Study of Self-Assembly of Boc-FF and Fmoc-FF Oligopeptides through Atomistic Molecular Dynamics Simulations

by

Despoina Tzeli

Advisor:
Prof. Vagelis A. Harmandaris

Master thesis committee:
Prof. Vagelis A. Harmandaris
Prof. Anna Mitraki
Prof. Ioannis N. Remediakis

Department of Materials Science & Technology, University of Crete, Greece

Heraklion 2015

Contents

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincerest gratitude to my supervisor Prof. Vagelis Harmandaris for his endless support in all my studies and research, his patience and invaluable assistance and guidance.

Furthermore, I am also very thankful to the other members of the supervisory committee: Prof. Anna A. Mitraki for the precious assistance and support she offered me at all levels and to Prof. I. Remediakis for his valuable comments.

I would also like to thank Dr. Anastassia Rissanou for our stimulating discussions and her significant guidance and her substantial help.

Moreover, I also thank all my colleagues from the Department of Materials Science & Technology and those from Prof. Vagelis Harmandaris's Group.

I would like to express my love and gratitude to my parents Solon and Tania by dedicating this thesis to them, and to thank them for her endless love and for their understanding throughout the duration of my studies.

Abstract

Diphenylalanine (FF) and chemically modified FF-peptides are very common peptides with many potential applications, both biological and technological, due to a large number of different nanostructures which they attain. Experimental observations indicate that the properties of FF peptide can be modulated by N-termini blocking amino acid changes, or conjugation to other chemical moieties. More specifically experimental findings on Boc-FF in a mixture of water and ethanol show a nucleation process in multi-steps, starting from nanospheres, which then undergo ripening and structural conversions to form the final supramolecular assemblies, depending on the concentration ratio of the two solvents. On top of that Fmoc-FF peptides have been observed to form hydrogel under physiological conditions.

The current work concerns a detailed study of the self-assembled structures of Fmoc-FF in an aqueous (H_2O) solution through molecular simulations. In more detail, atomistic Molecular Dynamics (MD) simulations of Fmoc-FF in water have been performed, using an explicit solvent model. The self -assembling propensity of Fmoc-FF in water is obvious. We studied structural properties of Fmoc-FF in water and a comparison with a system of diphenylalanine (FF) in the corresponding solvent was performed. In addition, temperature dependence studies were carried out. The simulation predictions were compared to experimental findings, which have shown that Fmoc-FF peptide forms hydrogel under physiological conditions. Good qualitative agreement between simulation and experimental observations was found. Simulations of Boc-FF peptides are in progress

To my parents

and my brother

1.INTRODUCTION

1.1: Basic knowledge about peptides

In the present work we focus on the study of peptides, which along with proteins, constitute an important family of biomolecules. **Peptides** are biologically occurring short chains of building units called amino acids which are linked by amide bonds, called also peptide bonds. The amide bonds are formed when the carboxyl group of one amino acid reacts with the amino group of another. The shortest peptides are dipeptides, consisting of just two amino acids joined by a single peptide bond, followed by tripeptides (three amino acids), tetrapeptides (four amino acids), etc. All peptides except the cyclic ones have a free N-terminal and a free C-terminal residue at the end of the peptide (as shown in Figure 1). [https://amit1b.wordpress.com/the-molecules-of-life/about/]. Two chemical groups are often used to chemically block free N-termini, the Fmoc group (*N*-fluorenylmethoxycarbonyl) and the tBoc group, (N-(t-butoxycarbonyl)-L-Phe-L-Phe-COOH).

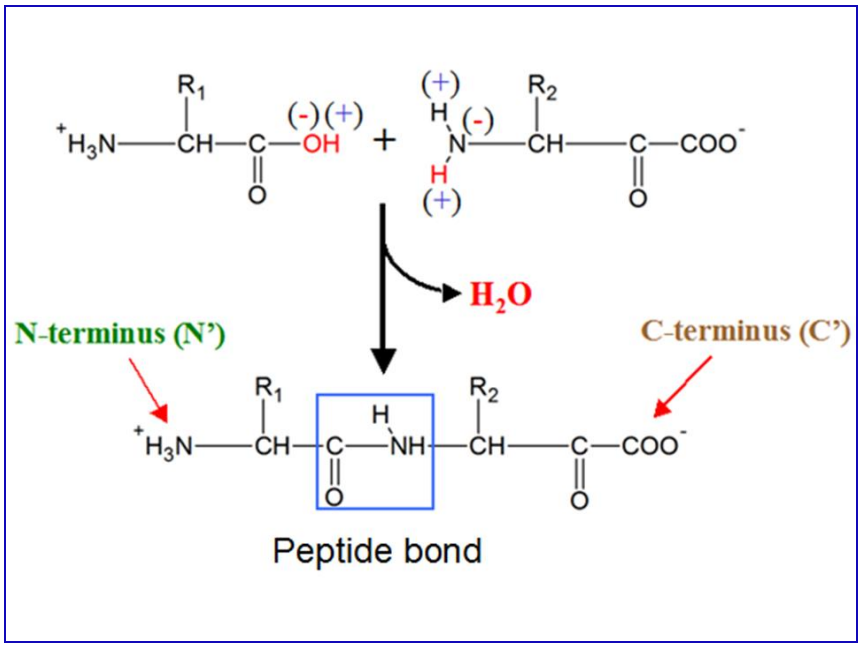

Figure 1. Schematic representation of the formation of a peptide bond. N – terminus, C-terminus and peptide bond are shown. [Obtained from: https://amit1b.wordpress.com/the-molecules-of-life/about/]

1.2: Introduction to Self- Assembly of FF-based Molecules

In the current study, we have focused our interest on diphenylalanine, FF-based molecules. The chemical type of diphenylalanine (FF) is given in Figure 2. Two

types of diphenylalanine peptides with blocked N-ternini, named as (*N*-fluorenylmethoxycarbonyl di-phenylalanine) Fmoc-FF (Figure 3) and (N-(t-butoxycarbonyl)-L-Phe-L-Phe-COOH) Boc-FF (Figure 4) in water are studied and comparisons with FF-water solutions are performed as well. The modification is based on the substitution of one hydrogen atom with the Fmoc group of atoms (circled region of Figure 3) and Boc group of atoms circled region of Figure 4).

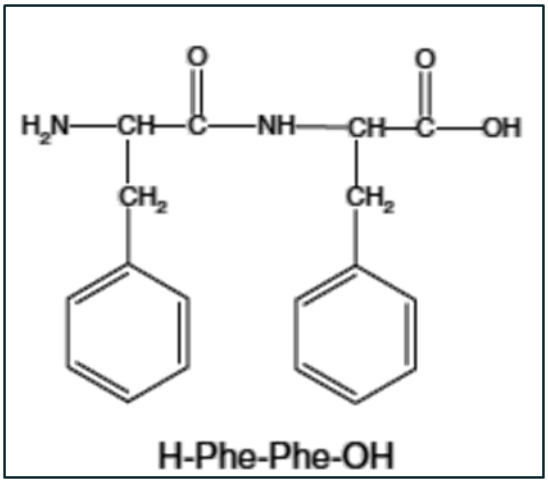

Figure 2. Atomistic structure of diphenylalanine (FF).

One especially intriguing feature which was observed by experimental observations on diphenylalanine peptide (FF) is that the same building block can self-assemble either into fibrillar, or spherical structures depending on conditions such as solvent and temperature.[1-5, 6] Additionally, it has been seen that the properties of FF peptide can be modulated by a simple chemical modification or amino acid changes.[1, 7-10] In more detail, experimental evidence of the polymorphism of these peptides is provided through various studies. An interesting work of Azuri et al.[11] has shown that the diphenylalanine peptide self-assembles to form nanotubular structures of remarkable mechanical, piezolelectrical, electrical and optical properties. Reches and Gazit[16] reported the structure of nanotubes formed by diphenylalanine as the core recognition motif of Alzheimer's β-amyloid polypeptide. The modification of short peptides with the group Fmoc results in a very efficient self-assembly propensity of these building blocks.[12] Another experimental study[13] has shown that the formation of Fmoc-phenylalanine hydrogel is a result of the collective action of different non-covalent interactions. In addition, Fmoc-conjugated alanine-lactic acid (Ala-Lac)

sequence self-assembles into nanostructures that gel in water. [14] Self-assembly in aqueous solution has been found for two Fmoc-tetrapeptides, as well. [15]

Besides experiments simulation studies provide precious insight towards understanding the mechanisms and the driving forces behind the various structures which are formed by these peptides.[17-20] In a combined simulation and experimental work Tamamis et al.[17] has demonstrated for the first time, the formation of planar nanostructures with β-sheet content by the triphehylalanine peptide (FFF). They characterize these structures using various microscopy and spectroscopy techniques. They have also obtained insights into the interactions and structural properties of FF and FFF nanostructures by molecular dynamics simulations of aqueous FF and FFF solutions. In the simulations an implicit solvent model was used and it was observed that the peptides form aggregates, which often contain open or ring-like networks, as well as elementary and network-containing structures with β-sheet characteristics.

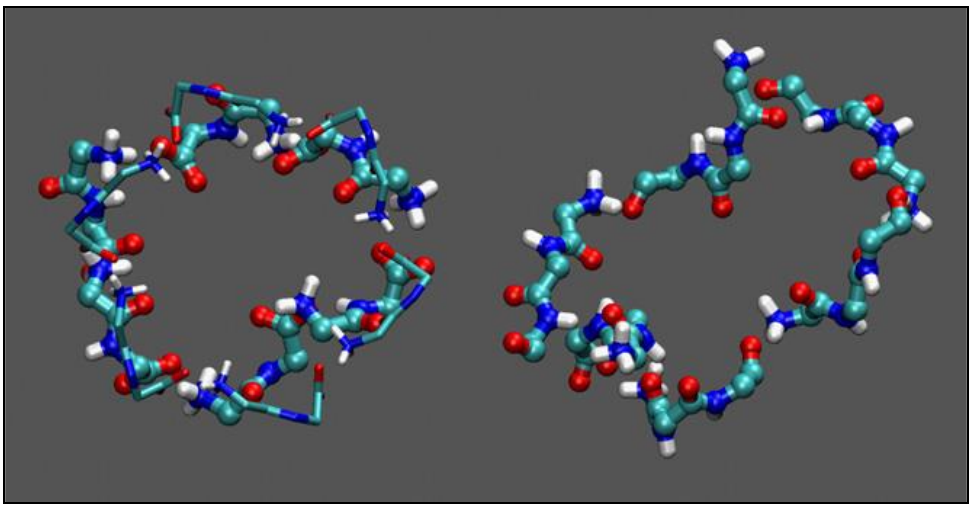

Figure 4 of ref. 17: Representative ring-like networks of six peptides observed in the FF (left) and FFF (right) 300 K simulations. Only the main-chain atoms of the networks are shown in licorice/CPK representation; other network atoms and the surrounding aggregate are omitted. For comparison, the hexagonal ring of the FF crystals (1,36,37) (backbone only, in thin licorice) is also superposed against the simulation ring.

Another combined study by Rissanou et al.[21] revealed the effect of aqueous and organic solvent, on the self-assembly of dialanine and diphenylalanine peptides. Atomistic Molecular Dynamics simulations were performed and an explicit solvent model was used. More specifically, the self-assembling propensity of FF in water was

obvious, while in methanol it was very weak. In addition, a detailed comparison of the behavior of dialanine and diphenylalanine peptide in the two different solvents, water and methanol, was presented. The effect of temperature on the properties of the system in both solvents was studied as well. Experimental data, coming from scanning electron microscopy techniques (SEM), were produced in the framework of that work and were in very good qualitative agreement with the results of the simulations.

Moreover, there are interesting simulation studies[18] in which solvent-free coarse-grained models were designed for dipeptides such that they correctly captured the conformational flexibility of molecules and reproduced the interaction between peptides in aqueous solutions. Furthermore, computational studies have revealed the assembly of Fmoc-dialanine (Fmoc-AA) molecules using molecular dynamics techniques. All simulations converged to a condensed fibril structure in which Fmoc groups stacked mostly in the center of fibril.[22] Another study has investigated pre-assembled aggregates modes of Fmoc-conjugated RGDS and GRDS (amphiphilic Fmoc-tetrapeptides) peptides using atomistic molecular dynamics simulations. The assembly of two peptides was dominated by the interactions among Fmoc units. [23]

1.3: Current Work : Study of Fmoc-FF peptides

In the following we further discuss the specific Fmoc-FF peptide which consists the heart of our work. Experimental observations indicate that the properties of FF peptide can be modulated by a simple chemical modification or amino acid changes or conjugation to other chemical moieties. [1, 7-10] For example, it is known that the usage of aromatic components in conjunction with peptides allows the formation of self-assembled structures with relatively small peptides by taking advantage of π-stacking interactions.[7, 24]

Furthermore, experimental studies[8] have shown that Fmoc-FF molecules form hydrogels under physiological conditions. This example and other closely related aromatic short peptide derivatives are known to form fibrous hydrogels that have found applications in biological sensing[25]and cell culture.[26,27] In ref [8] a number of spectroscopic techniques were applied to Fmoc-FF and a model was constructed based on the obtained data comprising a new nanocylndrical molecular architecture based on π-π interlocked β-sheets. Transmission electron microscopy

(TEM) and wide angle X-ray scattering (WAXS) were used to confirm the proposed model. Various experimental studies have shown than Fmoc-dipeptides were found to self assemble due to the combination of the hydrophobic and π-π interactions of fluorenyl moieties with hydrogen bonding of the peptidic components. Fmoc-FF molecules form transparent, homogeneous hydrogels under a number of different assembly conditions.[28-31]

The present thesis focus on the study of properties of Fmoc-FF (see Figure 3) in aqueous solutions through atomistic Molecular Dynamics simulations, using an explicitly solvent model. Our basic goal is to study the effect of the terminal group Fmoc on the self- assembly behavior of Fmoc-FF peptide and to perform detailed comparisons with FF aqueous solutions. Various structural properties have been studied, as well as the hydrogen bonds network between Fmoc-FF and water molecules. In addition, the temperature dependence of the formed structures is examined.

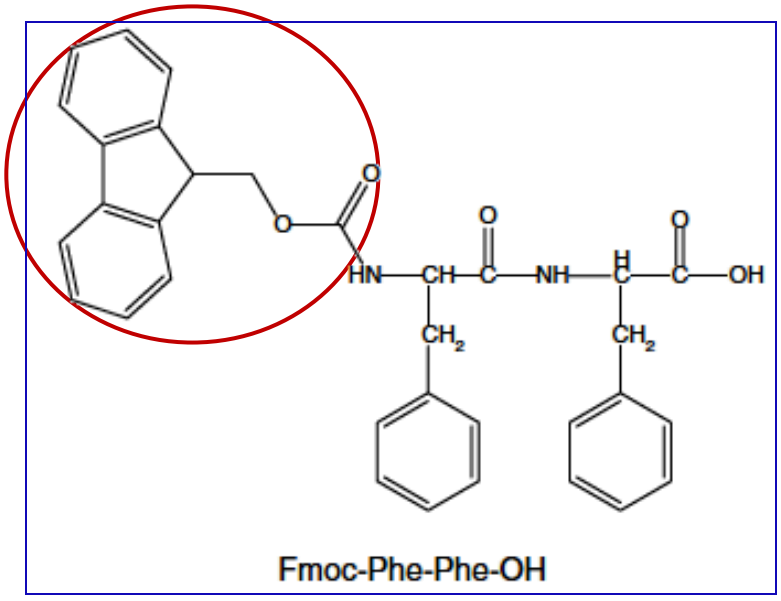

Figure 3. Atomistic structure of FMOC-diphenylalanine, FMOC-FF in water.

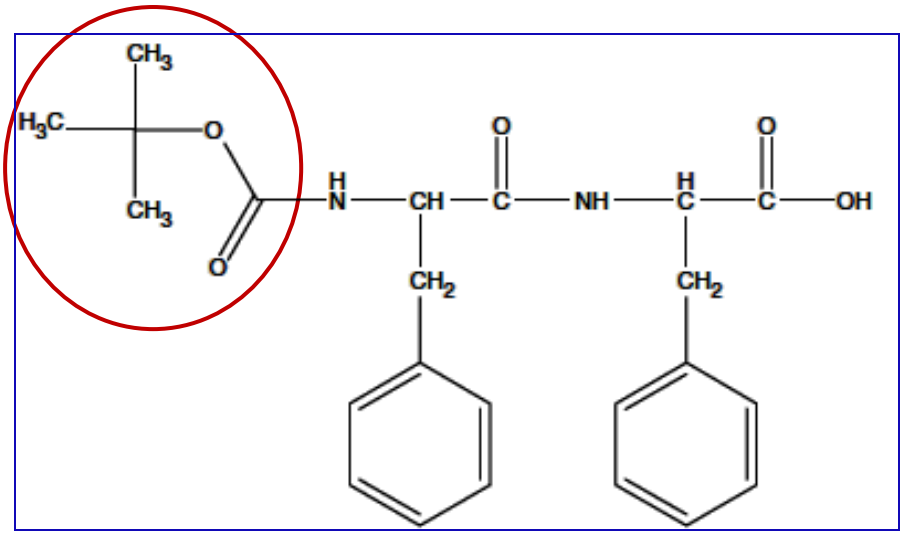

Figure 4. Atomistic structure of Boc-diphenylalanine, Boc-FF in water.

The study of the other chemically modified FF system, called Boc-FF (Figure 4) diphenylalanine, is in progress. Experimental findings on Boc-FF[32] in a mixture of water and ethanol show a nucleation process in multi-steps, starting from nanospheres, which then undergo ripening and structural conversions to form the final supramolecular assemblies, depending on the concentration ratio of the two solvents.

Concluding, we have to underline that the theoretical principles that govern self-assembly and polymorphism of these building blocks are currently unknown. Molecular simulations especially at the atomic level (i.e., atomistic molecular dynamics and explicit solvent models) are valuable tools in order to provide precious insight towards understanding and rational design of new generations of self-assembled nanostructures.

2.MOLECULAR DYNAMICS

2.1 Molecular Simulations

Experiments search for meaningful patterns in nature and theories model these patterns into mathematical language, which provides the predictive laws of nature. Molecular simulation attempts to closely imitate experiments on a real system using model potentials. In other words, molecular simulation uses microscopic properties of a system to calculate macroscopic variables. The following features of molecular simulations can be compared to experiments:

- As in real experiments, we have to prepare equilibrated samples under desired thermodynamic conditions.
- As in real experiments, we can measure the physical properties of the sample.
- Many different algorithms can reproduce the same physical properties within the limit of uncertainties.
- Because the simulator has access to complete information about the state of the model system, there are fewer restrictions on which properties can be measured.

1. Accordingly, information and insight can be gleaned from simulation that is not only easily obtainable by experiments. For example, a comparison of the model's phase behavior can be useful in helping to refine the model parameters.
2. Simulations can be used as test bed for theories.

The distinctive advantages of computer simulations over real experiments are that materials can be studied that are too expensive, too complicated, or too dangerous to be tackled by real experiments. The choice of algorithm is usually determined by factors such as desired thermodynamic conditions, expected thermophysical properties, computational efficiency, reproducibility of the experimental data, minimization of statistical fluctuation and ease of use.[35,38]

2.2 Equilibrium Molecular Simulations

The microscopic state of a system of N atoms enclosed in a fixed volume V is characterized, by a point in $6N$-dimensional phase space with q_i and p_i representing generalized coordinates and momenta, respectively. The temperature T, pressure p

along with *N* and *V* are the fundamental thermodynamic variables and a combination of such variables defines the overall thermodynamic state of that microscopic system. Statistical mechanics provides extracts and manipulates information from the microscopic state of a system and may correlate them with the thermodynamic variables. Molecular simulations calculate macroscopic properties of a system from interatomic or intermolecular details of interactions either from ensemble average (MC) or from time average (MD). The general idea behind MD is that if one allows a system of particles to evolve in time literally infinitely, that system will eventually pass through all possible configurations (atomic system) or conformations (molecular system). The feasibility of applying MD is thus related with accessible time scale which varies with the complexity of the system under consideration. In MC simulations the ensemble average of the measureable physical property is given by.

$$\langle X \rangle_{ensemble} = \int dp^N dq^N X\left(p^N, q^N\right)\rho\left(p^N, q^N\right) \qquad (2)$$

Here the integrand runs over all points in the phase space refers to the physical quantity of interest and $\rho(p^N, q^N)$ refers to the probability density of the ensemble defined by

$$\rho\left(p^N, q^N\right) = \frac{1}{Q} exp\left[\frac{-H\left(p^N, q^N\right)}{k_B T}\right] \qquad (3)$$

Where *H* is the Hamiltonian and *Q* the partition function. In MD simulations the time average of the measurable physical quantity *X* is given by:

$$\langle X \rangle_{time} = \lim\ t \to \infty\ \int_{t=0}^{t=\tau} X\left(p^N(t), q^N(t)\right)dt$$

$$= \frac{1}{M}\sum_{t=1}^{M} X\left(p^N, q^N\right) \qquad (4)$$

Here τ is the simulation time, M is the number of steps in the simulation[39] and

$$X\left(p^N(t), q^N(t)\right) = \text{instantaneous value of X at time t}$$

2.3 Molecular Dynamics Methodology

In molecular dynamics simulations, successive configurations of the system are generated by integrating the Newton's laws of motion. The procedure is to numerically integrate the equations of motion over such small time intervals or time steps that there is not a significant change in the velocity of any molecule during this time interval. During such a simulation, the estimates of structural, thermodynamic, transport, and dynamic properties are stored in each step by taking time averages. It should be noted that ***time averages*** are equivalent to ***ensemble averages*** for ***ergodic systems***. The equation below gives the motion of the particle of mass m_i within coordinate x_i with F_{xi} being the force exerted on the particle in that direction.

$$\frac{d^2x_i}{dt^2} = \frac{F_{x_i}}{m_i} \qquad (5)$$

MD simulations can give us dynamic information that MC simulations cannot. Among the first molecular dynamics calculation a hard sphere model was used i.e. spheres moved at constant velocity along straight lines between collisions (figure 1). The collisions were assumed to be perfectly elastic and it occurred when the sum of the radius of the spheres was equal to the sphere diameter.

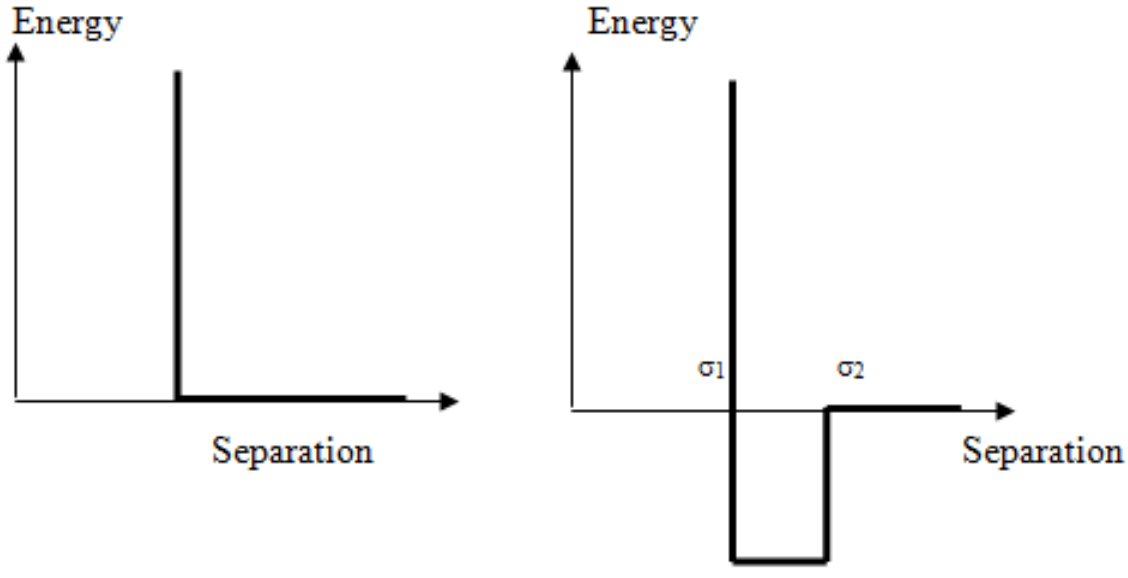

Figure 1: The Hard Sphere and Square well potential.[35]

Another common type known as the square well potential, the interaction energy between two particles is zero beyond a cut-off distance σ_2, infinite below a smaller cut off distance σ_1 and equal to certain value v_o between the two cut offs. The forces f_i acting on the atoms are derived from a potential energy U(**r**), where = (r_1; $r_2 \ldots r_N$) represents the complete set of *3N* atomic coordinates. The force on atom *i* can be computed directly from the derivative of the potential energy *U* with respect to the coordinates:

$$f_i = -\frac{d}{dr_i}\left(U\left(r_N\right)\right) \qquad (6)$$

The *U*(***r***) term represents an externally applied potential also called as *Force Field* [Allen and Tildesley, 1987] which is used for full periodic simulations of bulk systems.

2.4 Force Field

The extent to which a classical molecular simulation[35-40] accurately predicts thermophysical properties depends on the quality of the force field used to model the interactions in the fluid. The force field contains the following building blocks for the calculations of energy and force: (a) A list of atom types. (b) A list of atoms (c) Atom-typing rules. (d) Functional forms for the components of the energy expression. (e) Parameters for the functional terms. The force field[35] with functional form as given below:

$$U_{total} = \sum_{bonds} k_b\left(r-r_0\right)^2 + \sum_{angles} k_\theta\left(\theta-\theta_0\right)^2 + \sum_{dihedrals} k_\chi\left[1+\cos\left(n\phi-\delta\right)\right]$$
$$+ \sum_{impropers} k_\psi\left(\psi-\psi_0\right)^2 + \sum_{i=1}^{N-1}\sum_{j>1}^{N}\left\{\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}}\right\} \qquad (7)$$

The force field describing the molecule employs a combination of bonded interaction terms (first four terms in Eq. 7) such as bond distance r_0, bond angles θ_0, dihedrals δ and impropers ψ_0 (to maintain the planarity of the ring) which describe a part of the total potential U_{total} due to interactions between bonded atoms.

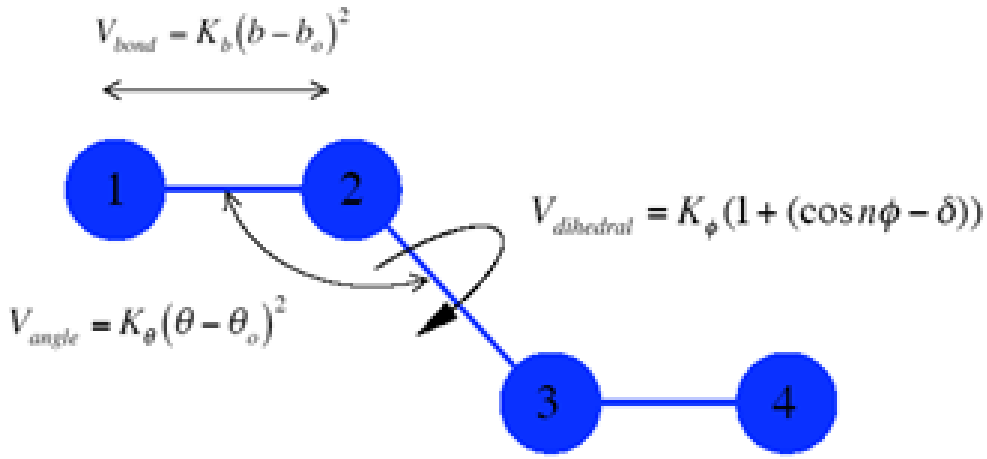

Figure 2: List of Intramolecular Energy terms

Together with the respective force constants k_b, k_θ, k_χ and k_ψ, they are together called as *force field parameters* (see Figure 2). The non-bonded terms include the van der Waals (5th term) better known as *Lennard Jones potential* and electrostatic interactions (6th term) between atoms.

2.5 Integrating Algorithms

For intermolecular interactions[35-40] the force on each particle changes as the position of the particle changes. It can also change if the position of any other particle also is disturbed. A many body problem rises with a continuous potential especially when taking the motions of all the particles together. Such equations of motion are integrated using a finite difference method. The finite difference approach uses continuous potential models which assume pairwise additivity. It basically generates the trajectories of molecules over a period of time. Here the integration on the Newton's law of motion and it broken down into number of small steps each of δt.The vector sum of interactions with other particles gives the total force on each particle at a particular time *t*. We can thus obtain the acceleration from these forces and further with the position and velocity at time *t*, we can calculate the position and velocities at time $t + \delta t$.Here the force is assumed to be constant between each time step. In this way the forces on the new positions are determined which are then again used to determine the positions and velocities at time $t + 2\delta t$.The various available

algorithms for integrating the equations of motion assume that the position and velocities can be approximated by Taylor series[35,39] i.e.

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \ldots\ldots \quad (9)$$

$$v(t+\delta t) = v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \frac{1}{6}\delta t^3 c(t) + \ldots\ldots$$

$$a(t+\delta t) = a(t) + \delta t b(t) + \frac{1}{2}\delta t^2 c(t) + \frac{1}{6}\delta t^3 d(t) + \ldots\ldots$$

$$b(t+\delta t) = b(t) + \delta t c(t) + \frac{1}{2}\delta t^2 d(t) + \frac{1}{6}\delta t^3 e(t) + \ldots\ldots$$

Here *v* is the velocity (derivative of position with time), *a* is the acceleration (the second derivative of position with time) and the *b* the third derivative. The Verlet Algorithm is the most common algorithm used for integrating the equations of motion. It uses the position of the previous step i.e $r(t-\delta t)$ and the positions and accelerations at time to calculate the new positions at time $r(t+\delta t)$ at time $(t+\delta t)$ .Writing out the equations for both the time steps we have:

Adding the two equations give:

$$r(t+\delta t) = 2r(t) - r(t-\delta t)\delta t v(t) + \delta t^2 a(t) \quad (12)$$

From equation 12, it is clear that the velocities are not appearing explicitly (see Figure 3 below). A simple method of obtaining the velocities is to divide or use the central difference formulae i.e. the position of the molecule at $(t-\delta t)$ and $(t+\delta t)$ by $(2t\delta)$. Thus we have

$$v(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} \quad (13)$$

In another alternative method one can use a half step method i.e

$$v\left(t+\frac{1}{2}\delta\right) = \frac{r(t+\delta t) - r(t)}{\delta t} \quad (14)$$

In MD the function determination require computation of the forces acting on each atom. In an MD simulation the force calculation consumes a large amount of the total CPU time (as much as 90%), and it is essential that it be performed no more than once per time step. Sometimes, in the force field calculation intramolecular bond terms are

not required in the potential energy function, because these bonds have very high vibration frequencies. Instead, the bonds are treated as constrained to have fixed length.

Families of force fields, such as AMBER [Cornell et al.,1995], CHARMM [Brooks et al.,1983] and OPLS [Jorgensen et al.,1996] are geared more to larger molecules (proteins, polymers) in condensed phases; their functional form is simpler and their parameters are typically determined by quantum chemical calculations combined with thermophysical and phase coexistence data.

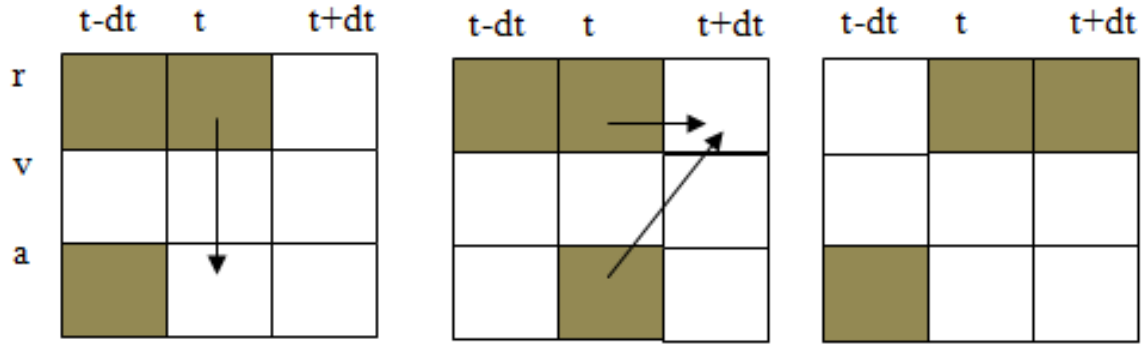

Figure 3: Flow of Verlet Algorithm (Here *t-dt*,*t* and *t+dt* refers to the previous, current and next position)(redrawn from the course notes of "*Molecular Theory and Modeling, 698D,1997*" by E.J.Maginn, Department of Chemical and Biomolecular Engineering, University of Notre Dame, US)

The *Verlet algorithm* has some disadvantages for e.g $\delta t^2 a(t)$ is added to the large terms of $2r(t) - r(t-\delta t)\,\delta tv(t)$.The lack of the velocity term makes it difficult since these are not available unless the position of the next step are computed[Leach,2001]. It is also not a self starting algorithm i.e. the new positions are available from the current step and the previous step. So at initial position i.e t=0 we have only one set of positions, so steps are required to compute the positions at $(t-\delta t)$. The equation 12 can be written as

$$r(-\delta t) = r(t) - \delta tv(0) - \delta tv(0) \qquad (15)$$

Some variations on Verlet algorithm such as *leap frog algorithm* (see figure 4 below) uses the following relations

$$r(t+\delta t) = r(t) + \delta t v\left(t+\frac{\delta t}{2}\right) \quad (16)$$

$$v\left(t+\frac{\delta t}{2}\right) = v\left(t-\frac{\delta t}{2}\right) + \delta t a(t) \quad (17)$$

The velocities $v\left(t+\frac{\delta t}{2}\right)$ are first calculated from the velocities at time $\left(t+\frac{\delta t}{2}\right)$ and the accelerations at time *t*. The positions at $r(t+\delta t)$ are computed using equation 8.

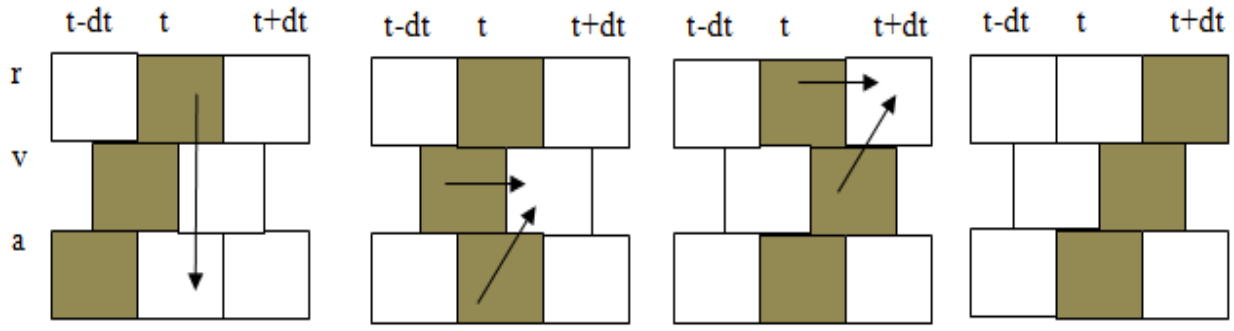

Figure 4: Flow of the Leap Frog Algorithm (here *r*, *v*, *a* refers to the position, velocity and acceleration)

Another variation namely the *velocity Verlet method* computes the positions, velocities and accelerations at the same time and also does not compromise on precision.

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) \quad (18)$$

$$v(t+\delta t) = v(t) + \frac{1}{2}\delta t\left[a(t) + a(t+\delta t)\right] \quad (19)$$

The algorithm is implemented in three steps which are:

1. Positions at (t+δt) are calculated according to equation 18
2. Now the velocities at time (t+δt/2) are computed using

$$v\left(t+\frac{\delta t}{2}\right) = v(t) + \frac{1}{2}\delta t a(t) \quad (20)$$

3. The forces now are computed using the current positions thereby giving α(t+δt). The velocities at time (t+δt) are thus obtained by

$$v(t+\delta t) = v\left(t+\frac{\delta t}{2}\right) + \frac{1}{2}\delta t a(t+\delta t) \qquad (21)$$

2.6 Periodic Box and Minimum Image Convention

For any size of the simulated system, the number of atoms N would be negligible as compared with the number of atoms contained in a macroscopic piece of matter (of the order of 10^{23}). Also the ratio between the number of surface atoms and the total number of atoms would be very large, causing surface effects to be dominant. For this *Periodic Boundary Condition*[35-40] (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.

The application of PBC allows us to simulate equilibrium bulk solid and liquid thermodynamic properties with a manageable number of atoms by eliminating surface effects. 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 neighbouring cell enters to replace the lost particle. As a result of applying PBC the number of interacting pairs increases enormously. This is because of each particle in the simulation box not only interacts with other particles in the box but also with their images. This problem can be handled by choosing a finite range potential within the criteria of *minimum image convention*. The essence of the minimum image criteria is that it allows only the nearest neighbours of particle images to interact. Assuming potential range to be short, a *minimum image convention* is adopted that each atom interacts with the nearest atom or image in the periodic array. In the course of the simulation, if an atom leaves the basic simulation box, attention can be switched to the incoming image. In practice, the mechanism of doing so is to use the potential in a finite range such that the interaction of two distant particles at or beyond a finite length can be neglected. This maximum length must be equal to or less than the half of the box length used in the simulation. A cutoff distance R_c (usually larger than the half the box length) or potential cut off is defined; when a particle is separated by a distance equal or larger

than R_c, two particles do not interact with each other (figure 5 and 6).This helps to avoid expensive force calculation. The time to examine all pair separations is proportional to the number of distinct pairs, ½ N(N-1) in an *N*-atom system. To avoid this a pair listing is made in which the potential cutoff sphere, or radius R_c, around a particular atom is surrounded by a `skin', to give a larger sphere of radius R_{list}. At the first step, a list is constructed of all the neighbors of each atom, for which the pair separation is within R_{list}. Over the next few MD time steps, only pairs appearing in the list are checked in the force routine. From time to time the list is reconstructed: it is important to do this before any unlisted pairs have crossed the safety zone and come within interaction range (Figure 7) .The value of *pairlist distance* should be chosen such that no atom pair moves more than pairlistdist − cutoff in one cycle

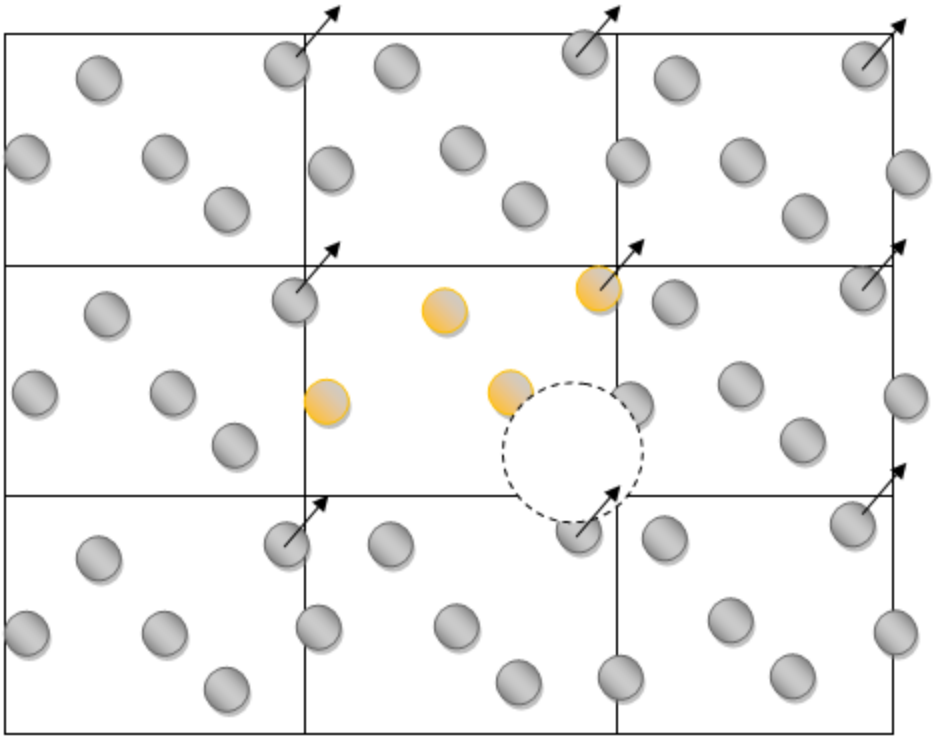

Figure 5: Periodic Box Condition (The dotted line circle indicates the cut off radius (R_c). The yellow colored cell indicates the central cell which is translated in x,y and z direction.

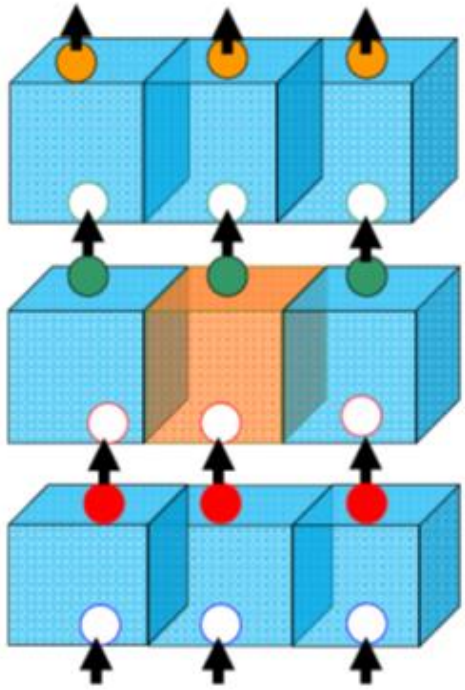

Figure 6 Periodic boundary conditions in a three dimensional view. The orange colour box is the central simulation box. All other boxes are the images of the original simulation box. The particles move in and out as shown with arrows.

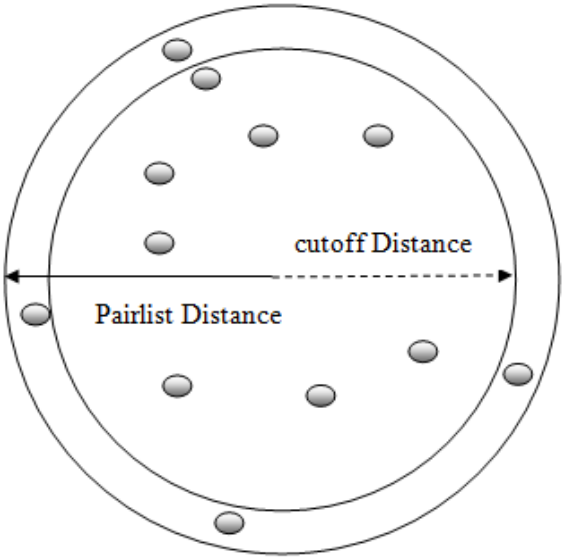

Figure 7: Depiction of the difference between the cutoff and the pair list distance. The oval shaped indicates atoms.

2.7 The Force Calculation

For a model system of N molecules we consider pairwise additive interactions. Considering the interaction between a particle and the nearest image of another particle we have have to compute $N\times(N\text{-}1)/2$ pair distances. So the force calculation is the most expensive in the MD simulation. There exits number of methods that reduces the evaluation of both short and long range forces in such a way that the time scales as N instead of N^2. We however display a simple algorithm for the force calculation. In the first step the distance between each pair of particle i.e *i* and *j* are calculated and stored in xr. A cut-off distance of r_c is chosen in such a manner so that the force computation is neglected for pairs which has distance more than this distance. This cut-off distance is usually half the distance of the periodic box (here denoted by *box*).Thus for the force calculation the distance between the particle *i* and the nearest image *j* should be less than *box*/2. The *nint*(x) rounds the integer to the nearest whole number[35-40] Thus the distance between I and the nearest image j is calculated via

$$xr=xr-box*nint(xr/box)$$

Finally a check is made for all the square of the distances rij2 with the square of distances rc2.The force is then computed by the following expression

$$f_x(r) = -\frac{d}{dx}(U(r))$$

$$f_x(r) = -\left(\frac{x}{r}\right)\frac{d}{dr}(U(r)) \qquad (24)$$

28 Integration of equation of motions

Once all the forces are available and computed we are in a position to integrate the newton's equation of motion[35-40]. A sample code is given below:

```
Subroutine integrate(f,en)
Sumv=0
Sumv2=0
do i=0,npart                          The MD loop starts
        xx=2*x(i)-xm(i)+detla**2*f(i)         The Verlet Algorithm
        vi=(xx-xm(i))/(2*delt)                Velocity Computation
        sumv=sumv+vi                          The velocity center of mass
        sumv2=sumv2_vi**2             Total Kinetic Energy
        xm(i)=x(i)                    Update previous position
        x(i)=xx                               Update current position
enddo
temp=sumv2/(3*npart)                          Compute the Instantaneous temperature
etot=(en+0.5*sumv2)/npart             Calculate the total energy per particle
return
end
```

Figure 11 Force calculation in MD (Adapted from D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications*, 2nd ed., Academic Press, London, 2001.)

Choosing a time step for the integrations is a balance between steps that are too small, so that the simulation to obtain accurate equilibrium thermodynamics averages will take too long; and steps that are too large leads to errors due to numerical integration. Due to this instabilities can arise as a result of an interaction (or collision) where the atom undergoes a large velocity change or even velocity reversal within the time step. Such instabilities could also lead to a violation of energy and linear momentum conservation. When simulating an atomic fluid the time step should be small as compared to the mean time between collisions. While simulating flexible molecules a useful guide is that the time step should be approximately one tenth the time of the shortest period of motion. Time steps of the order of femtoseconds (10^{-15} sec, one-quadrillionth of a second) are typical. The Table below suggests time steps for different kind of motion.

Systems	Types of Motion Present	suggested time step(s)
Atoms	Translational	10^{-14}
Rigid Molecules	Translational and rotation	5×10^{-15}
Flexible molecules, Rigid bonds	Translational ,Rotational, Torsional	2×10^{-15}
Felxible molecules, flexible bonds	Translational,Rotational,Torsional, Vibrational	10^{-15} or 5×10^{-16}

Table 1: Suggested time steps for different motions (Adapted from Leach,2001)

Lecture 15: Long Range Forces

2.9 Non Bonded Interactions van der Waals Interactions

A cut-off always introduces a discontinuity in both the force and potential energy near the cut-off zone. This creates a problem since energy conservation is a must. One of the alternative is to use a shifted potential where a constant term is subtracted from the potential at all points [Leach,2001]. For example

$$v^T(r) = v(r) - v_c \qquad r \le r_c$$
$$v^T(r) = 0 \qquad r > r_c$$

Here v(c) is the potential at cut-off radius. Another alternative is to use a switching function .This is a polynomial function of the distance with which the potential energy function is multiplied:

$$v^T(r) = v(r)S(r)$$

One such function can be written as:

$$v^T(r) = v(r)\left[1 - 2\left(\frac{r}{r_c}\right)^2 + \left(\frac{r}{r_c}\right)^4\right] \qquad (36)$$

It has the values of 1 at r=0 and 0 at r=r_c.

The non bonded van der Waals interactions are always truncated at the *cutoff* distance (R_c).The main option that affects van der Waals interactions is the *switching* parameter.

2.10 Temperature Control

Molecular Dynamics is usually performed at constant NVE ensemble. The results can be transformed into various ensembles which is possible in the limit of infinite system size. Thus it may be desired to conduct the simulation in NPT or NVT ensemble. A constant temperature simulation may be required if we wish to determine the behavior of the system that changes with temperature such as unfolding of a protein or glass formation[2,3,4,5]. The temperature of the system is related to the time average of the kinetic energy which is given by:

$$\langle K \rangle_{NVT} = \frac{3}{2} N k_B T \qquad (38)$$

The easiest way to scale the temperature is to rescale the velocities i.e if the temperature at time t is T(t) then the velocities are multiplied by a factor λ so that the associated change in temperature may be given as:

$$\Delta T = \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i(\lambda v_i)^2}{Nk_B} - \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i(v_i)^2}{Nk_B}$$

$$\Delta T = \left(\lambda^2 - 1\right)T(t)$$

$$\lambda = \sqrt{\frac{T_{new}}{T(t)}} \qquad (39)$$

Another alternative way to maintain the temperature is to couple the system to an external bath that is fixed at desired temperature. The bath acts as a source of thermal energy, removing or supplying heat from the system as appropriate. Here the velocities are scaled at each step in such a way that the rate of change of temperature is proportional to the difference in temperature between bath and system.

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{bath} - T(t)\right) \qquad (40)$$

τ is the coupling parameter whose magnitude defined the how closely the bath and system are coupled with each other.[35] So the change in temperature between successive steps is

$$\Delta T = \frac{\delta t}{\tau}\left[T_{bath} - T(t)\right]$$

.

Thus the scaling factor may be defined as

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left[\frac{T_{bath}}{T(t)} - 1\right] \qquad (41)$$

A large value of τ the coupling is weak, while a smaller value indicates strong coupling .As the value of the coupling constant reaches one time step,the algorithm reduces to a simple velocity scaling method. Based on previous work a coupling constant of 0.4 ps is deemed to be useful for a time step of 1 fs.

2.11 Noose Hover Dynamics

In both NVT and NPT ensemble the temperature needs to be maintained. Nosé dynamics[35,38] is a method for performing constant temperature dynamics. The approach in which an extra `thermal reservoir' variable is inserted into the dynamical equations is used to control temperature (Eq 12). This needs a timestep of 0.5fs with

the Verlet method in order to approach within 3% of the target temperature. The smaller the timestep, the closer it approaches the target temperature.

$$\dot{p}_i = f_i - \xi p_i$$

$$\dot{\xi} = \frac{\sum_{i\alpha} p_i^2 / m_i - NkT}{W} = \nu_T^2 \left[\frac{\sum_{i\alpha} p_i^2 / m_i}{NkT} - 1 \right] = \nu_T^2 \left[\frac{\tau}{T} - 1 \right] \qquad (42)$$

Here ξ is the friction coefficient which is allowed to vary in time. *W* is the thermal inertia parameter which is replaced by ν_T, a decay time for thermal fluctuations. *N* is the number of degrees of freedom. If $\frac{\tau}{T} > 1$, then the system temperature is hot, thus the friction coefficient ξ will increase and vice versa.

2.12 Langevin Dynamics

A molecular system in the real world is unlikely to be present in vacuum. The interaction of molecules causes friction, and the occasional high velocity collision will perturb the system. Langevin dynamics[35-40] attempts to extend MD to allow for these effects. This mimics the viscous effect similar to the procedure given above.

2.13 Pressure Control

The pressure is calculated by the viral theorem of Claussius. It is defined as the expectation value of the sum of products of the coordinates of the particles and the forces acting on them. This is written as:

$$W = \sum_{i=1}^{N} x_i \frac{dp}{dx_i} \qquad (43)$$

Here the term derivative term is the derivative of the momentum (*p*) along the *x* direction of the *i* th particle. According to the theorem this product is equal to -3Nk_BT.For an ideal gas where the interaction occurs only between the gas particles and the container the virial is equal to -3*PV*.However there are forces which occurs between the particles in a real gas and thus the total virial of a real system is equal to the sum of an ideal part i.e. -3*PV* and the interactions between particles:[35]

$$W = -3PV + \sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} \frac{dv(r_{ij})}{dr_{ij}} = -3Nk_BT \qquad (44)$$

The derivation of the real gas can be found from Leach, 2001. Here the term $\frac{dv(r_{ij})}{dr_{ij}}$ represents the force f_{ij} between particle i and j. We thus have the following expression for the pressure:

$$P = \frac{1}{V}\left[Nk_BT - \frac{1}{3}\sum_{i=1}^{N}\sum_{j=i+1}^{N} r_{ij} f_{ij}\right] \quad (45)$$

Constant pressure simulation requires periodic boundary conditions. Pressure is controlled by dynamically[39] adjusting the size of the cell and rescaling all atomic coordinates during the simulation.

$$K_v = \frac{1}{2} W \dot{V}^2 \quad (46)$$

Thus we have:

$$\frac{dP}{dt} = \frac{(P_d - P)}{t_p} \quad (47)$$

Here P_d is the desired pressure and t_p is the time constant for pressure fluctuations and $\beta = 1/kT$. At each step the volume of the cell is scaled by a factor of χ and the molecular centre of mass by a factor $\chi^{0.33}$.Thus we have:

$$\chi = 1 - \beta\frac{\delta t}{t_p}(P_d - P) \quad (48)$$

This method when combined with a method of temperature control creates trajectories in the NPT ensemble. For the NPT ensemble Hoover (1985) gave the following equations for pressure control.

$$\dot{s} = \frac{p}{mV^{1/3}} \quad (49a)$$

$$\dot{p} = f - (\chi + \xi)p \quad (49b)$$

$$\chi = \frac{\dot{V}}{3V} \quad (49c)$$

$$\dot{\chi} = \frac{(P - P_d)V}{t_p^2 kT} \quad (49d)$$

Here we have to specify the desired pressure P_d, time constant t_p, decay time ^{V_T} and instantaneous temperature of the piston. In addition, the damping coefficients $^{\xi}$ and desired temperature of the molecule T are also specified for maintaining the constant temperature. For bulk solvent/fluid the instantaneous pressure of a simulation cell will have mean square fluctuations of $kT/(V\beta)$ where β is the isothermal compressibility, having root mean square (RMS) of 100 bar for 10,000 atom bimolecular system. For more details see Ref. 35 and Lifshitz and Landau, 1985.

2.14 Our model & Simulation Details

System	Name	N-peptide	N-solvent	Total # of atoms	T(K)	c gr/cm^3
1	***Fmoc-FF in Water***	**16**	***16673***	***51011***	***300***	***0.0385***
2	*FF in Water*	*16*	*6840*	*21112*	*300*	*0.0385*

The number of Fmoc-FF peptides, the number of solvent molecules, the total number of atoms in the simulation and the temperature in K are included in Table 1. The concentration is equal to c = 0.0284 gFmoc/cm^3 solvent. The behavior of Fmoc-diphenylalanine (FF) was examined in an aqueous solvent. The results of these study have been compared with those of diphenylalanine (FF), which has been studied in previous work.[21] A direct comparison between the two peptides illustrates the effect of phenyl groups on various properties. The atomistic structures of Fmoc-FF are presented in Figure 1. Atomistic molecular dynamics (MD) simulations in the NPT statistical ensemble were performed using GROMACS code.[33] The pressure was kept constant at P = 1 atm, using a Berendsen barostat, while the stochastic velocity rescaling thermostat[30] was used to maintain the temperature value. All parameters for the description of intermolecular and intramolecular interactions were taken from the GROMOS53a6[33] force field. For the aqueous solutions the SPC model[33] for water was used. An all atom representation was applied except from CH methyl group of molecule's backbone and CH_2 which connects the backbone with the phenyl group,

which have been applied as united atoms. The time step was 0.5 fs and a cutoff of 10 Å for both electrostatic and non-bonded Van der Waals interaction was used. Bond lengths were constrained by means of "LINCS" algorithm. All systems were equilibrated and production runs of 80 ns were performed. The equilibration of the systems has been checked through two typical tests. The observation of the time evolution of the potential energy at different windows of time of the production run, as well as, the calculation of the radial distribution function, from data which correspond to different windows of time of the production run. Both quantities sue for equilibrated systems.

3.RESULTS

3.1 Potential of Mean Force Between Two Peptides

Our simulation results start by studying the interaction between two isolated peptides dissolved in water. This interaction can be quantified by calculating the potential of mean force (PMF) that describes the effective interaction between two molecules in a medium. We calculate the PMF between two Fmoc-FF molecules using the following procedure:

1. First, we kept the distance between the centers of mass (cm) of two molecules constant.
2. Second, we perform long simulations that allow the full sampling of phase space in this configuration.
3. Third, we repeat these simulations for a series of different center of masses (cm –cm) distances.
4. Finally, we calculated the PMF by integrating the mean force from an ensemble of configurations and we corrected it by adding an entropy term because of the cm-cm distance constraint (due to the rotation of the molecules around the cm's), through

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

where $U(r)$ is the PMF as a function of distance r, r_{max} is the maximum distance between the two molecules, beyond which $U(r)$, equals to zero, $F(r)$ is the mean force, and T is the temperature.

The potentials of mean force for Fmoc-FF and FF in water as a function of distance between the centers of mass for the two corresponding molecules are presented in Figure 1. With solid horizontal line, thermal energy (k_BT) is also shown. Starting with Fmoc-FF in water and comparing the PMF with the corresponding one for FF in water we observe that the depth of Fmoc-FF well is slightly larger. However, the basic feature is the drift of the peak at shorter distances which shows bigger attraction between Fmoc-FF peptides rather than FF peptides. Depending on the previous observation, one could expect much more coherent aggregates for the case of Fmoc-

FF peptides. The latter observation is in agreement with experimental results in which Fmoc-FF peptides form hydrogel under physiological conditions.

3.2 Potential of mean force, Pair radial distribution function and inverse radial distriburion function

It is useful to know how the free energy changes as a function of reaction coordinates, such as the distance between two atoms or the torsion angle of a bond in a molecule. When the system is in a solvent, the PMF incorporates solvent effects as well as the intrinsic interaction between the two particles. When the same two particles were brought together in the gas phase, the free energy would simply be the pair potential *u(r)*, which has only a single minimum. But the PMF between two particles in liquid oscillates with maximum and minimum. For a given separation *r*, between the two molecules, the PMF describes an average over all the conformations of the surrounding solvent molecules. Various methods have been proposed for calculating potentials of mean force. The simplest representation of the PMF is to use the separation r, between two particles as the reaction coordinate. The PMF is related to the radial distribution function using the following expression for the Helmholtz free energy:

$$A(r) = -k_B T \ln g(r) - const \quad (1)$$

The constant is chosen so that the most probable distribution corresponds to a free energy of zero. Unfortunately, the PMF may vary by several multiples of over the relevant range of the distance *r*. The algorithmic relationship between the PMF and the radial distribution function means that a relatively small change in the free energy (i.e. a small multiple of $k_B T$) may correspond to *g(r)* changing by an order of magnitude from its most likely value. The standard Monte Carlo or molecular dynamics simulation methods do not adequately sample regions where the radial distribution function differs drastically from the most likely value, leading to inaccurate values for the PMF. Moreover, we present the definition for pair radial distribution function (rdf) in order to comprehend this quantity. Pair radial distribution function $g(r)=g^2(r)$: gives the joint probability to find 2 particles at distance *r*,

$g(r) = \frac{1}{N^2}\sum_{i,j=1}^{N}\langle\delta(|r_{ij}| - r)\rangle$ and it can be calculated in experiments (such as like X-ray diffraction) and simulations.

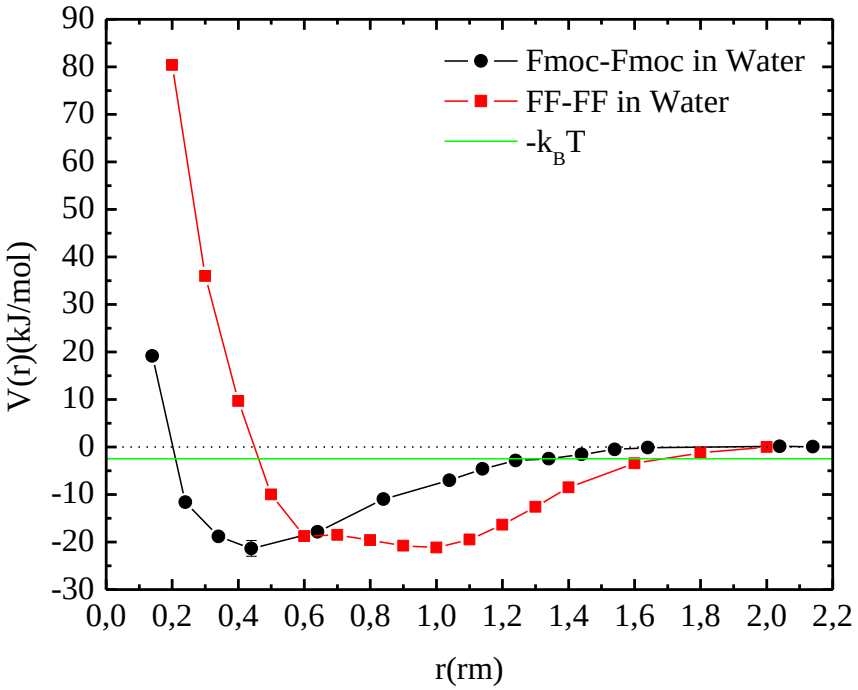

Figure 1. The PMF as a function of distance between the centers of mass of (a) Fmoc in water, (b) FF in water. Solid horizontal line corresponds to $-k_BT$, thermal energy.

In figure 2, we compare the PMF which have shown in figure 1 with the PMF coming from eq (1) which is the inverse *G(r)* of two Fmoc- FF peptides as well inverse *G(r)* of sixteen Fmoc- FF peptides in water. A first observation is that PMF (blue line) doesn't agree with the inverse g(r) of 16 peptides because the second one is extracted from a solution which contains **multiple peptides**, though it is calculated between pairs. On the other hand the PMF describes the effective interaction between **2 isolated peptides**.

Furthermore, we have also performed simulations wiyh a system of 2 peptides in water without constraints in motion, in order to compare the PMF (blue line) with inverse g(r) of two peptides (green line). Basic feature is that the potential minimum occurs at the same distance r = 0.44nm for both PMF (blue line) and inverse g(r) of two peptides (green line). One would expected, the well depth of pmf should be the same as that of the inverse g(r) of two peptides, which is not the case here. Moreover

for the later case the sampling concerns several distances without statistical weight, i.e. cases where the two peptides are at distances larger than the cut-off distance of the intermolecular (LJ) potential. So, it's evident that the calculation of PMF is much faster using the first method for its calculation. For the later case PMF has a concentration dependence whereas this does not happen in the first case.

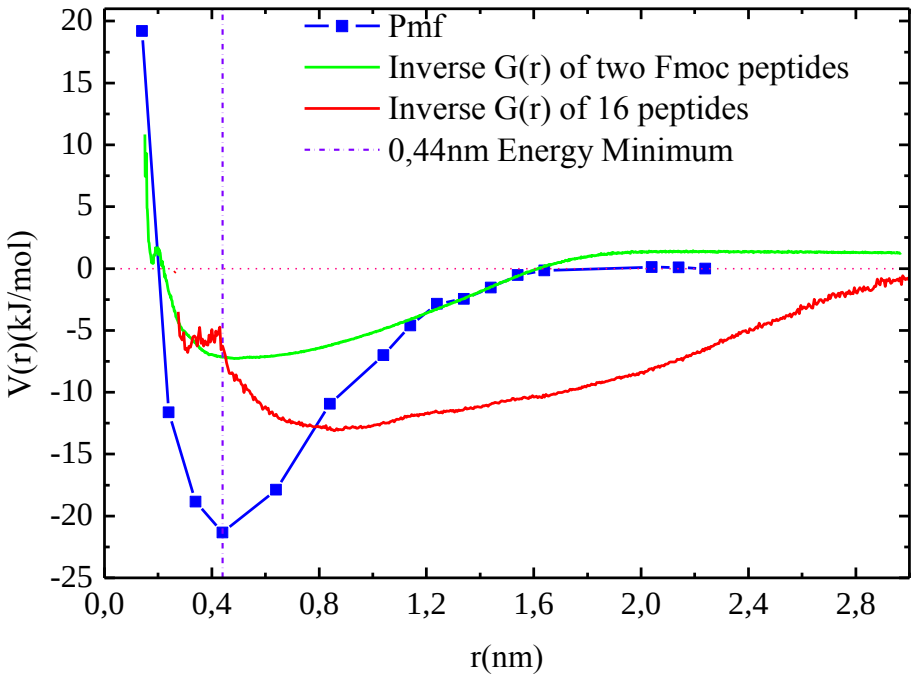

Figure 2. The PMF as a function of distance between the centers of mass of (a) Fmoc in water (blue line), (b) inverse *g(r)* of two Fmoc-FF peptides (green line), (c) inverse *g(r)* of sixteen Fmoc-FF peptides (red line). The vertical purple line is the potential minimum of Fmoc-FF at distance r = 0.44nm.

Structure of peptides in the level of molecule center-of-mass can be studied by calculating the pair radial distribution functions (rdf). A comparison of the rdf curves between Fmoc-FF and FF-FF[10] in water makes it clear that there is a stronger tendency for self-assembly of Fmoc-FF compared to FF-FF. First, we should note that the curve has been normalized, but in large distances, nominator equals to zero. The large peak of rdf in water in addition to the tail of the curve, which tends to zero for large distances, indicates the high probability of Fmoc-FF molecules to be close to one another and to exclude water molecules from their region.

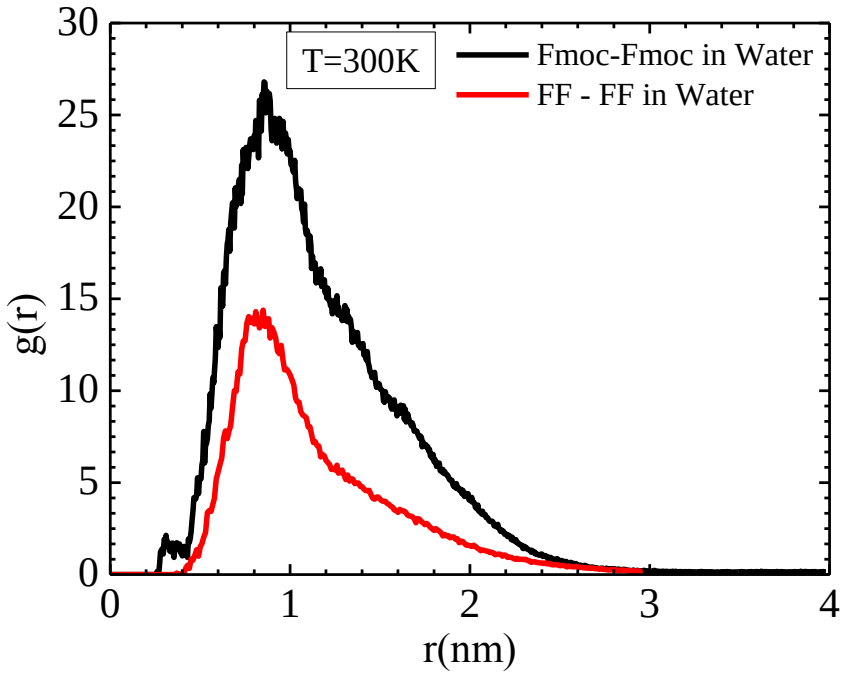

Figure 3. The pair radial distribution function (rdf) calculated for the centers of mass of peptides: Fmoc−FF (black line) and FF−FF (red line) in water at T = 300 K, c = 0.0284 gFmoc/cm^3 and c = 0.0385 gFF/cm^3 solvent.

Another significant feature is that here, we cannot see any difference in the position of the peak analogous to the one of Fig.1 for the minimum of PMF. This can be a result of the averaging over 16 peptides which stands for both cases, Fmoc-FF and FF-FF. The pair radial distribution functions between Fmoc-FF molecules and solvent molecules are presented in Figure 4. This figure provides supplemental information for the arrangement of Fmoc-FF peptides in water, which is consistent to the above discussion. Fmoc-FF molecules exclude water from their vicinity because they prefer to form self-assembled structures and as a result the values of rdf, at short distances, are much lower than one, whereas at higher distances they tend to unity.

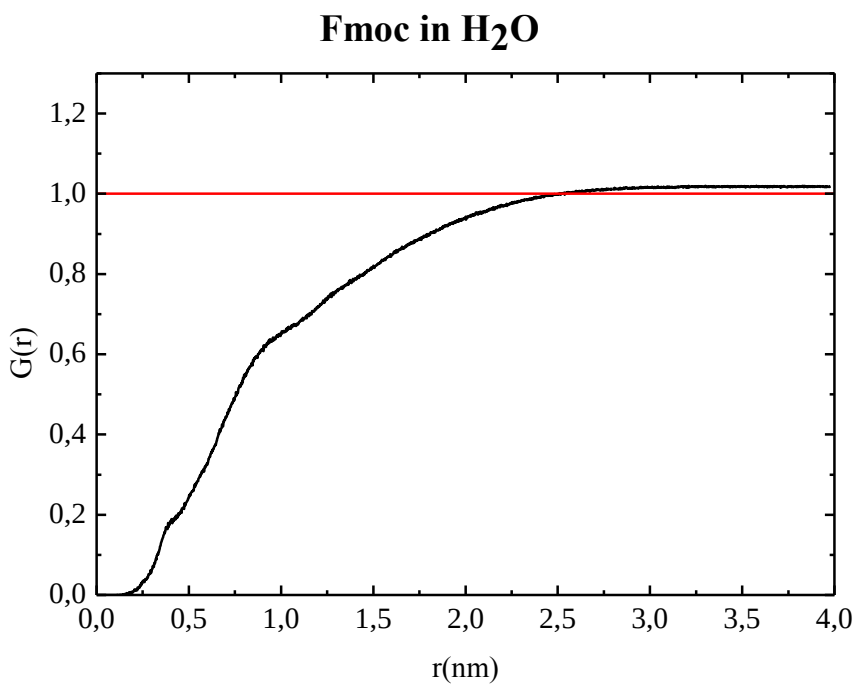

Figure 4. The pair radial distribution function (rdf) calculated for the centers of mass of peptides: Fmoc− H_2O at T = 300 K and 0.0284 gFmoc/cm^3 solvent.

3.3 Basic question: Are the systems thermodynamically equilibrated?

A basic and always questionable issue for a simulation is if we have succeeded thermodynamically equilibrium in order to analyze bulk properties of the Fmoc-FF peptides. For this reason, we checked the time dependence of g(r) at four time windows. It is evident from the following graphs (Figure 5 and snapshots shown in Figure 6) that at the time windows we examined our model system is equilibrated. All curves almost coincide with small differences at short distances which indicate a low probability of two peptides to approach each other very closely. This observation is in compliance with the snapshot we have seen above. The formation of one globule consisted of 16 peptides seem to be stable at least in the time frame of our simulation.

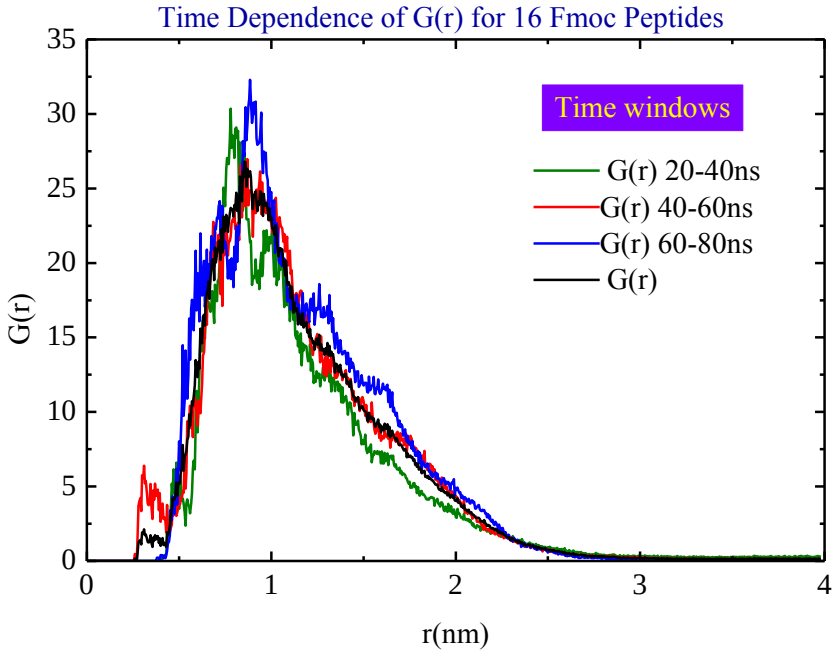

Figure 5. Time dependence of *G(r)* of sixteen Fmoc-FF peptides at four time windows (a) 20-40ns, (b) 40-60ns, (c) 60-80ns, (d) 20-80ns.

3.4 Snapshots

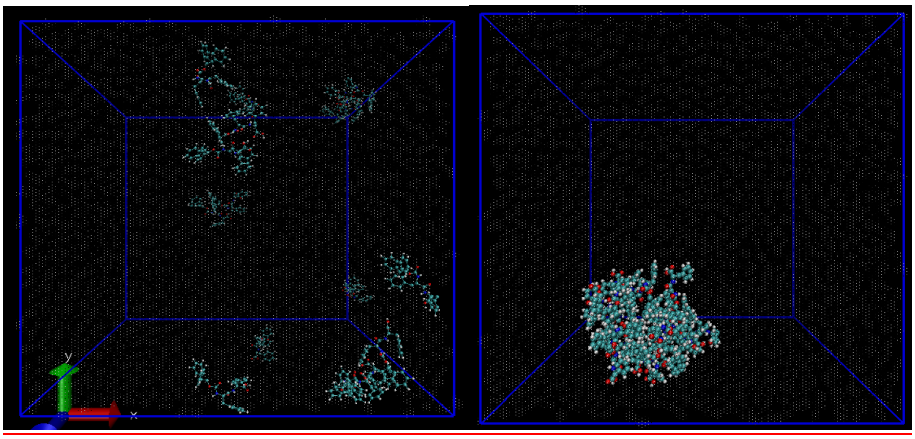

Figure 6. Snapshots of Fmoc-FF peptides during the simulation at T =300K.

3.5 Comformation Properties of Fmoc-FF peptides which are responsible for the behavior of these system.

In the previous section, we have calculated the pair radial distribution in water. This description is based on the molecular level, while a more detailed analysis, in atomic level, can based on the number of hydrogen bonds which are formed in the system.

Hydrogen Bonding

A hydrogen bond is a weak type of force that forms a special type of dipole-dipole attraction, which occurs when a hydrogen atom bonded to a strongly electronegative (donor) atom exists in the vicinity of another electronegative atom with a lone pair of electrons (acceptors i.e. O, C or N) . These bonds are generally stronger than ordinary dipole-dipole and dispersion forces, but weaker than true covalent and ionic bonds.

Intramolecular Hydrogen Bonds: intramolecular hydrogen bonds are those which occur *within* one single molecule. This occurs when two functional groups of a molecule can form hydrogen bonds with each other. In order for this to happen, both a hydrogen donor and an acceptor must be present within one molecule, and they must be within close proximity of each other in the molecule.

Intermolecular Hydrogen Bonds: intermolecular hydrogen bonds occur *between* separate molecules in a substance. They can occur between any *different* molecules as long as hydrogen donors and acceptors are present and in positions in which they can interact.

The most common way to characterize a hydrogen bond is to consider a geometric criterion, involving interatomic distances and angles.[21] In this study we use a standard geometric criterion originally used to investigate hydrogen bond networks in pure methanol solutions.[21] According to this, a hydrogen bond exists if three geometric conditions are satisfied simultaneously: $r(A...B) \leq 3.5$ Å, $r(A...H) \leq 2.6$ Å and angle $(A...B{-}H) \leq 30°$, where A and B are the electronegative atoms (i.e., N and O in our system) and H is the hydrogen. As a test case, we encountered the number of hydrogen bonds which are formed between solvent molecules (i.e., water−water) in the solutions of Fmoc-FF in water and FF in water correspondingly. The values were and correspondingly. In order to estimate the degree of destruction of the network of hydrogen bonds, which Fmoc-FF peptides cause to each solvent, we followed the

following procedure:

1. We encountered the hydrogen bonds, between all Fmoc-FF molecules.
2. We considered a sphere of radius equal to 1nm around the center of mass of each Fmoc-FF molecule and we counted the number of solvent molecules that lie in the region (creation of a list).
3. We counted the number of hydrogen bonds formed between Fmoc-FF peptides and solvent molecules, and the number of hydrogen bonds between solvent molecules of the list taking into account boundaries contribution.

Table 2 contains the corresponding results. The first column is the average number of hydrogen bonds per FF molecule, the average number of hydrogen bonds, which are formed within (intramolecular) and between (intermolecular) FF molecules, respectively. The last column contains the fraction of FF molecules that participate in more than one hydrogen bonds with other FF or solvent molecules in the list. Note that the data are independent of the radius of the sphere, as far as this radius is larger than about $2R_{G.}$ In order to further quantify the role of hydrogen bonds, we calculate the mean number of solvent molecules that are contained in the sphere of 1 nm radius, around one FF molecule; this is 83.13 for water and 52.44 for methanol. Then, the total number of hydrogen bonds around one FF in water is 2.44*54.01 + 8.32 = 211.16 though for 83.13 molecules of pure water this number would be equal to 3.44*54,01 = 285.97, which means a 39.4% decrease of water hydrogen bonds due to the presence of Fmoc–FF, whereas the decrease of hydrogen bonds for FF peptides equals to 26.2%.

Table 2: Average Number of Hydrogen Bonds between FF− FF, FF−Solvent, and Solvent−Solvent Molecules (Water, W, for FF in at c = 0.0385 grFF/cm^3 Solvent and average Number of Hydrogen Bonds between Fmoc− Fmoc, Fmoc−Solvent, and Solvent−Solvent Molecules (Water, W, for Fmoc in water and at c = 0.0284 gFmoc/cm^3)

<HB> FF – FF in H_2O/ FF mol.	<HB> Fmoc-Fmoc in H_2O/Fmoc mol.
0.36	**0.95**
Intra HB FF	Intra HB Fmoc
0.077	**0.40**
Inter HB FF	Inter HB Fmoc
0.278	**0.55**
<HB> (FF-W / FF mol.)$_{list}$	<HB> (Fmoc– W/ Fmoc mol.)$_{list}$
8.32	**1.65**
(W – W / W)$_{list}$	(W – W / W)$_{list}$
2.44	**2.08**

FmocFF-FmocFF Orientation

Another interesting issue is the way that Fmoc-FF molecules are positioned in water in terms of the preferable orientation of one peptide with regard to the orientation of another peptide which has a constant cm-cm distance from the first. For this reason, a number of simulation runs for a pair of Fmoc-FF peptides were performed for a series of different cm-cm constant distances. The preferable orientation of Fmoc-FF molecules is quantified by the dot product of the end to end vectors of the two molecules. The probability distribution of θ-value, $P(\theta)$, at different cm-cm distances, is presented in Figure 7. The main feature here is that peptides at longer distances from the potential well does not have any preferable orientation whereas, at the potential minimum dr = 0.44nm and at dr =0.34nm , the peptides tend to be antiparallel. For the short distances is expected that Fmoc-FF peptides would prefer

normal orientation in order to decrease strong repulsive interactions. Such a behavior was also observed for FF peptides in a previous study.[21]

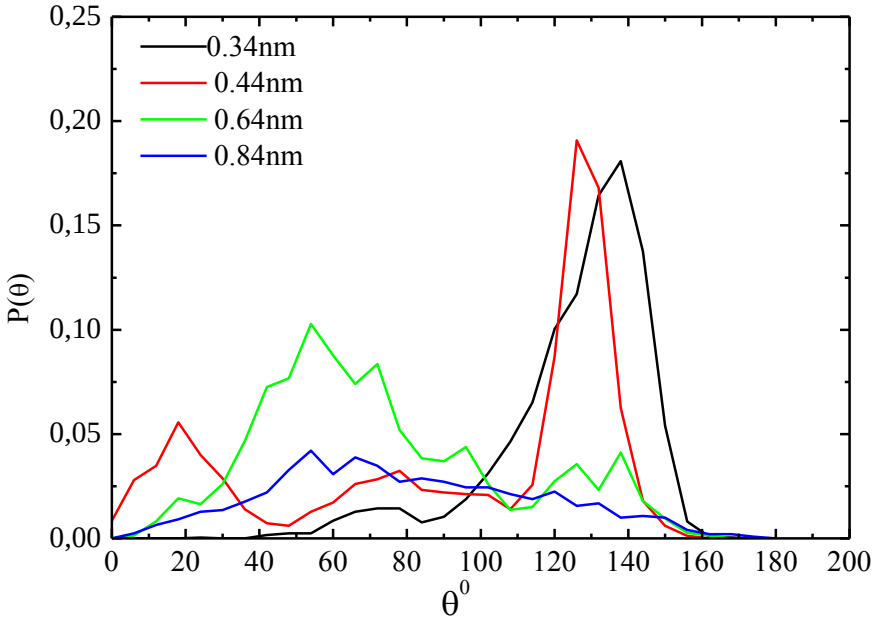

Figure 7. Probable orientations between a pair of Fmoc-FF peptides in water in terms of angles between their end to end vectors, at different cm-cm constant distances dr around the potential well.

3.6 Temperature Dependence of Fmoc-FF peptides.

Another interesting aspect concerns the effect of temperature on the structure of FF peptides in water. For this reason we have performed other three simulations at three different temperatures T = 315K (41.85ºC), T = 330K (56.85ºC), T = 345K (71.85ºC). We have chosen bigger temperatures than room temperature T = 300K in order to estimate the stability of the structures which are formed in water with temperature increasing. Starting with structure, the pair radial distribution function between Fmoc-FF molecules is presented in Figure 8, for four different temperatures. It is evident that there is not any considerable change in the shape of rdf curve for Fmoc-FF among four different temperatures. Strong fluctuations are observed which can be thought as a statistical result. Moreover, the radius of gyration of Fmoc-FF molecules $R_G = \sum_i \sqrt{\left\langle \frac{\sum_i m_i (r_i - R_{cm})^2}{\sum_i m_i} \right\rangle}$ in water in the whole range of temperatures has been

calculated at Figure 9. We found that R_G values remain constant within error bars at all temperatures studied here, which means that the size of Fmoc-FF peptides is independent of temperature.

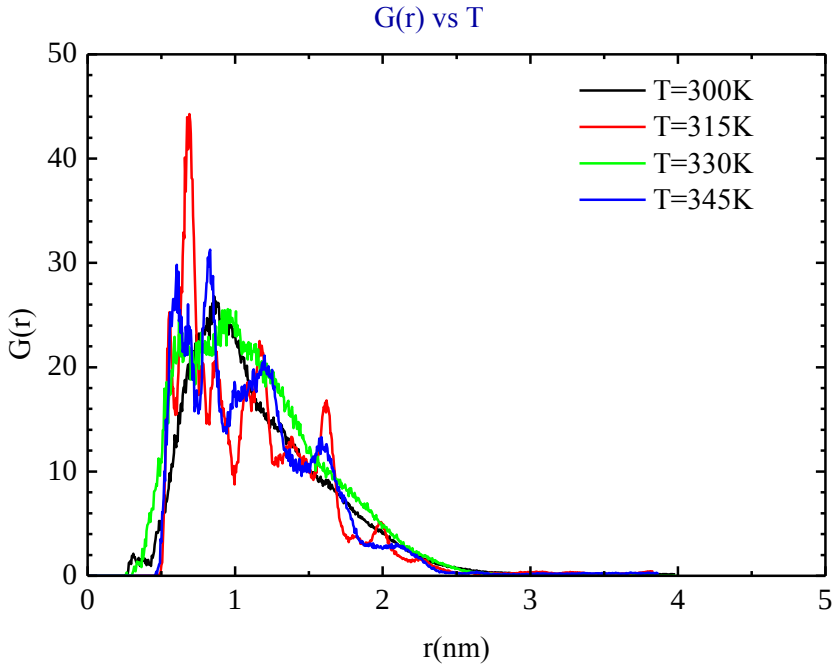

Figure 8. The pair radial distribution function (rdf) calculated for the cm of peptides at c = 0.0284 gFmoc/cm^3 solvent: Fmoc-FF in water, at T = 300, 315, 330, 345 K

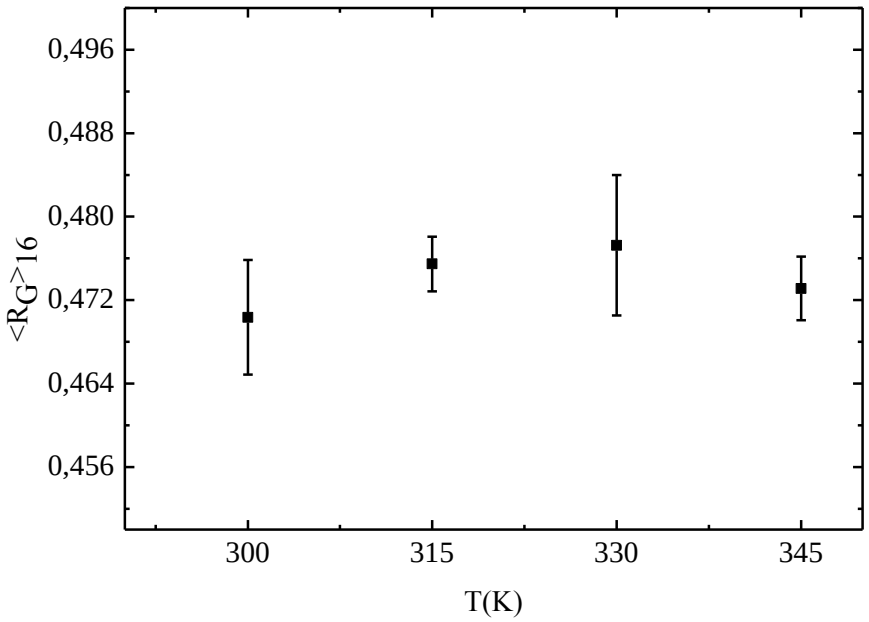

Figure 9. Mean radius of gyration (<R_G>) of Fmoc-FF peptides at four temperatures T = 300, 315, 330, 345K in water.

In addition to the above we observe bigger values than these of FF peptides (R_G = 0.254nm). This is expected because Fmoc-FF peptide is a bigger molecule than FF peptide. Interesting information for the structures, which are formed from Fmoc-FF peptides in water, is the size of the aggregates and the number of peptides which participate in these structures. Both of these quantities have been calculated and the way that they are affected by temperature has been examined. Aggregates are structures that are created and destroyed during the simulation and their size varies (i.e., number of Fmoc-FF that they contain), depending on temperature, because temperature rise induces larger energy fluctuations. In order to characterize an aggregate, we propose a definition of a quantity, similar to radius of gyration of a single molecule. In the present case this quantity can be thought as an effective "radius of gyration", (R^{eff}_g), and is based on the calculation of the center of mass of all Fmoc-FF peptides in our system, taking into account system's periodicity (i.e., minimum image convention). The effective radius of gyration, R^{eff}_g, of a cluster is a measure of the size of all FF molecules, even if no aggregates exist, because it takes into account the position of all Fmoc-FF peptides, either they have self-assembled or not. Using the position of the center of mass we applied the formula for R_g, where, the smaller the value of R^{eff}_g the more Fmoc-FF peptides constitute the aggregate. For a homogeneous system (where aggregates do not exist) a simple calculation leads to a result for the R^{eff}_g equal to the half of the simulation box (i.e., the distribution of R^{eff}_g is a δ-function around the center of the simulation box).

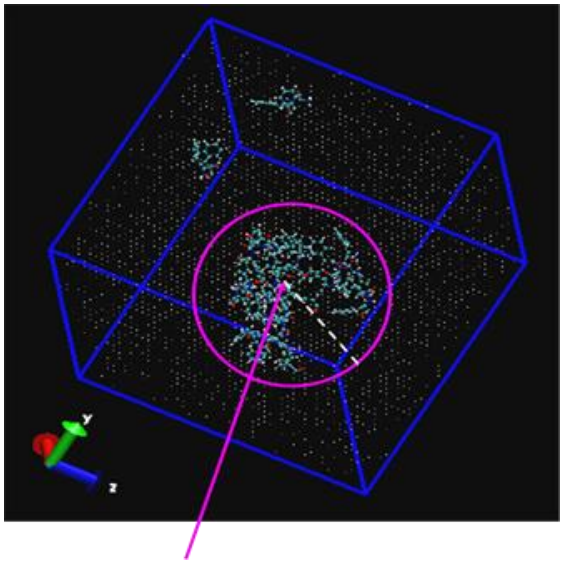

Figure 10. Effective radius of gyration, R^{eff}_g, radius on 2.5nm

The distributions of the R^{eff}_g's values at four different temperatures are presented in Figure 11. Curves of R^{eff}_g for Fmoc-FF peptides in water for all four temperatures does not present an specific trend. All curves have small width and they are centered at a value much lower than the half of the simulation box, 8/2 nm in the Fmoc-FF case.

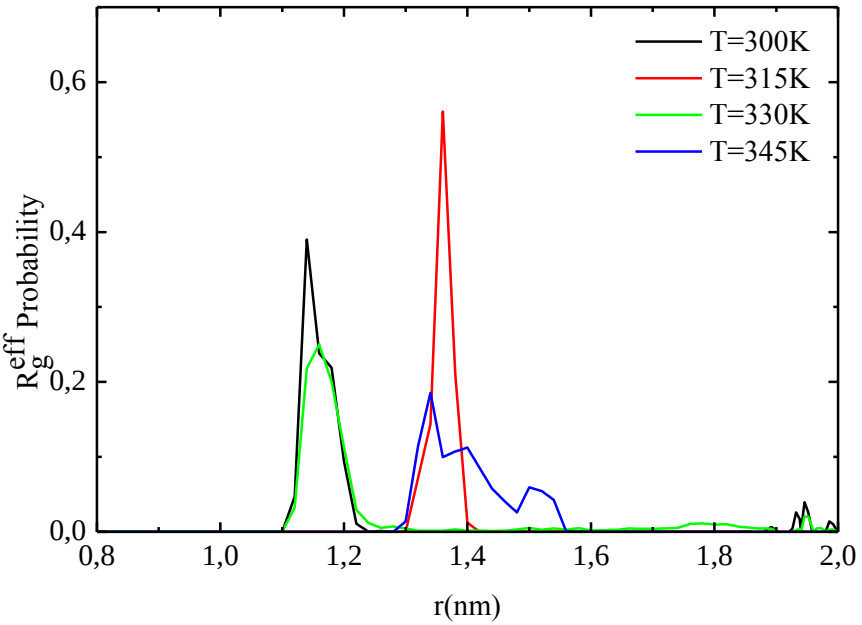

Figure 11. Effective radius of gyration for the system of (a) Fmoc-FF in water at T = 300K, 315K, 330K, 345K and at c = 0.0284 gFmoc/cm^3 solvent.

The above observation is again in agreement with the features of rdf curves (temperature dependence) and corroborates the existence of strong self-assembly of Fmoc-FF in water.

In the next stage we examined the stability of these aggregates as a function of time. Specifically, our goal was to check whether they are dissolved in the solvent during the simulation time. So, Figure 12 represents the time dependence of effective radius of gyration at four time windows. It is obvious that R^{eff}_g has not any significant time dependence during simulation. This result depicts a compact structure, i.e. a cohesive aggregate.

Furthermore, based on the above procedure the number of Fmoc-FF peptides in an aggregate can be calculated. We consider a spherical shell of an arbitrary radius of **2.5** nm around the calculated center of mass of the Fmoc-FF molecules and count the number of peptides whose center of mass lies in this region. Then the average number

over all the configurations is calculated. For a uniformly distributed solution a simple calculation gives that the number of molecules which lie in a sphere of radius (R_c) equal to 2.5 nm is almost 2, (i.e, for simulation box L = 8 nm and total number of Fmoc-FF molecules in the solution equal to 16: $(16/L^3) = [N/((4/3)\pi R^3_c)] \rightarrow N = 2.01$). In Figure 13, the average number of Fmoc-FF peptides in the spherical shell is presented as a function of temperature for Fmoc-FF in water. For all temperatures almost all peptides contribute to the formation of the aggregate (i.e. mean number is always between 15 and 16). This is another indication for the stability of the structure.

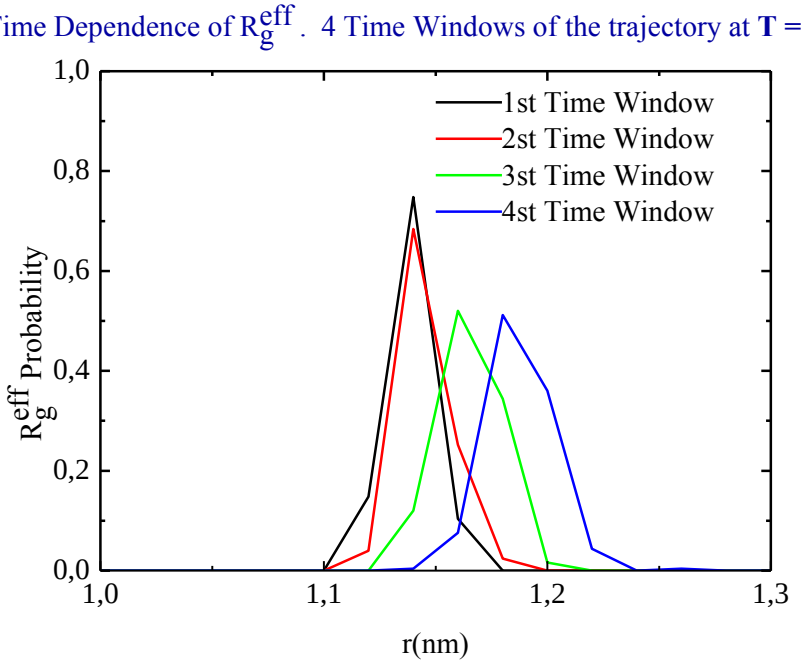

Figure 12. Time dependence of R^{eff}_g at four time windows at T =300K.

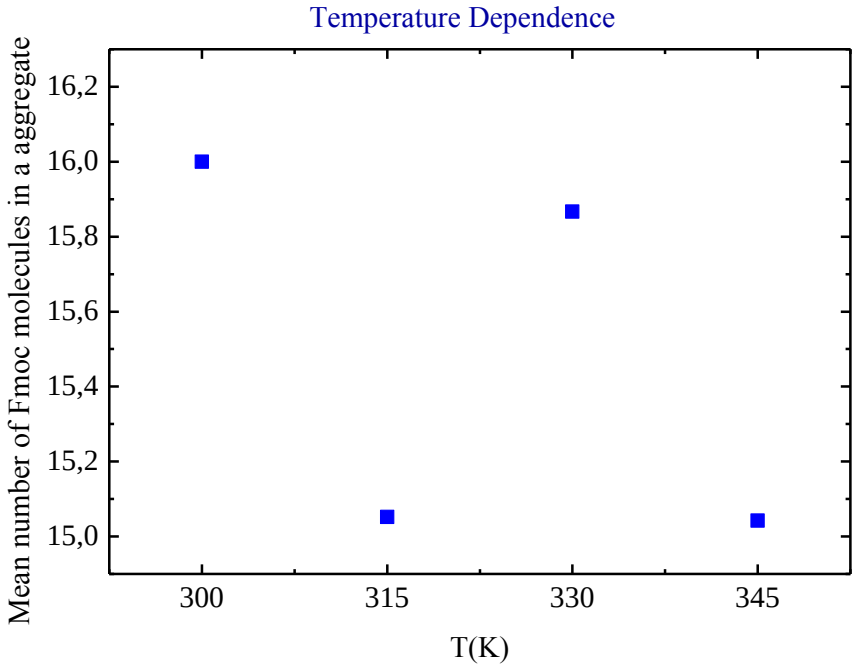

Figure 13. Average number of Fmoc-FF molecules in an aggregate as a function of temperature at c = 0.0284 gFmoc/cm^3 solvent for Fmoc-FF in water.

4.CONCLUSION-DISCUSSIONS

We have studied the self-assembly of Fmoc-diphenylalanine in aqueous solvent through detailed all-atom, molecular dynamics simulations, using an explicit solvent model. Clear evidence for the self-assembly of Fmoc-FF in water was found. The results were compared with these of diphenylanine (FF) which were found in a previous work.[21]

The potential of mean force between two Fmoc-FF molecules (Figure 1) constitutes the first evidence for the self-assembly of Fmoc-FF in water. There is a clear attractive part in the PMF curve for distances between the cm of the two peptides up to 2nm. This finding is further confirmed from the direct calculation of the pair radial distribution functions between Fmoc-FF molecules and between Fmoc-FF−solvent (FF−W) molecules (Figures 3, 4). There is an obvious attraction of Fmoc-FF in water which leads to the formation of aggregates. Furthermore an optical observation of snapshots of FF in water (Figure 6) supports our conclusion.

In addition, the radius of gyration of a single peptide, which is a measure of its mean size in a solution (R_g), is found to be slightly larger for Fmoc-FF than in FF in water. Atomistic simulations provide useful information for many unexplored issues, like the driving force of self-assembly in a specific solvent or the structure and dynamics in the atomic level.

Overall, our simulation work constitutes an extensive study of the above issues for modified FF in water. The arrangement of Fmoc-FF peptides in the self-assembled structures in water, tend to be antiparallel at the potential minimum dr=0.44nm and at dr =0.34nm, whereas for longer intermolecular distances does not have any specific orientation. Furthermore, the number of hydrogen bonds which are formed between Fmoc-FF peptides and FF−solvent molecules can be considered as a measure of self-assembly in atomic level. We have encountered hydrogen bonds and found that the number of hydrogen bonds between Fmoc-FF molecules is higher in water than the same case of FF peptides. Using these numbers and based on a simple calculation for the degree of destruction of the hydrogen bonds network due to the presence of Fmoc-FF, we show that hydrogen bonds constitute the major driving force for self-assembly and it is bigger than these of FF peptide.

The effect of temperature on the aggregate's formation was also explored. Our results indicate temperature independence of structure in the aqueous solution,

something that is in qualitative agreement with the experimental observations, as it was mentioned above as well. Furthermore the size of Fmoc-FF molecules seems to be unaffected by temperature and it is bigger than FF peptide, which is expected because of the bigger size of Fmoc-FF peptide molecule. Finally, the size of aggregates R_G ^{eff} indicates strong self-assembly and no specific temperature dependence for Fmoc-FF. This work is in progress aiming to the observation of the behavior of Boc-FF in aqueous solutions and in mixture of organic and inorganic solvents (i.e. methanol-water). A comparison between the 2 chemical modified versions of FF will highlight the effect of end groups.

Finally, current work concerns the application of the atomistic MD simulations on the Boc-FF molecules and the study of Fmoc-FF molecules at different solvents.

5. REFERENCES

1. *Mahler, A.; Reches, M.; Rechter, M.; Cohen, S.; Gazit, E. , Adv. Mater. 2006, 18, 1365−1370.*

2. *Reches, M.; Gazit, E. Science 2003, 300, 625−627.*

3. *(a) Görbitz, C. H. ,Chem. Commun. 2006, 14, 2332−2334. (b) Henrik, C.; Görbitz, C. H. Chem. Eur. J. 2007, 13, 1022−1031.*

4. *Görbitz, C. H. , Chem. Eur. J. 2001, 7, 5153−5159.*

5. *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. , Chem. Commun. 2004, 7, 1044−1045.*

6. *Raeburn, J.; Mendoza-Cuenca, C.; Cattoz, B. N; Little, M. A; Terry, A. E.; Cardoso, A. Z; Griffithsand, P. C; Adams, D. J, SoftMatter, 2015, 11 , 927–935.*

7. *Charalambidis, G.; Kasotakis, E.; Lazarides, Th.; Mitraki, A.; Coutsolelos, A. G., Chem. Eur. J. 2011, 17, 7213−7219.*

8. *Smith, A. M.; Williams, R. J.; Tang, C.; Coppo, P.; Collins, R. F.; Turner, M. L.; Saiani, A.; Ulijn, R. V. , Adv. Mater. 2008, 20, 37−41.*

9. *Orbach, R.; Adler-Abramovich, L.; Zigerson, S.; Mironi-Harpaz, I.; Seliktar, D.; Gazit, E. Biomacromole-cules 2009, 10, 2646−2651.*

10. *Orbach, R.; Mironi-Harpaz, I.; Adler-Abramovich, L.; Mossou, E.; Mitchell, E. P.; Forsyth, V. T.; Gazit, E.; Seliktar, D. Langmuir 2012, 28, 2015−2022.*

11. *Azuri, I.; Adler-Abramovich, L.; Gazit, E.; Hod, O.; Kronik, L., Am. Chem. Soc., 2014, 136, 963-969.*

12. *Tao, K.; Yoskovitz, E.; Adler-Abramovich, L.; Gazit, E., The Royal Society of Chemistry, 2015, 5, 73914.*

13. *Singh, V.; Snigdha, K.; Singh C.; Sinha, N.; Thakur, A. K., J. Soft Matter, 2015, 00, 1-3.*

14. *Eckes, K. M.; Xiaojia Mu; Ruehle, M. A.; Pengyu Ren, Suggs, L.J., Langmuir, 2014, 30, 5287-5296.*

15. *Castelleto, V.; Moulton, C. M. Hamley, C. G.; Hicks, M. R; Rodger, A.; Lopez-Perez, D. E; Revilla- Lopez, G.; Aleman, C., Soft Matter, 2011, 7, 11405.*

16. *Reches, M.; Gazit, E., Nano Lett. 2004, 4, 581−585.*

17. *Tamamis, P.; Adler.-Abramovich, L.; Reches, M.; Marshall, K.; Sikorski, P.L; Serpell, L.; Cazit, E.; Archontis, G., 2009, 96, 5020-5029.*

18. *Villa, A.; van der Vegt, N. F. A.; Peter, C. Phys. Chem. Chem. Phys. 2009, 11, 2068−2076. Villa, A.; Peter, C.; van der Vegt, N. F. A. , Phys. Chem. Chem. Phys. 2009, 11, 2077−2086.*

19. *Frederix, P. W. J. M; Ulijn, R. V.; Hunt, N. T.; Tuttle, T. Virtual , J. Phys. Chem. Lett. 2011, 2 (19), 2380−2384.*

20. *Guo, C.; Luo, Y.; Zhou, R.; Wie, G., ACS Nano 2012, 6, 3907−3918.*

21. *Rissanou, A .N. ; Georgilis, E. ; Kasotakis, E. ; Mitraki, A.; Harmandaris, V. ,Journal of Physical Chemistry B, 2013, 117(15), 3962-3975.*

22. *Mu, X. ; Eckes, K. M. ; Nguyen, M. M. ; Suggs, L. J. ; Ren, P. , Biomacromolecules, 2012, 13(11), 3562-3571.*

23. *Lopez-Perez, D. E. ; Revilla-Lopez, G. ; Hamley , I.W ; Aleman, C., Soft Matter, 2013, 9, 11021.*

24. *Tamamis, P. ; Kasotakis, E. ; Mitraki, A. Archontis, ; G., J. Chem. Phys. B, 2009, 113, 15639-15647.*

25. *Zhang, Y. ; Gu, H. W. ; Yang, Z. M. ; Xu, B., J. Am. Chem. Soc. 2003, 125(45), 1869.*

26. *Jayawarna, V. ; Ali, M.; Jowitt, T.A. ; Miller, A. E. ; Saiani, A. ; Gough, J. E. ; Ulijn, R. V. , Adv. Mater. 2006, 18(5), 611.*

27. *Mahler, A. ; Reches, M.; Rechter, M. ; Cohen, S. ; Gazit, E., Adv. Mater. 2006, 18(11), 1365.*

28. *Raeburn, J.; Mendoza-Cuenca, C.; Cattoz, B. N; Little, M. A.; Terry, A, E,;, Cardoso, A. Z.; Griffiths, P. Z., Adams, D. J, J. Soft Matter, 2015, 11, 927.*

29. *Tang, C.; Smith, A. M.; Collins, R. F; Ulijn, R. V.; Saiani A., J. Langmuir, 2009, 25(16), 9447-9453.*

30. *Truong, W. T.; Yingying S.; Gloria D.; Braet, F.;Thordarson, P., J. Biomaterials Scence, 2014, J. Biomater., Sci., 2015, 3, 298.*

31. *Tang, C.; Ulijn, R. V.; Saiani, A., Langmuir, 2011, 27, 14438-14449.*

32. *Levin, A; Mason,T. O; Adler-Abramovich, L.; Buell, A. K; Meisl, G.; Galvagnion, C.; Bram, Y; Stratford, S. A; Dobson, C. M; Knowles, T. P. J Gazit, E, Nature commun. 5, 5219 (2014).*

33. *Ostenbrink, C.; Villa, A.; Mark, A. E.; Van Gusteren,W. F, J. Comp. Chem 25, 1656-1676 (2004).*

34. *Kumar, R.; Schmidt, J. R; Skinner, J. L, J. Chem. Phys. 126, 204107 (2007).*

35. *Frenkel, D. and B. Smith, Understanding Molecular Simulations: From Algorithms to Applications. Academic Press, New York, 1996.*

36. *Allen, M.P. and D.J. Tildesley, Computer simulation of liquids. Oxford Science, Oxford, 1987.Doi, M. and S.F. Edwards, The Theory of Polymer Dynamics. Claredon: Oxford, 1986.*

37. *Ciccoti, G. and W.G. Hoover, Molecular Dynamics Simulations of Statistical Mechanics Systems. Proceedings of the 97th Int. Enrico Fermi School of Physics, North Holland, Amsterdam, 1986.*

38. *Harmandaris, V.A., Atomistic Molecular Dynamics Simulations of Polymer Melt Viscoelasticity. Ph.D. Thesis, University of Patras, 2002.*

39. *Harmandaris, V.A., et al., Molecular Dynamics Simulations of Polymers, Chapter in Book Simulation Methods for Polymers. Marcel Dekker, New York, 2004.*

40. *Goldstein, H., Classical Mechanics. Addison-Wesley, MA, 1980.*