

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Structural Bioinformatics for Protein Engineering

Davi S. Vieira, Marcos R. Lourenzoni, Carlos A. Fuzo,
Richard J. Ward and Léo Degrève
*University of São Paulo - FFCLRP, Department of Chemistry
Ribeirão Preto, São Paulo
Brazil*

1. Introduction

Proteins are amongst the most abundant and functionally diverse macromolecules present in all living cells, and are linear heteropolymers synthesized using the same set of 20 amino acids found in all organisms from the archaebacteria to more complex forms of life. By virtue of their vast diversity of amino acid sequences, a single cell may contain a huge variety of proteins, ranging from small peptides to large protein complexes with molecular weights in the range of 10^3 - 10^6 Daltons. Consequently, proteins show impressive diversity in biological function which includes enzyme catalysis, signal transmission via hormones, antibodies, transport, muscle contraction, antibiotics, toxic venom components and many others. Among these proteins, it may be argued that enzymes present the greatest variety and specialization, since virtually all the chemical reactions in the cell are catalyzed by enzymes (Aehle, 2004).

Due to their biotechnological and biomedical potential, enzymes and antimicrobial peptides have been the focus of extensive research efforts. This interest has been driven by a growing market for enzymes for various applications such as food supplements for humans and animals, enzymes to replace organochlorine compounds in some industrial processes, production of biofuels, vaccines, drugs, etc. (Aehle, 2004). Furthermore, the applications of enzymes in health, food, energy, materials and environment is likely to increase.

The elucidation of the three-dimensional structure of a protein is a fundamental step to understand its biological function. The 3D structure can be obtained by experimental techniques such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and more recently by cryo-electron microscopy. As of mid-May 2011 there were more than 73,000 protein structures deposited in the Protein Data Bank (PDB - www.rcsb.org/pdb), of which approximately 90% have been determined by X-ray diffraction, 9% by NMR and 1% by other techniques.

Despite the availability of established experimental methodologies for the determination and analysis of three-dimensional structures of proteins to elucidate their structure-activity relationships, there are inherent limitations which can restrict their application. X-ray diffraction determines the spatial electron density distribution in the biomolecule and

thereby provides the Cartesian coordinates of all atoms in the protein, except hydrogens. NMR spectroscopy is limited to low molecular weight proteins, generally smaller than 30kDa, although with the introduction of TROSY-based experiments (transverse relaxation-optimized spectroscopy) the molecular weight limitation has been extended to approximately 800kD (Fernández & Wider, 2003). Cryo-electron microscopy is a low-resolution technique that is a powerful tool for the determination of the structure of large proteins or protein complexes, such as virus structures (Mancini *et. al.*, 2000).

Conformational changes and dynamics are key processes to understand biological mechanisms and functions. Many biological processes involve functionally important changes in the protein 3D structure, however techniques for structure determination do not describe (or describe poorly) protein dynamics. Examples of the importance of protein dynamics in structural biology include the conformational changes associated with protein folding, catalytic functions of enzymes and signal transduction. In this context, bioinformatics techniques, such as molecular dynamics (MD) simulation, together with homology and statistical analysis of protein structures have become useful tools in structural biology not only for understanding protein function (van Gunsteren *et. al.*, 2008), but also for protein engineering (Vieira & Degrève, 2009). Growing interest in bioinformatics techniques and advances in experimental methodologies have led to many successes in engineering enzyme function (Aehle, 2004).

The purpose of this chapter is to introduce the use of MD simulations in a biological context to target proteins of biotechnological interest, such as xylanases, antibody/antigen interactions and antimicrobial peptides. Using relevant models we discuss the use of MD simulations to study properties such as enzyme thermostability, protein conformational changes and functional movements, intra and intermolecular interactions through hydrogen bonding analysis, protein-protein interaction and modelling of peptide-membrane interactions. The main goal is to demonstrate how the three-dimensional structure, function, interaction potential and dynamics are highly correlated and together comprise a valuable data set to explore the broad research field of structure/function relationships of proteins.

2. An introduction to molecular dynamics simulations

Molecular dynamics (MD) simulation is one of the main computational methodologies which can provide detailed microscopic information at an atomic and molecular level for a variety of systems. The behaviour of matter can be understood by the structure and dynamics of its constituent atoms or atomic groups, and is considered as an N -body problem. Due to the lack of general analytical algorithms, a solution to the classical N -body problem is exclusively a numerical task. Therefore, the study of matter at this level requires the computational resources to allow the investigation of the movements of individual particles or molecules, and the MD approach aims to reproduce the properties of a real system using a microscopic model system. The continual increase in computational power enables the investigation of increasingly large model systems, and consequently to approach broader and more complex biological questions. MD simulation is not limited to biological and biochemical systems, and MD has a broad range of applications in material science, physics, chemistry and engineering (Leach, 2001).

MD simulations are applied to systems that are in states in which quantum effects are disregarded, which imposes a limit on the types of problem that can be addressed. For example, chemical reactions and transitions between different electronic states present a quantum nature due to their dimensional scale and their high energies, and cannot be treated by MD. MD simulation employs the classical equations to describe the motion of atoms and molecules as a function of the time under determined conditions (pH, temperature and pressure) taking into account an intermolecular interaction potential. Hybrid methodologies (Quantum Mechanics/Molecular Mechanics, QM/MM) are capable of treating quantum phenomena in MD simulations to describe, for example, chemical reactions using Valence Bond MD. Other well established hybrid methods are Carr-Parrinello and *ab initio* MD (Leach, 2001), however these are limited by the intensive computing power required.

The use of MD simulation allows the study of the temporal evolution of molecular motions, generating a series of complete atomic coordinates (ie. Cartesian x,y,z coordinates for each atom) that permit the prediction of microscopic and macroscopic properties of the system from Statistical Mechanics analysis (McQuarrie, 1976). Starting from a pre-defined atomic arrangement that defines the initial configuration of the system, the individual atoms move under the influence of their intermolecular potentials. Given the atomic positions and velocities at a given time (t), the resulting forces can be calculated for each particle in the system permitting calculation of the atomic positions and velocities at a later time ($t+dt$). This procedure generates molecular trajectories for the whole system over the total simulation time.

The choice of the number of atoms in a system depends on the properties that are to be studied with the available computational capacity, and must be representative of the macroscopic real system. The current computational power with present methods allows simulations of systems constituted by up to 10^8 atoms. Recently, a 1-millisecond all-atom MD simulation of a folded protein has been reported (Shaw et. al., 2010). The equilibrium properties of the system are computed as time averages over given time intervals that must be somewhat longer than the corresponding event observed at the atomic scale. Temperature, radial pair distribution functions, potential and kinetic energies are examples of fast convergence properties, while surface tension, diffusion coefficient and pressure are properties of slow convergence. An important characteristic of MD simulation is that information on the system is gained on a time scale that is often not easily obtained by experimental approaches (Rapaport, 2005).

2.1 Molecular dynamics and statistical mechanics

Statistical mechanics deals with ensemble averages, and is well suited to analysis of the data produced by MD methods. During MD simulations, the mechanical, thermal and chemical thermodynamic variables must be fixed and simulations are usually conducted under constant NPT , NVT , NVE or μNT conditions. (where N is the number of atoms, P the pressure, T the temperature, V the volume, E the internal energy and μ the chemical potential). An ensemble is a large set of replicas generated under the defined external conditions (McQuarrie, 1976), and in the NVT (or canonical) ensemble, N , V and T are fixed external parameters and therefore the instantaneous temperature (obtained from the kinetic energy) must be controlled to a fixed value by a thermostat algorithm, which multiplies the instantaneous atomic velocities by a coefficient defined by the difference between instantaneous and desired temperatures. Several algorithms can be used to control the

temperature including the Berendsen, Nosé-Hoover and V-rescale thermostats (Allen & Tildesley, 1992; Rapaport 2005). *NVT* and *NPT* ensembles are widely used in thermal treatments such as annealing in order to promote thermal denaturation of proteins in unphysical conditions (Rocco et. al.; 2008). *NPT* or isobaric-isothermal ensembles present better characteristics than the canonical ensemble, since the lack of the thermodynamic information about the equilibrium conditions are circumvented by the adjustment of the pressure to the imposed external conditions. This allows the volume to fluctuate by scaling the dimensions of the system by a coefficient defined by the instantaneous and desired pressures.

The equilibrium average of a given property A is expressed by:

$$\langle A \rangle_{ensemble} = \frac{\iint A(\vec{p}^N, \vec{r}^N) e^{-\beta E(\vec{p}^N, \vec{r}^N)} d\vec{p}^N d\vec{r}^N}{Q} \quad (1)$$

where \vec{p}^N and \vec{r}^N are the momenta and positions of the N particles at time t , $E(\vec{p}^N, \vec{r}^N)$ is the sum of the potential, $U(\vec{r}^N)$, and kinetic energy, $K(\vec{p}^N)$, and $\beta = 1/kT$. The denominator term, Q , is the partition function. For a system of N identical particles the partition function for the canonical ensemble is:

$$Q_{NVT} = \frac{1}{N!} \frac{1}{h^{3N}} \iint e^{-\beta E(\vec{p}^N, \vec{r}^N)} d\vec{p}^N d\vec{r}^N \quad (2)$$

where h is the Planck's constant. The partition function is central to the statistical mechanics analysis since it contains all the information to determine the macroscopic thermodynamics properties, which are crucial for correlating the microscopic and macroscopic levels (McQuarrie, 1976). Generally, the partition function cannot be computed since it depends on the interatomic forces and positions.

The most fundamental axiom of statistical mechanics, the ergodic hypothesis, states that the ensemble average equals the time average.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \quad (3)$$

This theory is based on the fact that if one allows the system to evolve indefinitely in time it will pass through all possible energy states, therefore the MD simulation should generate enough configurations in time to ensure that the equality is satisfied. The average in eq.(1) is an average over all the possible positions of the particles. During a MD simulation, the averages are calculated over the configurations obtained as a function of time so that in accord with the ergodicity hypothesis:

$$\langle A \rangle_{time} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^{\tau} A[\vec{p}^N(t), \vec{r}^N(t)] dt \approx \frac{1}{M} \sum_{m=1}^M A_m(\vec{p}^N, \vec{r}^N) \quad (4)$$

where M is the number of $A_m(\vec{p}^N, \vec{r}^N)$ values produced by the MD simulation.

A variety of thermodynamic properties can be calculated from the configurations generated by the MD simulations. Comparison with experimental results is an important way to check

the accuracy of the simulation and the validity of the model and method. Simulation techniques can also be used for prediction of thermodynamic properties for which there is no experimental data. For example, the internal energy can be calculated as follows:

$$\langle E \rangle = \frac{1}{M} \sum_{i=1}^M E_i \quad (5)$$

and the heat capacity at constant volume, C_V , as given by the internal energy fluctuation is obtained from:

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{kT^2} \quad (6)$$

where $\langle E^2 \rangle$ and $\langle E \rangle^2$ can be calculated at the end of simulation.

Heat capacity is an important quantity to investigate phase transitions, because it shows a characteristic dependence upon temperature. A first-order phase transition shows an infinite heat capacity at the transition temperature and a discontinuity for a second-order phase transition. The heat capacity can also be calculated from a series of simulations at different temperatures followed by the differentiation of the energy with respect to the temperature.

2.2 Molecular interactions and equations of motion

The main influence of a molecular simulation is the potential energy, which determines how the particles of the system interact. The simplest model of a particle is a sphere, and the simplest model of interaction is that between pairs of atoms. Such simple models are capable of reproducing two main features of real systems, (i) the resistance to compression (short range) and (ii) the capability of keeping atoms together in a condensed phase since atoms are attracted to each other over longer ranges (Rapaport, 2005). MD techniques use analytical expressions for the energy potential functions that are obtained from experimental data or quantum calculations. These analytical expressions are generally effective pair interaction potentials that allow the molecular systems to exhibit the correct characteristics of real systems (Maitland et. al., 1981).

The most widely used of these interaction potentials between monoatomic molecules was initially proposed to reproduce structural properties of liquid argon (Maitland & Smith, 1971), written in the form of the Lennard-Jones (LJ) potential:

$$U(r_{ij}) = 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (7)$$

where r_{ij} is the distance between atom i and j , $-\varepsilon$ is the depth of the potential well and σ the radius of the particles. An intermolecular potential energy must reproduce the attractive (long-range) and repulsive (short-range) behaviours, and in the case of LJ potential the repulsive and attractive behaviours are represented by the first and second terms respectively. More complete expressions describing the total intra and intermolecular potential energies for complex molecules will be described in section 2.5.

The intermolecular interactions are generally cut off at the separation distance r_c , i.e.

$$U(r_{ij}) = 0 \quad r_{ij} \geq r_c \quad (8)$$

The force between the atoms i and j is given by:

$$\vec{F}_{ij} = -\frac{\partial U(r_{ij})}{\partial r_{ij}} \quad (9)$$

Special care must be taken with the treatment of the long range interactions, since the distance cut-off may introduce a discontinuity in the potential energy and the force near the given value. This artefact creates problems due to the requirement for energy conservation in MD simulations, and efficient mathematical approaches have been suggested to circumvent this problem, such as the use of switching functions (Allen & Tildesley, 1992).

The coordinates of the atoms of the system are obtained by solving the equation for Newton's second law;

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = m_i \vec{a}_i = \vec{F}_i = \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^N \vec{F}_{ij} \quad (10)$$

where \vec{F}_i is the force acting on the atom i of mass m_i and acceleration \vec{a}_i . Each pair of atoms need only be identified once because $\vec{F}_{ij} = -\vec{F}_{ji}$ (Newton's third law).

The differential equation (10) is solved by integration using standard numerical techniques, such as finite difference or predictor-corrector. Finite difference methods are widely employed in MD simulation to integrate the equation of motion, for example the Verlet and leap-frog algorithms (Allen & Tildesley, 1992).

The Verlet algorithm is based on positions $\vec{r}(t)$ and accelerations $\vec{a}(t)$. The positions at times $t \pm dt$ are obtained by Taylor expansion of $\vec{r}(t)$:

$$\vec{r}_i(t + dt) = \vec{r}_i(t) + \vec{v}_i(t)dt + \vec{a}_i(t)\frac{dt^2}{2!} + \dots \quad (11)$$

$$\vec{r}_i(t - dt) = \vec{r}_i(t) - \vec{v}_i(t)dt + \vec{a}_i(t)\frac{dt^2}{2!} - \dots \quad (12)$$

where $\vec{v}_i(t)$ is the velocity of the atom i , dt is the time step and, $\vec{r}_i(t + dt)$ and $\vec{r}_i(t - dt)$ are the later and earlier positions relative to $\vec{r}_i(t)$. By addition of equations (11) and (12) and taking into account equation (10), we obtain the equation for advancing the positions:

$$\vec{r}_i(t + dt) = 2\vec{r}_i(t) - \vec{r}_i(t - dt) + \frac{dt^2}{m_i} \vec{F}_i(t) \quad (13)$$

Although the velocities are not necessary to compute the trajectories, they are useful to estimate the kinetic energy and hence the total energy of the system. The atomic velocities can be calculated from:

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + dt) - \vec{r}_i(t - dt)}{2dt} \quad (14)$$

Other integration algorithms are available (Allen & Tildesley, 1992). The numerical integration of the motion equations encompasses numerical approximations so that the time step, dt , must be chosen to be at least one order of magnitude smaller than the high-frequency motions and generally of the order of a few femtoseconds. The high-frequency motions, such as bond stretching involving hydrogen atoms that can impair the use of $dt \approx 1fs$, are avoided by maintaining constant bond lengths, since generally these bonds are not interesting and have a negligible effect on the overall stability of the system.

2.3 Initial configuration of the system

The MD simulations must be capable to sample the regions of the phase space. A MD simulation must be insensitive to the initial state, therefore any appropriate initial state is permissible. A simple method is to build the system from a random distribution of particles or from a regular lattice obeying the numerical density (N/V) of the real system. The initial velocities can be also assigned in different ways, either randomly or by fixing the value based on temperature (Maxwell-Boltzmann distribution). Velocities are adjusted to ensure that the center of mass of the system is at rest. In MD simulation of biomolecules, the initial structure must be known and the solvent molecules and ions are added to fulfil the system and neutralize it. Initial protein structures can be found in the PDB data bank or using specific methods of bioinformatics that predict and validate three-dimensional structures.

The set-up of the simulation box is the important first step in the definition of the system. The limits of the simulation box should not produce interface effects in homogeneous systems, and for this reason periodic boundary conditions (PBC) are applied to extend the system to the thermodynamic limit. PBC consists of building replicas of the around of the main box in order to obtain a system that mimics a macroscopic thermodynamic system. All the particles in the replica move identically so that the motion of the atoms is not limited by the walls of the box. During the course of a simulation, when a molecule moves in the central box, all its images move in the same manner (Allen & Tildesley, 1992). In other words, a system which is bounded but free of physical walls can be simulated by the application of PBC.

Five shapes of cells produce periodic images: the cube, the parallelepiped, the truncated octahedron, the rhombic dodecahedron and the hexagonal prism. The choice of the simulation box should be made based on the geometry of the system of interest, for example a globular protein is suitable to be simulated in a cubic or dodecahedron periodic cell, but inadequate for a parallelepiped.

2.4 Controlling the simulations

Many changes occur in the structure of the system during a simulation process. As the initial structure undergoes modifications as a function of time, the MD data allow examination of the changes at each step and consequently, the evaluation of the equilibrium conditions. The equilibrium condition, *i.e.* small fluctuations near mean values of a given property, is fundamental to control the MD simulations. Stabilization conditions can be identified by accompanying determined properties that are calculated during the simulation. Many properties can be used to control the MD simulations such as the energy profiles, the

stability of the temperature, etc. After establishing that the system is stabilized, the calculation phase can be initiated in which the properties of interest can be computed.

In a system constituted by a protein in water, the stability is established by a delicate energetic balance involving intramolecular and intermolecular interactions and kinetic energy. Despite the importance of the intramolecular interactions, the stability of biomolecules in general is strongly influenced by protein-water interactions. The solvent molecules make significant contributions to the conformational changes of the protein, and at equilibrium the protein/solvent interactions are completely relaxed. An important tool to detect equilibrium conformations in MD simulation of proteins is the Root Mean Square Deviation (RMSD), which is defined by;

$$RMSD(t) = \sqrt{\frac{\sum_{i=1}^n |\vec{R}_i - \vec{r}_i(t)|^2}{n}} \quad (15)$$

where n is the number of atoms (frequently only the C_α atoms) used in the calculation. $|\vec{R}_i - \vec{r}_i(t)|$ is the distance between the atom i at time t and the same atom in the reference structure. The reference structure can be the experimentally determined initial structure or any other structure, for example the last structure generated by the simulation. The RMSD calculation is made after the superposition of the entire structures or may be applied to more limited groups of atoms or residues. The RMSD provides information about the structural stability as a function of time and the changes of the structures during the simulation process. Figure 1 shows two different cases of the RMSD of a protein which is not stabilized even after 100.0 ns as compared with the RMSD of a protein that reaches equilibrium in only 5.0 ns. These results illustrate the important point that the equilibration time is inherent to each protein, to its environment and to the quality of the initial three-dimensional structure.

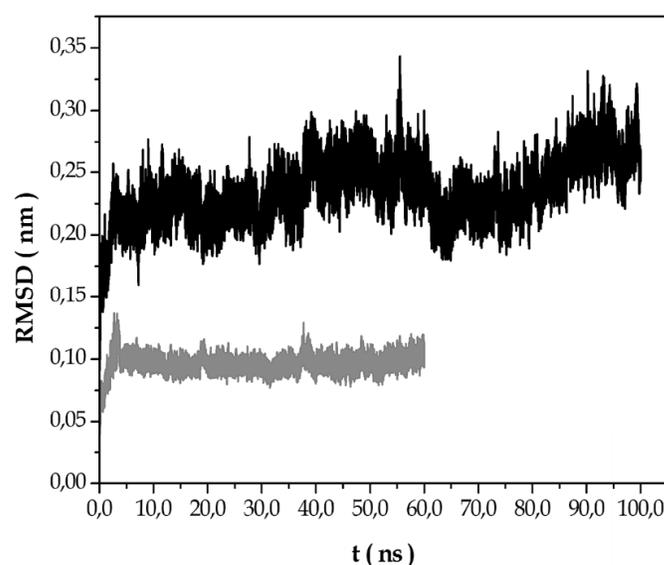


Fig. 1. Root mean square deviations (RMSD) for two different MD simulations. The black line corresponds to a protein not stabilized in 100.0 ns and the gray line to the protein that shows a fast equilibration.

2.5 Force fields and molecular simulations software packages

The Force Field in MD simulations is the name given to the set of parameters that define the potential interaction energies. These parameters are adjusted to fully describe the molecular behaviour and its interactions. Intra- and intermolecular potential interactions can be expressed in terms of relatively reduced number of types of components. The interaction between bonded atoms can be subdivided into bond stretching, angle bending and dihedral-angle bending (torsion) components, while the interactions between non-bonded atoms are by electrostatic and van der Waals interactions. The full expression can be described by:

$$V_{bonds} = \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 \quad (16)$$

$$V_{angles} = \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \quad (17)$$

$$V_{torsions} = \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] \quad (18)$$

$$V_{non-bonded} = \sum_{i=1}^N \sum_{j=i+1}^N \left[\left(\frac{A_{12}}{r_{ij}^{12}} - \frac{B_6}{r_{ij}^6} \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \quad (19)$$

$$V(\vec{r}^N) = V_{bonds} + V_{angles} + V_{torsions} + V_{non-bonded} \quad (20)$$

$V(\vec{r}^N)$ denotes the total potential energy as a function of the positions \vec{r} of N particles or atoms. Bonds and angles are modelled by harmonic potentials (eq. 16 and 17) that give the change in energy as $x_i (x \equiv l, \theta)$ deviates from the reference value, $(l_{i,0}, \theta_{i,0})$. The third term, eq. 18, is the torsional contribution that evaluates how the energy changes in the vibrations of structures of four atoms (planar, like aromatic rings or angular such as in the sp^3 hybridization) and finally, the fourth term is the interaction between non-bonded atoms, including van der Waals and electrostatic contributions, as calculated between the non-bonded pairs of atoms in the same or different molecules. The terms A_{12} and B_6 in eq. 19 are the Lennard-Jones parameters which consider the size of the molecules and the distance between atoms at zero energy potential.

Electrostatic interactions are important long-range forces that influence protein stability and present difficulties in their treatment in MD simulations, and are integral parts of MD simulation software packages. Some methods can be cited as the most important for protein systems, such as the Reaction Field method and the Ewald Sum. Excellent books are available for consultation on this issue (Allen & Tildesley, 1992; Leach, 2001). Several Force Fields are available and all of them are, or at least should be equivalent, independently of the methodology used to compute their parameters.

An increasing number of force fields and MD software packages are available, many of which are free and can be obtained from the web sites of the research group responsible for their development. MD software packages have been employed in research in theoretical

chemistry and biology, protein engineering and industry (pharmaceutical and materials). Some of the most important and referenced MD packages are listed below, and all are well documented and many offer parallelization options.

GROMACS (Groningem Machine for Chemical Simulation): Currently available in version 4.5.3 (2011) at the website www.gromacs.org. It refers to both the MD software package and the GROMACS Force Field that is updated like other force fields. It is parameterized to several solvent, proteins, nucleic acids and carbohydrates. *AMBER (Assisted Model Building with Energy Refinement)*: refers to both the Force Field and the MD program specially developed for biomolecules, and can be obtained at symbolic price. *CHARMM (Chemistry at Harvard Macromolecular Mechanics)*: is also a Force Field and MD program which although is not fully available for academic purposes, can be obtained from Accelrys, Inc through a commercial license. *NAMD (Not Another Molecular Dynamics)*: has as its main feature the high degree of scalability in multi-processed computational systems, up to around one hundred processors with no loss performance. *Q Program*: is a MD software package designed for free energy calculations in biological systems. It is fully available for academic purposes from the Åqvist group website (Åqvist, 2007). Other MD simulation packages include TINKER, LAMMPS, AMMP, BOSS, CERIU2, CPMD, INSIGHT2, ORAC and OPLS.

3. Biotechnological applications

3.1 Xylanases

Xylanases (EC 3.2.1.8) are enzymes produced by a wide variety of bacteria and fungi that hydrolyze the 1,4- β -D-xylosidic linkages of xylans, a mixed family of plant cell wall polysaccharides that are abundant in nature. The enzymatic hydrolysis of xylan has attracted considerable biotechnological interest with applications in the food engineering, bio-ethanol fuel production and cellulose pulp industries (Beg et al., 2001). Ongoing efforts have been directed toward both the identification of new xylanases and the improvement of the catalytic properties of existing enzymes with the goal of enhancing their compatibility for industrial applications.

Based on aminoacid sequence similarity and three-dimensional structural homology, xylanases are classified as family GH10 or GH11 hydrolases. GH11 xylanases present compact globular three-dimensional structures with a conserved scaffold comprised of a single α -helix and two extended pleated β -sheets forming a jellyroll fold. The major surface feature is a long cleft that spans the entire molecule and which contains the active site. The overall shape of GH11 xylanases resembles a right-hand, and as shown in Figure 2 the various structural features have accordingly been denominated as fingers, palm, thumb, cord and helix regions.

The globular structure and relatively small size (~200 amino acids) of the GH11 xylanases are good models for molecular simulation which can be used to investigate important physicochemical and biochemical properties, such as thermostability.

Thermostable proteins are of biotechnological interest because they can be employed in industrial processes which use high temperatures, such as cellulose pulp bleaching. Several strategies have been attempted with the aim of enhancing thermostability of proteins, including the introduction of disulfide bridges (Wakarchuk et al., 1994), increasing the hydrophobic contacts by addition of aromatic interactions (Georis et al., 2000), optimization of

the electrostatic surface potential (Torrez et al., 2003), and optimization of intramolecular and protein/solvent hydrogen bonding (Viera & Degrève, 2009). Due to the important enthalpic contribution of hydrogen bonding in biological systems, hydrogen bond optimization is a particularly interesting strategy to improve protein thermostability. Formally, hydrogen bonds are rigorously evaluated by quantum mechanics methodologies using the Morokuma Analysis that deconstructs the energy of hydrogen bonds into five components: electrostatic, polarisation, exchange repulsion, charge transfer and mixing (Morokuma, 1977). In contrast, molecular simulation obtains a reasonable estimative of hydrogen bond occurrence and stabilization by evaluating geometric criteria. For example, intramolecular hydrogen bonds are identified by interactions with a donor-receptor distance shorter than 0.24 nm, the (A-H...O) angle larger than 110° ($A \equiv N$ or O) and the fractional occurrence larger than 20% of the total simulation time. The intermolecular hydrogen bonds can be detected by two criteria; (i) based on the radial distribution functions, $g(r)$, and (ii) distribution of the interaction energies between the protein atoms and the solvent, $F(E)$, (Degrève et al., 2004).

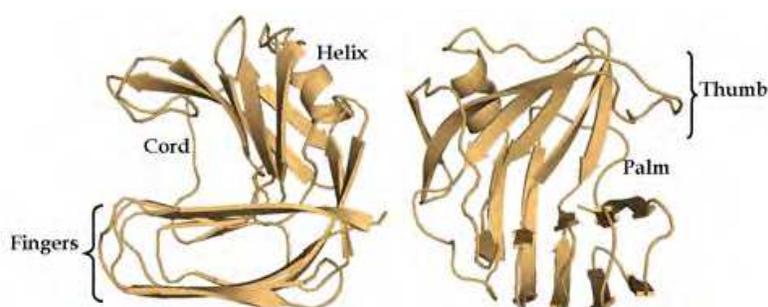


Fig. 2. Two different views of three-dimensional structure of the GH11 xylanase from *Bacillus subtilis*, PDB code 1XXN. The regions based in the right-hand analogy are indicated.

The dynamics and energetic properties of MD simulations are well suited to monitor interatomic distances and angles over time, and can characterize all atom pairs involved in hydrogen bonds, hydrophobic clusters or salt bridges. Figure 3 show the hydrogen bonding network in a three-dimensional structure of the *Bacillus subtilis* xylanase from a MD simulation identified by the geometric criteria describe above.

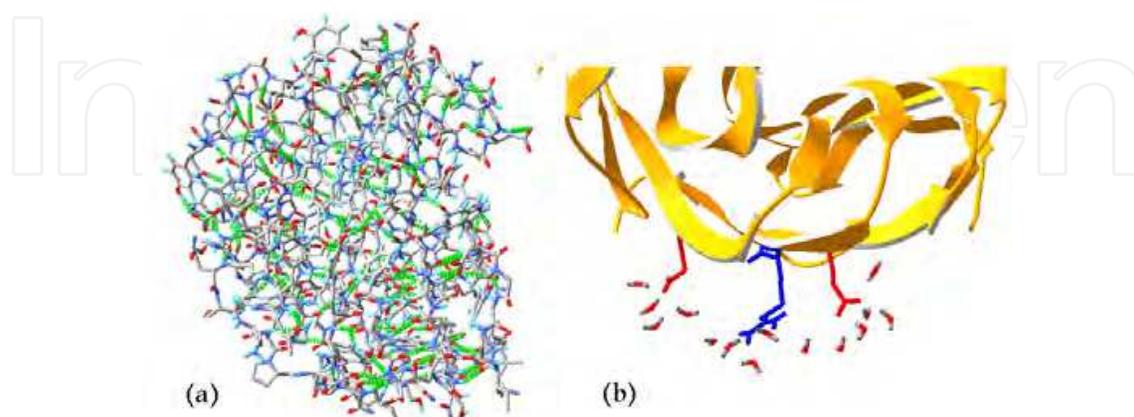


Fig. 3. (a) A snapshot at 10.0ns of GH11 xylanase from the MD simulation in canonical ensemble at physiological pH. The intramolecular hydrogen bonding network is showed as the green dotted lines. (b) The intermolecular hydrogen bonding interactions of the first solvation shell of two Glu residues (red) and one Arg residue (blue).

Intermolecular hydrogen bonds may be identified by considering the first solvation shell of a given residue. As shown in Figure 3(b), the $g(r) \equiv g_{AW}(r)$ function must display a peak in an appropriate region, approximately 2\AA , defined by the nature of the protein atom. Furthermore, the $F(E)$ should display a peak, or a shoulder, in the attractive energy region (negative energies) as observed in the example shown in Figure 4.

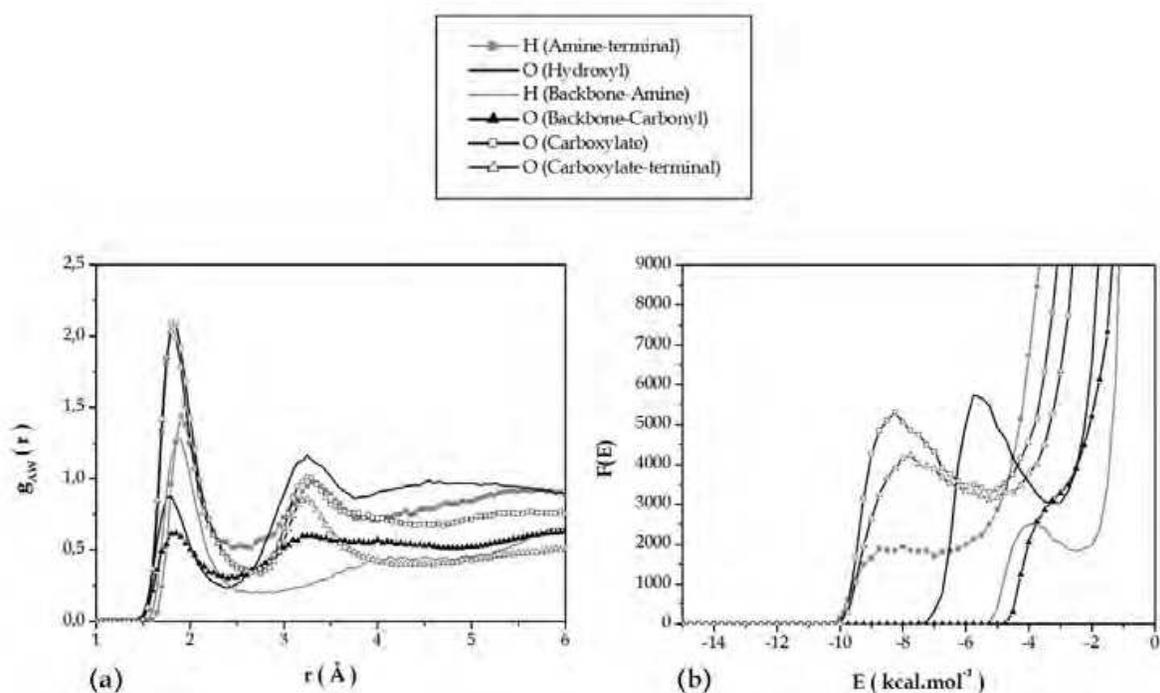


Fig. 4. (a) Radial distribution functions $g_{AW}(r)$, where A is the protein atoms and W the water molecule. (b) Distribution of the interaction energies between atoms of protein and the water molecules. Protein atoms are defined in the legend.

As shown in Figs 4(a) and 4(b), each atom or a group of atoms has a defined influence on the protein energetics. These interactions are unique to each protein, and provide an individual map of the protein enabling us to make predictions with respect to the thermostabilization mechanism adopted by proteins. Recently, Vieira & Degrève have published an MD simulation study of a thermophilic-mesophilic pair of GH11 xylanases, in which a difference of 20°C in thermostability could be interpreted in terms of different patterns of intra and intermolecular hydrogen bonds exhibited by these proteins (Vieira & Degrève, 2009).

MD studies by Vieira & Degrève (2009) estimate the total interaction potential energy for the thermophilic xylanase to be around -600kcal.mol^{-1} more stable than its mesophilic counterpart, and approximately 20% of this difference is due to intermolecular hydrogen bonds energies. Although the backbone of the thermophilic xylanase is more rigid than the mesophilic enzyme (as determined by the intramolecular hydrogen bonds and salt bridges), the increased thermostability seems to be a consequence of the greater degree of solvation. Although both xylanases are highly solvated, the intermolecular hydrogen bonding network in the thermophilic enzyme is energetically more attractive than those at mesophilic counterpart. In other words, the highly solvated surface characterized by optimization of

strong protein-water intermolecular hydrogen bonds is clearly a thermostabilization factor. The main residues which contribute to the thermostability of thermophilic xylanase were located in the finger region and have been targeted for site-directed mutagenesis. This insight as to a major contribution to protein thermostability would not be available from the analysis of a structure determined by X-ray crystallography, in which water molecules are poorly resolved. Therefore, MD simulations can complement the experimental studies in the development and of design process of new proteins.

3.2 Antibody engineering

Antibodies or immunoglobulins (Ig) are globular proteins that play an important role in protecting the host organism against infectious diseases. The primary function of an antibody is to bind molecules that are foreign to the host (antigens), and understanding the interactions between an antibody and its antigen is important for the design and development of new or improved antibodies. The host typically produces several types of antibody that have similar structures, consisting of four polypeptides (two heavy chains and two light chains) that are linked by disulphide bonds to form a molecular complex. Five heavy chain isotypes are known, giving rise to five different Ig classes (IgG1-4, IgA1-2, IgD, IgM, IgE), each with a distinct function in humans. Each class of heavy chain can combine with one of two light chain isotypes (kappa and lambda). As a result of these molecular combinations, antibodies differ in size, electric charge, amino acid composition and carbohydrate content (Roitt, 1997).

The most common class of IgG antibodies present a basic structure of two light chains (each of which has two domains) and two heavy chains (each having four domains). The two N-terminal domains of the heavy chain are linked by disulphide bonds to the two domains of the light chains to form a "Y"-shaped structure. The N-terminal domain of both the heavy and the light chain show pronounced amino acid sequence (and are appropriately called the variable domains, VL and VH), while the other domains are called constant domains (CL, CH1, CH2 and CH3) (Alberts et al, 1989). The variable domains have three regions of hypervariability in the amino acid sequence and are termed Complementarity Determining Regions (CDRs). These regions are responsible for the recognition and binding of the antibodies to a specific antigen, and determine the affinity and specificity to the antibody-antigen interaction. The "Y"-shaped antibody structure can be divided into three fragments: Fv (variable fragment), Fab (antigen binding fragment) and Fc (crystallization fragment). An engineered antibody may be obtained from separate segments of Fv heavy and light chains joined by a flexible peptide linker to form a single-chain Fv (scFv).

Molecular modeling by homology uses experimentally determined protein structures to predict the three-dimensional conformation of a protein with a similar amino acid sequence whose structure is unknown. Molecular modeling can be used with reasonable success if the amino acid sequences of the known and the unknown proteins share at least 40% identity, and within these limits more than 90% of the atoms in the main chain can be modeled with an accuracy of about 0.1 nm (Sanchez & Šali, 1997). The program *Modeller* (Šali & Blundell, 1993) can be used for homology modeling in conjunction with the programs *PROCHECK* (Laskowski et al., 1993) and *VERIFY_3D* (Luthy et al., 1992) that can be used to assess the

quality of the generated structures. These tools can be used to model both the variable regions of antibodies and antigen.

Starting from the three-dimensional structures of antibody and antigen, the next step is to identify the epitopes on the antigen molecule that are recognized by the antibody. Docking techniques can be used to explore the regions of best fit between antibody and antigen, and the more information about the epitopes better the final solution generated by the docking program. For identification and optimization of regions of the antibody that enter in contact with the antigen, the use of available structural, biological and biochemical information is essential to obtain the most reliable results. Furthermore, the docking approach should not simply focus on solving a problem of rigid body fitting of geometric shapes between the two components, but should recognize that interatomic forces in the context of complex dynamics can give rise to structural changes that can maximize the contacts at the interface of the antibody-antigen complex.

Starting from predictions of antibody-antigen interfaces derived from docking studies, MD simulation can be used for the evaluation of the structures of the proposed antibody-antigen complexes. MD is a powerful tool that can be used to study the structural movements of antibody and antigen and to monitor interatomic interactions in a physiological aqueous environment. With this strategy, a new scFv (Ab) for a particular antigen (Ag) has been developed at the Brazilian company, Cientistas Associados (Cientistas Associados, 2008). It has been demonstrated that the binding between Ag and Ab is dynamic, with structural variations observed in both molecules, which can be improved by changing the contact specificity through the mutation of key residues at the Ag | Ab interface.

Crotoxin is a phospholipase A2 neurotoxin, and is a major component of the venom of the rattlesnake *Crotalus durissus terrificus*, and the lethal effects are due to blockage of neuromuscular transmission. Crotoxin is comprised of two non-identical subunits (CA and CB), which separately present low toxicity. The CA subunit (acidic protein) and CB (basic phospholipase) spontaneously associate in a 1:1 complex, giving rise to the particularly lethal neurotoxin (Faure et al., 1991). Studies reported that three scFv could bind crotoxin (Cardoso et al., 2000). These scFvs recognized distinct epitopes close to CB, however only one scFv was able to effectively neutralize the activity induced by crotoxin.

In order to study this antibody | antigen complex, the structures of crotoxin and scFv were subjected to homology modelling. The template for the crotoxin was the phospholipase A2 from *Crotalus atrox*, which has 50% identity when compared to CB. In the case of a template for the scFv, a search was made of the "framework" (BLAST) and the structure of the CDRs with highest resolution and structural identity was selected. The Ag | Ab complex was obtained with the program HEX (Ritchie, 2003), and since the binding site of the antibody (CDR) is known through epitope mapping (Choumet et al. 2003), the crotoxin was in the correct orientation in relation to the CDRs.

Four MD simulations were performed with the GROMACS software package using a time step of 2 fs at 298 K in the NVT ensemble. A cut-off radius of 1.3 nm was applied. The simulated systems were electrically neutralized by the addition of counter ions and the simulation boxes were designed to avoid overlap of the cut-off radius through the sides of the boxes. The systems include a free Ab in solution (System 1), a free Ag in solution (System 2), Ag | Ab complex (System 3) and mutate Ab in Ag | Ab complex (System 4). In

System 3, the antigen and antibody are initially separated by 0.7 nm. The initial structure of the Ab in the System 4 was mutated in four positions as compared to the Ab in System 3 to enhance the Ag|Ab interaction.

The Ag|Ab complex was formed by superposition of the final simulated antibody and antigen structures by docking. In the System 3, formation of the complex was monitored through the intermolecular interaction potential (IIP) between the Ag and Ab calculated during the simulation (Figure 5).

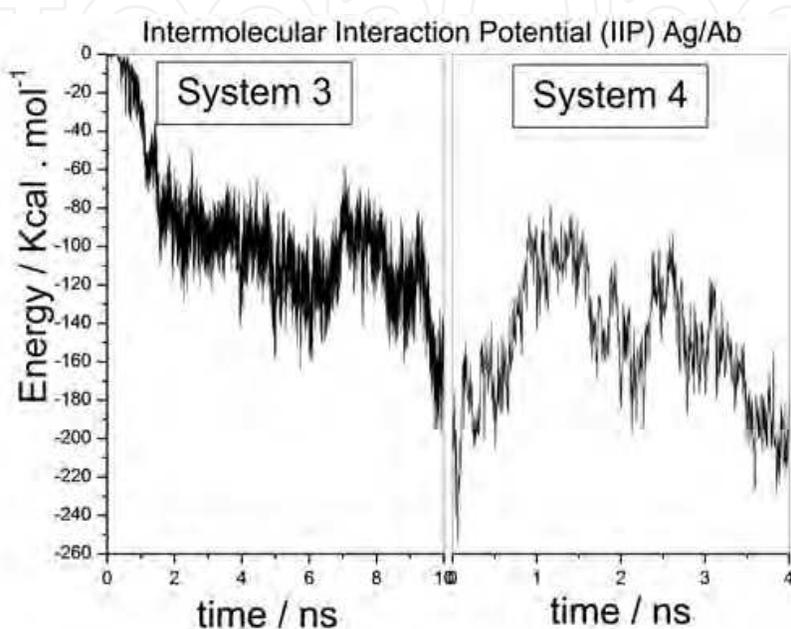


Fig. 5. Intermolecular Interaction Potential between the antigen and the antibody. A) System 3 and B) System 4 (see text for details).

The IIP shows that when Ag and Ab are separated by 0.7 nm (Figure 6a, time = 0 ns), the potential is almost zero and there is no effective interaction. Time around 1.0 ns and IIP = -60 kcal. mol⁻¹ marks the start of the Ag|Ab complex formation (Figure 6b), and from that moment the attractive interaction at the Ag|Ab interface increases, and the potential decreases to approximately -170 kcal.mol⁻¹ in the equilibrium structure at the end of the simulation (Figure 6c). The Ag|Ab complexes obtained from MD simulation are presented in Figure 7. It is evident that the Ag|Ab interface can be extended since the Tyr and Thr residues are distant from the antigen.

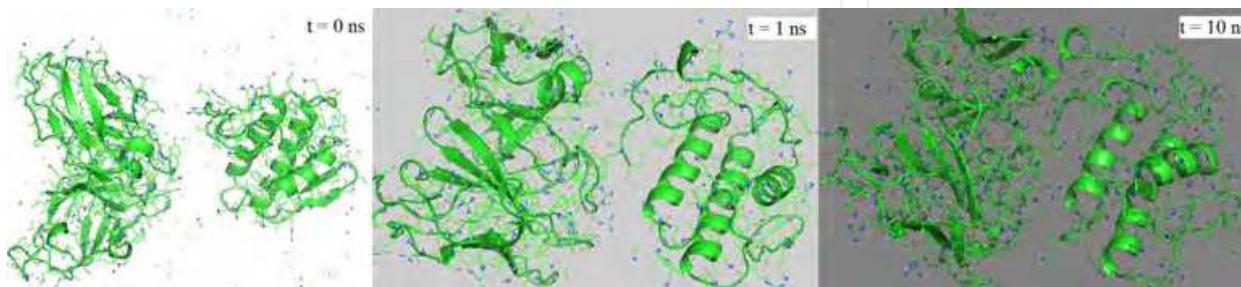


Fig. 6. Configurations of System 3 complex formation between the antibody (left) and antigen (right), at times 0, 1 and 10 ns.

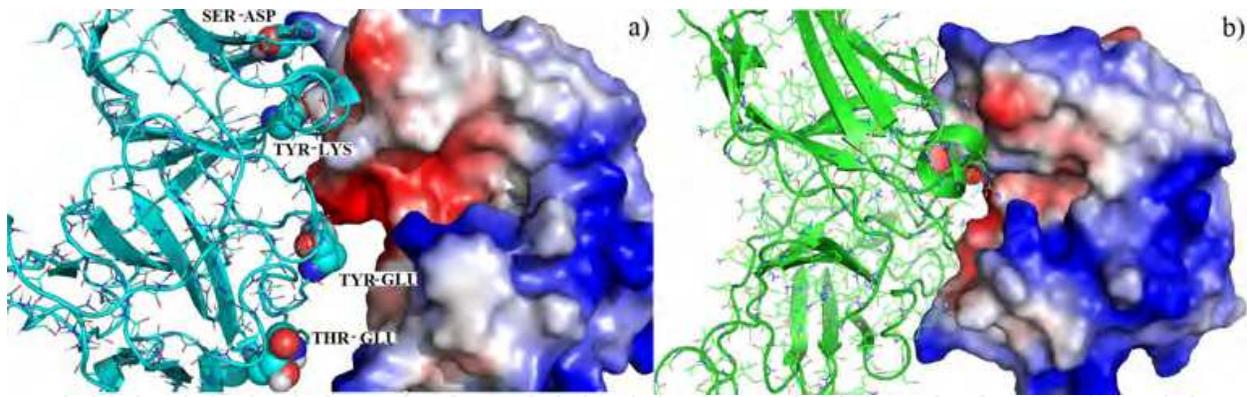


Fig. 7. Final configurations of Systems 3 and 4 from MD simulations, showing the Ag|Ab complex (left and right molecule respectively). Potential isocontours shown at +1 (kT/e) (in blue) and -1 (kT/e) (in red) for permittivities equal to 2 (solute) and to 78.5 (solvent). a) System 3 and b) System 4.

To study the contribution of electrostatic and van der Waals interactions of individual residues to the stability of the interface, four amino acids in the antibody were mutated with the aim of enhancing the Ag|Ab interaction. The substitutions were chosen to optimize the interface by choosing amino acids with opposite charges to those observed in the potential isocontours of the antigen. Figure 8 shows the residues mutated; Thr/Glu, Tyr/Lys, Tyr/Asp and Glu/Ser (System 4). Except the Tyr/Glu mutation, all other residues show a strong residue complementarity, and the polar Ser and Thr were replaced by negatively charged residues.

The most significant results of reduction of the IIP were observed for Thr/Glu and Ser/Asp mutations, which resulted in the addition of two charged residues. IIP for Thr/Glu and Ser/Asp mutations presented a reduction of approximately -10 and -20 kcal.mol⁻¹, respectively. Panels C and G, in Figure 8, show that there was a local change in the potential isocontours of the antigen in System 4. The final structures collected for the System 4 show that the complex is more compact, with a more complementary interface in relation to System 3 (Figures 7a and 7b).

The phenolic side chain of the Tyr residues have both aromatic and polar characteristics, which result in different interaction potential. The electrostatic potential is predominant in the hydroxyl group interaction, while in the aromatic ring the intermolecular interaction is predominantly van der Waals. In the case of Tyr/Glu mutation, the IIP increased as a result of repulsive interaction on the site. In System 3, the IIP of two Tyr residues are attractive, approximately -20 kcal. mol⁻¹, a sufficient energy for Tyr hydroxyl group to form hydrogen bonds. In this context, the mutations and Tyr to Lys and Tyr to Glu proved to be inefficient, since the IIP for the Lys is similar to that of Glu and Tyr shows positive IIP as a result of a local repulsion.

We conclude that uncharged residues at the Ag|Ab interface, such as Tyr residues (System 3), may result in more attractive interactions in relation to charged residues when in a favorable conformation with a good fit, showing that van der Waals forces play an important role in interface stabilization. In addition, charged residues on the antibody interact with antigen residues over longer distances, strengthening the binding between antigen and antibody. Therefore, the formation of Ag|Ab complex is mediated by residues

whose interactions are dominated by van der Waals interactions and which are important for specificity (antigen recognition), while the charged residues are important for the enhancing affinity (binding). A MD strategy can therefore be used in antibody engineering for the evaluation of the proposed changes both locally to evaluate each residue for specificity, and globally to evaluate the effect of residue changes on affinity.

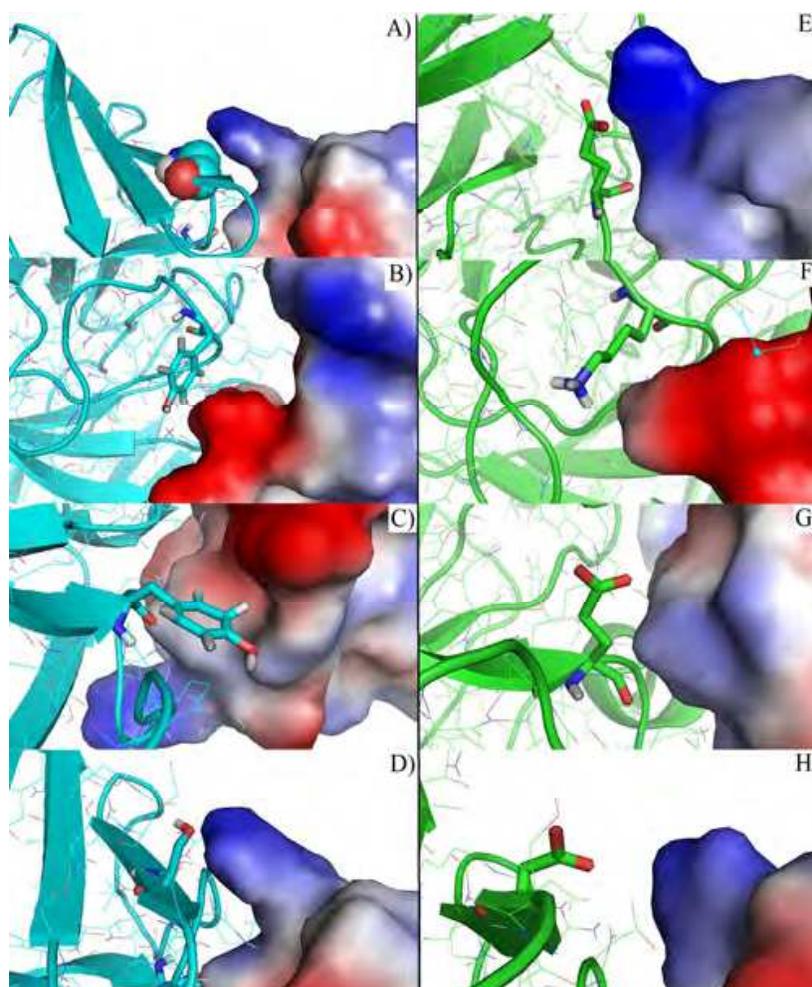


Fig. 8. Final configurations from MD simulations of Systems 3 and 4, showing the local environments of each residue of the antibody (left) compared to the antigen (right), represented as potential isocontours. Panels A, B, C, and D show the antibody residue in System 3 and panels E, F, G and H show the mutated residues in System 4.

3.3 Antimicrobial peptides

Antimicrobial peptides (AMP) play an important role in the innate immune defense system in all organisms and display activity against a wide variety of microorganisms. AMP are generally comprised of fewer than 50 amino acid residues, and are characterized both by an overall positive charge due the presence of multiple lysine and arginine residues and by a large proportion (at least 50 %) of hydrophobic residues. The Antimicrobial Peptide Database (APD) (Wang & Wang, 2004) contains more than 1750 AMPs, despite this experimental effort, knowledge with respect to the structures of these peptides remains limited since 3D structures are available for only 229 (13% of the total) of these AMPs. Therefore, it is evident that the

investigation of AMPs structure and dynamics will yield knowledge that is important for the understanding of the mode of action of these biomolecules.

During the last decade, the accumulation of a large body of experimental data has demonstrated that AMPs act by predominantly affecting the integrity of cell membranes through their interaction with phospholipids. The cytoplasmic membranes of multicellular organisms and bacteria have distinct lipid compositions, and the specificity of AMPs activity is therefore determined not only by the physico-chemical properties of the peptide but also by the composition of the target cell membranes. An understanding of the interactions between AMP and cell membrane at a molecular level is therefore of importance for the development of AMPs as therapeutic agents, which ideally would have a potent antimicrobial activity with low toxicity against host cells. The cytoplasmic membrane of the Gram negative bacteria *Escherichia coli* contains 70 to 80% neutral lipids with phosphatidylethanolamine (PE) head groups, 20 to 25% with phosphatidylglycerol (PG) head groups, which imparts a negative charge to the membrane, and other lipids that are present in smaller quantities (Dowhan, 1997). In eukaryotic cells, the extracellular leaflet of the cytoplasmic membrane is composed predominantly of lipid with phosphatidylcholine (PC) head groups (Mateo et al., 2006). It is crucial that these differences in lipid composition between eukaryotic and bacterial cell membranes be taken into account for investigations with respect to the toxic and antimicrobial activities of AMPs.

AMPs with altered amino acid sequences have been synthesized with the aim of decreasing toxicity and increasing cell selectivity, for example, the indolicidin (Selsted et al., 1992). Indolicidin (IND) is a peptide containing 13 amino acid residues (ILPWKWPWWPWR-NH₂) and contains a large portion of tryptophan residues (39%). The IND is amidated at its C-terminus, displays activity against a broad range of microorganisms however is toxic to lymphocytes and erythrocytes. Two mutants of IND, a Pro/Ala mutant (CP10A ILAWKWAWWAWWR-NH₂) and the mutant CP11 (ILKKWPWWPWRK-NH₂) that contains two extra positive charges, have been the target of several functional studies (Zhang et al., 2001; Halevy et al., 2003). This set of peptides is of interest as a model system since the toxic activity of the CP10A mutant is higher against human erythrocytes when compared to the native IND, while the CP11 has the lowest hemolytic activity among the three peptides (Halevy et al., 2003). Conversely, the antimicrobial activity is greatest for the CP11 and lowest for the CP10A (CP11>IND>CP10A) (Zhang et al., 2001). The functional properties of these mutants is therefore influenced by the physico-chemical characteristics of the peptide, and provides an excellent model system to study the peptide | membrane interaction using MD simulations.

MD studies were carried out with the IND and the two mutants CP10A and CP11 were used as model peptides to study the interaction of the peptides with two membrane models: one containing dipalmitoylphosphatidylcholine (DPPC) with 64 lipids in each layer, and the other containing a mixed bilayer with 64 lipids in each layer formed by 75 % of dipalmitoylphosphatidylethanolamine (DPPE) and 25 % of dipalmitoylphosphatidylglycerol (DPPG). Both bilayer systems were solvated and electrical neutrality of the mixed bilayer was maintained by the addition of one sodium ion for each DPPG molecule. Six systems were studied by introducing one of the peptides (IND, CP10A or CP11) into the bulk solution within a distance of about 4 nm from the center of the mass (CM) of one of the bilayers (DPPC or DPPE/DPPG). The simulations were initiated by energy minimization using the steepest

descent algorithm, in order to eliminate bad contacts and undesirable forces. For each system, an initial simulation was performed for 0.5 ns applying restrictions to the peptides and bilayer atomic coordinates to equilibrate the systems using a $dt = 0.5$ fs. The restrictions were removed and the systems were initially simulated by 30 fast heating and cooling cycles of simulated annealing (Fuzo et al., 2008) with $dt = 2$ fs. In the next stage, the systems were submitted to 20 ns NpT simulations with dt of 2 fs, with 50 ns total simulation time for each system. Full details of the simulation parameters are given in Fuzo & Degrève (2009, 2011).

At the start of the simulations, the peptides were positioned in the solvent phase in order to allow a free interaction with the bilayer. During the simulation the peptides diffused through the solvent toward the membrane, before associating with the bilayers as observed in Figure 9. Differences were observed when the peptides were inserted into the two types of bilayers, and the mean position of the peptide CMs in relation to the bilayer CMs, given by a value d as measured along the axis perpendicular to bilayer plane, and which is lower when the peptide is inserted into the bilayer.

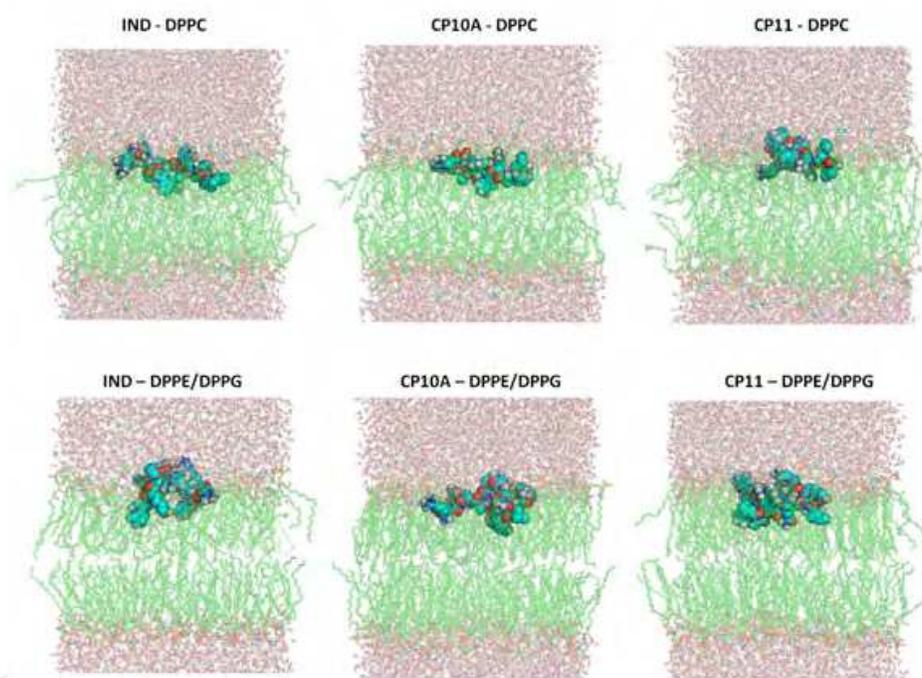


Fig. 9. Final configurations of systems containing peptides inserted into DPPC (top) and DPPE/DPPG (bottom) bilayers.

Table 1 shows that in the case of DPPC bilayers, all d values were less than the average position of the phosphorus atoms of DPPC (1.8 nm) in the order $CP11 > IND > CP10A$, showing that the CP10A peptide inserted almost to the centre of the bilayer, the CP11 peptide remained close to the interface of the bilayer with the aqueous phase. The behavior of the IND peptide was intermediate between the two extremes. In the DPPE/DPPG systems the peptides also inserted into the bilayer giving d values that were lower than the phosphorous atoms, however it was observed that the order of insertion of the peptides was altered (ie. $CP11 > CP10A > IND$) as compared to the DPPC bilayer. In all cases, the peptide that inserts deeper into a given bilayer is that which has the highest activity against a cell membrane of the same composition, thus in the systems studied with the DPPC bilayer the

peptides that show a deeper insertion are those that have the highest hemolytic activity (Halevy et al., 2003), whereas for the DPPE/DPPG systems those peptides with higher antimicrobial activity were more deeply inserted (Zhang et al., 2001). This correlation between membrane activity and peptide insertion may have an important impact for the design of new AMPs, since more effective mutants can be designed rationally based on the information obtained from MD simulation.

Peptide	d (nm)		N_{HB}	
	DPPC	DPPE/DPPG	DPPC	DPPE/DPPG
IND	1.43 ± 0.06	2.02 ± 0.05	24.4 ± 1.6	$(13.8 \pm 1.8)/(4.1 \pm 0.8)$
CP10A	1.24 ± 0.08	1.84 ± 0.08	20.8 ± 1.6	$(8.1 \pm 1.5)/(10.0 \pm 1.6)$
CP11	1.61 ± 0.07	1.70 ± 0.06	25.7 ± 2.1	$(12.5 \pm 1.5)/(4.0 \pm 1.0)$

Table 1. Distances between the CM of peptides and bilayers along the axis perpendicular to bilayer plane (d), and the number of hydrogen bonds between the peptides and bilayers (N_{HB}).

Two important differences were observed between the interactions of the peptides and the two different bilayers. The first was a greater number of hydrogen bonds (N_{HB}) between the peptides and oxygen atoms of the DPPC bilayers (Table 1), which was due to a stretched conformation of the phosphatidylcholine groups of the DPPC molecule (Figure 10A) in which the distance between the CM of the phosphate and choline groups was 0.47 ± 0.06 nm, which compares with the phosphatidylethanolamine groups in the **DPPE/DPPG bilayer** where the distances between the phosphate and amine groups was 0.37 ± 0.01 nm. The proximity between the amine and phosphate groups in the DPPE molecules makes causes greater difficulty for the positively charged groups in peptides to approach the lipid headgroup phosphate group due to electrostatic repulsion with the positively charged amino group, thereby hindering the formation of HBs between the peptides and oxygen atoms. The second difference was the presence of cation- π interactions between choline groups of the DPPC and the side-chains of some Trp residues of the peptides. The cation- π interactions were evaluated by determining the radial distribution functions between the CM of the tryptophan side chains and nitrogen atoms of the choline groups, where the presence of the first peak (D_{max}) at distances smaller than 0.5 nm and a number of choline groups greater than one (N_{Chol}) is indicative of cation- π interactions. These cation- π interactions were found with 1 to 2 choline groups for some tryptophan residues for all three peptides (as shown in Table 2). Examples of the proximity of the choline headgroups for each peptide are shown in Figures 10C, 10D, and 10E for IND, CP10A, and CP11, respectively. The cation- π interactions observed between the choline groups and the tryptophan residues do not occur in lipid bilayers containing PE due the geometry of this group, in which the amine and phosphate groups are closer, thereby impeding the approximation of the tryptophan side chains to the amine groups to form cation- π interactions.

The simulations of the three peptides have shown that cation- π interactions between the choline headgroup and the tryptophan make an important contribution to the recognition of eukaryotic membranes by these peptides, thereby indicating that this type of interaction must be considered when designing new AMPs.

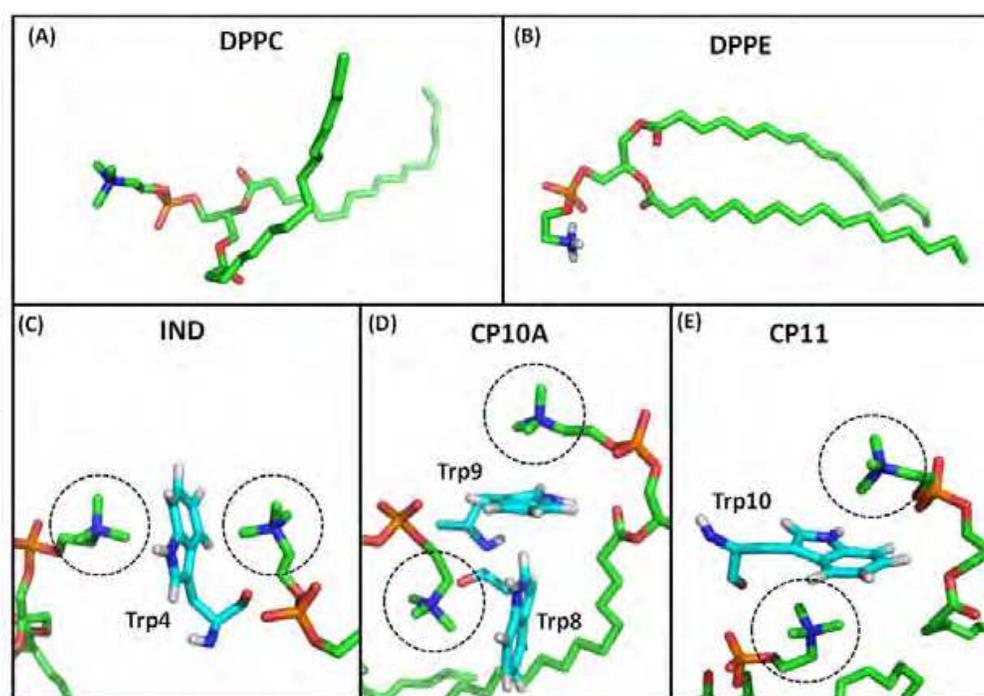


Fig. 10. Typical geometries of DPPC (A) and DPPE (B) molecules obtained from simulations. In (C), (D), and (E) are shown some examples of choline group neighbors of Trp residues for peptides IND, CP10A, and CP11, respectively. Choline groups are circled by a dotted line.

Residue	N_{Chol}	D_{max} (nm)
IND		
Trp4	2.0	0.46
Trp8	1.0	0.44
CP10A		
Trp8	1.0	0.48
Trp9	1.0	0.46
Trp11	1.1	0.48
CP11		
Trp5	1.0	0.46
Trp7	2.0	0.48
Trp10	2.1	0.45

Table 2. Parameters extracted from radial distribution functions between the CM of Trp side chains and nitrogen atoms of choline groups.

4. Trends in structural bioinformatics

One of the most important tools in the theoretical studies of bio(macro)molecule is the MD simulation. MD calculates the time dependent properties of a (bio)molecular system. Current MD simulations of biomolecules are directed to investigate structural and dynamics properties of systems composed of very large number of atoms over timescales

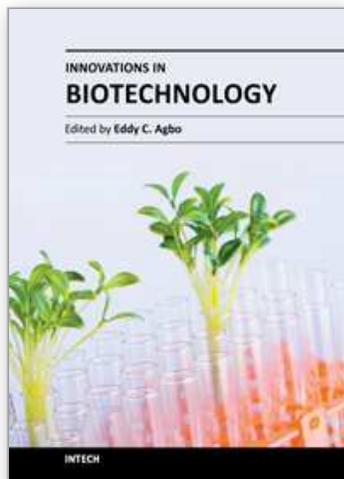
of many nanoseconds to microseconds (case of peptides folding). The continued growth of computer power and the advent of new algorithms like GPU-accelerated calculations are closely related to the more realistic approach of the molecular systems that allows chemical processes in condensed phases to be studied in an accurate and unbiased way. For example, MD trajectories can be generated with forces obtained from accurate electronic structure details (*ab initio* MD), as well as the improvement of the calculation efficiency when dealing with the complexity of biological problems. In this context we can cite a pharmacological interesting system, the case of quaternary structures of virus proteins and their interactions with active biological molecules, and the protein-carbohydrate interactions in glycoproteins. Future bioinformatics oriented studies will likely provide us with valuable tools for validating, improvement, prediction and development of a variety of biotechnological applications.

5. References

- Aehle, W. (2004). *Enzymes in Industry – Production and Applications*, John Wiley, ISBN 352-729-592-5, New York.
- Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K. & Watson, J.D. (1989). *In: The Immune System*. (Robertson, M., Ed.), Garland, New York.
- Allen, M. P. & Tildesley, D. J. (1992). *Computer Simulation of Liquids*, Clarendon Press, ISBN 978-019-855645-9, Oxford.
- Åqvist, J. (April 13 2007). The Åqvist Group, May 28 2011, Available from: <<http://xray.bmc.uu.se/~aqwww/>> .
- Beg, Q. K.; Kapoor, M.; Mahajan, L.; Hoondal, G. S. (2001). Microbial xylanases and their industrial applications: a review. *Appl Microbiol Biotechnol.*, 56, April 27, pp. 326-338 ISSN 0175-7598.
- Cardoso, D. F.; Nato, F.; England, P.; Ferreira, M. L.; Vaughan, T. J.; Mota, I.; Maize, J. C.; Choumet, V. & Lafaye, P. (2000). Neutralizing human anti crotoxin scFv isolated from a nonimmunized phage library. *Scandinavian Journal of Immunology*, 51, pp.337-344, ISSN 1365-3083.
- Choumet, V.; Faure, G.; Robbe-Vincent, A.; Saliou, B.; Mazié, J. C. & Bon, C. (2003). Immunochemical analysis of a snake venom phospholipase A2 neurotoxin, crotoxin, with monoclonal antibodies. *Molecular Immunology*, 29, pp. 871-882, ISSN 0161-5890.
- Cientistas Associados, (2008) *Intelligent System for developing antibodies based on the structure of the antigen (AbEvo)*, last accessed 28/07/2011, Available from: <<http://www.cientistas.com.br/projetos/projects.htm>>.
- Degrève, L.; Brancaleoni, G. H.; Fuzo, C. A.; Lourenzoni, M. R.; Mazzé, F. M.; Namba, A. M. & Vieira, D. S. (2004). On the role of water in the protein activity. *Brazilian Journal of Physics*, 34, March 2004, pp. 102-115, ISSN 0103-9733.
- Dowhan, W. (1997). Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annual Review of Biochemistry*, 66, July 1997, pp. 199-232, ISSN 0066-4154.
- Faure, G.; Guillaume, J. L.; Camoin, L.; Saliou, B. & Bon, C. (1991). Multiplicity of acidic subunit isoforms of crotoxin, the phospholipase A2 neurotoxin from *Crotalus durissus terrificus* venom, results from posttranslational modifications. *Biochemistry*, 32, pp. 8074-8083, ISSN 0006-2960.

- Fernández, C. & Wider, G. (2003). TROSY in NMR studies of the Structure and Function of Large Biological Macromolecules. *Current Opinion in Structural Biology*, 13, October 2003, pp. 570-580 ISSN 0959-440x.
- Fuzo, C.A. & Degrève, L. (2011). Study of the antimicrobial peptide indolicidin and mutants in eukaryotic modelled membrane by molecular dynamics simulations. *Molecular Physics*, 109, January 2011, pp. 289-300, ISSN 1362-3028.
- Fuzo, C.A. (2009). *Molecular simulation studies of peptide/bilayer systems: application to structure/activity relationship of the indolicidin and mutants*. Doctoral Thesis, University of São Paulo, Brazil.
- Fuzo, C.A.; Castro, J.R.M & Degrève, L. (2008). Searching the global minimum of a peptide/bilayer potential energy surface by fast heating and cooling cycles of simulated annealing. *International Journal of Quantum Chemistry*, 108, June 2008, pp. 2403-2407, ISSN 1097-461x.
- Georis, J.; Esteves, F. L.; Lamotte-Brasseur, J.; Bougnet, V.; Devreese, B.; Giannotta, F.; Granier, B. (2000). An additional aromatic interaction improves the thermostability and thermophilicity of a mesophilic family 11 xylanase: structural basis and molecular study. *Protein Sci.*, 9, September 8, pp. 466-475 ISSN 0961-8368.
- Halevy, R.; Rozek, A.; Kolushevav, S.; Hancock, R.E.W. & Jelinek, R. (2003). Membrane binding and permeation by indolicidin analogs studied by a biomimetic lipid/polydiacetylene vesicle assay. *Peptides*, 24, November 2003, pp. 1753-1761, ISSN 0196-9781.
- Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, W. R.; Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79, July 15, pp. 926-938 ISSN 0021-9606.
- Laskowski, R. A.; Macarthur, M. W.; Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26, pp. 283-291, ISSN 1600-5767.
- Leach, A. R. (2001). *Molecular Modelling: Principles and Applications*, Prentice Hall, ISBN 058-238-210-6, Harlow, England.
- Luthy, R.; Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with threedimensional profiles. *Nature*, 356, p.83-85, ISSN 0028-0836.
- Maitland, G. C. & Smith, G. C. (1971). The Intermolecular Pair Potential of Argon. *Molecular Physics*, 22, January 1971, pp. 861-868 ISSN 1362-3028.
- Maitland, G. C.; Rigby, M., Smith, E., B., & Wakeham, W. A. (1981). *Intermolecular Forces : Their Origin and Determination*, Clarendon Press, ISBN 019-855-611-x, Oxford.
- Mancini, E. J.; Clarke, M.; Gowen, B. E.; Rutten, T. & Fuller, S. D. (2000). Cryo-Electron Microscopy Reveals the Functional Organization of an Enveloped Virus, Semliki Forest Virus. *Molecular Cell*, 5, February 2000, pp. 255-266 ISSN 1097-2765
- Mateo, C.R.; Gómez, J.; Villalaín, J. & Ros, J.M.G. (2006). *Protein-Lipid Interactions: New Approaches and merging Concepts*. Springer-Verlag, ISBN 3-540-28400-1, Germany.
- McQuarrie, D. A. (1976). *Statistical Mechanics*, Harper and Row, New York.
- Morokuma, K. (1977). Why do Molecules Interact? The Origin of Electron Donor-Acceptor Complexes, Hydrogen Bonding and Proton Affinity. *Accounts of Chemical Research*, 10, August 77, pp. 294-300 ISSN 0001-4842..
- Rapaport, D. C. (2005). *The Art of molecular Dynamics Simulation*, Cambridge University Press, ISBN 052-182-568-7, Cambridge, England.

- Ritchie, D.W. (2003). Evaluation of protein docking predictions using *Hex* 3.1 in CAPRI rounds 1 and 2, *Proteins: Structure, Function, and Bioinformatics*, 52, pp. 98-106, ISSN 1097-0134.
- Rocco, A. G.; Mollica, L.; Ricchiuto, P.; Baptista, A. M.; Gianazza, E.; Eberini, I. (2008). Characterization of the Protein Unfolding Processes Induced by Urea and Temperature. *Biophysical Journal*, 94, November 2010, pp. 2241-2251 ISSN 0006-3495.
- Roitt, I. M. (1997). *Essential immunology*, Oxford: Blackwell Science, Malden, MA.
- Šali A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234, pp.779-815, ISSN 0022-2836.
- Sánchez, R. & Šali A. (1997). A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, 1, pp. 50-58.
- Selsted, M. E.; Novotny, M. J.; Morris, W. L.; Tang, Y. Q.; Smith, W. & Cullor, J. S. (1992). Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *Journal of Biological Chemistry*, 267, March 1992, pp. 4292-4295, ISSN 1083-351x.
- Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y. & Wriggers, W. (2010) Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*, 330, October 2010, pp. 341-346 ISSN 0036-8075.
- Torrez, M.; Schultehenrich, M.; Livesay, D. R. (2003) Conferring Thermostability to Mesophilic Proteins through Optimized Electrostatic Surfaces. *Biophys J.*, 85, November 1, pp. 2845-2853 ISSN 0006-3495.
- van Gunsteren, W. F.; Dolenc, J. & Mark, A. E. (2008). Molecular Simulation as an Aid to Experimentalist. *Current Opinion in Structural Biology*, 18, February 2008, pp. 149-153 ISSN 0959-440x.
- Vieira, D. S., Degrève, L. (2009). An Insight into the Thermostability of a Pair of Xylanases : The Role of Hydrogen Bonds. *Molecular Physics*, 107, February 26, pp. 59-69 ISSN 0026-8976.
- Wakarchuk, W. W.; Sung, W. L.; Campbell, R. L.; Cunningham, A.; Watson, D. C.; Yaguchi M. (1994). Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bonds. *Protein Eng.*, 7, July 17, pp. 1379-1386 ISSN 0269-2139.
- Wang, Z. & Wang, G. (2004). APD: the Antimicrobial Peptide Database. *Nucleic Acids Research*, 32, January 2004, pp. D590-D592, ISSN 1362-4962.
- Zhang, L.; Rozek, A. & Hancock, R.E.W. (2001). Interaction of cationic antimicrobial peptides with model membranes. *Journal of Biological Chemistry*, 276, July 2001, pp. 35714-35722, ISSN 1083-351x.



Innovations in Biotechnology

Edited by Dr. Eddy C. Agbo

ISBN 978-953-51-0096-6

Hard cover, 474 pages

Publisher InTech

Published online 17, February, 2012

Published in print edition February, 2012

Innovations in Biotechnology provides an authoritative crystallization of some of the evolving leading-edge biomedical research topics and developments in the field of biotechnology. It is aptly written to integrate emerging basic research topics with their biotechnology applications. It also challenges the reader to appreciate the role of biotechnology in society, addressing clear questions relating to biotech policy and ethics in the context of the research advances. In an era of interdisciplinary collaboration, the book serves an excellent indepth text for a broad range of readers ranging from social scientists to students, researchers and policy makers. Every topic weaves back to the same bottom line: how does this discovery impact society in a positive way?

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Davi S. Vieira, Marcos R. Lourenzoni, Carlos A. Fuzo, Richard J. Ward and Léo Degève (2012). Structural Bioinformatics for Protein Engineering, Innovations in Biotechnology, Dr. Eddy C. Agbo (Ed.), ISBN: 978-953-51-0096-6, InTech, Available from: <http://www.intechopen.com/books/innovations-in-biotechnology/structural-bioinformatics-for-protein-engineering>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen