Report

Molecular Dynamics

M. Bitar[1,*], M. Offman[1]

[1]Department for Bioinformatics and Computational Biology, TUM, Boltzmannstraße 3 - Garching

ABSTRACT

Motivation: Molecular Dynamics simulations have been used for more than 3 decades to study the molecular motion in the most diverse systems. In Biology, a major use for this computational technique is the examination of protein dynamics on the atomic level. This report will address basic elements of a Molecular Dynamics algorithm and the simulation of biological systems. Additionally, a few examples of the use of the results produced by this approach will be briefly described.

1 INTRODUCTION

The first Molecular Dynamics (MD) simulation of a biological macromolecule was carried out in 1977 by McCammon [1]. It was a 9.2 ps long simulation of the Bovine Pancreatic Trypsin Inhibitor (BPTI), a small, highly stable protein structure that has been widely used as a case-study for new developments in MD [2]. This was the first consistent analysis of a protein molecule as a dynamical system, although the concept of proteins as flexible entities was already present. As Richard Feynman addressed in his famous book *The Feynman Lectures on Physics* more than 10 years before this first MD simulation of a protein, "All things are made of atoms and everything that living things do can be understood in terms of the jigglings and wigglings of atoms" [3]. Following this elucidation, in 1955 Landerstrom-Lang waas able to measure the solvent accessibility of protein residues through his innovative technique called Hydrogen-Deuterium Exchange [4]. This was probably the first experimental evidence of a dynamical protein world as predicted by Feynman.

Around the decade of 1970, biologists were already familiar with protein sequencing techniques and experimental determination of protein structures. The first protein to be sequenced was the bovine insulin, in 1951 by Frederick Sanger [5], while the first protein structures were determined by X-ray crystallography less than 10 years after that [6, 7]. These were the first ingredients for the development of an *in silico* approach for observation of protein dynamics.

The first use of MD simulations was described 1957 through the work entitled *Phase Transitions for a Hard Sphere System*, by B.J. Alder and T. E. Wainwright [8], who were studying the interactions between solid spheres. Subsequently to the application of the new developed technique in this simple system, there was a gap of almost 20 years before its application to a representation of a real system. In 1974, the MD simulation of a liquid water system was performed [9], opening new possibilities to the use of MD in biology. Today, MD simulations serve as a powerful tool for elucidating the atomic-level behavior of proteins [10].

Since McCammon published the results of the first MD simulation of a protein, the computational environment available for such calculations has become more and more unlimited. As mentioned, the first simulation of BPTI was only approximately 10 ps long. As addressed by Karplus and McCammon, during the years that followed this first simulation, a wide range of motional phenomena were investigated through MD simulations of proteins [2]. Last year, the same system (the BPTI protein) was the subject of a 100000000 times longer simulation, that could only be accomplished with the use of a specific-purpose supercomputer [10]. Other long simulations were performed in the past years, with special note to a 10 ms long simulation of a fast-folding protein domain, that needed three months to be finished [11]. Although, in general the average length of a simulation ranges around 10 ns nowadays [2].

To give the reader an idea about which biological processes can be analyzed through currently possible MD simulations, the time frame for some molecular movements are listed in the table below (table 1).

Molecular Movement	Time Range
Local Motions	
Atomic fluctuations	10^{-15} to 10^{-1} s
Sidechain motions	
Loop motions	
Rigid-Body Motions	
Helix motions	
Domain motions	10^{-9} to 1 s
Subunit motions	
Large-Scale Motions	
Helix-Coil transitions	
Association of molecules	10^{-7} to 10^{4} s
Folding and unfolding	

Table 1: Time range for common molecular movements as listed by the Swiss Institute of Bioinformatics [12].

2 METHODOLOGY

Three basic ingredients are required for the programing of a MD simulation algorithm, namely a description of the system, a description of the system's initial conditions and a defined Force Field. To describe the system, usually some features are necessary, for example the identities and coordinates of all atoms that it contains, as well as their mass, charge and radius; a description of the bonds between these atoms; detailed information about the

solvent molecules that take part in the system. Additionally, to set the initial conditions of the system to be simulated, the initial velocities for each atom and the temperature in which the system should be simulated are also needed, as well as a description of the boundary conditions to be placed. The third element mentioned, the force field, is perhaps the most complex of the three and will be described in more detail, although it is not the aim of the present report to deeply access each term that it contains.

The first observation to be made is that biological systems are often too complex to be described in the atomic level by Quantum Physics. This issue is addressed in the manual of an important MD program (Gromacs) in the following statement: “Ideally, the Schrödinger equation describes the properties of molecular systems with high accuracy. Unfortunately, anything more complex than the equilibrium state of a few atoms cannot be handled at this level. Thus, approximations are necessary.” [13]. The approximations to which the authors are referring to include, as mentioned, that Classical Mechanics is employed instead of Quantum Physics, that the movement of the electrons is ignored (and to validate this approximation one can explore the Born-Oppenheimer hypothesis) and that the force fields are approximated, to list a few.

Still in the subject of approximations, related to the fact that the simulation is classical, atoms are represented as solid spheres, with a van der Waals radius, and their mass and charge is considered to be concentrated in their center (point charge and point mass). Every two atoms that are bonded together can be represented as two spheres (as described above) connected by a spring that obeys Hooke's Law. Therefore a protein, as a molecule composed by several atoms, is described as a specific set of spheres and springs.

To analyze the molecular movement of such set, it is of course required to address the position of each atom through time. A simple equation (1) can determine the coordinates of a particle in a specific time step, given the velocity of the particle, its position in a former time step and the time between both moments, as follows.

$$X_1 = X_0 + v\,t \tag{1}$$

Where X_i represents the coordinates of the particle in time step i, v stands for its velocity and t is the time between the steps. As the coordinates of the particle in the former time step are assumed to be known and the length of the time step can be defined, one is left with only one variable to be taken care of. The velocity of the particle. Which can be addressed in the following approach.

$$F = ma = m\frac{dv}{dt} \qquad v = \int \frac{F}{m}\,dt \qquad F = -\frac{dU}{dx} \tag{2}$$

(3)

$$X_1 = X_0 + \frac{\int \frac{-dU}{dx}\,dt}{m}\,t \tag{4}$$

In the set of equations (2), the new introduced variables are m, the mass of the particle, a its acceleration and the force F acting on it. It therefore comes to a problem of defining this force, which is detailed in the equations (3), leading to the need of quantifying the energy of the system (U), as it becomes more clear when equations (1), (2) and (3) are combined in to equation (4).

The Force Field is therefore defined by terms that added together yield the total energy of the system. Several terms can be defined, arising from contributions of different features and interactions, but some elements are basic to the structure of a Force Field and these are represented in the figure below (figure 2).

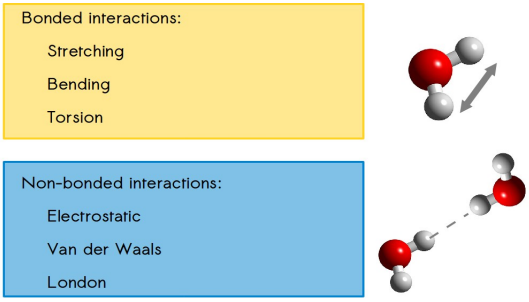


Figure 2: The interactions that define the Force Field are listed and classified in Bonded and Non-bonded, according to the relation between the atoms to be considered. The Hooke's Law is able to model the Bonded interactions, while for the Non-bonded interactions more than one potential is used. The eletrostatic component is represented by the Coulomb's potential and other contributions (such as van der Waals and London forces) can be modeled by the Lennard-Jones potential.

According to the former mentioned Hooke's law and considering two bonded atoms as spheres connected by a spring with force constant k, the following general equation can model the Bonded interactions between atoms.

$$F = -kx \qquad \frac{dU}{dx} = -F = kx \qquad \int dU = \int kx\,dx \qquad U = \frac{1}{2}kx^2 \tag{5}$$

Where x is to be defined. Three different cases can be mentioned concerning Bonded interactions: the stretching, the bending or the torsion of a bond. These will be presented individually.

$$U_{stretching} = \frac{1}{2}\sum_{i=1}^{N_1} k_i^s (l_i - l_i^0)^2 \qquad U_{bending} = \frac{1}{2}\sum_{i=1}^{N_2} k_{ij}^b (\theta_{ij} - \theta_{ij}^0)^2 \tag{6}$$

The set of equations (6) describes the stretching (the variation on bond length) and the bending (the variation on a three-atom bonds angle). For the stretching, x is defined as the difference in bond length between two time steps, while for bending, this variable represents the difference in the angle defined by two bonds

between the time steps. These features are illustrated in the figure that follows (figure 3).

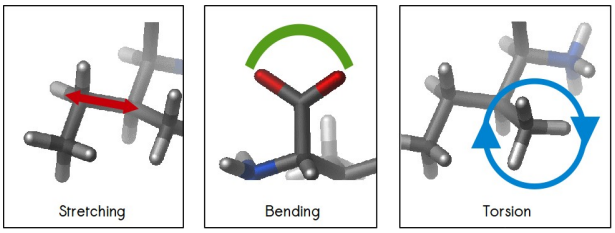


Figure 3: Representation of the three features of Bonded interactions.

The torsion potential is modeled by equation (7) and the specific variables are listed. Together, the terms represented in (6) and (7) are the first part of the Force Field, addressing all the basic Bonded interactions.

$$U_{torsion} = \frac{1}{2}\sum_{i=1}^{N_3} V_n[1+\cos(n\omega-\gamma)] \tag{7}$$

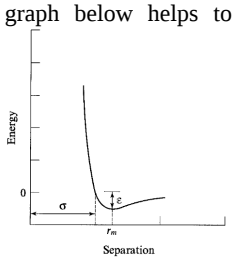
ω – torsion angle
γ – phase factor
V_n – barrier

The Non-bonded interactions, on the other hand, are usually modeled by the Coulomb and Lennard-Jones potentials, showed in (8) and (9). While the Coulomb potential takes into account the charged atoms (with charge q and seprated by distance r) and the related electrostatic interactions, the Lennard-Jones potential models induced interactions, such as van der Waals.

$$U_{Coulomb} = \sum_{i,j=1}^{N_4} \frac{q_i\, q_j}{r_{ij}} \tag{8}$$

$$U_{Lennard-Jones} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{9}$$

For the Lennard-Jones potential, the graph below helps to understand where the variables come from.



Together, all terms discussed here give rise to the general equation of the Force Field (10) and (11). Although this terms can be differently modeled and new terms can be described and incorporated to this basic formulation, it is a good idea of how the equation looks like and where the terms come from.

$$U = U_{bending} + U_{stretching} + U_{torsion} + U_{Coulomb} + U_{Lennard-Jones} \tag{10}$$

$$U = \frac{1}{2}\sum_{i=1}^{N_1} k_i^s (l_i - l_i^0)^2 + \frac{1}{2}\sum_{i=1}^{N_2} k_{ij}^b (\theta_{ij} - \theta_{ij}^0)^2 + \frac{1}{2}\sum_{i=1}^{N_3} V_n[1+\cos(n\omega-\gamma)] + \sum_{i,j=1}^{N_4} \frac{q_i\, q_j}{r_{ij}} + 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{11}$$

This brief explanation concludes the description of the basic elements required for the construction of an algorithm to conduct MD simulations. Below there is a short workflow (figure 4) describing the main steps of a general-purpose MD protocol. All steps can be adapted to better fulfill the needs of a specific analysis and additional steps can be incorporated to drive the simulation in a specific exploration of the system.

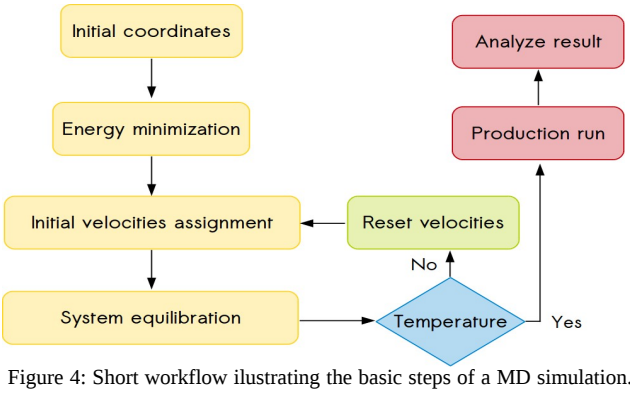


Figure 4: Short workflow ilustrating the basic steps of a MD simulation. Initial conditions and temperature value are supplied by the user and the system is simulated according to those and the Force Field of choice. Once the temperature is constant and the system is in equilibrium, the production part (the actual MD simulation) takes place and a trajectory is produced.

Since the first introduction of MD simulations, several programs have been developed to perform such task. Nowadays, one can choose between a diverse set of different softwares, with a variety of unique features. To name the most widely used programs, we can cite: CHARMM (Chemistry at HARvard Macromolecular Mechanics) which was the first to be developed, in 1982 [14], Gromacs (GROningen MAchine for Chemical Simulations) which is currently the most widely used [15], NAMD (Not (just) Another Molecular Dynamics program) [16] and Amber (Assisted Model Building with Energy Refinement) [17].

3 APPLICATIONS

In this session, some applications of the MD technique will be shortly presented. The reader should keep in mind that these examples do not cover the extremely wide range of applications for MD simulations. This is probably a good moment to mention as well that MD simulations in Biology are not restricted to protein motion analysis. In principle, all biological molecules can be explored through MD simulations and the systems to be studied by this approach are as diverse as one can imagine. Unfortunately, the computational environment available is not always suitable for simulations of complex systems or for long simulations of rather simple systems. Common applications in the protein world are concerned with accessing structural stability, conformational

changes, folding, molecular recognition, biological ion transport, among others.

Here, we will highlight three different cases of application of MD simulations to Biological systems.

1) Molecular Dynamics Study of Aquaporin-1 Water Channel in a Lipid Bilayer [18]

In this work, from 2001, the authors modeled the water channel protein aquaporin-1 in a lipid bilayer using MD simulations. The analysis included the interactions between the protein and the membrane, as well as the interactions between monomers within the protein. Structural features of the channel that are important for its biological function were also investigated, including the diffusion of water molecules through the channel. The simulations revealed a single file of water molecules being formed inside the channel.

The authors observe that the protein exhibited significant fluctuations during the simulations, specially by the side chains. This could be due to low accuracy in the crystal structure for side chain structures, although the overall folding of the protein is properly described. Several amino acids interact with lipids from the bilayer, part of those forming hydrogen bonds.

After 1 ns of simulation of an initially 'empty' channel (absent of water molecules in its interior), several water molecules were already placed along the protein pore. To better address the arrangement of the water molecules inside the protein structure, the authors generated another initial condition for the system, with the protein pore already entirely occupied by water molecules. After a short time, the water molecules rearranged forming a single file in which each molecule was bound to the next through a hydrogen bond (figure 5). This single file was subsequently disrupted along the second phase of the simulation, where the protein structure is free to move. This was probably due to the intensity of the structural movement.

Quoting the authors, we can conclude that the calculations revealed a spontaneous diffusion of water molecules from the bulk into the interior parts of the channel and demonstrated in some cases the formation of continuous water files inside the monomers.

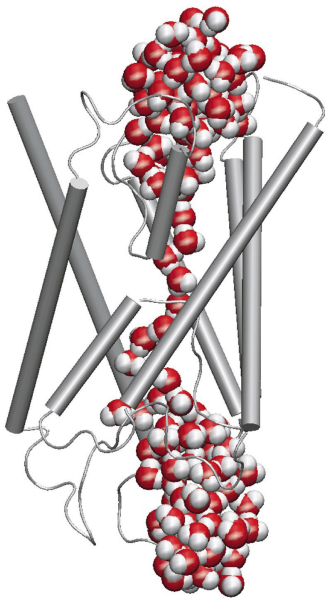


Figure 5: (From the original paper [18]) Snapshot of water molecules inside the channel at the end of the 1 ns equilibration (phase I) in which the protein was fixed.

2) Molecular Dynamics Study of Unbinding of the Avidin-Biotin Complex [19]

In this work, the authors explore relatively short (40-500 ps) MD simulations that induce the unbinding of biotin from avidin. Biotin is a water soluble B-complex vitamin (B7) that functions as a coenzyme in the metabolism of fatty acids and leucine and in gluconeogenesis. Biotin binds to the tetrameric protein avidin, in one of the strongest protein-ligand interactions (K_d in the order of 10^{-15}) [20]. Binding and unbinding of ligands to proteins is an essential biochemical process characterized by the transition between equilibrium states. To simulate the unbinding of biotin from avidin, the authors applied external forces to the system, mimicking the mechanism of micromanipulation through atomic force microscopy cantilevers. The results show a variety of dissociation pathways, the key residues involved in the protein-ligand binding and the spatial range for the binding of biotin into avidin. The simulations of enforced unbinding between these molecules revealed stepwise slips, involving changes in protein-ligand hydrogen bond patterns and hydrophobic contacts as well as a barrier characterized by a loop capping the binding pocket. The unbinding of the ligand was a multistep process in which each step corresponds to the rupture of a network of hydrogen bonds formed between biotin and polar residues at the binding pocket. As the authors themselves recognized, this was the stablishment of a new methodology for exploring protein-ligand binding, unbinding and recognition.

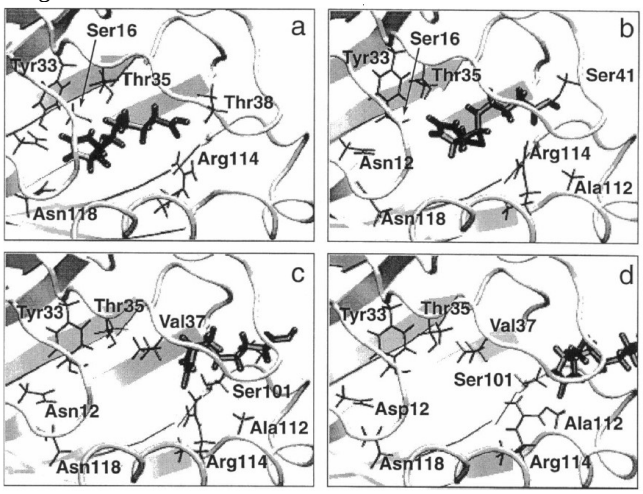


Figure 6: (From the original paper [19]) Contacts of biotin with polar residues in the process of unbinding through the MD simulation: (a) t = 0 ps, (b) t = 195 ps, (c) t = 372 ps and (d) t = 495 ps.

4 CONCLUSIONS

MD simulations provide powerful tools for explore the panorama of the conformational energy available for a specific molecular system. The rapid growth of computational capacity together with technological progress characterizes the current moment as an important historical point to simulations in the structural biology field. The conformational dynamics of proteins is an intrinsic molecular characteristic that cannot be efficiently treated through experiments. To accomplish such task, several computational methods were developed since the first protocol, applied in 1977 (see references 14-17, for examples). Recent theoretical advances are repeatedly leading MD simulations to a more refined state, allowing the

technique to deal with more complex systems. MD is capable of providing details for atomic movement through time, describing structural properties of the system under investigation and enabling the analysis of its biophysical characteristics.

ACKNOWLEDGEMENTS

I would like to acknowledge the essential work of Marc, the supervisor of this work and the precious help of all supervisors, specially Christian, Edda and Andrea (and Tanya), who provided great insights and criticism upon this work and a very helpful overall feedback on the seminar. Additionally, I would like to thank the students who attended the seminar and contributed with a very substantial discussion and feedback. Last but not least, thanks to Burkhard Rost for the support and for providing me with the opportunity to be part of this group.

REFERENCES

1. J.A. McCammon, B.R. Gelin, M. Karplus. *Dynamics of folded proteins*. Nature, 267:585-267, 1977
2. M. Karplus, J.A. McCammon. *Molecular dynamics simulations of biomolecules*. Nature Structural Biology (Review), 9:646-652, 2002
3. R.P. Feynman. *Feynman Lectures on Physics*. Addison-Wesley, 1965
4. K. Linderstrom-Lang. *Deuterium exchange between peptides and water.* Chemistry Society (London). 2:1-20, 1955
5. F. Sanger, H. Tuppy. T*he amino acid sequence in the phenylalanyl chain if insulin 1. The identification of lower peptides from partial hydrolysates.* Biochem. Journal. 49:463-481, 1951
6. H. Muirhead, M.F. Perutz. *Structure Of Hemoglobin: A Three-Dimensional Fourier Synthesis of Reduced Human Haemoglobin at 5.5 Å Resolution.* Nature 199:633-638, 1963
7. J.C. Kendrew, G. Bodo, H.M. Dintzis, R.G. Parrish, H. Wyckoff, C. Phillips. *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature. 181:662-666, 1958
8. B.J. Alder, T.E. Wainwright. *Phase transition for a hard sphere system.* Journal of Chemical Physics. 27:1208-1209, 1957
9. F.H. Stillinger, A. Rahman. *Improved simulation of liquid water by molecular dynamics.* Journal of Chemical Physics. 60:1545-1557, 1974
10. D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers. *Atomic-level characterization of the structural dynamics of proteins.* Science. 330:341-346, 2010
11. P.L. Freddolino, F. Liu, M. Gruebele, K. Schulten. *Ten-Microsecond molecular dynamics simulation of a fast-folding WW domain.* Biophysical Journal. 94:L75-L77, 2008
12. Swiss Institute of Bioinformatics. *Theory of Molecular Dynamics Simulations.* Website (http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html)
13. D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen and H.J.C. Berendsen, *Gromacs User Manual version 4.5*, www.gromacs.org 2011
14. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus. *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry 4:187-217, 1983
15. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl. *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of Chemical Theory and Computation. 4:435–447, 2008
16. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten. *Scalable molecular dynamics with NAMD.* Journal of Computational Chemistry. 26:1781-1802, 2005
17. D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. *AMBER 11*. University of California, San Francisco, 2010
18. F. Zhu, E. Tajkhorshid, K. Schulten. *Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer.* FEBS Letters. 504:212-218, 2001
19. S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, K. Schulten. *Molecular dynamics study of unbinding of the avidin-biotin complex.* Biophysical Journal. 72:1568-1581, 1997
20. O.H. Laitinen, V.P. Hytonen, H.R. Nordlund, M.S. Kulomaa. *Genetically engineered avidins and streptavidins.* Cellular Molecular Life Science. 63:2992–3017, 2006