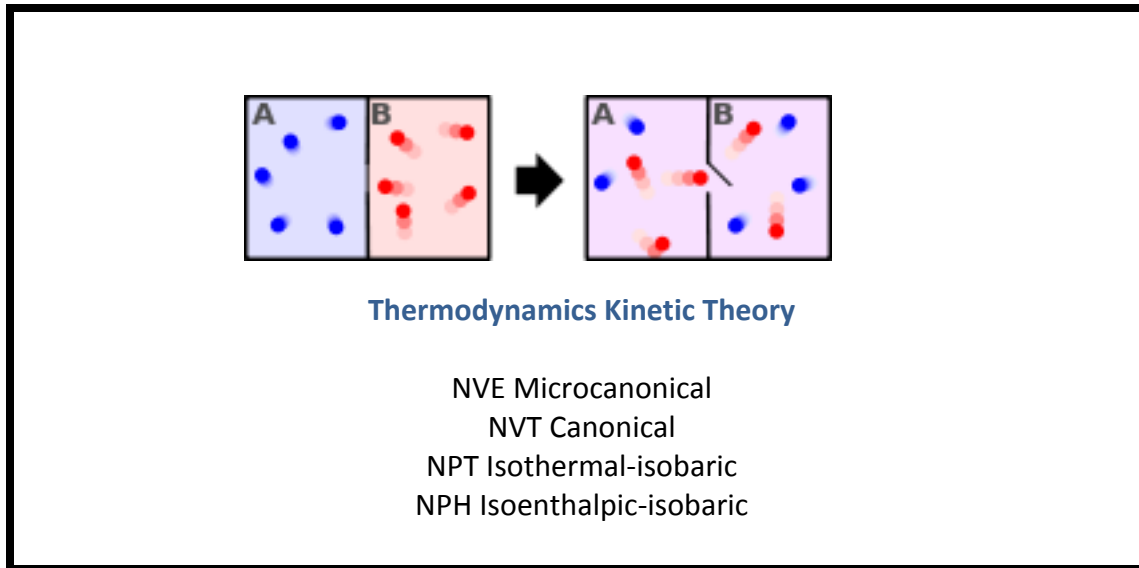
Figure 18. Statistical ensembles [24]

---

d. The Berendsen Thermostat-Barostat

Berendsen proposed a way for performing MD simulations without the need to use an extended Lagrangian as other Thermostats such as *The Nose-Hoover Thermostat*, by coupling a system into a temperature and/or pressure bath.[21-23] The algorithm corrects slowly the initial given temperature T_o according to the following equation:

$$\frac{dT}{dt} = \frac{T_o - T}{\tau_T}$$

which means that a temperature deviations decays exponentially with the time constant τ.

Correspondingly the equation for pressure coupling:

$$\frac{dP}{dp} = \frac{P_o - P}{\tau_p}$$

T and P are the instantaneous values of temperature and pressure calculated from the momenta and configuration of the system.

The solution of these equations forces velocities to be scaled at every time step by factors χ_T and χ_P respectively by:

$$\chi_T = (1 + \frac{dt}{\tau_T}(\frac{T_o}{T} - 1))^{1/2}$$

$$\chi_P = 1 - b_T \frac{dt}{\tau_P}(P_o - P)$$

where b_T : the isothermal compressibility of the system.

This method is simple but suffers from the fact that the phase space probability density it defines, does not conform to a specific statistical ensemble.

e. Periodic Boundary Conditions
Every simulated system has a number of atoms much less than a macroscopic piece of matter of the order 10^{23}. In addition, to perform a simulation it would be very expensive having such a big system. For this Periodic Boundary Condition (PBC) is used in which particles are enclosed in a box, and the box is replicated to infinity by rigid translation in all the three Cartesian directions, completely filling the space.[25, 26]

Figure 19 describes the periodic boundary conditions in a simulation box. The basic idea behind the PBC is that if an atom moves in the original simulation box, all its images move in a concerted manner by the same amount and in the same fashion. The computational advantage of this method is that we need to keep track of the original image only as representative of all other images. As the simulation evolves, atoms can move through the boundary of the simulation cells. When this happens, an image atom from one of the neighboring cell enters to replace the lost particle.
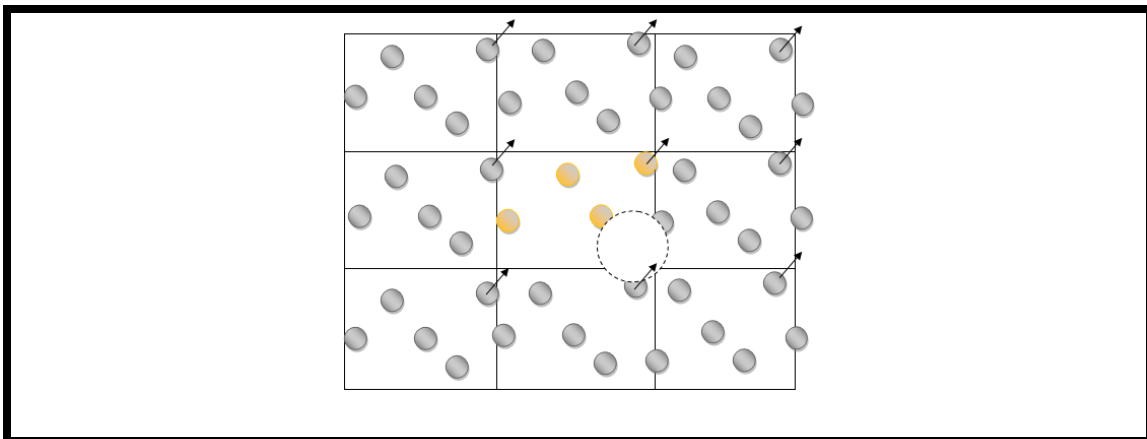
Figure 19. Periodic Box Condition. The dotted line circle indicates the cut off radius (R_c). The yellow colored cell indicates the central cell which is translated in x, y and z direction. [25]

f. Coordinates and velocities
Necessary to start the MD run is to set coordinates and velocities of all particles of the system. The box size is also required. Box size and shape is determined by three vectors **b1**, **b2**, **b3**, which represent the three basis vectors of the periodic box. Run starts when time is set to **t=t0** and coordinates at t=t0 are known. Initial atomic

velocities $v_i, i=1, ...,3N,$ if they are unknown, can be generated at a given temperature T by the following formula;[23]

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}} \exp - \left(\frac{m_i}{2kT}\right)$$

where k: Boltzmann's constant

Figure 10 describes a typical Maxwell-Boltzmann velocity distribution.
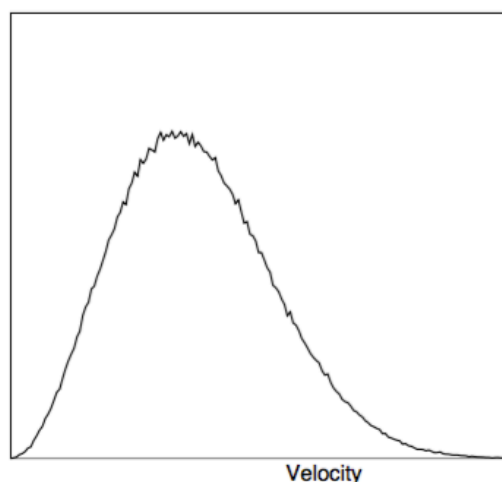
Figure 10. A Maxwell-Boltzmann velocity distribution, generated from random numbers. [23]

g. Topology and Forcefield

The description of molecular interactions, bonded and non-bonded is performed using potential functions. [23]

Forcefield is a semi-empirical way to mimic the interactions between atoms and molecules. It consists of a functional form with adjustable parameters, which depend on the chemical nature of the molecules studied. Two force fields may utilize the same functional form, but have different parameter sets.

g.1 Non-bonded Interactions

The part of the potential energy $\mathcal{U}_{non-bonded}$ representing non-bonded interactions between atoms is split into 1-body, 2-body, 3-body, ... terms:

$$\mathcal{U}_{non-bonded}(\boldsymbol{r}^N) = \sum_i u(\boldsymbol{r}_i) + \sum_i \sum_{j>i} v\,(\boldsymbol{r}_i, \boldsymbol{r}_j)$$

The $u(\boldsymbol{r})$ term represents an externally applied potential field or the effects of the

container walls. It is usual to concentrate on the pair potential $v(\mathbf{r}_i, \mathbf{r}_j) = v(\mathbf{r}_{ij})$ and neglect 3-body or higher order interactions.

g.1.1. The Lennard-Jones interaction
This potential is the most commonly used form for computing the non-bonded interactions between two atoms (Van der Waals):

$$V_{\mathcal{L}\mathcal{J}}(r_{ij}) = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}}$$

Figure 11 describes the LJ potential as a function of distance r.
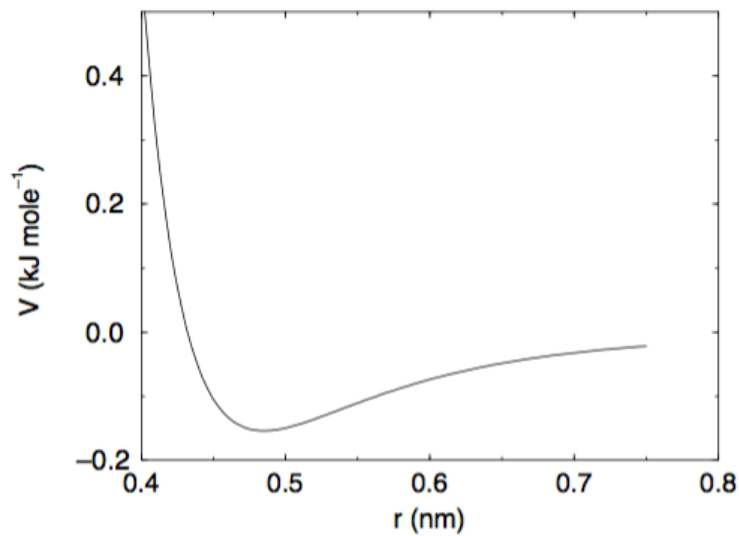
Figure 11. The Lennard-Jones interaction. [23]

The force derived from this potential is:

$$F_i(r_{ij}) = \left(12\frac{C_{ij}^{(12)}}{r_{ij}^{13}} - 6\frac{C_{ij}^{(6)}}{r_{ij}^{7}} \right) \frac{r_{ij}}{r_{ij}}$$

The LJ potential may also be written in the following form:
$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right)$$

where ϵ and σ are two parameters which define the type of combination rule in the topology of the force field.

There are three types of combination rules in order to form the parameter matrix of the non-bonded LJ parameters:

Type 1:
$$C_{ij}^{(6)} = \left(C_{ii}^{(6)} \, C_{jj}^{(6)} \right)^{1/2}$$

$$C_{ij}^{(12)} = \left(C_{ii}^{(12)} \, C_{jj}^{(12)} \right)^{1/2}$$

Type 2:
$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj})$$

$$\epsilon_{ij} = (\epsilon_{ii} \, \epsilon_{jj})^{1/2}$$

And there is a geometric average of the above:

Type 3:
$$\sigma_{ij} = (\sigma_{ii} \, \sigma_{jj})^{1/2}$$

$$\epsilon_{ij} = (\epsilon_{ii} \, \epsilon_{jj})^{1/2}$$

For this work it is chosen the second type of combination rule.

g.1.2. Coulomb Interaction
The Coulomb interaction between two particles is estimated by the formula:

$$V_c(r_{ij}) = f \, \frac{q_i q_j}{\varepsilon_r r_{ij}}$$

where $f = \frac{1}{4\pi\varepsilon_0} = 138.935\,485$.

The force derived from this potential is:

$$F_i(r_{ij}) = f \, \frac{q_i q_j}{\varepsilon_r r_{ij}^2} \frac{r_{ij}}{r_{ij}}$$

Figure 12 depicts the potential of Coulomb interactions as a function of distance r.
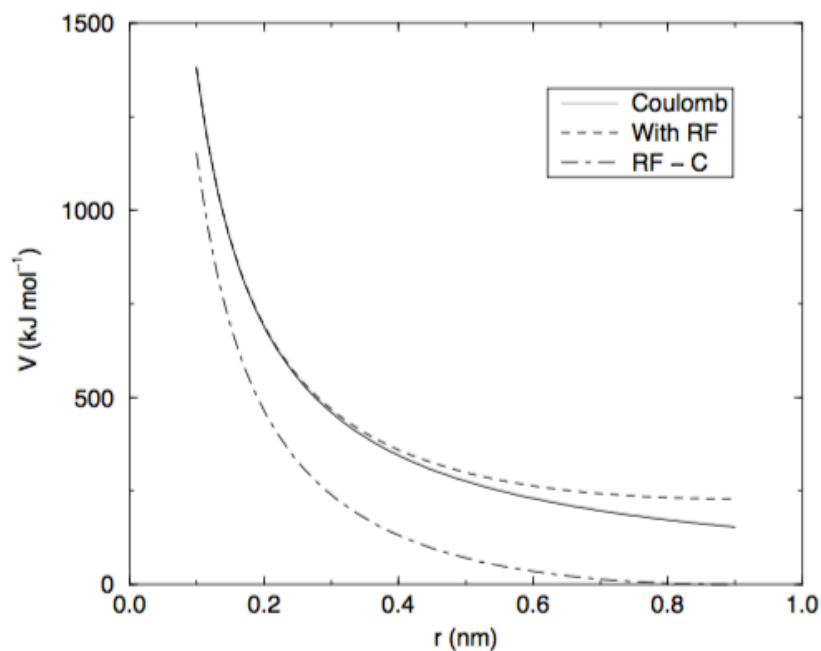
Figure 12. The Coulomb interaction (for particles with equal signed charge) with and without reaction field. In the later case $\varepsilon_r = 1$, $\varepsilon_{rf} = 78$ and $r_c = 0.9$ nm. The dot-dashed line is the same with dashed line, exept for a constant. [23]

Electrostatics' Methods of Coulomb Interactions

g.1.2.1. PME
The total electrostatic energy of N particles and their periodic boundary images is equal to:

$$V = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_{z*}} \sum_i^N \sum_j^N \frac{q_i\, q_j}{r_{ij,\mathbf{n}}}$$

$(n_x, n_y, n_z) = \mathbf{n}$ is the box index vector, and the star * indicated that terms with $i = j$ should be omitted when $(n_x, n_y, n_z) = (0,0,0)$. The distance $r_{ij,\mathbf{n}}$ is the real distance between the charges and not the minimum-image. This sum is conditionally convergent, but very slow. Ewald summation was first introduced as a method to calculate long-range interactions of the periodic images in crystals. The idea is to convert the single slowly converging sum into two quickly converging terms and a constant term:

$$V = V_{\text{dir}} + V_{\text{rec}} + V_0$$

$$V_{\text{dir}} = \frac{f}{2} \sum_{i,j}^N \sum_{n_x} \sum_{n_y} \sum_{n_{z*}} q_i q_j \frac{erfc(\beta r_{ij,\mathbf{n}})}{r_{ij,\mathbf{n}}}$$

22

$$V_{\text{rec}} = \frac{f}{2\pi V} \sum_{i,j}^{N} q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_{z*}} \frac{\exp\left(-\left(\frac{\pi \boldsymbol{m}}{\beta}\right)^2 + 2\pi i \boldsymbol{m} \cdot (\boldsymbol{r}_i - \boldsymbol{r}_j)\right)}{\boldsymbol{m}^2}$$

$$V_0 = -\frac{f\beta}{\sqrt{\pi}} \sum_{i}^{N} q_i^2$$

where β is a parameter that determines the relative weight of the direct and reciprocal sums and $\boldsymbol{m} = (m_x, m_y, m_z)$. In this way we can use a short cut off (1 nm order) in the direct space sum and a short cut off in the reciprocal space sum. Unfortunately the computational cost of the reciprocal part of the sum increases as N^2 or $N^{3/2}$ and it is therefore not realistic for use in large systems.

PME stands for Particle Mesh Ewald summation and it is used to calculate long-range interactions. It was proposed by Tom Darden to improve the performance of the reciprocal sum. Instead of directly summing wave vectors, the charges are assigned to a grid using interpolation. The grid is then Fourier transformed with a 3D FFT algorithm and the reciprocal energy term obtained by a single sum over the grid in k-space.

The potential at the grid points is calculated by inverse transformation, and by using interpolation factors we get the forces on each atom.

The PME algorithm scales as $N \log(N)$, and is a fast on medium to large systems.

When we use it, the short-range Coulomb potential must be modified, and is given by:

$$V(r) = f \frac{erfc(\beta r_{ij})}{r_{ij}} q_i q_j$$

where β is a parameter that determines the relative weight between the direct space sum and the reciprocal space sum and $erfc(x)$ is the complementary error function.

g.1.2.2. Cut-off
The coulomb interaction is estimated for Cut-off electrostatics method, and also for LJ interactions, by the following equation:

$$V(r) = \begin{cases} f \sum_{i,j}^{N} \frac{q_i q_j}{r_{ij}} & , r \le r_{cut} \\ 0 & , r > r_{cut} \end{cases}$$

23

where q_i , q_j the atom charges
   r_{ij} the distance between of i , j atoms
   f the coulomb force
   r_{cut} the given cut-off distance
   N the total number of atoms.

When using this method, we assume a sphere of a given radius which is the cut-off distance r_{cut}. All interactions inside the sphere are counted on the total potential. Above this radius nothing is taken on account, like it does not happen. Cut-off electrostatics is a fast method for force and potential estimation and is often preferred for MD simulations. Though, it does not reflect the real force field because it skips many interactions that may affect the results.

g.1.2.3. Reaction-Field
For homogenous systems the coulomb interaction can be calculated by the following formula, by assuming a constant dielectric environment beyond the cut-off r_c with a dielectric constant ε_{rf} :

$$V_{crf} = f \frac{q_i q_j}{\varepsilon_r r_{ij}} \left[1 + \frac{\varepsilon_{rf} - \varepsilon_r}{2\varepsilon_{rf} + \varepsilon_r} \frac{r_{ij}^3}{r_c^3} \right] - f \frac{q_i q_j}{\varepsilon_r r_c} \frac{3\varepsilon_{rf}}{2\varepsilon_{rf} + \varepsilon_r}$$

which is equal to zero when the cut-off is r_c . For charged cut-off spheres, this corresponds to neutralization with a homogenous background charge. For simplicity, the equation can be written as:

$$V_{crf} = f \frac{q_i q_j}{\varepsilon_r} \left[\frac{1}{r_{ij}} + k_{rf} r_{ij}^2 - c_{rf} \right]$$

with

$$k_{rf} = \frac{1}{r_c^3} \frac{\varepsilon_{rf} - \varepsilon_r}{(2\varepsilon_{rf} + \varepsilon_r)}$$

$$c_{rf} = \frac{1}{r_c} + k_{rf} r_c^2 = \frac{1}{r_c} \frac{3\varepsilon_{rf}}{(2\varepsilon_{rf} + \varepsilon_r)}$$

For large ε_{rf} the k_{rf} goes to $r_c^{-3}/2$, while for $\varepsilon_{rf} = \varepsilon_r$ the correction vanishes. The force derived from this potential reads:

$$\boldsymbol{F_i}(\boldsymbol{r}_{ij}) = f \frac{q_i q_j}{\varepsilon_r} \left[\frac{1}{r_{ij}^2} - 2k_{rf} r_{ij} \right] \frac{r_{ij}}{r_{ij}}.$$

g.2. Bonded interactions

These interactions regard to bond, angles, proper and improper dihedrals of atoms in the system.

$$\mathcal{U}_{intramolecular} = \mathcal{V}_{bond} + \mathcal{V}_{angle} + \mathcal{V}_{proper} + \mathcal{V}_{improper}$$

where

$$\mathcal{V}_{bond} = \frac{1}{2}k_{ij}^{b}\,(r_{ij} - b_{ij})^2$$

describes the harmonic potential of the bond stretching between two covalently bonded atoms i and j (Figure 13).
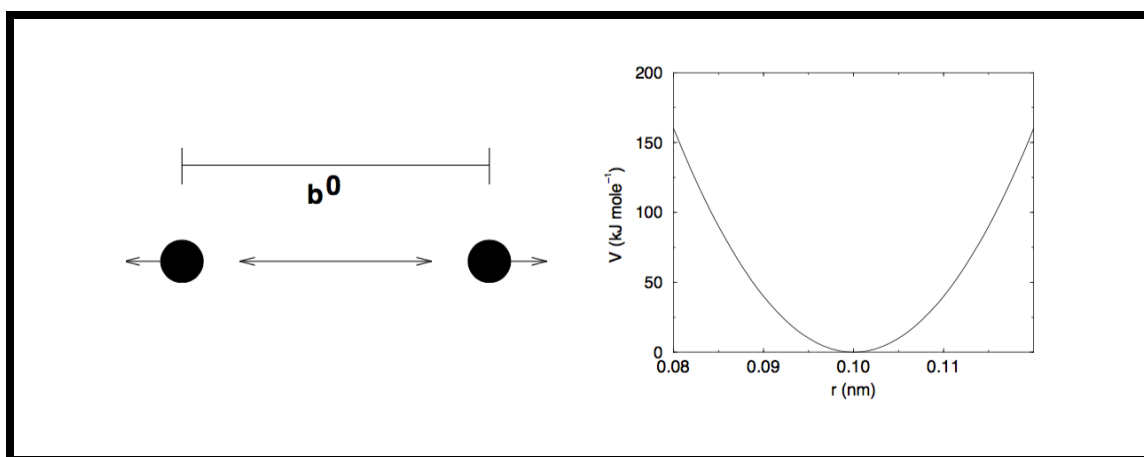
Figure 13. Bond stretching (left) and its potential (right). [23]

$$\mathcal{V}_{angle} = \frac{1}{2}k_{ijk}^{\theta}\,(\theta_{ijk} - \theta_{ijk}^{0})^2$$

which describes the harmonic potential of the bond-angle vibration between three atoms i, j, k and their angle θ_{ijk} (Figure 14).
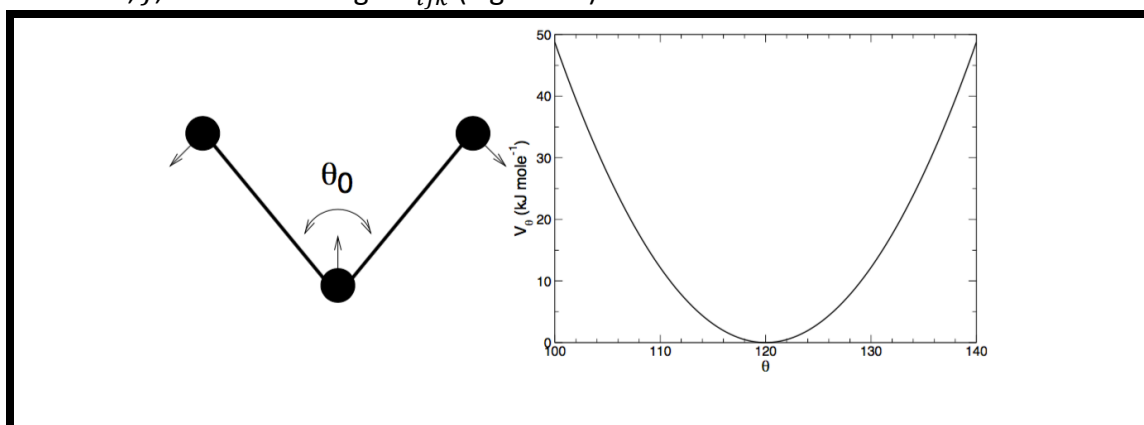
Figure 14. Angle vibration (left) and its potential (right). [23]

$$\mathcal{V}_{proper} = k_{\varphi}\left(1 + \cos\left(n\varphi - \varphi_s\right)\right)$$

which describes the potential of proper dihedral interactions where φ is the angle between ijk and the jkl planes, with i and l on the same side i.e. zero corresponding to the *cis* configuration (Figure 15).
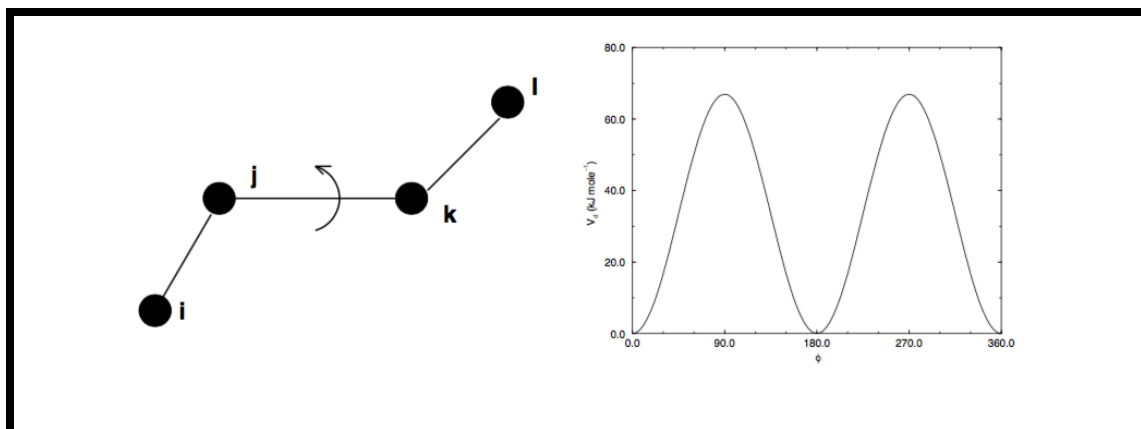
Figure 15. Proper dihedral angle (left) and its potential (right). [23]

$$\mathcal{V}_{improper} = \frac{1}{2}\,k_{\xi}(\xi_{ijkl} - \xi_0)^2$$

which describes the improper dihedrals that are meant to keep planar groups, such as aromatic rings, planar or prevent molecules from flipping over their images (Figure 16).
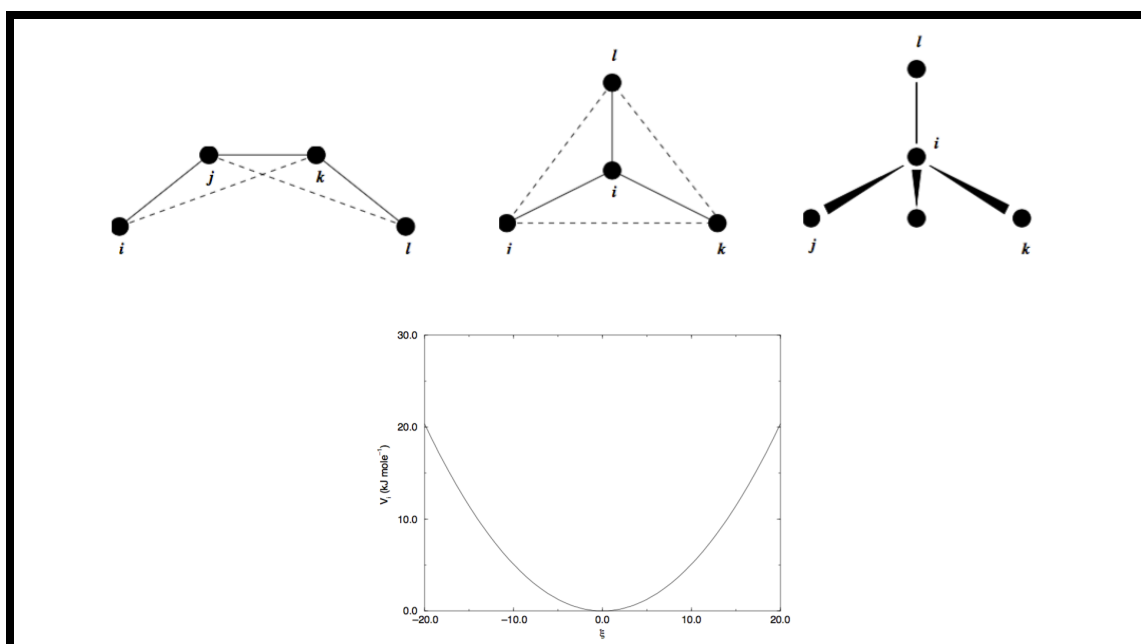
Figure 16. Improper dihedral angles (up) and its potential (down). The improper dihedral angle is defined as the angle ξ between planes (i, j, k) and (j, k, l). [23]

h. Neighbors
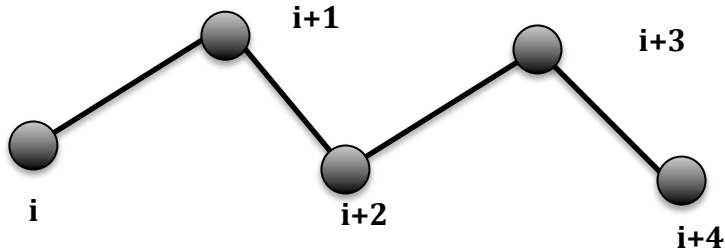
Figure 17. Atoms in a chain

Atoms within a molecule that are close by in a chain, atoms that are bonded or linked by one or two atoms are called *first neighbors, second* and *third* respectively. [23] Reactions between atoms i, i+1, i+2 are mainly quantum mechanical and cannot be modeled by a LJ potential. These interactions are modeled by:

o   a harmonic bond term or constraint for **i, i+1** atoms
o   a harmonic angle term for **i, i+1, i+2** atoms

The first and second neighbors of atom i are excluded from the LJ interaction. Atoms **i+1**, **i+2** are called exclusions of atom **i.**
For third neighbors, the normal LJ repulsion is sometimes so strong which means that if it's applied to a molecule, the molecule might break or deform.
Figure 17 describes schematically a typical neighborhood of atoms in a chain.

i. Cut-offs

The force calculation in molecular dynamics is a $O(N^2)$ problem. To reduce it, a cut-off distance is applied for the non-bonded force calculations. This means that only particles within the certain distance are interacting in the simulation and any interaction between atoms separated by more than the cutoff distance is ignored. This method reduces the error to $O(N)$, a quite acceptable quantity, and is generally accepted as being sufficiently accurate for Van der Waals forces, which decay to zero as the distance increases. Electrostatic forces fall off much more slowly with distance. So a simple truncation at the cutoff distance may cause errors. Approximations have been proposed to modify the electrostatic potential so that the forces approach zero at the cutoff distance or be exactly zero. [23]

An alternative approach is to fully account for the long-range component of the electrostatic interactions, as Ewald summation techniques. This kind of techniques involves splitting the electrostatic interactions into a quickly decaying near component that can be calculated for all atom pairs within a fixed cutoff and a long-range component that can be more efficiently calculated using other methods.

Two sort of parameters are used for treatment of cut-offs, depending on the method, as they are listed in Table 1:

Type		Parameters
Coulomb	Cut-off	r_c, ε_r
	Reaction field	r_c, ε_{rf}
Van der Waals	Cut-off	r_c

Table 1. Parameters for two basic different functional forms of the non-bonded interactions. [23]

III. Aim of this study

The choice of a proper force field for the description of molecules is always a big issue in simulations. One of the more demanding case is that of biomolecules. In this case the development of a new force field appropriate for the specific molecule is the most accurate method.

However this demands expensive ab initio calculations, which are not always easy to be performed. For this reason usually biomolecules (i.e. peptides) are described through existing force fields, which have been developed for the description of similar entities. Parameters for bonded and non-bonded interactions as well as charge distribution can differentiate the calculated results with the last being one of the most important factors.

This is exactly the focus of the current study on FF peptides in THF solvent. Our work highlights the effect of the force field on the structural and conformational characteristics of diphenylalanine peptides in THF. More specifically small changes of the charge distribution seem to play a crucial role on our findings. On top of that issues concerning the method of calculation of electrostatic interactions (i.e. PME, cut off, reaction field) as well as, the accuracy of calculations (i.e. cut off distance for non-bonded interactions) are examined.

It is important to note here that even in experiments, factors like charge distribution are sensitive to the way that sample is prepared, the experimental technique and the conditions under which measurements are performed (i.e. pressure, temperature). For example the existence of a co-solvent could potentially affect the initial charge distribution or even continuous charge transfer could happen under certain conditions.

Formed structures are very sensitive to there parameters and this is what the current study tries to address from the point of view of the simulations.

IV. Simulated systems and models

1. Systems

The systems studied in this work are listed in Table 2.

System	Name	FF-type	N-peptide	N-solvent	no. atoms	r_{cutoff} (nm)	Electrostatic method	Time (ns)
1	FF in THF	Non-polar	16	1912	25448	0.9	Cut-off	100
2	FF in THF	Non-polar	16	1912	25448	1.0	Cut-off	100
3	FF in THF	Non-polar	16	1912	25448	1.0	PME	30
4	FF in THF	Non-polar	16	1912	25448	1.3	Cut-off	30
5	FF in THF	Non-polar	16	1912	25448	1.3	Reaction field	100
6	FF in THF	Polar	16	1899	25279	1.3	Reaction field	100
7	FF in THF	Polar	48	5696	75824	1.3	Reaction field	100
8	FF in THF	Non-polar	48	5760	76656	1.3	Reaction field	100
9	FF in THF	Non-polar	2	767	10045	1.3	Reaction field	30
10	FF in THF	Non-polar	2	768	10058	1.3	Cut-off	50
11	FF in THF	Polar	2	763	9993	1.3	Reaction field	100
12	FF in THF	Polar	2	763	9993	1.3	PME	100
13	FF in THF	Polar	2	763	9993	1.0	PME	100
14	FF in THF	Polar	2	763	9993	1.0	Cut-off	30
15	FF in water	Polar	16	6837	21103	1.0	Cut-off	50
16	FF in methanol	Non-polar	32	6552	20840	1.0	Cut-off	50
17	FF in mixture of 50% water & 50% methanol	Non-polar	16	4026	12670	1.0	Cut-off	50

Table 2. Systems studied

Details such as the number of peptides, the number of the solvent molecules, the number of total atoms of the simulated systems is included in the table above. The concentration of systems 1-8 is equal to 0.035 gFF/cm^3 solvent. Two atomistic structures of FF are used; a polar and a non-polar, as it is explained below. There is a slight difference in the terminal groups of the FF in those two structures.

The molecular dynamics simulations were performed by the GROMACS-5.0.4 code. All systems were simulated in the isothermal-isobaric (NPT) statistical ensemble, temperature was set at T = 300 K using a Berendsen thermostat and the pressure was kept constant at P = 1 atm using a Berendsen barostat. The integration time step was 0.002 ps and a range of different cut-off distances (0.9, 1.0 and 1.3 nm) for both electrostatic and non-bonded interactions was tested, to compare and see how it affects the results. Bond lengths were constraint by the LINCS algorithm. Total time of the simulation production was 100 ns for the already equilibrated systems. Equilibration runs have been preceded. Periodic boundary conditions have been used in all three dimensions.

Equilibration of all systems was tested through the simulation by checking some parameters. Temperature, pressure and energy have small fluctuations around their constant value when system is well equilibrated. Figure 20 presents these three quantities for a system of FF$_{polar}$ charge distribution, reaction field electrostatics and cut off distance equal to 1.3 nm (system no. 6).
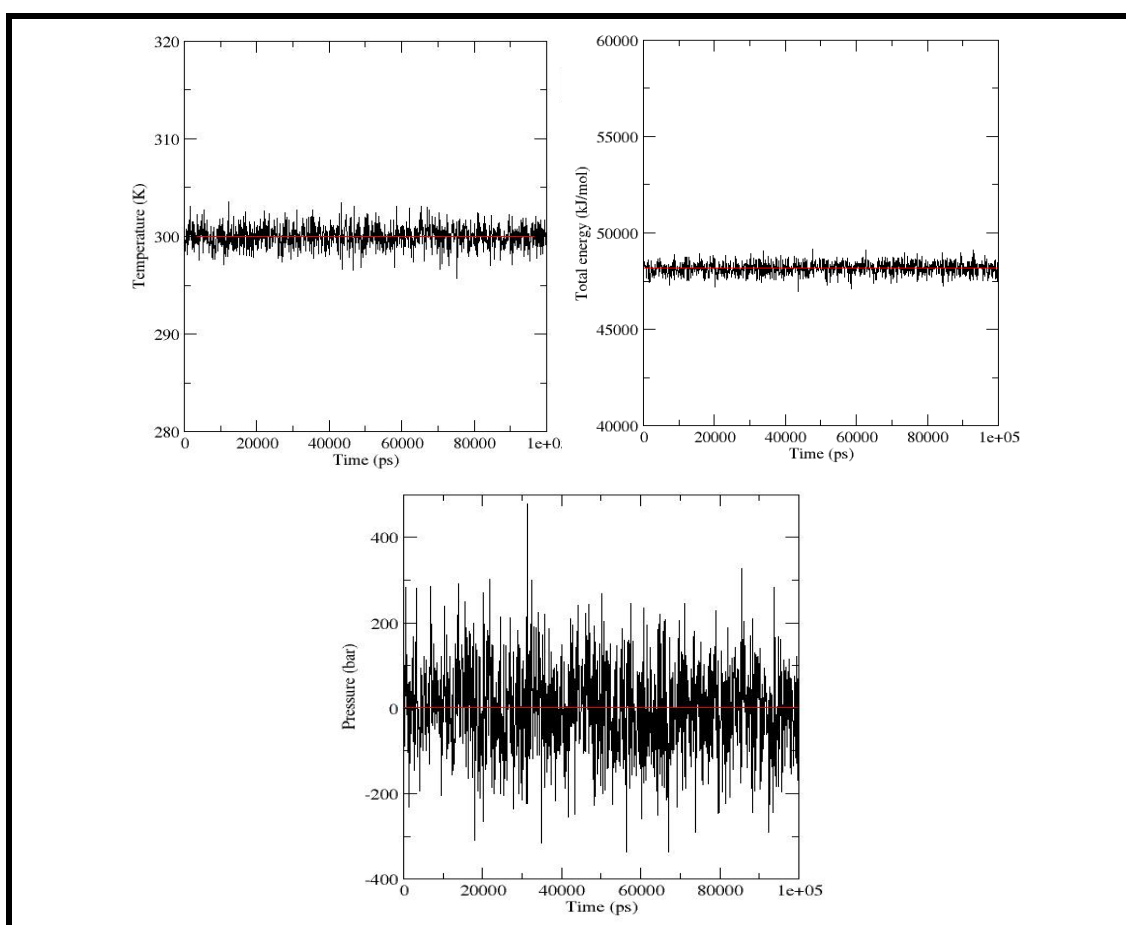

Figure 20. Temperature, pressure and total energy have values between a constant number in a well-equilibrated system.

2. Running time

The actual time of the run depends on several parameters. The simulation method, the method of calculation electrostatic interactions and the cut off distance are factors that affect the real time of the simulation. For instance, cut off electrostatic method is faster than PME and reaction field, because it takes into account interactions up to a specific distance and not from the whole system. Cut off distance has the same physical meaning; bigger cut off distance gives a bigger radius, so more interactions between the atoms are counted.

The time-step for integration (*dt*) of the equations of motion regulates the acceleration of the run. For example, a time step of order 0.001 ps is slower than 0.01 ps.

Additionally, a large system with many atoms and molecules needs more time to be simulated than a smaller one.

The technical characteristics of the computer can reduce the required time. Ideally running in *N* processors makes the simulation *N* times faster. In practice, this can only be achieved for a small number of processors. The use of GPU (graphics processing unit) together with a CPU (central processing unit) accelerates in a satisfactory level the time required for the run. A CPU consists of a few cores optimized for sequential serial processing while a GPU has a massively parallel architecture consisting of thousands of smaller, more efficient cores designed for handling multiple tasks simultaneously. In Figure 21, there is a representation of CPU and GPU.
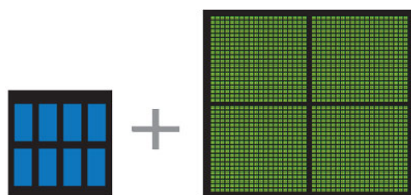
Figure 21. CPU (left) has multiple cores and GPU (right) has a thousand of cores to process parallel workloads efficiently. [27]

The characteristics of the computer machine we did the simulation are the following:
- 32 Gbit (4X QDR) Infiniband networked cluster
- Root node:
    8 x Intel(R) Xeon(R) CPU E5606 @ 2.13GHz - 12 GB RAM
- Compute nodes:
    4 x 12 cores - 2 x Intel(R) Xeon(R) CPU X5650 @ 2.67GHz - 16GB RAM
    2 x 12 cores - 2 x Intel(R) Xeon(R) CPU E5-2430 0 @ 2.20GHz - 16GB RAM
    2 x 12 cores - 2 x Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.60GHz - 32GB RAM
    5 x 24 cores - 2 x Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz - 32GB RAM

Table 3 contains a typical average time required for MD run.

Simulation time	*dt*	no. nodes	no. processors	no. atoms	Real time
1 ns	1 fs	1	12	25448	55 min
1 ns	2 fs	1	12	25448	31 min

Table 3. Average time of a simulation according to *dt* and number of atoms.

3. Tetrahydrofuran model (THF)

THF is a commonly used solvent in chemistry; numerous solutes are studied experimentally in tetrahydrofuran. In order to take more information through molecular simulations, a good force field is required.

In the literature, three types of models have been found for THF. The first is a united-atom force field developed by Briggs et al., the second is a united-atom force field developed by Helfrich et al. from the parameters of the AMBER force field, and at last, the third is a full-atom force field developed by Faller et al.

Different parameter sets have been created, using Lennard-Jones potential and charges from the references above, and some other sets have been optimized by hand or by a simplex algorithm in order to reproduce the density and heat of vaporization of liquid THF.

Three conformations of the THF have been found in the literature, as it is shown in the Table 4.[28] They are called envelope, twist and planar. The first two are the most stable for this ring. In addition, ab initio chemical calculations have been performed (HF/6-31G**; B3LYP/6-31G**) in order to check the energies of the conformations. In the planar conformation, the oxygen and carbons are co-planar and the torsions between these atoms are fixed to 0°.

The energy difference between the envelope and twist conformations is small (0.5-2 kJ.mol^{-1}), with the twist model having a little bit lower energy.

In this project it is chosen the twist conformation and a full atomistic model of an improved model of Faller R. [29] as Plathe G. describes it.[28]

In Figure 22, it is presented the chemical structure of THF. Tables 5 and 6 contain the THF model parameters for bonded and non-bonded interactions.
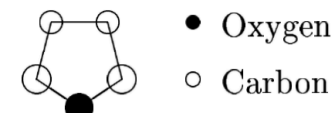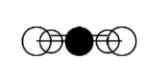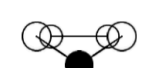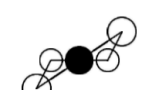
	Rel. Energy (kJ/mol)	
• Oxygen ○ Carbon	HF	B3LYP
planar	14.27	16.69
envelope	1.80	0.55
twist	0	0

Table 4. Possible conformations of tetrahydrofuran. [28]

Chemical structure of Tetrahydrofuran
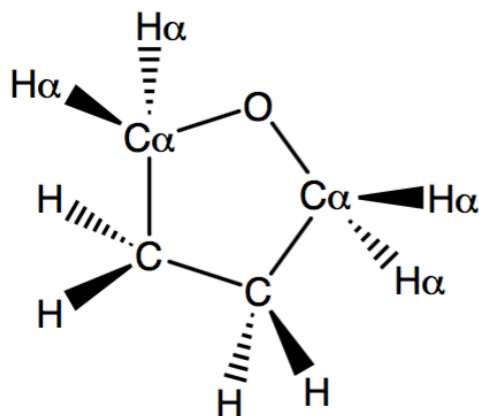
Figure 22. Chemical structure of THF. [28]

Non-bonded parameters

Type	mass (u)	q (e)	σ (nm)	ε (kJ/mol)
C	12.011	0.061	0.295	0.290
Ca	12.011	0.231	0.295	0.290
H	15.999	0	0.240	0.170
Ha	1.008	0	0.240	0.170
O	1.008	-0.584	0.300	0.628

Table 5. Bonded parameters of THF. [28]

Bonded parameters

Parameter		Value	
BONDS	**(constraints)**	**Length (nm)**	
O - C_a		0.1409	
C_a - C		0.1527	
C - C		0.1529	
C - H		0.1086	
C - H_a		0.1086	
C_a - H		0.1086	
C_a - H_a		0.1086	
ANGLES		$k_\phi = 450$ (kJ.mol^{-1}.rad^{-2})	
		ϕ_0 (deg)	
O - C_a - C		106.1	
O - C_a - H_a		109.0 / 109.3	
C_a - C - C		101.4	
C_a - O - C_a		111.2	
C - C_a - H_a		113.2 / 111.0	
C_a - C - H		110.4 / 112.8	
C - C - H		110.4 / 113.7	
H_a - C_a - H_a		108.2	
H - C - H		108.1	
DIHEDRALS		**Periodicity p=3**	
		ϕ_s (deg)	k_p (kJ.mol^{-1})
C_a - C - C - C_a		216.2	5.0
C - C - C_a - O		149.9	5.0
C - C_a - O - C_a		191.3	3.0
C_a - O - C_a - C		192.8	3.0
O - C_a - C - C		149.0	5.0

Table 6. Bonded parameters of THF.[28]

4. Diphenylalanine model

The model of the diphenylalanine peptide is taken from the GROMOS53a6 force field.[30] All intermolecular and intramolecular interactions are described in this model. All atoms of the peptide are represented on a full-atomistic model except for the CH methyl group of molecule's backbone and CH_2 which connects that backbone with the phenyl group, that have been applied as united.

In this project, they have been studied two atomistic structures of the FF peptide, as they are presented in Figure 23. The first is a polar structure as when FF is diluted in aqueous solvent such as water. The second structure is a non-polar, as it is when FF is diluted in organic solvent i.e. methanol.

Both structures are presented in Figure 23 left and right respectively and the man difference is the transfer of one hydrogen from $NH3^+$ terminal group to O^- forming hydroxyl (OH^-) group.
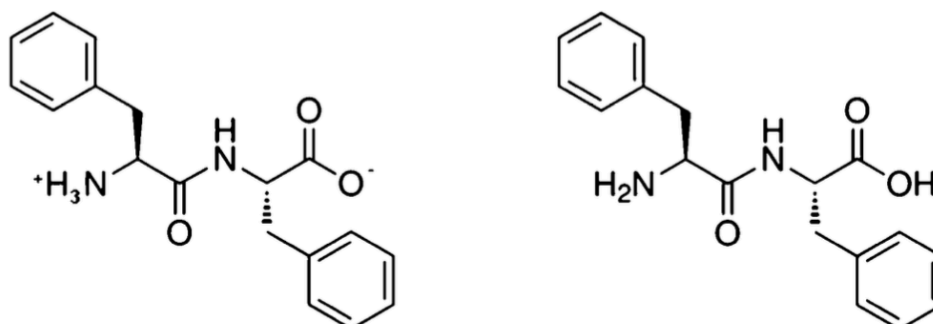
Figure 23. Polar FF (left) and non-polar FF (right).[30]

Corresponding snapshots of FF model molecules are presented in Figure 24. Up is the polar and down the non-polar charge distributions of FF. Tables 7 and 8 contain the FF model parameters for bonded and non-bonded interactions. Except for these parameters, the model of FF we used on this study includes pairs of some atoms of diphenylalanine that have been excluded from specific intramolecular interactions, but are not shown here.

36
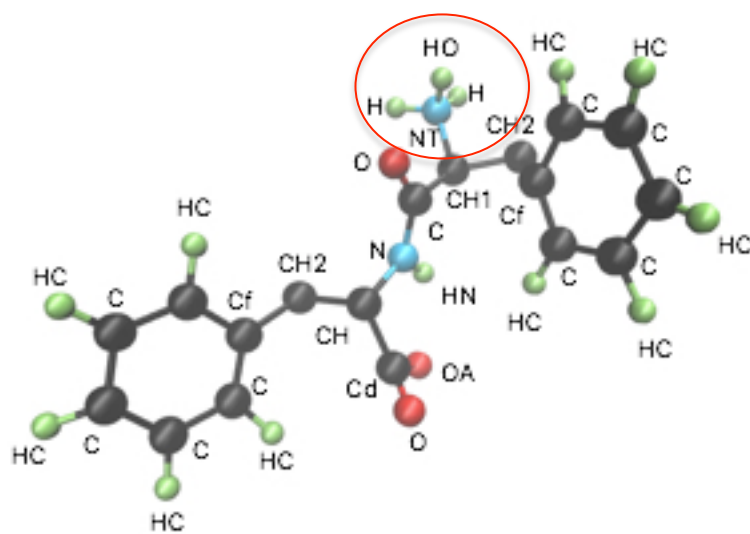
POLAR

NON POLAR

Figure 24. Polar (up) and non-polar (down) charge distribution of FF.

Non-bonded parameters

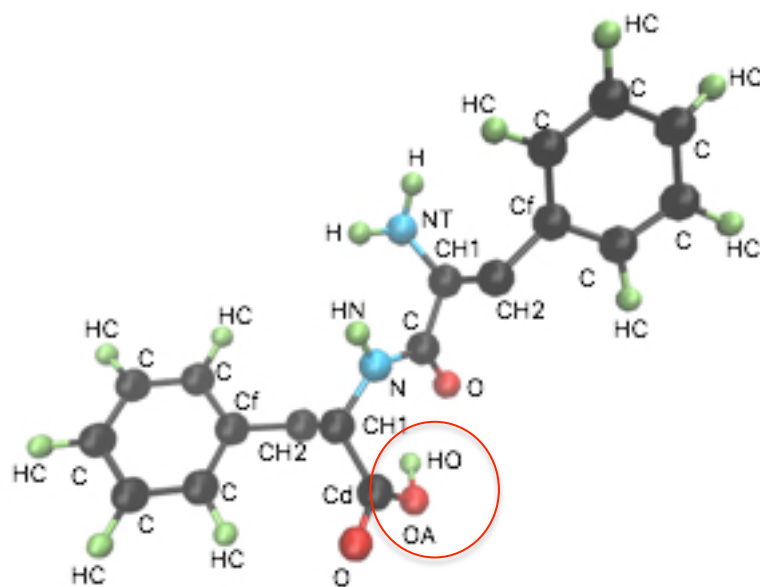| type	mass (u)	q$_{FF polar}$ (e)	q$_{FF non-polar}$ (e)	σ$_{FF polar}$ (nm)	σ$_{FF non-polar}$ (nm)	ε$_{FF polar}$ (kJ/mol)	ε$_{FF non-polar}$ (kJ/mol)
H	1.008	0.248	0.440	0	0	0	0
HN	1.008	0.310	0.310	0	0	0	0
HC	1.008	0.140	0.140	0.237	0.236	0.118	0.118
HO	1.008	0.248	0.408	0	0.240	0	0
NT	14.0067	0.129	-0.660	0.313	0.357	0.639	0.293
CH1	13.019	0.127	-0.220	0.501	0.501	0.094	0.094
CH2	14.035	0	0	0.407	0.407	0.410	0.410
CF	12.011	0	0	0.358	0.358	0.277	0.277
C	12.011	-0.140	-0.140	0.358	0.358	0.277	0.277
Cd	12.011	0.140	0.330	0.358	0.358	0.277	0.277
O	15.9994	-0.450	-0.450	0.276	0.276	1.279	1.279
N	14.0067	-0.310	-0.310	0.313	0.313	0.639	0.639
CH	13.019	0	0	0.501	0.501	0.094	0.094
OA	15.9994	-0.635	-0.288	0.262	0.295	1.725	1.279

Table 7. Non-bonded parameters for FF. [31]

Bonded parameters

Bonds	Length (nm)
H - NT	0.1000
NT – CH1	0.1470
CH1 – CH2	0.1530
CH2 – CF	0.1530
CF – C	0.1390
C – HC	0.1090
C - C	0.1390
CH1 – C	0.1530
C – O	0.1230
C – N	0.1330
HN – N	0.1000
N – CH	0.1470
CH – CH2	0.1530
CH – Cd	0.1530
Cd – OA	0.1360
OA – HO	0.1000
HO - NT	0.1000

ANGLES	ϕ_0 (deg)		k_ϕ (kJ.mol^{-1}.rad^{-2})	
	$FF_{non-polar}$	FF_{polar}	$FF_{non-polar}$	FF_{polar}
H - NT - H	109.50	109.50	380.0	380.0
H - NT - CH1	109.50	109.50	425.0	425.0
NT - CH1 - CH2	109.50	109.50	520.0	520.0
NT - CH1 - C	109.50	109.50	520.0	520.0
CH1 - CH2 - CF	111.00	111.00	530.0	530.0
CH1 - C - O	121.00	121.00	685.0	685.0
CH2 - CH1 - C	109.50	109.50	520.0	520.0
CH2 - CF - C	120.00	120.00	560.0	560.0
CF - C - HC	120.00	120.00	505.5	505.5
CF - C - C	120.00	120.00	560.0	560.0
C - C - HC	120.00	120.00	505.5	505.5
C - C -C	120.00	120.00	560.0	560.0
C - CF -C	120.00	120.00	560.0	560.0
CH1 - C - N	115.00	115.00	610.0	610.0
C - N - HN	123.00	123.00	415.0	415.0
C - N - CH	122.00	122.00	700.0	700.0
O - C - N	124.00	124.00	730.0	730.0
HN - N - CH	115.00	115.00	460.0	460.0
N - CH - CH2	109.50	109.50	520.0	520.0
N - CH - Cd	109.50	109.50	520.0	520.0
CH - CH2 - CF	111.00	111.00	530.0	530.0
CH - Cd - O	121.00	117.00	685.0	635.0
CH - Cd - OA	121.00	117.00	685.0	635.0
CH2 - CH - Cd	109.50	109.50	520.0	520.0
O - Cd - OA	124.00	126.00	730.0	770.0
Cd - OA - HO	109.50	-	450.0	-
HO - NT - H	-	109.50	-	380.0
HO - NT - CH1	-	109.50	-	425.0

DIHEDRALS	ϕ_s (deg)		k_p (kJ.mol^{-1})		periodicity	
	$FF_{non-polar}$	FF_{polar}	$FF_{non-polar}$	FF_{polar}	$FF_{non-polar}$	FF_{polar}
H - NT - CH1 - C	0	0	3.77	3.77	3	3
NT - CH1 - CH2 - CF	0	0	5.92	5.92	3	3
NT - CH1 - C- N	0	0	1.00	1.00	6	6
CH1 - C - N - CH	180	180	33.50	33.50	2	2
C - N - CH - Cd	180	180	1.00	1.00	6	6
HN - CH - CH2 - CF	0	0	5.92	5.92	3	3
N - CH - Cd - OA	0	0	1.00	1.00	6	6
CH - Cd - OA - HO	180	180	16.70	1.00	2	2
CH1 - CH2 - CF - C	0	0	1.00	1.00	6	6
CH - CH2 - CF - C	0	0	1.00	1.00	6	6
CF - C - C - C	0	0	167.42309	167.42309		
C - C - C - C	0	0	167.42309	167.42309		
C - C - CF - C	0	0	167.42309	167.42309		
CF - C - C - CH2	0	0	167.42309	167.42309		
C - C - CF - HC	0	0	167.42309	167.42309		
C - C - C - HC	0	0	167.42309	167.42309		
C - CF - C - HC	0	0	167.42309	167.42309		
CH1 - NT - C - CH2	35.2643	35.2643	334.84617	334.84617		
C - CH1 - N - O	0	0	167.42309	167.42309		
N - C - CH - HN	0	0	167.42309	167.42309		
CH - N - Cd - CH2	35.2643	35.2643	334.84617	334.84617		
Cd - CH - OA - O	0	0	167.42309	167.42309		

Table 8. Bonded parameters for FF. [31]

40

V. Definition of measured quantities

a. Potential of Mean Force between two peptides (PMF).
The calculation of potential of mean force for molecular systems is defined as the potential of an average force over all the configurations of a molecular system. It was first introduced by Kirkwood in 1935 and constitutes a main quantity for computational studies of macromolecular systems. [32]

PMF, $V(r)$, is given by the following mathematical formula:

$$e^{-\beta V(r)} = -k_B T \ln g(\boldsymbol{r})$$

where $\beta = \dfrac{1}{k_B T}$, T is the temperature, k_B is the Boltzmann constant and $g(r)$ is the pair radial distribution function.

The study of PMF gives information on the interactions between two molecules in a medium. In order to analyze these interactions, two molecules are isolated and then the solvent is added in the system.

In this study, two diphenylalanine peptides were solved in tetrahydrofuran solvent (see 9-13 systems). Figure 25 presents a pair of FF peptides at a constraint distance of cm-cm.
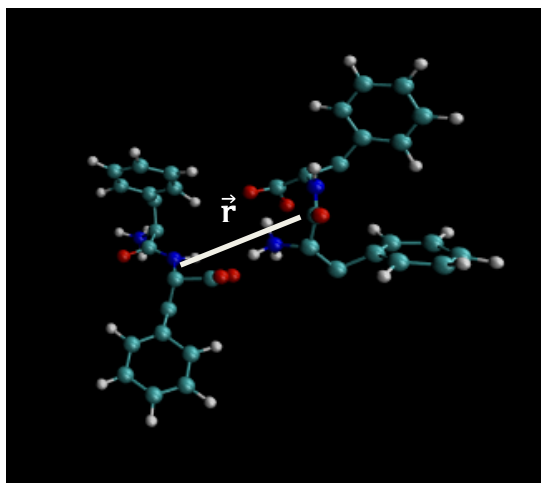

Figure 25. Snapshot of FF - FF at r=0.8 nm.

Methodology of PMF
1. Two FF peptides are put in the simulation box.
2. THF solvent is added in the system.
3. MD simulations are performed for a few nanoseconds, until the system is   well equilibrated.
4. The centers of mass (CM) of each peptide are held at fixed distance.
5. Long simulations are performed with the constant distance of CM.
6. The previous simulations are repeated for a series of different CM distances, usually from 0.3 nm to 2.0 nm.

PMF is calculated by the integration of mean force over the total configurations of the simulation and corrected by an entropy term, due to the constraint distance of centers of mass. [30] In these runs, the two molecules are held at fixed distance and the average constraint force is the negative of the mean force. After repeating the constraint forces over all configurations with different cm-cm distance, the integration of the force is the potential of mean force. During this method, two masses, which do not interact, are pulled apart by an entropic force, because larger volume particles in phase space are sampled at larger constraint length r.[33] The formula of the entropic force is:

$$\frac{2k_BT}{r} = -\frac{d}{dr}\left[-k_BT\log(4\pi r^2)\right]$$

where k_B is the Boltzmann constant, T is the temperature and r is the distance of centers of mass.

The PMF is obtained by the integration of the force between the two molecules backwards from infinity and adding the entropy term to this constraint force. In simulations, the integration starts at a given distance r_{max} and backwards.

Here we compute the mathematical model of PMF through the mean force $F(r)$ as:

$$U(r) = \int_{r_{max}}^{r} F(r)\, dr - 2k_BT \ln r$$

where $F(r)$ is the mean force depending on distance r
$\quad\quad\quad r_{max}$ is the maximum distance of $CM - CM$
$\quad\quad\quad k_B$ is the Boltzmann constant
$\quad\quad\quad T$ is the temperature.

The equation $U(r)$ is a function of distance r. In long distances over r_{max} the mean force is weakening, so $U(r)$ is equal to zero. The entropy term of the equation contains the system's thermal energy k_BT.

We performed a series of simulations to calculate the potential of mean force by comparing three different terms, as it is described in systems 9-13:
  i.   the type of diphenylalanine; FF_{polar} and $FF_{non\text{-}polar}$
  ii.   the electrostatic method; Reaction field, Cut off and PME.
  iii.   the cut off distance; $r_{cut\,off}$ = 1.0 and 1.3 nm.

b. Radial distribution function (rdf).

The structure of peptides can be studied by the radial distribution functions (rdf). Information in the level of center of mass is found in order to understand the behavior of peptide and solvent in the system. Rdf is useful to describe the structure of liquids in general. The radial distribution function is of interest for two reasons: First, neutron and X-ray scattering experiments on simple fluids, and light scattering experiments on colloidal suspensions, yield information about $g(r)$. Second, $g(r)$ plays a main role in theories of the liquid state. Results of experiments for $g(r)$ can be compared with theoretical predictions.

The measure $g(r)$ is the ratio between the average number density $\rho(r)$ at a distance r from a given atom and the density at a distance r from an atom of the system at the same total density (Figure 26).

It determines the numbers of atoms expected to find at a distance r from a specific atom, i.e. the total number of atoms at a distance between r + dr. We consider a shell of thickness r from the center, so the volume is:

$$\Delta V = \frac{4}{3}\pi(r+dr)^3 - \frac{4}{3}\pi r^3$$
$$= \frac{4}{3}\pi r^3 + 4\pi r^2 dr + 4\pi r(dr)^2 + \frac{4}{3}\pi(dr)^3 - \frac{4}{3}\pi r^3$$
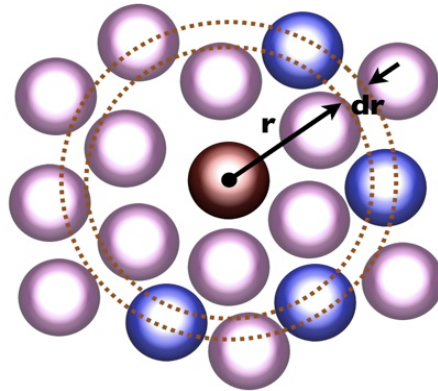$$\approx 4\pi r^2 dr.$$

Figure 26. Radial distribution function. [34]

g(r) between two particles A, B is defined in the following equation: [23]

$$g_{AB}(r) = \frac{\langle\rho_B(r)\rangle}{\langle\rho_B\rangle_{local}} = \frac{1}{\langle\rho_B\rangle_{local}} \frac{1}{N_A} \sum_{i\in A}^{N_A}\sum_{j\in B}^{N_B} \frac{\delta(r_{ij}-r)}{\Delta V}$$

where $\langle\rho_B(r)\rangle$ the particle density of type B averaged over all spheres around particles A with radius r_{max}

N_A the number of particles of A

In a simulation box with periodic boundary conditions r_{max} is smaller or equal to half the box size.

The radial distribution function in liquids has small peaks at short distances and moves around a constant value at longer distances. At distances less than the atomic diameter, $g(r)$ is zero due to the strong repulsive forces. When peaks are observed, $g(r)$ exceeds 1 because there are attractive forces between the particles. As r grows, the forces weaken and the probability to find two particles around our radius is the same, and that is why $g(r) = 1$. In general, any deviation of $g(r)$ from unity reflects correlations between the particles of the system due to the intermolecular interactions.

So, g(r) is a conditional probability to find another particle at r distance away from the origin.

c. Radius of gyration.

The shape and size of a molecule are essential properties for understanding its general behavior and thus, the solution in various substances. These characteristics also give useful information about hydrodynamic and transport properties and constitute a bridge among different fields of science including chemistry, biology and engineering.

In order to investigate the size of particles, we introduce a measure, which is called Radius of gyration. We assume a sphere of radius R_{cm} around the center of mass of the molecule and count the distances r of each atom of the molecule from the center of mass (Figure 27). The atoms' masses are given.

The mean size of a molecule is estimated by the following equation: [30]

$$\langle R_g \rangle = \sqrt{\left\langle \frac{\sum_i m_i (r_i - R_{cm})^2}{\sum_i m_i} \right\rangle}$$

where R_{cm} is the radius of center of mass

r_i is the distance of each atom from the center of mass
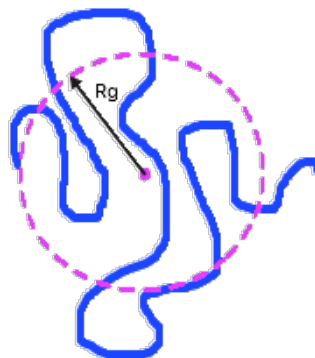
m_i is the mass of every atom.

Figure 27. Radius of gyration of a polymer

VI. Results

As it was mentioned previously, the answer to the question of which is the most appropriate force field for the description of a solution of peptides is difficult. In the following a classification of various parameters of the force field has been performed. We start with the results concerning the potential of mean force.

1. PMF

1.1. Effect of charge distribution of the FF peptide on the intermolecular interactions (Polar- Non Polar).

A possible charge transfer during the solution of the peptide in the solvent can dramatically alter the results either of the simulation or experiment in the laboratory. Figure 28 describes the main difference between the PMF curves of systems 9 and 11. Both systems have been simulated with the same parameters but with different charge distribution. In system 9, which in the following will be called Non-polar, the end of FF peptide has a carboxyl group that prevents the peptides to make aggregates. The black open-dotted curve represents the $FF_{non\ polar}$ - $FF_{non\ polar}$ potential. No strong attraction is present, as the line tends to zero from distance 1.0 nm between the two centers of mass of the peptides. At smaller distances (r=0.3 - 0.8 nm) positive values are observed due to the repulsion forces of the molecules. The non-polar FF in THF has no attraction well even in short distances. The other curve, red with closed dots, represents the FF_{polar} - FF_{polar} potential. In system 11, FF polar has an oxygen instead of a carboxyl as FF non polar, and stronger interactions are performed. The deep well from r=0.4 nm to r=1.2 nm shows the attractive interactions that are performed at the respective distances of the peptides' centers of mass. The charge distribution affects strongly the FF - FF potential and so, its solution in the solvent. It is not clear which charge distribution is the ideal to describe these interactions, and it is still under further investigation.

Error bars at PMF curve of FF polar in Figure 28 are 1- 31% of the actual values. The points in the well have the larger values of the error bars (i.e. r = 0.5 - 1.1 nm).
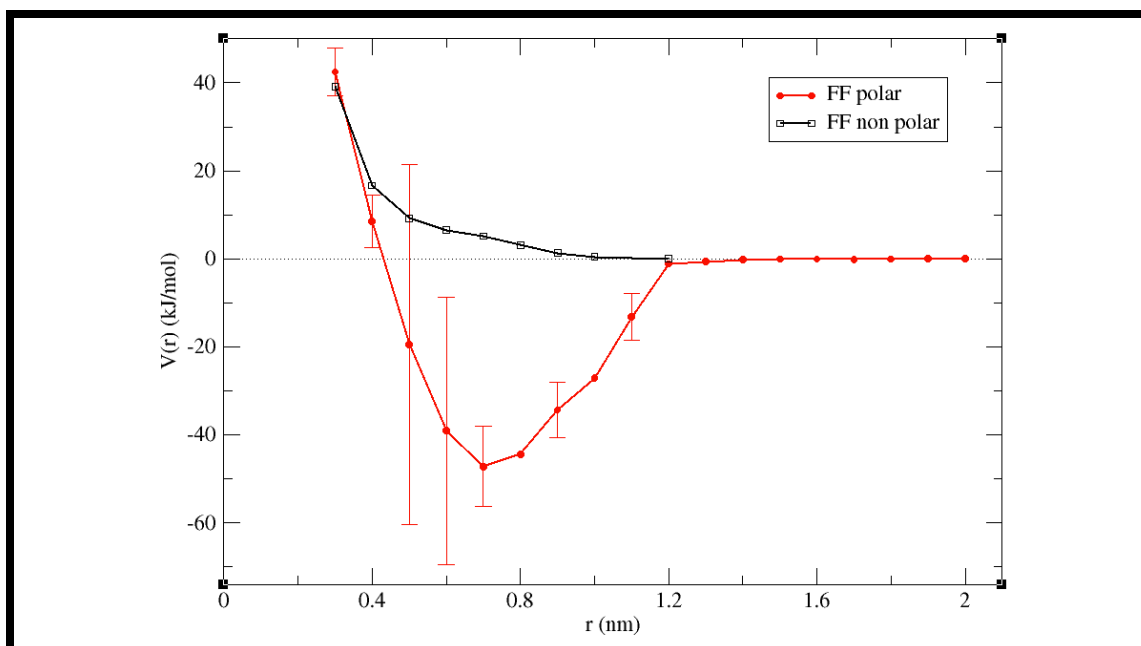
Figure 28. The PMF as a function of distance between the centers of mass of FF with non-polar structure (FF$_{non-polar}$) in THF (open symbols) and with polar structure (FF$_{polar}$) in THF (closed symbols) by the electrostatic method of Reaction field and r$_{cutoff}$ = 1.3 (Systems 9, 11).

1.2. Effect of method

The accuracy of the calculations of the various interactions is a very important issue. For this reason we performed a series of comparisons concerning the method of electrostatic interactions. The effect is large for polar while for non polar, the method of calculation of electrostatic interactions does not play important role as depicted in Figure 29. Slight differences are shown mainly at small cm-cm distances. Moreover, the non-polar PMF has not special differences between Reaction field and Cut off method as it is shown at Figure 30. On the other hand, the polar structure has deeper well at distances r=0.4-1.0 nm with the PME and Cut-off methods, where the effect is stronger. This information is important because PME compared to Reaction field is a more accurate method. When the centers of mass of FF are far from the other, PMF goes to zero. The two methods Cutoff and PME give similar results for PMF, with small expected differences. A large deviation is observed though Reaction Field method, which needs further investigation.
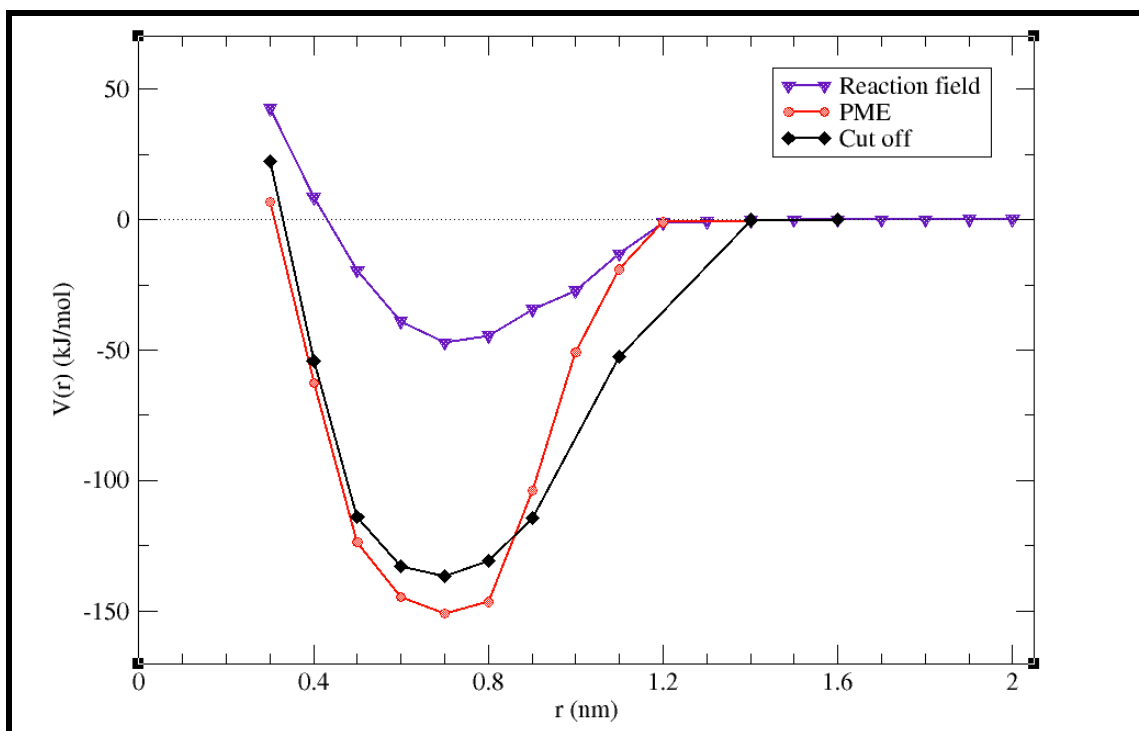
Figure 29. The PMF as a function of distance between the centers of mass of FF with polar structure (FF$_{polar}$) in THF and r$_{cutoff}$ = 1.3 nm by the electrostatic methods of Reaction field (triangle symbols), PME (circle symbols) and Cut off (diamond symbols).(Systems 11, 12, 13).
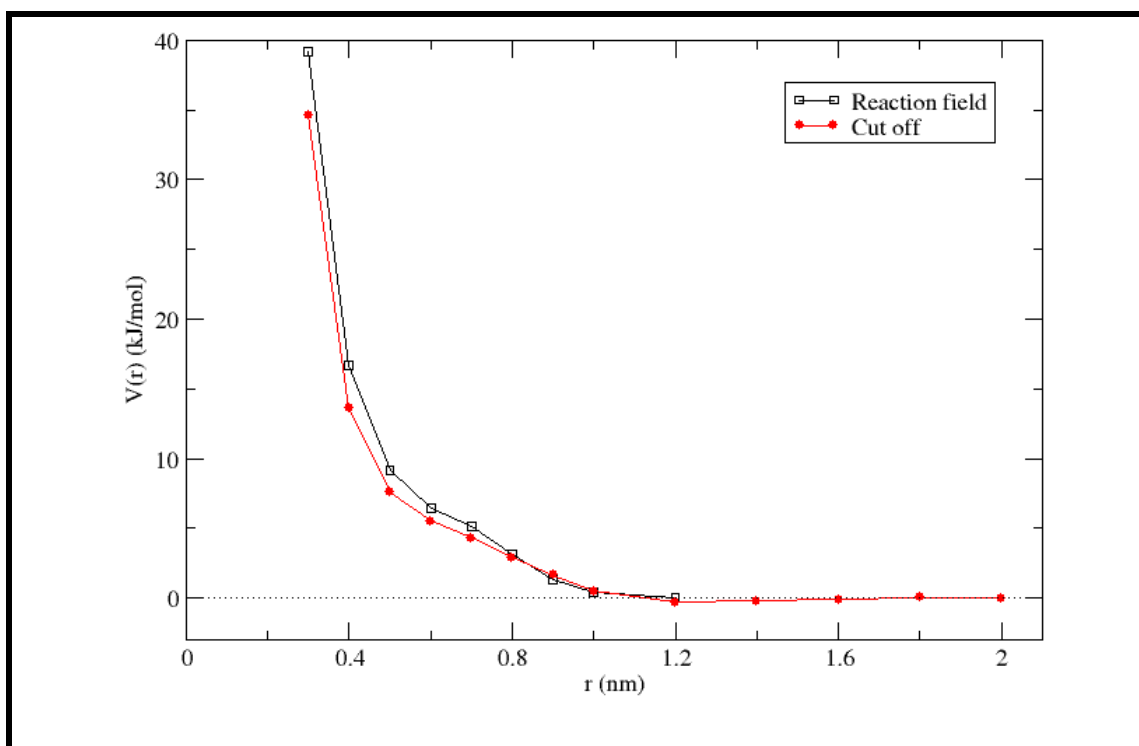

Figure 30. The PMF as a function of distance between the centers of mass of FF with non-polar structure (FF$_{non-polar}$) in THF and r$_{cutoff}$ = 1.3 nm by the electrostatic methods of Reaction field (open symbols) and Cut off (closed symbols) (Systems 9, 10).

1.3. Effect of cut off distance using the same method.

The value of r_{cutoff} seems to play also a role on PMF. From Figure 31, it is obvious that PMF's calculation by the same method (i.e. cut off, reaction field, PME), same FF type but change of cut off distance from 1.0 to 1.3 nm affects the attraction of cm-cm in small distances. The important difference is identified between 0.6 – 1.0 nm distance, where r_{cutoff} = 1.0 nm presents stronger attraction. The larger the cut off distance, the closer to reality the results are and that because interactions are performed all over the simulation box and not only up to a certain distance r_{cutoff} and beyond which are zeroed. However, it costs in time setting a big cut off distance, because there are more interactions to be calculated.
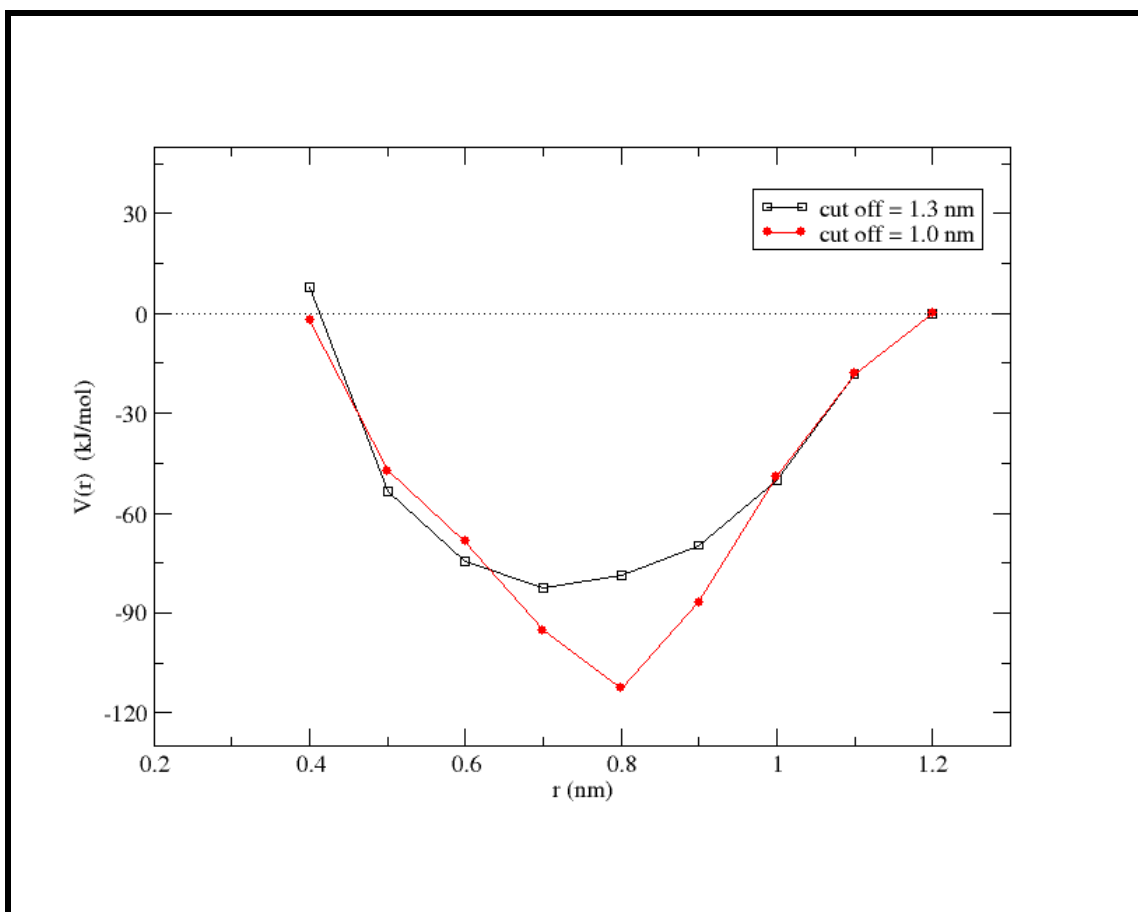

Figure 31. The PMF as a function of distance between the centers of mass of FF with polar structure (FF$_{polar}$) in THF by the electrostatic method of PME and r_{cutoff} = 1.3 nm (open symbols) and r_{cutoff} = 1.0 nm (closed symbols) (Systems 10, 13).

2. SOLUTION FF-THF

In the following solutions, FF peptides in THF solvent are examined. We start with structural and conformational properties describing the radial distribution function (rdf). Characteristic snapshots of FF in THF are presented in Figure 32.
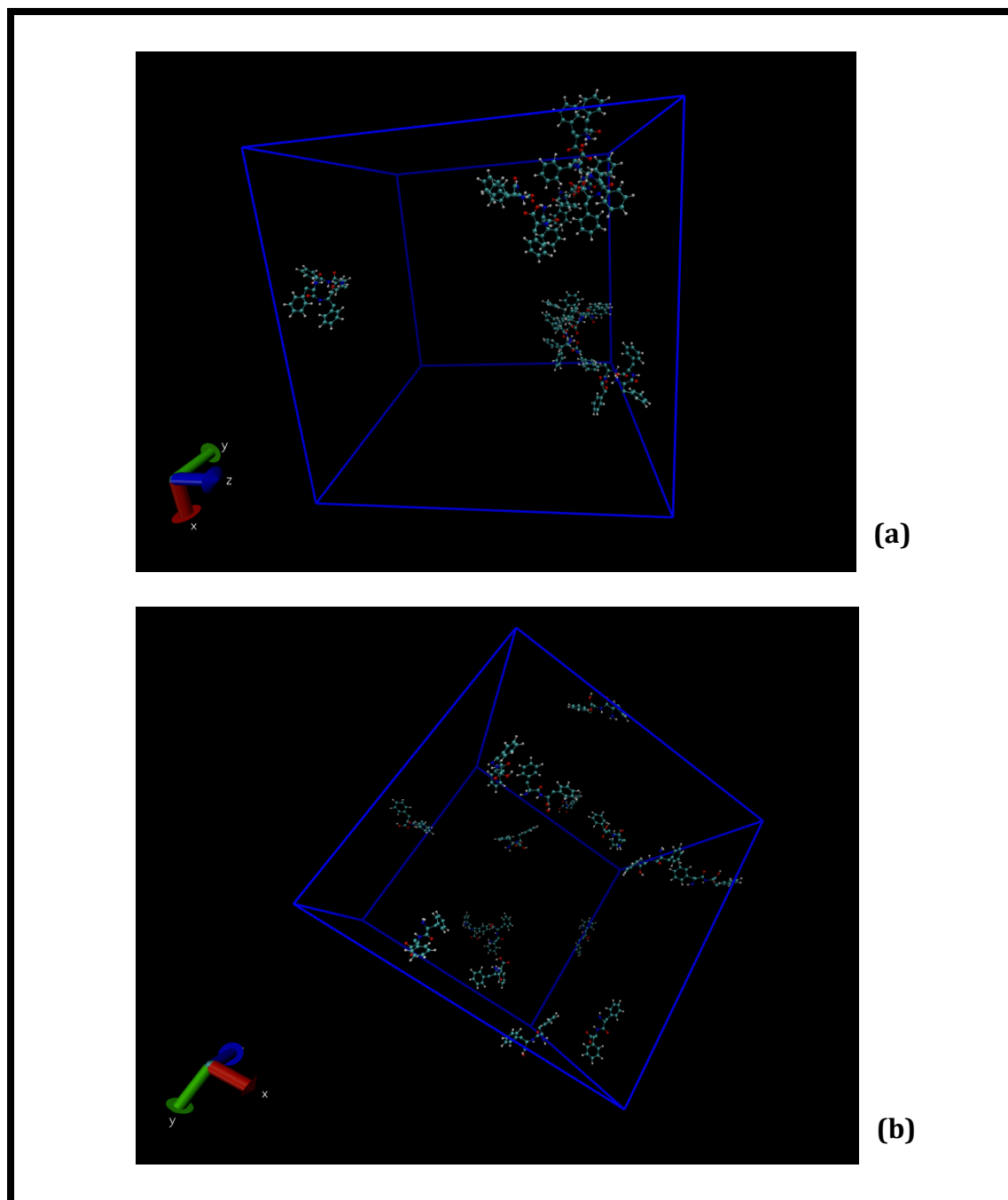

(a)

(b)

Figure 32. Snapshots from MD simulations of the solution of (a) FF polar and (b) FF non polar in THF solvent. Only the peptides are shown.

In Figure 32 is presented the solution of 16 diphenylalanine peptides in tetrahydrofuran solvent in their simulation boxes. In the first picture (a) there is the polar charge distribution of FF. The peptides have the tendency for self-assembly and create aggregates. On the other hand, the next figure shows the non-polar charge distribution of FF and its solution in THF in the simulation box. The second picture (b) is the snapshot of FF non polar. There are not any structures between the peptides but they can be found in various positions into the box.

To get information about the solvent effect on diphenylalanine solution in tetrahydrofuran solvent, we estimated the radial pair distribution function between two FF peptides and peptide-solvent.

In this work, the rdf of systems 1 – 8 are studied between:
   i) peptide–peptide
   ii) peptide–solvent.

In order to take information about the effect of electrostatic method, the cut off distance, the structure of FF and the effect of solvent, there are comparative figures of rdf curves.

Analyzing the peptide-peptide curves of Figure 33 (a), a steep high peak is presented in system of FF (polar) in THF. At r=0.7 nm, FF approach each other (black curve), and as distance grows, $g(r)$ tends to zero. FF self-assembles and performs aggregates for its polar charge distribution, and excludes the molecules of solvent away. This is obvious from the (b) picture of Figure 33, as the red curve which presents the $g(r)$, is smaller than 1 at near distances and after $r = 2$ nm goes to unity. The physical importance is that FF prefers to be near another FF at short distances and pull away the THF molecules, and at longer distances there is an homogenous distribution of the molecules. In contrast, at bigger distances $r > 1.5$ nm, FF performs repulsive forces for other FF peptides. Interesting information we take from this measurement is that THF may not be a good solvent for the peptide with the polar charge distribution, and from literature is written that phenyl groups are responsible for the hydrophobic behavior of diphenylalanine. [30]

As for THF as solvent for diphenylalanine, we saw that the radial distribution function varies depending on the charge distribution of the peptide.
In Figure 33 (a), $g(r)$ of FF non-polar - FF non-polar in THF has no peak at any distance r. Subsequently $g(r)$ tends to unity. According to FF-FF rdf curve, the FF-THF curve supplies the previous results for FF polar - FF polar $g(r)$ curve. When peptides get near, tetrahydrofuran molecules are pulling away ($0.6 \leq r < 1$ nm). At longer distances FF and THF coexist without performing strong interactions. Similarly, the $g(r)$ of FF non-polar - THF curve is complementary to the FFnon-polar - FF non-polar curve of Figure 33 (a) and shows almost no tendencies between FF and THF molecules.

The main difference of the force fields used on this work is mounted on observing the change between FF polar and FF non polar in the same solvent, which is

tetrahydrofuran. In the first case, diphenylalanine forms aggregates and has a tendency for self-assembly. But FF non polar has a completely different behavior, which is shown in the two snapshots of this section of the chapter (Figure 32), and seems to be solved good in THF solvent.
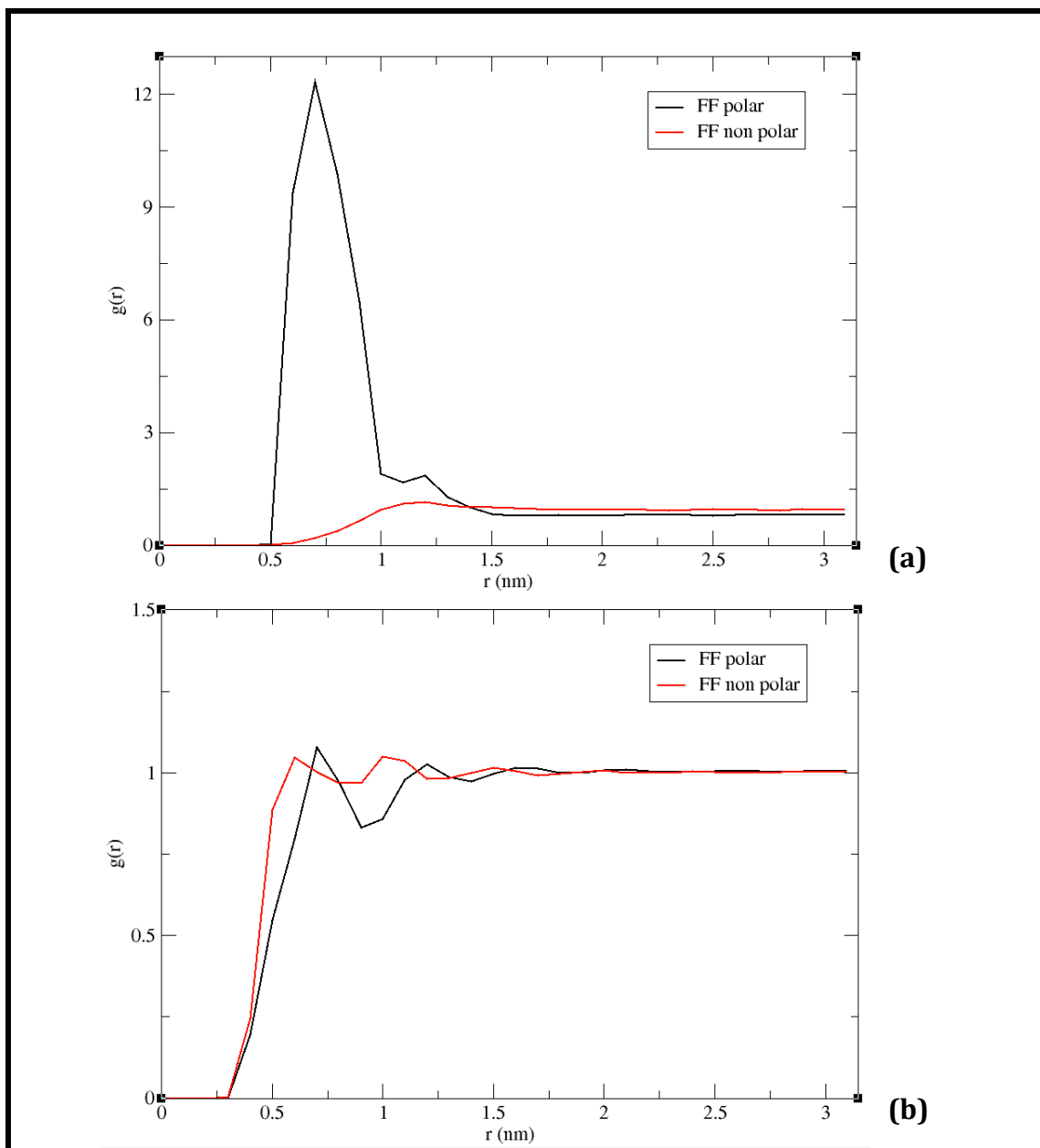

Figure 33. Pair distribution functions of (a) FF-FF and (b) FF-THF for FFpolar (black) and FFnon-polar (red) charge distribution (Reaction field, $r_{cut-off}$=1.3nm).

3. RADIUS OF GYRATION

The mean size of an FF molecule was found by calculating the radius of gyration of the FF peptide. Its size is equal to $0,3927 \pm 0,029$ in THF.

To estimate the radius of gyration $\langle R_g \rangle$ of the peptide, we found the average size within all peptides and their configurations (trajectories). After we estimated the standard deviation too.

In the following table (Table 9), we list the values of $\langle R_g \rangle$ according to the system studied.

System	$\langle R_g \rangle$
1 $FF_{non-polar}$ Cutoff, r_{cut}=0.9nm	$0,3981 \pm 0,034$nm
2 $FF_{non-polar}$ Cutoff, r_{cut}=1.0nm	$0,4027 \pm 0,032$nm
3 $FF_{non-polar}$ PME, r_{cut}=1.0nm	$0,3959 \pm 0,033$nm
4 $FF_{non-polar}$ Cutoff, r_{cut}=1.3nm	$0,4090 \pm 0,030$nm
5 $FF_{non-polar}$ Reaction field, r_{cut}=1.3nm	$0,4000 \pm 0,033$nm
6 FF_{polar} Reaction field, r_{cut}=1.3nm	$0,3506 \pm 0,013$nm

Table 9. Average of $\langle R_g \rangle$ of FF solution in THF with error bars

It seems that the method for electrostatics (Cut off, Reaction field, PME) as well as the cut off distance (r_{cut}) do not play important role in the radius of gyration as all values are similar within error bars. A small difference is observed between the polar and non-polar charge distribution where the peptide seems to attain a smaller size $\langle R_g \rangle$ and this may be attributed to the different allocation of charge.

4. ORIENTATION

In order to explore the conformational properties of FF peptides into a solvent, a series of measurements about the orientation of the peptides were performed. First we need to define what orientation is. This term describes the way that a molecule is positioned in relation to another molecule. Calculations were performed on pairs of peptides that the distance of their centers of mass was kept constant.

More specifically, the code that quantifies the orientation of the molecules, does the inner product of the end-to-end vectors of the two molecules and after finds the angle that is formed between the vectors.[30]

To consider the orientation of FF, a number of simulation runs for a pair of peptides in THF solvent was performed for different cm-cm constant distances (see Systems 9 and 11 in chapter 3). Both FF with polar and non-polar charge distribution were applied in the runs so that information about how the charge distribution of a molecule could possibly affect its preferable position with respect to another molecule.

Figure 36 presents the probability distribution *P(θ)* of θ-value for FF polar (a) and FF non-polar (b) in THF solvent. Various cm-cm distances are shown with different colors.

The main characteristic of the orientation of FF polar (a) is the preference to orient antiparallel at short distances (0.4-0.6 nm). This is due to both electrostatic interactions as well as strong hydrogen bonding, which is discussed in the following. The tendency of antiparallel orientation exists up to $\cong$1nm but angles have moved to lower values ($\cong$120°). The other distances have no specific orientations and a random orientation is observed. Figure 37 describes the preferences of FF to orient.

On the other hand, FF non polar (b) prefers a vertical orientation at short distances like 0.4nm due to strong electrostatic repulsions. A 0.6nm an almost parallel orientation seems to be preferable. Longer cm-cm distance reduces where the intermolecular interactions between FF peptides are reduced, lead to various random conformations.

It is worth to note here the difference with FF non polar and FF polar at the same short cm-cm distances (0.4-0.6 nm). This is interpreted due to the polar charge of FF polar, which has stronger interactions. The peptides are forced to have standard conformations due to hydrogen bonding as well.

In Figures 34 and 35, characteristic snapshots of antiparallel and vertical orientations are presented for FF polar and FF non-polar at short distances. When distance is growing, the orientation is random for both charge distributions of FF, as it is depicted in Figure 37.
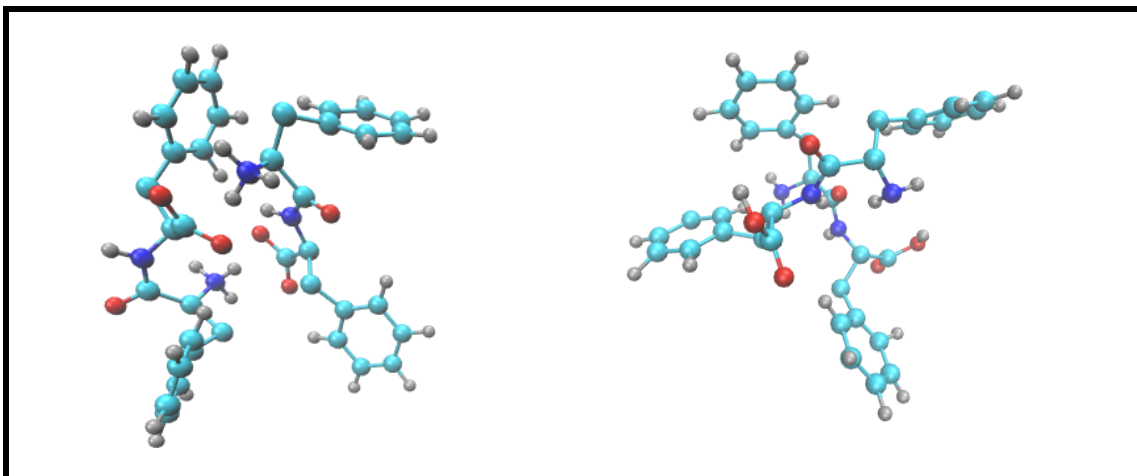
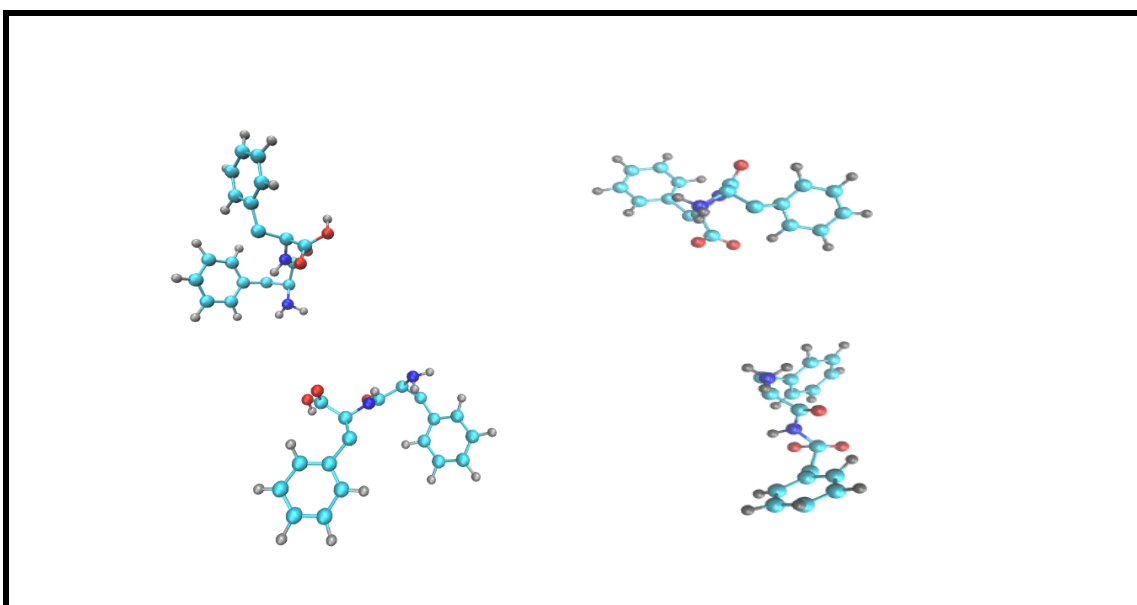**Figure 34.** Snapshot from cm-cm distance dr=0.4nm of FF polar (left) and FF non-polar (right).

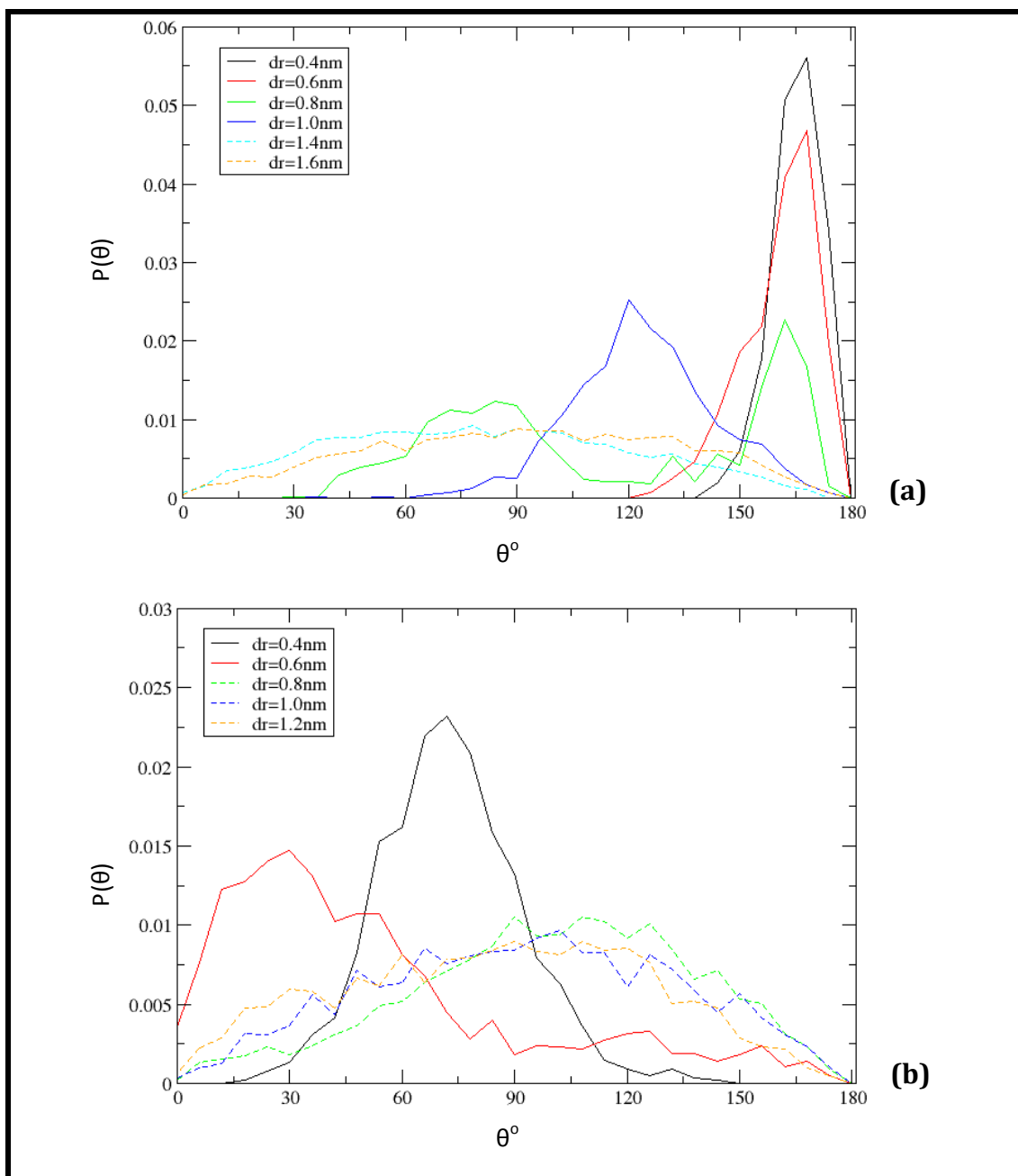**Figure 35.** Snapshot from cm-cm distance dr=1.2nm of FF polar (left) and FF non-polar (right).

Figure 36. Orientations between a pair of (a) FF polar and (b) FF non polar peptides in THF solvent at different cm-cm distances (Systems 9, 11).

5. HYDROGEN BONDS

Another measure of self-assembly of FF peptides in a solvent is by observing the hydrogen bonds that are formed in the system. This study is based on atomic and not on molecular level, like the pair distribution functions.

Firstly, we need to define what a hydrogen bond is. It is an attractive interaction between two electronegative atoms that occurs when a hydrogen, bound to one of

them, is attracted by the other. This is characterized by a geometric criterion (1), which contains angles and distances of the three particles. [35]
A hydrogen bond exists between two particles A and B if:

  a) r(A…B) ≤ 3.5 A
  b) r(A…H) ≤ 2.6 A
  c) angle (A…B - H) ≤ 30°

The atoms A and B in the systems are the electronegative atoms N and O in FF, O in THF and H the peptide's hydrogens.

The methodology applied in order to estimate the number of hydrogen bonds between the molecules of the system was by counting the number of all hydrogen bonds between peptides and between peptides and solvent molecules for all trajectories. For peptide's-peptide's calculation, we made a code, which counts all the electronegative atoms of FF (i.e. N, O, H), and puts them in a list. After it executes the criterion (1) and checks if a hydrogen bond exists. The same thing is done for peptide's-solvent's calculation. In this case, we considered a sphere of radius equal to 1 nm around the center of mass of each FF molecule and for the number of THF molecules that lie in this region, we counted the number of hydrogen bonds which are formed between FF and THF molecules. In other words, we found the number of hydrogen bonds that exist around an FF molecule at distance 1nm considering a sphere around the peptide's center of mass, and do this for all FF molecules and trajectories. At last we divide our results with the number of conformations and the number of peptides in order to find an average number of hydrogen bonds around one peptide. Figure 37 describes a hydrogen bond formed between two water molecules.
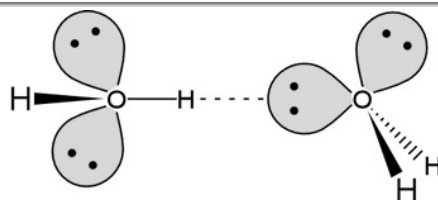

Figure 37. Example of hydrogen bond between two water molecules. [36]

Three measurements of hydrogen bonding were performed: between THF-THF, THF-FF and FF-FF molecules. The system studied in this section was FF_{polar} in THF solvent with reaction field method and cut off distance equal to 1.3 nm (System 6).

i. THF-THF
No hydrogen bonds are formed between the solvent despite the fact that there are eight hydrogens in each molecule. The solvent is an organic ring and no interatomic interactions exist. Consequently, zero number of hydrogen bonds is found and there is not any structures or networks of THF in the system.

ii. THF-FF
As it was mentioned previously, the hydrogen bonds of solvent-peptide are formed between the first and the hydrogen of the last. N and O of the peptide are the

donors, because diphenylalanine's H is bounded to them, and O of tetrahydrofuran is the acceptor. In order to estimate the degree of the structure of hydrogen bond networks around an FF, we did the following procedure. A sphere of radius 1 nm was taken around each FF molecule and it was calculated how many solvent molecules are situated inside it. In the following, the number of hydrogen bonds that are performed between FF and solvent molecules are calculated and was equal to 2.96.

iii. FF-FF

Hydrogen bonds are formed between peptides in THF solvent. As it has been shown from snapshots during the simulation, FF_{polar} peptides have a tendency for self-assembling. The number of hydrogen bonds per FF with another FF was found equal to 0.27. Both intermolecular and intramolecular hydrogen bonds are performed within FF molecules and head-to-tail bounds as well.

Table 10 contains the corresponding results for FF with polar charge distribution.

molecules	<HB>
FF-FF/FF in THF	0.27
(THF-FF/FF)$_{List}$ in THF	2.96
THF-THF/THF	0

Table 10. Average number of hydrogen bonds between FF-FF, FF-solvent and solvent-solvent.

6. COMPARISON WITH OTHER SOLVENTS

We studied the behavior of FF in THF solution at previous sections. It is also important to compare our observations to other studies such us the solution of FF peptide in different solvents. In the following, we compare results from FF solution in organic and aqueous solvent by the potential of mean force, the radial distribution functions, the preference of orientation of two peptides at a given distance of cm-cm and the number of hydrogen bonds that are performed. Information about FF diluted in water and methanol was taken from previous studies by Rissanou et al. [30]

6.1. PMF

At first, we observe completely different curves of potential of mean force between organic and aqueous solvent, as they are described in Figure 38 from older studies. PMF of FF in water presents an attractive well at distances from 0.6 nm to 1.6 nm and tends to zero at longer distances. The attraction of FF molecules is due to the hydrophobic phenyl groups of the peptides. On the other hand, diphenylalanine peptides have no strong attractive interactions in methanol, as the PMF curve has no negative values on y-axis. It is also presented the thermal energy k_BT with green

color in Figure 38. Additionally the choice of solvent is very important for the results. For instance, FF_{polar} in THF solvent is much deeper (V_{max}= 50 kJ/mol) than in aqueous solvent (V_{max} = 22 kJ/mol), as it is obvious by the comparison of Figures 38 and 39.
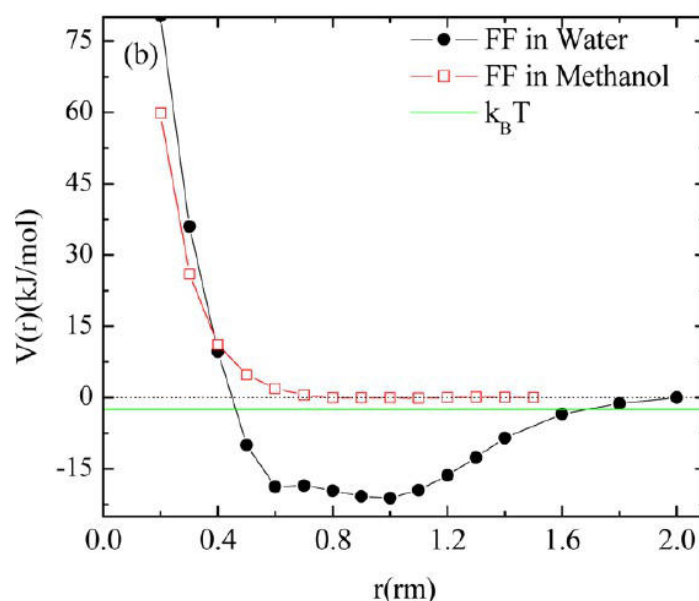

Figure 38. PMF as a function of distance of cm-cm of FF peptides. FF in water solvent is presented with black and in methanol with red. [30]

PMF of FF in THF differs from the one charge distribution to the other and this is depicted in Figure . The PMF of FF with non-polar charge distribution is alike the PMF curve of FF in methanol as it is shown in Figure . Similarly, FF with polar charge distribution behaves like when FF is solved in water, because there is an attractive well at both cases.
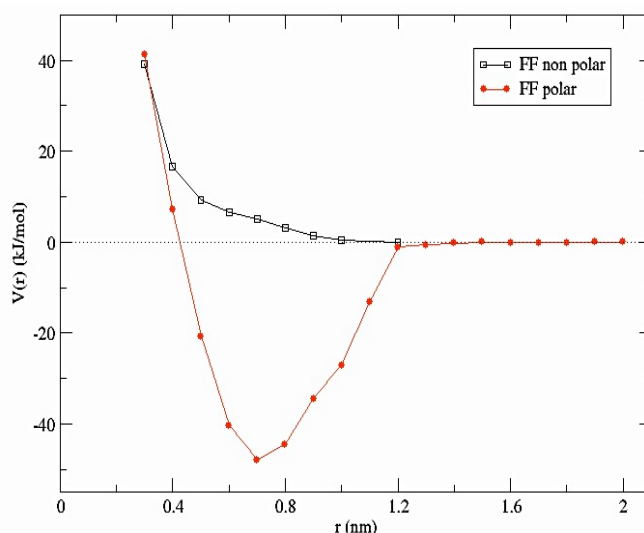

Figure 39. PMF as a function of distance of cm-cm of FF peptides. Solution of FF in THF solvent. With red is FF polar and with black FF non-polar.

6.2. Radial distribution function

Another method to compare the behavior of FF is by the calculation of radial distribution functions. Pair distribution function depends on the solvent of the system and the interactions between solvent and solute. The effect of organic and aqueous solvents affects the rdf curves.

In Figures 40 and 41, it is depicted the $g(r)$ of FF in three different solvents; THF, water and methanol. From previous studies we have already calculated the $g(r)$ of FF in methanol and water. For comparison reasons, the same cut-off distance and the same method for electrostatics have been used, i.e. r_{cut} = 1 and Cut-off method except for Figure 41, where the FF in THF system was performed with reaction field electrostatics and r_{cut} = 1.3 nm. We note that the choice of electrostatics between cut-off and reaction field does not differentiate the results, as it stated previously, for FF$_{non-polar}$.

The FF's charge distribution is non-polar in methanol and polar in water at both Figures 40-41. Although, in Figure 40, FF polar is solved in THF, where peaks are observed, and FF non-polar is described in the Figure 41 and has no peaks at all. This comparison was done to highlight the contrast among polar and non-polar charge distribution, or the effects of using a different model on the outcome results.

The rdf curves of FF-FF in methanol and FF-methanol in methanol (non-polar charge distribution) show no peaks or steepness. At Figure 40 (a), the blue line which represents the $g(r)$ of FF-FF in methanol, is under 1 at short distances, which is justified due to the intermolecular interactions, and after tends to unity. Almost the same effect is observed at $g(r)$ of FF-methanol (Figure 40 (b)), confirming that organic solvents like methanol dissolve well FF molecules.
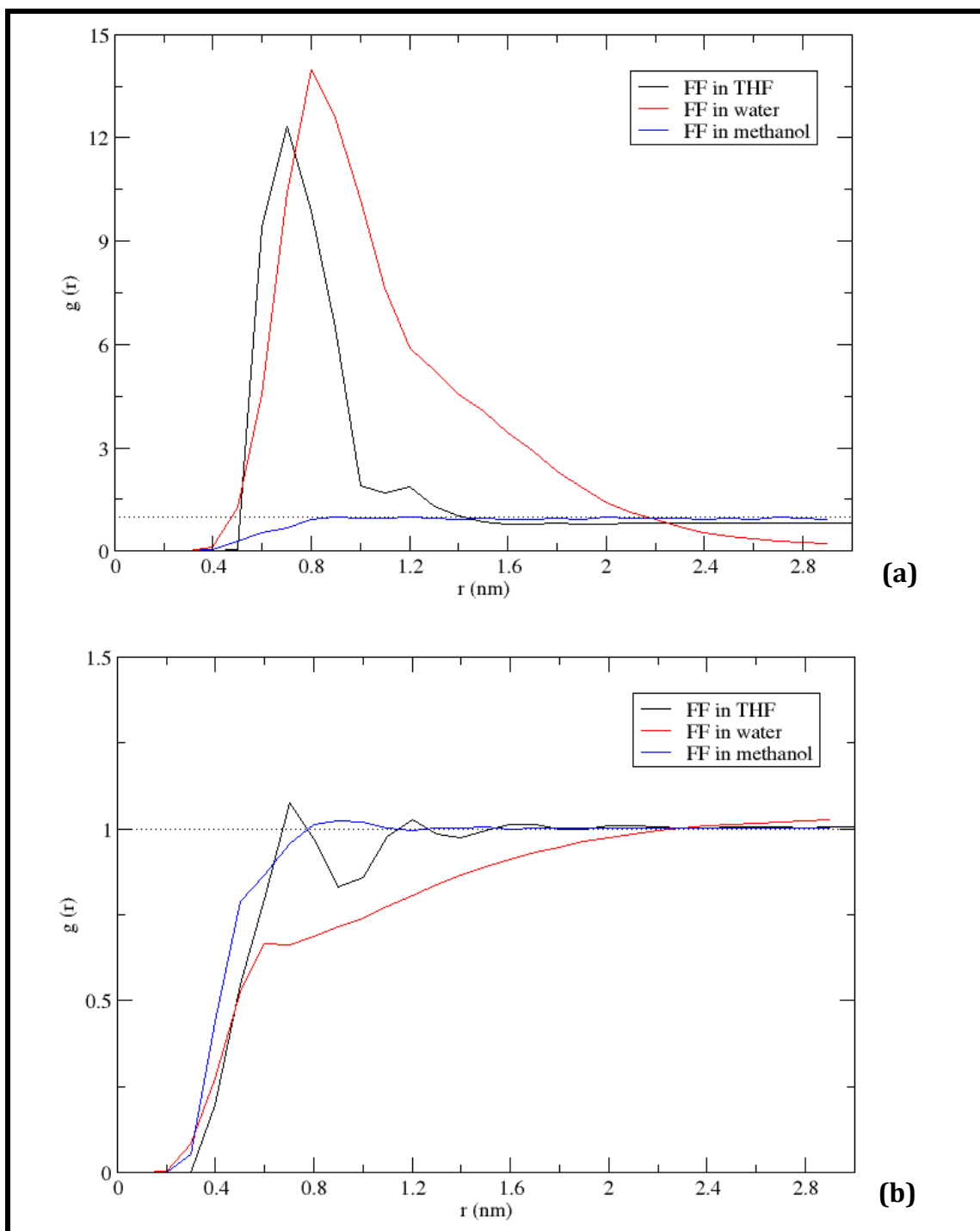the FF non polar -THF curve is like methanol's too.

Figure 40. Radial distribution functions of (a) FF-FF and (b) FF-solvent in three different solvents. FF polar is used in THF and water system and FF non-polar for methanol system (Systems 6, 14, 15).
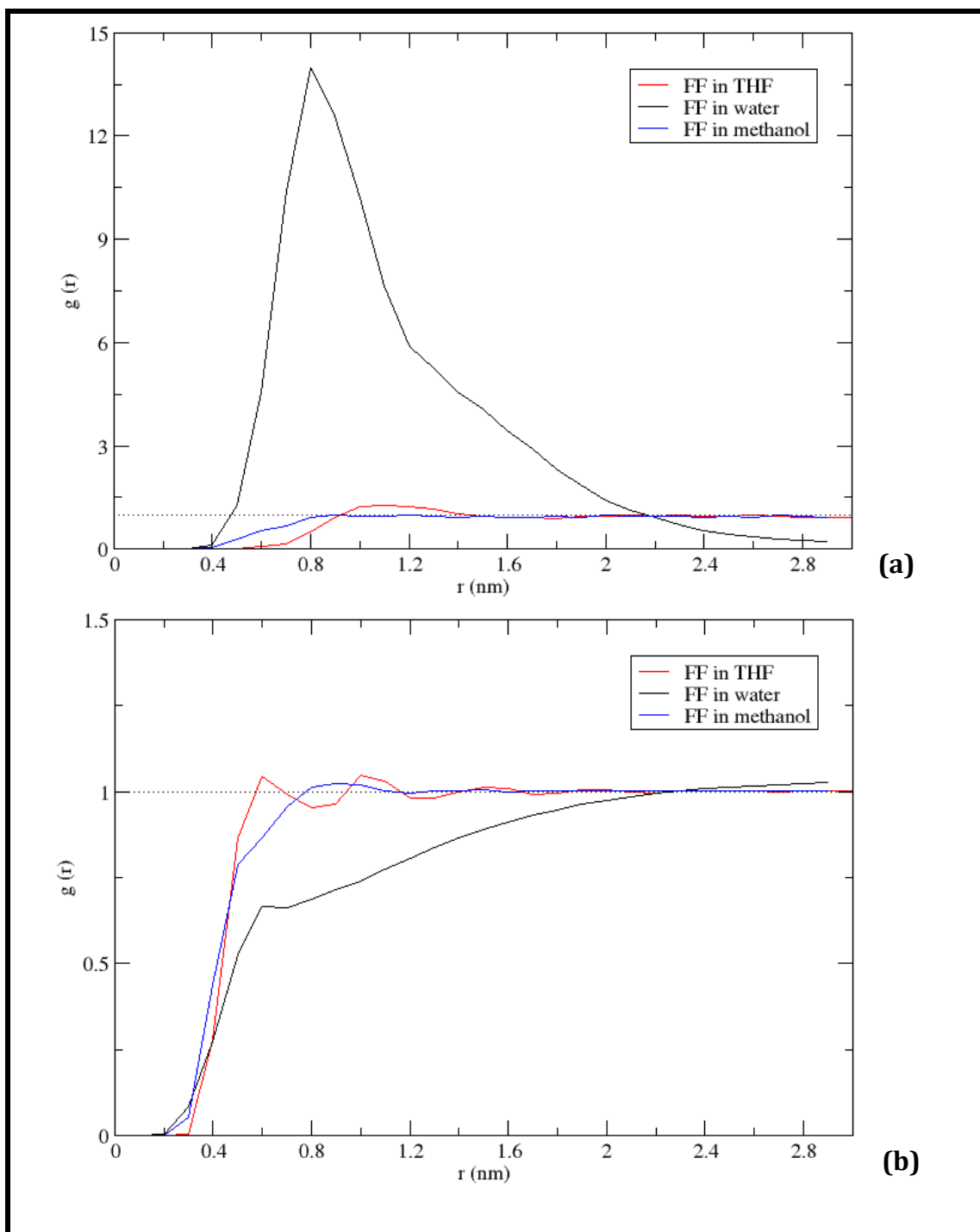
Figure 41. Radial distribution functions of (a) FF-FF and (b) FF-solvent in three different solvents. FF non-polar is used in THF and methanol systems and FF polar for water system (Systems 2, 14, 15).

6.3. Radius of gyration

The mean size of FF_{polar} peptides in THF solvent was found equal to $0{,}3927 \pm 0{,}029$, while it is 0.361 ± 0.06 nm in water and 0.395 ± 0.08 nm in methanol, which is known from previous studies. FF behaves almost the same way as in methanol, as the two sizes are very close. So THF can be considered a better solvent for FF than water, just

like methanol. Table 11 contains the estimated Rg of diphenylalanine in three different solvents; methanol, water and THF. The Rg of FF$_{non-polar}$ in THF is coming from system 5 and from FF$_{polar}$ from system 6.

System	$\langle R_g \rangle$
FF in methanol	0,395 $\pm$ 0,08 nm
FF in water	0,361 $\pm$ 0,06 nm
FF$_{non-polar}$ in THF	0,400 $\pm$ 0,03 nm
FF$_{polar}$ in THF	0,3506 $\pm$ 0,013 nm

Table 11. Average of $\langle R_g \rangle$ in three different solvents.

6.4. Orientation

Concerning the preference of FF peptides to orient, it is clear that the role of solvent is determinant. For this reason we appose the orientation of FF in THF solvent and in organic and aqueous solvent. Figure 42 presents the preferable orientation of FF in water and methanol and Figure 36 the same but in THF solvent. We note that FF$_{non-polar}$ has a vertical orientation to another FF, and FF$_{polar}$ an antiparallel at short distances and above them a random. In Figure 42 (a), the peptides prefer to orient antiparallel at short distances, like FF$_{polar}$ in THF, due to the electrostatic interactions at the charged end groups. In contrast, there is a vertical orientation at distance of 0.4 nm of peptides cm-cm for FF in methanol (Figure 42 (b)). Almost the same orientation is preferred by FF$_{non-polar}$ in THF.

The similarities between the systems of FF in methanol and FF$_{non-polar}$ in THF (and FF in water and FF$_{polar}$ in THF respectively) are due to the charge distribution, which is the same for both type of peptides. This is why we observe the same orientation characteristics. Figure 42 describes the orientation of the peptide, which has already been studied by Rissanou et al.[30]
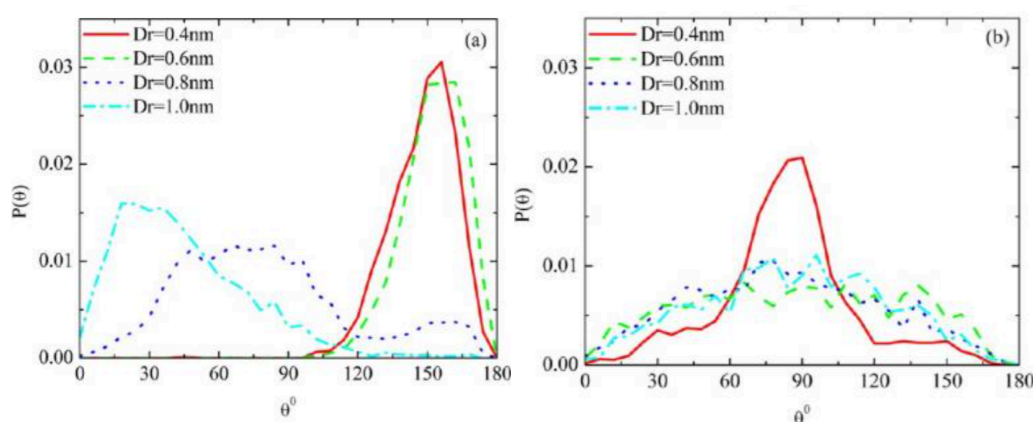

Figure 42. Probable orientations between a pair of FF peptides in (a) water and (b) methanol, at different cm-cm constant distances (Dr).

6.5. Hydrogen bonds

Comparing the number of hydrogen bonds that FF forms with different solvents, is a measure that gives important data about the effect of solvent in the peptide's structures. Table contains the number of hydrogen bonds that are formed between FF-FF and FF-solvent in three different systems; solution of FF in methanol, in water and in tetrahydrofuran.

Concerning the hydrogen bonds between peptide-peptide, the results show that more hydrogen bonds are formed when the solvent is water, after comes THF and at last with less hydrogen bonds in organic solvent. This result comes in agreement with the tendency for self-assembly of FF in water. [30]

As for hydrogen bonds between peptide-solvent, the most are formed between FF-water and the least between FF_{polar}-THF. A possible explanation of the fact that FF forms less hydrogen bonds with THF is that there is only one acceptor, the oxygen of THF. Additionally, it is observed more limited approaching between FF and THF because of the fact that the structure of THF is more extended than the one of water and methanol, and because THF is an almost planar ring. Water and methanol have at least three acceptors/donors, so the possibility to form hydrogen bonds is bigger. Table 12 describes our results for hydrogen bonding.

molecules	<HB>
FF-FF/FF in THF	0.27
FF-FF/FF in water	0.36
FF-FF/FF in methanol	0.20
$(THF-FF/FF)_{List}$ in THF	2.96
$(water-FF/FF)_{List}$ in water	8.32
$(methanol-FF/FF)_{List}$ in methanol	4.41

Table 12. Average number of hydrogen bonds between FF-FF and FF-solvent for FF in THF, water and methanol.

VII. Conclusions and perspectives

Conclusions

The focus of this study was to highlight the effect of the force field on the structural and conformational properties of diphenylalanine peptides in tetrahydrofuran. The choice of an appropriate force field is always an important issue in simulations, because this constitutes a more proper method for exporting accurate results.

We have studied the effect of charge distribution of diphenylalanine in tetrahydrofuran solvent through molecular dynamic simulations. Results prove that polar and non-polar charge distributions of the same peptide differ radically even in the same solvent.

The potential of mean force between two FF peptides shows the attraction for polar charge distribution, which is absent for the non-polar case. Additionally, the calculation of pair radial distribution functions confirms the aggregates of FF polar in THF and the fact that no structures are performed when FF non-polar is solved.

Snapshots and optical observation support the above conclusion. Another clue is the estimation of radius of gyration. The mean size of FF in the solution is found to be slightly larger for the non-polar charge distribution of FF than the polar one, which is also due to the effect of charge.

Concerning the arrangement of FF peptides in THF, it depends again on the charge distribution. It was found antiparallel for polar charge distribution at short distances, because of the electrostatic interactions between the end charged groups and the hydrogen bonding, and more random at the other distances. FF with non-polar charge distribution prefers vertical orientation at short distances and random at longer.

Hydrogen bonding is another issue studied at the current work. FF forms hydrogen bonds with another FF in its solution in THF. The number of hydrogen bonds we found between FF and THF could be due to the fact that only one atom of THF forms hydrogen bond as acceptor.

Comparing the solution of FF in THF with other solvents (i.e. organic and aqueous), we emphasized the differences between specific quantities, which were the potential of mean force (PMF), the radial distribution functions (rdf) between FF-FF and FF-solvent, the radius of gyration, the preference of peptide's orientation and the hydrogen bonding. Our results for THF were more likely with those to methanol solvent when $FF_{non-polar}$ was applied, and with those to water, when FF_{polar} was used. So the charge distribution is the feature we need to examine more in order to understand its effect in the simulation systems.

Perspectives

Mixtures of solvents have been simulated in a variety of percentages of organic and aqueous (50%-50%, 35%-65%, 20%-80% of mixed methanol and water). The charge distribution in these systems is a more complicated issue. Preliminary results with the use of non-polar and polar charge distribution extracted from some test runs and show totally a different behavior.

By the calculation of the radial distribution function and the observation of snapshots during the simulation, peptides tend to self assemble due to the presence of water but the aggregates are spoiled soon because of the organic solvent which dissolves them.

This study is under further investigation and more information should be gathered about setting the charge distribution in a force field according to the solvent.

References

1. Marie E. Oskarsson, Johan F. Paulsson, Sebastian W. Schultz, Martin Ingelsson, Per Westermark and Gunilla T. Westermark. In Vivo Seeding and Cross-Seeding of Localized Amyloidosis; a Molecular Link between Type 2 Diabetes and Alzheimer Disease. *The American Journal of Pathology*, 2015, 185: 834-846.
2. http://www.pebiosciences.com/diphenylalanine-peptide.html
3. Gazit E. and Cherny I. Amyloids: not only pathological agents but also ordered nanomaterials. *Angew. Chem. Int. Ed.,* 2008, 47, 4062.
4. Song Y. J., Challa S. R., Medforth C. J., Qiu Y., Watt R. K., Pena D., Miller J. E., van Swol F.,Shelnutt J. A. Synthesis of peptide-nanotube platinum-nanoparticle composites. *Chem. Commun.* 2004, 7, 1044-1045.
5. Reches M. and Gazit E. Formation of closed-cage nanostructures by self-assembly of aromatic dipeptides. *Nano Let.* 2004, 4, 581-585.
6. Gazit E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J.* 2002, 16, 77-83.
7. Yan X., Zhu P., Li J. Self-assembly and application of diphenylalanine-based nanostructures. *Chem Soc Rev.*, 2010 Jun; 39(6):1877-90.
8. Alexandre Borges, Joao M. M. Cordeiro. Detailing the structure of liquid THF based on an EPSR study. *Computational chemistry*, 2015, 3, 1-7.
9. Ren, Y., Chai, H., Goetz, M. and Kinghorn, A.D. (2013) A Cytotoxic Decahydronaphthalenylpropenal Derivative and Tetrahydrofuran Lignans from the Stems of *Cameraria latifolia*. *Tetrahedron Letters*, **54**, 4854-4858. http://dx.doi.org/10.1016/j.tetlet.2013.06.109
10. Chi, N.-J., Liu, Y. and Aisa, H.A. (2012) Tetrahydrofuran Lignans and Flavonoids from *Artemisia absinthium*. *Che- mistry of Natural Compounds*, **48**, 666-667. http://dx.doi.org/10.1007/s10600-012-0342-x
11. Silva-Filho, A.A., Costa, E.S., Cunha, W.R., Silva, M.L.A., Nanayakkara, N.P.D. and Bastos, J.K. (2008) *In Vitro* Antileishmanial and Antimalarial Activities of Tetrahydrofuran Lignans Isolated from *Nectandra megapotamica* (Laur- aceae). *Phytotherapy Research*, **22**, 1307-1310. http://dx.doi.org/10.1002/ptr.2486
12. Royo, V.A., Santos, F.F., Souza, V.A., Pereira, A.C., Silva, R., Vinhólis, A.H.C., Donate, P.M., Silva, M.L.A., Albu- querque, S. and Bastos, J.K. (2003) Biological Activity Evaluation of Dibenzylbutyrolactone Lignans Derivatives against *Leishmania braziliensis*. *Brazilian Journal of Pharmacognosy*, **13**, 18-21.
13. Souza, V.A., Silva, R., Pereira, A.C., Royo, V.D., Saraiva, J., Montanheiro, M., Souza, G.H.B., Silva, A.A., Grando, M.D., Donate, P.M., Bastos, J.K., Albuquerque, S. and Silva, M.L.A.E. (2005) Trypanocidal Activity of (−)-Cubebin De- rivatives against Free Amastigote forms of *Trypanosoma cruzi*. *Bioorganic & Medicinal Chemistry Letters*, **15**, 303- 307. http://dx.doi.org/10.1016/j.bmcl.2004.10.079
14. Neto, A.G., Silva, A.A., Costa, J.M.L.C., Vinholis, A.H.C., Souza, G.H.B., Cunha, W.R., Silva, M.L.A.E., Albu- querque, S. and Bastos, J.K. (2004) Evaluation of the Trypanocidal and Leishmanicidal *in Vitro* Activity of the Crude Hydroalcoholic Extract of *Pfaffia glomerata* (Amarathanceae) Roots. *Phytomedicine*, **11**, 662-665. http://dx.doi.org/10.1016/j.phymed.2003.06.005
15. https://commons.wikimedia.org/wiki/File:Tetrahydrofuran-3D-balls.png
16. Mary Jane Shultz and Tuan Hoang Vu. Hydrogen bonding between water and THF relevant to clathrate information. *J. Phys. Chem. B*, 2014, 10.1021.
17. Alexandre Borges, Joao M. M. Cordeiro. Detailing the structure of liquid THF based on an EPSR study. *Computational chemistry*, 2015, 3, 1-7.

18. Daniel D. Bowron, John L. Finney and Alan K. Soper. The structure of liquid tetrahydrofuran. *J. Am. Chem. Soc.,* 2006, 128, 5119-5126.
19. B. J. Alder, T. E. Wainright. Studies in Molecular Dynamics I. General Method. Lawrence Radiation Laboratory, University of California, Livermore, California (1959). *J.Chem.Phys* Vol.31, no.2.
20. A. Rahman. Correlations in the Motion of Atoms in Liquid Argon. Phys. Rev. **136**, A405 (1964).
21. Vagelis Harmandaris and Vlasis Mavrantzas. Chapter XX in Book "Simulation Methods for Polymers", Edited by M.J. Kotelyanskii and D.N. Theodorou, Marcel Dekker, New York, 2004**.**
22. Daan Frenkel and Berend Smit. Understanding molecular simulation: From algorithms to applications, *Academic Press,* 2002, ISBN 0-12-267351-4.
23. M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 5.0.4*, www.gromacs.org (2014)
24. https://en.wikipedia.org/wiki/Statistical_ensemble_(mathematical_physics)#/media /File:Increasing_disorder.svg
25. Michael P. Allen, Introduction to molecular dynamics simulation. Computational soft matter: From synthetic polymers to proteins, lecture notes, Norbert Attig, Kurt Binder, Helmut Grubmuller, Kurt Kremer (Eds.), John von Neumann Institute for computing, Julich, NIC Series, Vol. 23, ISBN 3-00-012641-4, pp. 1-28, 2004.
26. NPTEL, Chemical Engineering, Molecular Simulations in Molecular Engineering, Periodic Box and Minimum Image Convention. http://www.nptel.ac.in/courses/103103036/15
27. NVIDIA. http://www.nvidia.com/object/what-is-gpu-computing.html
28. Severine Girard and Florian Muller-Plathe. Molecular dynamics simulations of liquid tetrahydrofuran: On the uniqueness of force fields. 2001.
29. Roland Faller, Heiko Schmitz, Oliver Biermann and Florian Muller-Plathe. Automatic parameterization of force fields for liquids by Simplex optimization. *J. Comput. Chem*., Vol. 20, No. 10, 1009-1017, 1999.
30. Anastassia Rissanou, Evaggelos Georgilis, Emmanouil Kasotakis, Anna Mitraki and Vagelis Harmandaris. Effect of solvent on the self-assembly of dialanine and diphenylalanine peptides. *J. Phys. Chem. B* 2013, 117, 3962-3975.
31. Chris Oostenbrink, Alessandra Villa, Alan E. Mark and Wilfred F. van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem*. 25, 2004, 1656-1676.
32. Benoit Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* 91, 275-282, 1995.
33. Berk Hess, Christian Holm and Nico van der Vegt. Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Phys. Chem.,* 2006*,* 124, 164509.
34. http://matdl.org/matdlwiki/index.php/Image:Rdf_schematic.jpg
35. R. Kumar, J. R. Schmidt and J. L. Skinner. Hydrogen bonding definitions and dynamics in liquid water. *J. Phys. Chem.,*2007, 126, 204107.
36. http://vignette3.wikia.nocookie.net/ellesmere-chemistry/images/8/82/H-bond-water.gif/revision/latest?
37. M. P. Allen and D. J. Tildesley, Computer Simulation of Liquids, Published by Clarendon PressISBN 10: 0198556454.