Structure and Self-Assembly of Biomolecules through Molecular Dynamics Simulations

M. Arnittali[1,*], A.N. Risssanou[2], V. Harmandaris[1,2]
[1]Department of Mathematics and Applied Mathematics, University of Crete, Heraklion, Greece
[2]Institute of Applied and Computational Mathematics, Foundation for Research and Technology-Hellas, Heraklion, Greece
(*maria.arnittali@gmail.com)

ABSTRACT
In the field of bio-inspired materials, the understanding of physicochemical determinants that underlie peptide self-assembly and protein folding are fundamental steps for the formation of nanostructured, biologically functional materials. Nowadays, molecular dynamics simulations have become a powerful technique to address the latter challenges. Here we present a detailed simulation study concerning the structural and dynamical properties of nanostructured biomolecular systems. The first part of our work concerns the study of a very common peptide diphenylalanine (FF). The results reveal a strong self-assembling propensity of FF in water. The second part of our work concerns the modeling of two proteins in the native state, Rop and RM6 proteins. Our findings show that both proteins have stable native states.

INTRODUCTION
In the field of bio-inspired materials, the non-covalent self-assembly of relatively simple peptide based molecules and proteins have gained increasing attention for the formation of nanostructures, all with nanoscale order[1]. Moreover, polypeptide self-assembly has been often associated with human medical disorders. Understanding the physicochemical determinants that underlie peptide self-assembly and protein folding problem is a fundamental step, in view of the rational design or redesign of already existed nano-building blocks for biotechnological and biomedical applications.
Proteins are complex biological macromolecules and their functionality is obtained when they fold up in their native state[2-6,15]. The building blocks of proteins are the amino acids. Different sequences of amino acids can lead to different structures of proteins. In addition, the stability of a protein is determined by interactions (i.e hydrogen bonding, electrostatics). The observation of the physical and chemical properties of proteins constitute the focus of many experimental and computational studies.

MODELS AND METHODOLOGY
Molecular Dynamics (MD) simulations become nowadays a powerful technique, in particular for understanding the physical basis of the structure of biological molecules[1]. We use all-atom MD simulations to explore conformational and structural properties of peptides and proteins. MD generates a trajectory by numerical integration of classical equations of motion. The produced trajectory contains all information, necessary for our analysis. In addition, we can observe the time evolution of the molecular structure of proteins, their fluctuations and calculate all energetic interactions. Atomistic MD simulations provide information in microscopic level which can be used for the calculation of macroscopic properties.
Our work concerns the modeling of small biological molecules, such as peptides, as well as proteins, where the self-assembly propensity and the conformational properties are studied through all-atom Molecular Dynamics (MD) simulations using an explicit solvent model. Our study is divided in two parts.
The first part of our work concerns the self-assembly of a very common but of particular interest peptide that is diphenylalanine (FF) in aqueous solution[1]. In the second part two proteins in the

native state, are studied under specific (physiological) conditions. The small homodimeric Rop protein, which is a paradigm of a canonical 4-a-helical bundle, and its loopless mutation RM6 are simulated in aqueous solution. The stability of their native state is examined.
The atomistic structure of FF in water is presented in Figure 1. Simulation snapshots (in cartoon representation) of the 3D conformation of the two proteins Rop and RM6 are depicted in Figure 2 (a) and (b) respectively.

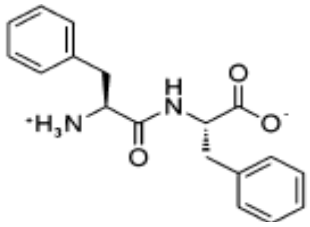

Figure 1. *Chemical type of diphenylalanine in water.*

.

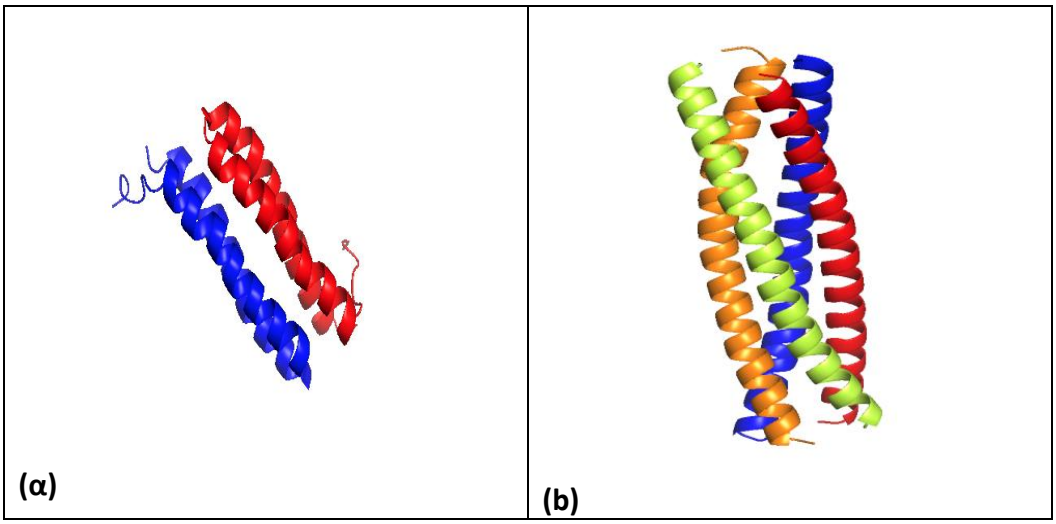

Figure 2. *Cartoon representation of Rop (α) and RM6 (b) proteins.*

Atomistic molecular dynamics (MD) simulations were performed using Gromacs software package[19]. The gromos53a6 force field (ff) is used for the description of all the parameters for intermolecular and intramolecular interactions[17]. An all atom representation was applied except from the non-polar hydrogen atoms of CH/CH2 groups which are treated as united atoms. We used the spc/e water model for the aqueous solution. All the simulations were performed in the isothermal-isobaric (NPT) statistical ensemble. The pressure was controlled constant at P=1 atm using Berendsen barostat, while the temperature was maintained at T=300K using velocity rescaling thermostat.

RESULTS AND DISCUSSION

a) FF in Water

In Figure (3) snapshots of the initial (α) and the final (b) configurations of FF peptides in aqueous solution are presented. In the initial configuration FF peptides are distributed randomly in the simulation box. The self-assembly of FF molecules is observed after ~100ns of the simulation time.

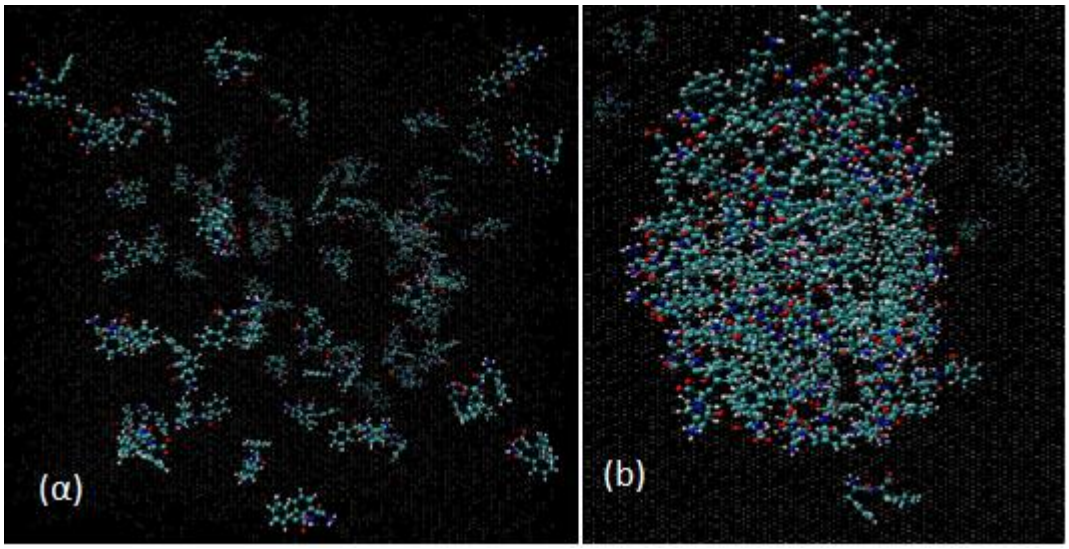

Figure 3. *Snapshot of 74 FF peptides in water solution at 300K (a) initial, (b) final configuration. Water molecules are represented as ghost-molecules for clearer visualization.*

In order to analyze this behavior the interaction between peptides solvated in water is quantified through a potential of mean force (PMF), which describes the effective interaction between two isolated molecules in a medium[1]. We keep the distance between the centers of mass (com) of two FF molecules constant and perform long simulation runs that allow an appropriate sampling of phase space. We repeat simulations for a series of different com-com distances. PMF is calculated by integrating the mean force from an ensemble of configurations and is corrected by adding an entropy term because of the com-com distance constraint, through:

$$U(r) = \int_{r_{max}}^{r} F(r)dr - 2k_B T lnr \quad (2)$$

where, $U(r)$ is the PMF as a function of the distance r, r_{max} is the maximum distance between the com of FF molecules. $F(r)$ is the mean force, k_B is the Boltzmann's constant and finally T is the temperature.. $U(r)$ tends to zero beyond the r_{max}.

Figure (4) shows the PMF of FF in water. We observe that it is repulsive at short distances, an attractive well is formed for com-com distances between 0.5 and 1.4 nm, and it becomes zero at longer distances. This attraction is responsible for the self-assemble propensity of peptides.

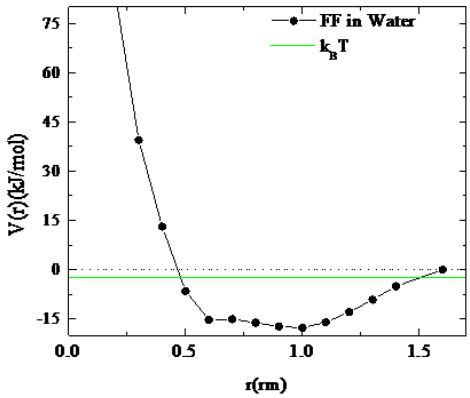

Figure 4. *The potential of mean force (PMF) as a function of distance between the centers of mass of FF in water at 300 K. Solid horizontal lines correspond to the* $k_B T$*, thermal energy.*

The mean size of the FF aggregate in water is qualified by the radius of gyration (Rg), Eq. (3)

$$Rg = \sqrt{\sum_{i=1}^{N} m_i r_i^2 / \sum_{i=1}^{N} m_i} \quad (3)$$

where m_i is the mass of the atom i and r_i is the distance of atoms i from the center of mass of the protein [1-2]. For FF aggregate Rg ~1.9nm.

b) <u>Proteins</u>

We calculate the root mean square deviation (RMSD) as an indicator of the accuracy of our proteins models[2,13]. This could also ensure that our systems are in equilibrium. The RMSD is a measure of the difference between two structures. Each configuration from the trajectory file is compared to the reference structure which is the initial configuration.
The RMSD is defined by the following formula:

$$RMSD = \sqrt{\sum_{i=1}^{N} m_i(\boldsymbol{r}_i - \boldsymbol{r}_{ref})^2 / \sum_{i=1}^{N} m_i} \quad (3)$$

where m_i is the mass of the atom i, $\boldsymbol{r}_i$ is the coordinates of atom i at a certain instance and $\boldsymbol{r}_{ref}$ represents the coordinates of atom i at its reference state.
The smaller the deviation, the more spatially equivalent the two compared structures. Ideally, it should be zero, but measurement errors and other variations cause deviation.
Each configuration from a trajectory file is compared to a reference structure. In our study, the reference conformation is the initial configuration of the simulation. All the measurements of the RMSD are done based only on the Ca atoms (Ca is the backbone carbon before the carbonyl atom in the amino acids). The RMSD of all Ca backbone atoms with respect to the reference structure as a function of time is shown in Figure (5).

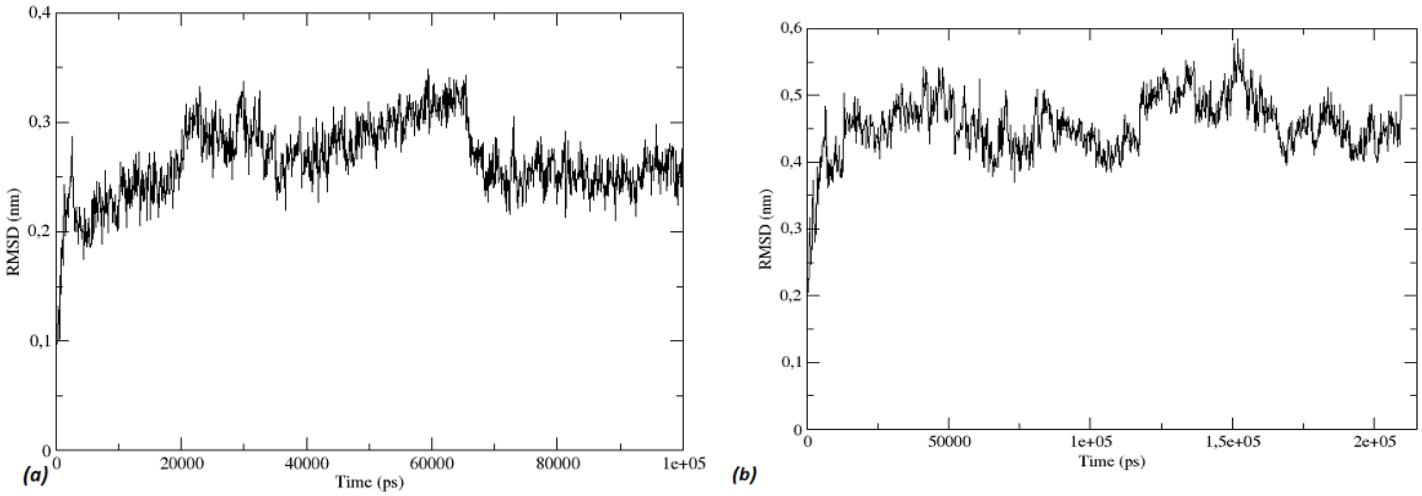

Figure 5. *Root mean square deviation (rmsd) plot showing the rmsd based on Ca atoms as a function of time for the simulations for Rop (a) and RM6 (b) in water.*

In Table 1 is presented the Rg values of the two proteins. We computed the Rg for the whole protein. We observe that the Rg of RM6 is greater than Rop protein. This is expected, due to in Rop protein we have only two chains and the two α-helices are connected with a loop, whereas in RM6 which is a tetramer we have 4 chains where we removed 5 amino acids from the loop and the two α-helices are in an almost linear line.

Table 1.*The average value of Radius of gyration of our proteins/peptides*.

Systems	Measured Quantities	Rg (nm)
Rop in water	Rop	1.51 ±0.004
RM6 in water	RM6	2.520±0.026

The conformation of amino acids in proteins can be described by two torsion angles ϕ and ψ[8-10]. The phase diagram of the two torsion angles is known as the Ramachandran plot. There are two major allowed regions of residue conformation in the Ramachandran plot. These are the α-region and β-region and match to the two major classes of secondary structure: the α-helix and β-sheet respectively. If the majority of residues is in the α-region the generated structure is the α-helix.

In Figure (6) (α),(b), we present the Ramachandran plot for the combinations of (ϕ-ψ) angles for Rop and RM6 proteins respectively. The Ramachandran plots are produced by PROCHECK tool[16]. and validates the backbone structure of proteins. As we can notice in both diagrams there are squares and triangles. All the residues of Rop and RM6 identified by squares with exception Gly residue, which is shown with triangles. Each black symbol represents the conformation of the main chain of one residue of the protein. The darker the color the more favorable the (ϕ-ψ) combination.
As we can see observe the majority of points are clustered in the area which is represented with red (marked with A) color.

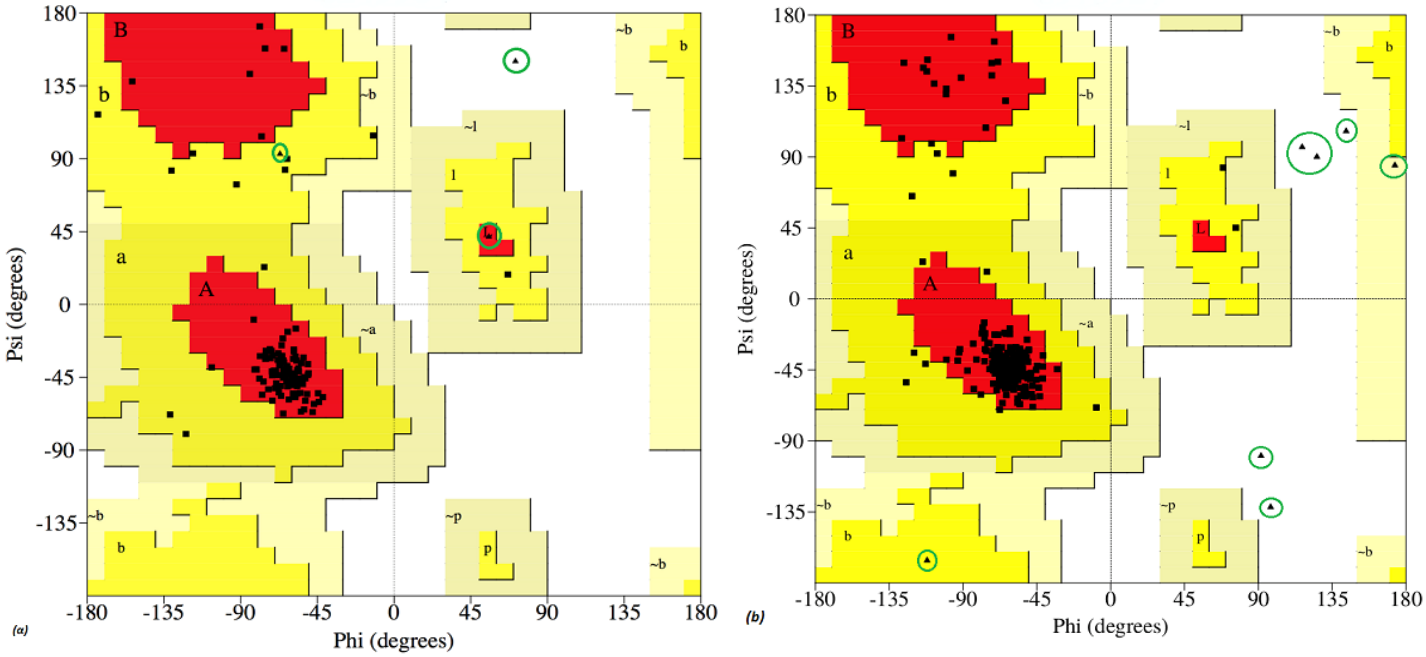

Figure 6. *Ramachandran plot of (φ-ψ) angles of Rop (α) and RM6 (b) proteins.*

This result excuses the fact that both proteins attain α-helices conformations.

CONCLUSIONS

We performed all-atom Molecular Dynamics simulations for two biological systems: a) diphenylalanine peptides in water and b) proteins in aqueous solution.

Self-assembly of FF in water was observed in our simulations. The potential of mean force between FF molecules constitutes the first evidence for attraction between FF peptides. This attraction leads to the formation of an aggregate.

We have also studied the Rop protein and its loopless mutation RM6, in terms of the stability of their native state, in aqueous solution. The analysis of the root mean square deviation and the radius of gyration with respect to time[2] indicate that the two structures are stable. Comparing the mean size of the two proteins (i.e., their radius of gyration) we observe that RM6 protein is bigger than Rop protein. Finally, both proteins are clustered in the region of α-helices according to the Ramachandran plot.

ACKNOWLEDGEMENTS

We acknowledge support of this work by the project "Advanced Research Activities in Biomedical and Agro alimentary Technologies" (MIS 5002469) which is implemented under the "Action for the Strategic Development on the Research and Technological Sector", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Co-financed by Greece and the European Union

BIBLIOGRAPHY

[1] A.N. Rissanou, E. Georgallis, A. Mitraki, V. Harmandaris. J. Phys. Chem. B. 117 (2013) 3965-75.
[2] M. Arnittali. M.Sc. Thesis. Department of Mathematics and Applied Mathematics. UoC (2018).
[3] D.W. Banner, M. Kokkinidis, D. Tsernoglou. J. Mol. Biol. 196 (1987) 657-675.
[4] M. Ambrazi. Phd Thesis, UOC, ITE (2012).
[5] W. Eberle, A. Pastore, C. Sander, P. Rösch. J. Bio. NMR 007. 1 (1991) 71-82
[6] M. Amprazi, D. Kotsifaki, M. Providaki, . Lee. J. Chem. Eng. Data 41 (2014) 497-499.
[7] A.L. Morris, M.W. MacArthur, E.G. Hutchinson, J.M. Thornton. Wiley-Liss, INC. 12 (1992) 345-364.
[8] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan. J. Mol. Biol. 7 (1963) 95-9.
[9] J. Janin, S. Wodak, M. Levitt, B. Maigret. J. Mol. Biol. 125 (1978) 357-386.
[10] R.J. Anderson, Z. Weng, R.K. Campbell, X. Jiang. Proteins, 60, (2005) 679-89.
[11] S. Hovmöller, T. Zhou, T. Ohlson. Acta Cryst. D58 (2002) 768-776.
[12] S.A. Hollinsworth, P.A. Karplus. BioMol. 1 (2010) 271-283.
[13] I. Kufareva, R. Abagyan. Methods Mol. Biol. 857 (2012) 231-257.
[14] M. Yu. Lobanov, N.S. Bogatyreva, O.V. Galzitskaya. J. Mol. Biol. 42 (2008).
[15] A.J. Mulholland. J.R. Soc Interface 5. 3 (2008), 169-172.
[16] http://servicesn.mbi.ucla.edu/PROCHECK
[17] L.D. Schuler, X. Daura, W.F.V. Gunsteren. J. Comp. Chem. 22 (2001) 1205-1218.
[18] http://www.gromacs.org/