Thesis for doctoral degree (Ph.D.) 2006

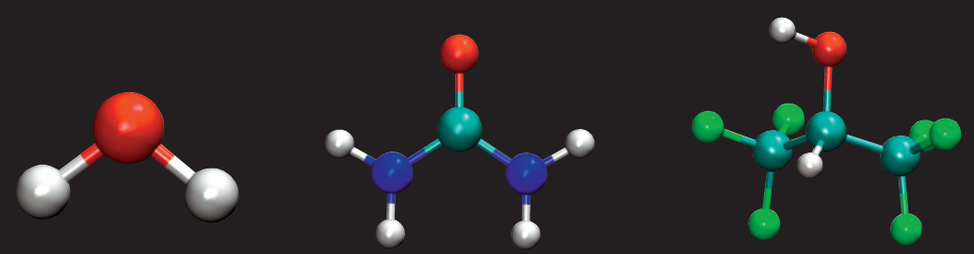
Environmental Influence on α-Helical Peptides

Ana Caballero Herrera

Karolinska Institutet

Thesis for doctoral degree (Ph.D.)
2006

Environmental Influence on α-Helical Peptides



Ana Caballero Herrera

Karolinska Institutet

From the Department of Biosciences and Nutrition
Karolinska Institutet, Stockholm, Sweden

ENVIRONMENTAL INFLUENCE ON α-HELICAL PEPTIDES

Ana Caballero-Herrera

Karolinska Institutet

Stockholm 2006

All previously published papers were reproduced with permission from the publisher.
Published and printed by
REPROPRINT AB
Stockholm 2006
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

ISBN 91-7140-949-1

ABSTRACT

Most proteins at physiological conditions fold into a native functional three-dimensional conformation. The stability of the Native state is a balance between different interactions within the protein and with the solvent. The role of the solvent is therefore important for the stability of proteins and determinant for their three-dimensional structure and function.

Secondary structure elements such as α helices are often found in the nucleation core structures proposed by the nucleation-diffusion model for protein folding. Understanding their stability and their folding/unfolding mechanism can shed light on the protein folding problem.

This thesis is devoted to the study of α-helical peptides in different environments, where they show very different behavior.

Proteins unfold at high denaturant concentrations, such as urea or guanidinium. Their solubility increases with denaturant concentration. The particular properties of the urea-aqueous solutions are still not well understood. Urea dimerization has been pointed out as a fundamental factor for the thermodynamical behavior of urea-water solutions. In the first work, we calculate urea molecule parameters consistent with the CHARMM force field and TIP3 water model taking in account the urea dimerization feature. Simulations of 2M and 8M urea contained significant amount of cyclic dimers with very favorable interactions as well as small amount of head-to-tail dimers, as found in crystals Comparisons with potentials obtained from *ab initio* calculations indicate the non adequacy of the parameters obtained with this methodology. Furthermore, the DFT results indicate that urea may adopt a planar conformation in solution as it does in crystals.

The urea parameters previously obtained were used to perform MD simulations of two peptides: a C-peptide analogue and a variant (same amino acid composition but different sequence) in 8M urea-water solution and in water at different temperatures. The results indicated that the stability of the C-peptide arises from electrostatic side chain-side chain interactions. In contrast, the second peptide was unstable in all solvents. In the presence of urea the ability of water molecules to form hydrogen bonds to the peptide as well as their life time is increased. The rotation of the water molecules is reduced on the peptide surface. In addition, urea formed long lived hydrogen bonds to the peptide and accumulated in excess on the peptide surface.

MD simulations of the putative α-helical transmembrane (TM) spanning domain of two glycoproteins E1 and E2 of the Semliki Forest virus wild type and mutants were performed. The two helices pack in left handed two-stranded rope fashion. The residues that occupy the important positions of the heptad (***abcdefg***) were identified. Success of packing was a compromise between small and medium residues at E1 and E2 interfaces that allow hydrogen bond formation. Defects on packing were predicted. The method can be extrapolated to sketch TM helical packing in other alphaviruses.

The dynamical properties of the peptide hormone motilin (22 amino acids) were studied. The MD simulations in water and HFIP coincide with the NMR and fluorescence anisotropy decay (FAD) results. However, the choice of the vectors under study is very important to get agreement. NMR relaxation parameters and spectral density functions were also obtained from the simulations which evidenced the α-helical character of the peptide but also its high flexibility even in the HFIP-water solution. The results point out the difficulty for using NMR dynamical analysis with flexible peptides.

LIST OF PUBLICATIONS

I. **Urea Parametrization for Molecular Dynamics Simulations**
A. Caballero-Herrera and L. Nilsson,. *Mol. Struct.: THEOCHEM,* 2006; **758** (2-3): 139-148

II. **Effect of Urea on Peptide Conformation in Water: Molecular Dynamics and Experimental Characterization.**
A. Caballero-Herrera, K. Nordstrand, K. D. Berndt, and L. Nilsson; *Biophys. J*., 2005; **89**(2): 842 - 857

III. **Molecular Dynamics Simulations of the E1/E2 Transmembrane Domain of the Semliki Forest Virus**
A. Caballero-Herrera and L. Nilsson; *Biophys. J*., 2003; **85**(6): 3646-3658

IV. **Dynamics of the Flexible Peptide Motilin by NMR, Fluorescence Anisotropy and Molecular Dynamics Analysis**
A. Caballero-Herrera, T. Massad, P. Damberg, A. Gräslund and L. Nilsson, 2006 (Manuscript)

CONTENTS

LIST OF ABBREVIATIONS

ABNR	Adopted Basis Newton-Raphson
AMBER	Assisted Model Building with Energy Refinement
ASA	Accesible Surface Area
BLYP	Becke gradient-exchange correction and Lee-Yang-Parr correlation Functional
CHARMM	Chemistry at HARvard Molecular Mechanics
CD	Circular Dichroism
DFT	Density Functional Theory
D	Denatured State
ENCAD	Energy Calculation and Dynamics
GpA	Glycophorin A
GROMOS	GROningen Molecular Simulation
HF	Hartree-Fock
HFIP	Hexafluoroisopropanol
J(ω)	Spectral Density Function
IR	Infrared Spectroscopy
MD	Molecular Dynamics
MO	Molecular Orbital
N	Native State
NOE	Nuclear Overhauser effect
NMR	Nuclear Magnetic Resonance
OPLS	Optimized Parameters for Liquid Simulations
PBC	Periodic Boundary Conditions
PME	Partcle Mesh Ewald
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SCF	Self Consistent Field
SD	Steepest-Descent
SFV	Semliki Forest Virus
STO	Slater Type Orbitals
TM	Transmembrane

To my father

You cannot teach a man anything,
you can only help him to find it for himself.
Galileo Galilei

1 INTRODUCTION

Proteins at physiological conditions fold into a unique three-dimensional conformation. This three-dimensional structure, in turn determines the function of the proteins. The information necessary for proteins and peptides to fold is encoded in their sequence. However, the mechanism by which they fold is still not clearly understood and its importance is becoming more critical due to the fact that protein misfolding is known to be involved in many human diseases. In the last few decades, experimental and theoretical studies together have advanced the understanding of protein folding. From this synergetic effort, a unified model describing protein folding called the *diffusion-collision* model (Islam et al., 2004) has emerged. This model proposes that proteins are formed by an initial creation of a folding core or nucleus that diffuses towards the native state. Secondary structure elements such as α helices are key elements of this folding model. They are commonly found at the core of the intermediate compact structures but also they have been found in the denatured state (Religa et al., 2005). Understanding the mechanism of α helical folding/unfolding is therefore important as a fundamental approach to address the protein folding problem.

The main features of the mechanism of the protein folding process have been obtained using simplified lattice models of proteins. These models explain how folding can be achieved by polypeptides in a reasonable time, solving the *Levinthal paradox*. In fact polypeptides do not have to search their native state by a random search, but only through a reduced fraction of the free energy hypersurface. This is possible because folding is biased towards the native state, the global minimum of the free energy landscape. The achievement of a correct folding depends on the appropriate balance between the effective energy that favors the native state and the configurational energy that favors the denatured state. The use of molecular dynamics (MD) simulations, which give atomic description of the motion as a function of time, is necessary to study the folding process of a specific polypeptide. However, due to the conformational sampling problem these studies are mainly limited to the analysis of small peptides.

Also the role of the solvent is of crucial importance since the stability of the native conformation is a balance between different interactions within the protein and with the solvent. In fact, the hydrophobic effect is one of the main driving forces in protein folding. A change in the protein environment leads to a conformational change in the protein. For instance, adding high amounts of denaturant such as urea to the solvent induces protein denaturation and the use of fluorinated cosolvents or insertion of the protein in lipid bilayers, in most cases favors α-helical formation.

Although urea is widely used as denaturant agent the molecular mechanism by which urea induces unfolding is still not clear. Two main theories have been proposed to explain the properties of urea induced denaturation. The *indirect mechanism* and the *direct mechanism*. In the first one urea acts as "water structure breaker" whereas in the second protein denaturation is induced by the direct action of urea. The first part of this work is focused on the understanding of this mechanism as well as on the process of unfolding of α-helical peptides. A new urea force field was developed for this purpose.

This thesis is also devoted to the study of alpha helical peptides in low dielectric environments and in HFIP-water solution to address some issues in peptide stability, structure and dynamics with emphasis on the role of the environment.

2 PROTEIN FOLDING AND STABILITY

2.1 Protein Folding

Proteins fold into unique highly structured functional states. The information for folding is completely coded in the sequence of the protein. However, the understanding of the protein folding process is still one of the biggest challenges in molecular biology. The important role it could play for assignment of structural and functional properties to the vast amount of protein primary structures available from the different genomics projects is evident. But also the understanding of folding/unfolding processes per se is fundamental for developing novel therapeutic strategies that deal with human diseases associated with misfolding such as cystic fibrosis and in some case to amyloid formation such as Alzheimer.

In nature proteins fold rapidly to their native structure (the lowest energy conformation), however if proteins should search for their native state among all the possible conformations theoretically available, they would never reach the native state in a finite time. This was called the *'Levinthal paradox'* . Levinthal solved the paradox adducing that proteins fold following a "folding pathway", instead to through a random search. The folding pathway perspective or *sequential micropath perspective* is known as the "classical" view of folding. The "new view" which appeared more recently, takes into account the concept of averaged effective energy surfaces or energy landscapes. Here a configuration, in particular the denatured state, is not a unique state but an ensemble of states and ensemble averages are needed. Therefore it is also known as *the ensemble perspective*. In this perspective, the effective energy surface of a polypeptide will bias the folding towards the native state leading to a rapid reduction of the accessible configurational space, thereby reducing the configurational entropy (Dobson et al., 1998; Dinner et al., 2000; Onuchic and Wolynes, 2004). The balance between the decrement in effective energy which favors the native state, and conformational entropy, which favors the denatured state, is what ultimately will determine if a protein folds in a finite time.

The biggest theoretical advances in protein folding have arisen from lattice models with simplified potentials together with statistical mechanics. Although these methods can not explain how a specific protein folds, they outline the general features of protein folding. In fact theories about a unified theory of folding has emerged from these models, sometimes called diffusion-collision model (Karplus and Weaver, 1994) or other groups call it nucleation-collapse mechanism(Daggett and Fersht, 2003). Processes of folding/unfolding of small peptides with all atom representation for the peptide and implicit solvent, have also been simulated as in the case of a three stranded β-sheet peptide(Cavalli et al., 2002).

These theories suggest that there are many ways to reach the native state. Moreover, they propose a folding funnel where proteins fold falling energetically down to the end of an "energetic funnel landscape", the native state, as balls roll down on a irregular funnel. Furthermore, the entropy is reduced when rolling down since when the protein advance toward the native configuration there are fewer available conformations(Brooks, 2002).

Also some theories make use of the analogy with chemical reactions of small molecules(Dobson et al., 1998; Dinner et al., 2000), where the reactant is the denatured state and the product the native state. In contrast to the small molecules chemical reactions in this case configurational entropy plays an essential role. Although it is not simple to define the reaction coordinates, a useful reaction coordinate is the fraction of native-like contacts that allows visualizing the funnel–like nature of the protein folding

process: The energy and the number of conformations accessible to the polypeptide decreases as the number of native contacts increases. The conclusion from these simplified models of proteins is that a random polypeptide chain collapses rapidly into an ensemble of compact species. The hydrophobically driven collapse drastically reduces the number of accessible configurations. This collapse can be disorganized with no or few native-like contacts, or compact, also called molten globules, with high amount of probable native-like contacts. The former kind of collapse has large configurational entropy whereas the second has low configurational entropy. After the collapse the polypeptide evolves searching for the transition state (Daggett and Fersht, 2003), from which the native state is reached immediately. For the first case of compact states this is a slow process where the polypeptide fluctuations have to lead to the formation of the nativelike contacts. In the second case the structural preferences and formation of main chain hydrogen bonds (α helices and β strands) will give a "folding core" or "nucleus" that will serve as foundation for the assembly of the rest of the polypeptide chain. However, pseudo-stable incorrect structures (intermediate states) can be formed and the structure has to go back into its search of the native state slowing down the process of folding. Furthermore, to prevent misfolding the penalty for nonnative like interactions increases through out the folding process. Only the final stage of folding seems to need the side chain interactions for specific side chain compact packing, which indeed is of critical importance to differentiate between the native state and misfolded structures.

This form the basis of the unified folding mechanism.

Experiments such as NMR, CD and IR together with protein engineering can contribute to the verification of this theory, to find transition and intermediate states, and to understand the folding processes of specific proteins. Additionally, the complementation of these experiments with MD simulations can give a more complete and detailed picture of the folding processes.

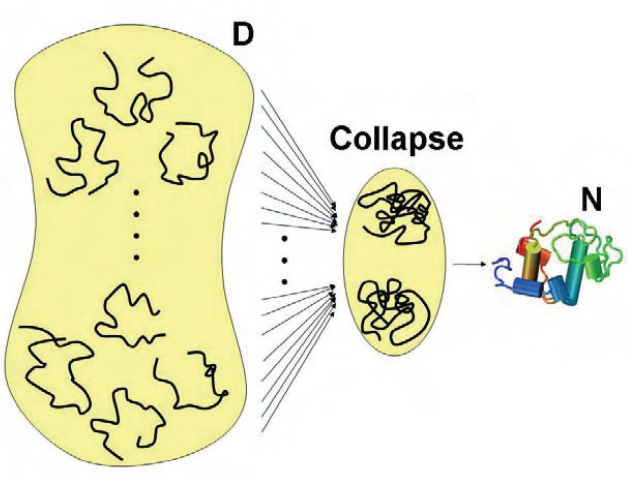


Figure 1. Protein folding mechanism.

2.2 Protein Stability

The study of the stability of proteins and peptides is based in Thermodynamics, and since this is related with macroscopic quantities and we want to know about microscopic interactions we will need also to use Statistical Mechanics. This section is based on a recent review by Lazaridis and Karplus (Lazaridis and Karplus, 2003)

From the point of view of statistical mechanics any state of the protein is an ensemble of microstates where its average properties, the ensemble averages, are consistent with a set of specific macroscopic properties. And from now, every time we will speak about a state or conformation it has to be understood that we are referring to the ensemble of microstates to which it belongs.

The native state (N) of biomolecules is only marginally more stable than the unfolded state. The reason may be because the need of a fast response due to changes in environmental conditions that biomolecules constantly experience in the biological processes. Moreover, it is still unclear if the native state is the most stable state under physiological conditions or if it is the most kinetically accessible state.

For our purposes we have to assume the first hypothesis. If the second case would happen that would mean that in nature there are energetic barriers impossible to surmount. Therefore the ergodic hypothesis will not be fulfilled, i.e. there will be conformations that will never be accessed and the sampling of the configurational space will not follow a Boltzmann distribution. In fact, the implication of the assumption of ergodic hypothesis is that the biological macromolecules will reach equilibrium in their environments in a finite time.

Thus, the native state at physiological conditions is the global minimum of the energy landscape of the biomolecule. The energy landscape of a biomolecule is the hypersurface defined by the effective energy in the configurational space. The shape of this hypersurface will determine the thermodynamics and dynamics of the system.

If we assume that the Hamiltonian H is additive and that it can be decomposed in intramolecular interactions, biomolecule-solvent interactions and solvent-solvent interactions $H=H_{mm}+H_{mw}+H_{ww}$, and that we have a system where the number of particles, the temperature T and the volume V are constant, i.e. the canonical ensemble, the Helmholtz free energy A can be written as a function of the canonical partition function Q:

$$A = -k_B T \ln Q \tag{1}$$

The effective energy, W, of the system is the intramolecular energy of the biomolecule and the equilibrium solvation free energy. The solvation entropy is included in this term too.

$$W = H_{mm} + \Delta G^{slv} \tag{2}$$

with

$$\Delta G^{slv} = -k_B T \ln \left\langle \exp(-\beta H_{mw}) \right\rangle_w \tag{3}$$

$\beta=1/k_BT$ and the ensemble average taken over the solvent.

Also the configurational entropy can be written as function of the probability of one state having internal coordinates q:

$$S^{conf} = -k_B \int p(q) \ln p(q) dq \tag{4}$$

The Helmholtz free energy can then be expressed as:

$$A = A^w + A^{id} + < W > -TS^{conf} \tag{5}$$

Where the first term is the free energy of the pure solvent, the second is the ideal contribution due to translation and rotation of the biomolecule, the third is the average effective energy and the last is the configurational entropy of the biomolecule.

The native state is the distribution of configurations of the biomolecule p(q) that minimizes the free energy of Eq. (5) under physiological conditions, and the probability distribution is a Bolzmann distribution.

For the folding/unfolding equilibrium it is the difference in free energy what is of interest, and if we assume that the two first terms in Eq. 5 are similar for the folded and unfolded states we get

$$\Delta A = \Delta \langle H_{mm} \rangle + \Delta \langle \Delta G^{slv} \rangle - T \Delta S^{conf} \quad \textbf{(6)}$$

or for the Gibbs free energy

$$\Delta G = \underbrace{\Delta \langle H_{mm} \rangle + \Delta \langle \Delta G^{slv} \rangle}_{\Delta \langle W \rangle} - T \Delta S^{conf} + P \Delta V \quad \textbf{(7)}$$

For processes with small volume changes the PΔV term is often negligible, making the Helmholtz and Gibbs free energies approximately equal.

The stability of the protein is the result between two opposite tendencies: the effective energy favors the native state, while the conformational entropy favors the denatured state. We can estimate the depth of the well of the native state in the energy landscape of the protein from Δ<W>. Since the entropic cost of localizing the protein in the well is supposed to be a few hundred kcal/mol, and the free energy of the native state is only about 5-10 kcal/mol lower than that of the denatured state, the depth of the well is also around a few hundred kcal/mol.

Using an extended conformation to represent the unfolded state, Lazaridis et al. (Lazaridis et al., 1995) have studied the different intramolecular contributions to the enthalpy of a typical unfolding process (not thermally induced). Using the thermodynamic cycle of Figure 2, it can be observed that in vacuum the free energy of the folding/unfolding transition has no contribution from solvation. The vertical reaction just corresponds to the insertion of the native and denatured states in the solvent. The evaluated increment of enthalpy of the cycle upon unfolding on four proteins was found to be 900-1500 kcal/mol. From this the main contribution (60-75%) arose form the van der Waals interactions, whereas the electrostatic interactions contributed around 25-35%. It was also corroborated that the contribution from the bonded interactions is negligible. Surprisingly, they found that the most important contribution to the van der Waals interactions comes from the nonpolar-polar interactions. This fact indicates that the interactions in the interior of a protein can be more complex than simply the result from hydrophobic interactions.

Moreover, the contribution to the enthalpy of solvation from nonpolar and polar groups is also calculated. The results show that the polar groups give a negative contribution to the enthalpy of unfolding and the nonpolar groups make the opposite. These values together with previously calculated entropies upon unfolding also of nonpolar and polar groups(Makhatadze and Privalov, 1995) give an estimation of the free energy of unfolding in solvent. In this case the contribution from the nonpolar groups is the most relevant. This is in agreement with the classical theory that asserts the hydrophobic effect as the main driving force in protein folding. In addition it is concluded that polar groups are responsible for protein solubility as well as for "introducing specificity in the low free energy conformation that makes up the native protein". In fact, what differentiates the

native state N from the rest of low free energy conformations is that these last ones have poor polar interactions. Furthermore, a quite compact denatured state D is proposed although less tight than the native state. In this denatured state a significant amount of hydrogen bonds are formed. However, the non bonded interactions of the denatured state can more easily be broken upon an increment of temperature than in the native state. The increment of enthalpy with T can explain the experimental high values of the heat capacity observed in unfolding processes, in contrast to the widely extended theory that attribute those high capacity values to the total exposure of the nonpolar groups to the solvent(Lazaridis and Karplus, 2003).

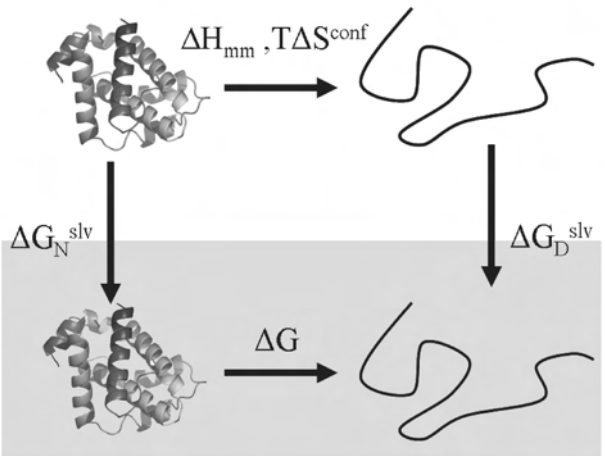


Figure 2. Thermodynamic cycle of protein unfolding, vacuum and solvent (grey)

2.3 Peptide Stability

The thermodynamics and kinetics of oligopeptides forming only secondary structure, polypeptides with some tertiary interactions and proteins with a hydrophobic core are very different. This can be in part rationalized by the different dimensionalities of the free enthalpy hypersurfaces of these tree different types of biopolymers(Daura, 2006). Moreover, it has even been suggested that the mechanism of α-helix and β-sheet formation can be essentially different(Kritzer et al., 2005). Therefore it can be that protein folding is not the simple extrapolation of the peptide folding mechanism.

The study of the folding/unfolding process on peptides is more feasible due to a more suitable sampling of the configurational space, i.e. the sampling of the equilibrium state, with peptides than proteins. Actually there exist simulations of peptides on the μs order thanks to the greater computational resources, the use of implicit solvent representations and the use of new sampling techniques as replica-exchange MD simulations(Garcia and Sanbonmatsu, 2001). These recent simulations of peptides folding/unfolding equilibrium suggested that the number of accessible conformational states can be described by a scale-fee network(Rao and Caflisch, 2004; Daura, 2006), which implies that a few "hubs" (conformations to which it is easy to get from a big number of conformations during the folding/unfolding process) dominate the total connectivity of the network in a hierarchical manner(Ravasz and Barabasi, 2003; Rao and Caflisch, 2004) And that the configurational space resembles a small-world network (Daura, 2006)which in this context means that the path length between any two conformations or nodes visited

during the folding/unfolding process is relatively short(Albert et al., 1999; Jeong et al., 2000)

Moreover it has been shown that these few hubs are low-lying minima of the free energy on the free energy landscape(Rao and Caflisch, 2004). This network analysis of protein folding has advantages over the study of free energy landscapes to localize transition states but also more details can be obtained about the non-native like conformatins. In addition, conformational space networks also inform about the dynamic connectivity(Rao and Caflisch, 2004).

The intramolecular bonded and non bonded interactions as well as solvent-solvent interactions were found to favor peptide folding at all temperatures whereas the non-bonded peptide-solvent interactions are unfavorable for folding. What was interesting in the Daura's work is the observation of the importance of the solvent-solvent interactions (not pure water) to folding which is of similar magnitude as the intramolecular nonbonded interactions(Daura, 2006).

It is now clear how important the determination and accurate description of the unfolded state is for the characterization of physical, chemical and biological properties of polypeptides and proteins, as well as for the study of their mechanism of folding/unfolding. The most common methods that induce protein denaturation are the substantial increment of the temperature and the addition of high amounts of denaturant to the solution. These two procedures are valid both experimentally and for simulations. Although computationally, the use of denaturants is restricted to MD simulations with explicit solvent. Several MD simulations of proteins in urea-aqueous solutions have been reported. In general, these works point out the difficulties of obtaining full denaturation in the simulations.

3 SIMULATIONS OF BIOMOLECULAR SYSTEMS

Although now there is a large amount of information about the different events that take place in living organism, this information is mainly descriptive and qualitative. Since biological systems and in particular proteins and peptides should obey the natural laws of physics, it is through physical sciences and especially thermodynamics that the question about why these events occur the way they do can be addressed.

Simulations try to mimic nature, using as basis physical laws for obtaining a full physical understanding about how different specific events occur. The simulations ultimate goal is also to predict the biomolecules behavior when the conditions at which they occur are changed in order to understand and control the structure/function relation ship.

Since the simulations got in to the field of protein science thirty years ago (McCammon et al., 1977), they have undergone a rapid development(Karplus and McCammon, 2002). Actually their accuracy and complementation with experimental results have helped to explain important phenomena that was impossible to observe at the moment by experimental procedures(Karplus and McCammon, 2002). The water passage through an aquaporin water channel(de Groot and Grubmuller, 2001), ion transport in a potassium channel(Aqvist and Luzhkov, 2000) or the pathway between the open and closed conformations of the bacterial chaperonin GroEl(Ma et al., 2000) are some examples. On the other hand there is a constant need of experiments for validation of these methods.

Simulations can principally be performed for three different types of applications: (a) the refinement or determination of structures form experimental data. Such simulations focus mainly on the sampling of conformational space (b) the description of structural and motional properties as well as determination of thermodynamic parameters from the system at equilibrium. In this case, appropriate conformational sampling as well as the need of every state being weighted by the Boltzmann factor is required (c) the study of the biomolecules dynamics, which needs both appropriate conformational sampling and Boltzmann weighting but also the order in which the processes occur is of critical importance. For the two first cases molecular dynamics or Monte Carlo simulations can be used. However for the third case only molecular dynamics (MD) simulations can give such a time dependent description.

MD is a widely used simulation method for computing numerically the atomic positions along time of biomolecules. From analysis of the trajectories structural and thermodynamical quantities can be obtained. Since this system comprises a large amount of atoms, especially if the solvent is treated explicitly, only a classical description of the dynamics can give simulations of reasonable time length. This methodology uses a force field, which is a simplified expression of the potential energy of the system as a function of the atomic coordinates in conjunction with a set of parameters obtained from fitting to experimental data or *ab initio* calculations. Newton's equations of motion are iteratively solved to obtain the trajectories. This approach treats the atoms as soft spheres centered at the nuclear positions. However, the electronic contributions are neglected explicitly although the empirical parameters include electronic effects implicitly.

On the other hand there are processes that can not be studied classically such as bond breakage, chemical reactions, electron transfer, etc. In these cases a quantum mechanics treatment of the problem is necessary. Although this method is very accurate the disadvantage is their computational cost. For this reason, only isolated small molecules are studied by quantum mechanics calculations. It is also possible to use both methods together, and treat the small region were the phenomenon of interest takes place with

QM and the rest of the system by molecular mechanics. These are called MM/QM simulations.

3.1 Empirical Force Fields: Molecular Mechanics

Force fields are an analytical approximation to the true quantum mechanics Born-Oppenheimer surface in which the motions of electrons and nuclei are decoupled and the energy of an atom can be considered only as a function of the nuclear coordinates. Empirical force fields are constituted by two parts: a set of equations that relate the coordinates of the system and the energy of the system; and a set of parameters that have to be used in those equations(Mackerell, 2004). The term *Molecular Dynamics* refers to the motion of the system on this potential energy surface.

The most widely used form of the potential energy function, is known as Class I additive potential energy functions. It consists of a series of terms dependent on internal coordinates and pair-wise interaction terms. The interactions are divided into internal or intramolecular and nonbonded interactions. In the intramolecular part, the contribution to the energy from bonds, angles, improper angles (out-of-plane) and Urey-Bradley (UB) are treated harmonically, whereas the energetic contribution from the dihedral or torsional rotations has a sinusoidal form. Also, there are two more terms arising from the nonbonded interactions. The first describes the van der Waals (vdW) interactions, by a Lennard-Jones (LJ) (12-6) potential and the second is a coulombic term describing the electrostatic interaction energy between two charged particles. In turn, the LJ potential contains an attractive term, proportional to r^{-6} that dominates at large distances. This term is known as London-dispersion term and gives account of the induced dipole-dipole interactions between atoms due to the instantaneous asymmetrical charge distribution of the electronic clouds around the atoms. The other part, with r^{-12} dependence, is a repulsive part that dominates the interaction at short distances. This term is related to the Pauli exclusion principle about the inability of two electrons of occupying the same wave function simultaneously.

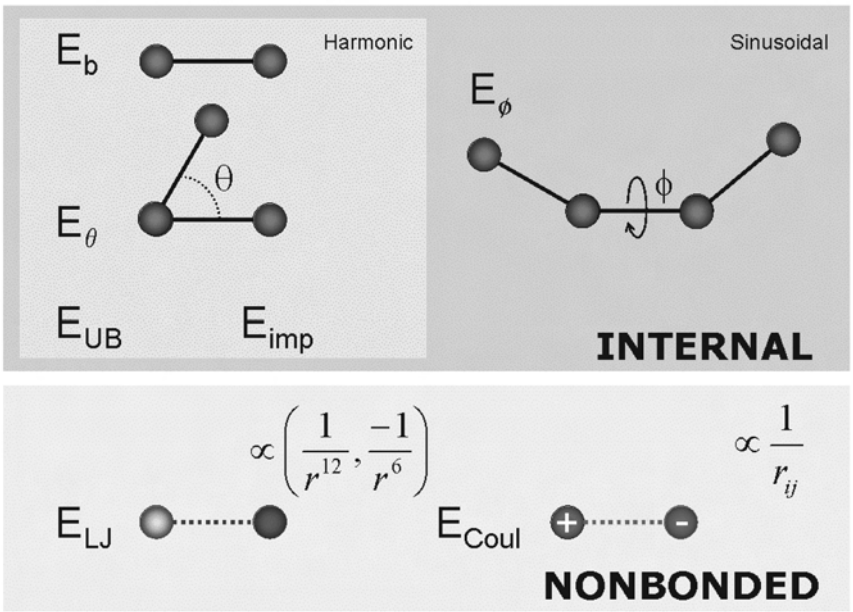


Figure 3. Different types of energy terms of the force field

The mathematical expression of this force field is:

$$E = E_{INT} + E_{NonBond} \tag{8}$$

$$E = E_b + E_\theta + E_\phi + E_{impr} + E_{UB} + E_{LJ} + E_{Coul} \tag{9}$$

$$\begin{aligned} E(\vec{r}) = & \sum_{bonds} K_b (b - b_o)^2 + \sum_{angles} K_\theta (\theta - \theta_o)^2 + \sum_{dihedrals} K_\phi (1 + \cos(n\phi - \delta)) + \\ & + \sum_{impropers} K_\varphi (\varphi - \varphi_o)^2 + \sum K_{UB} (\rho - \rho_o)^2 + \\ & + \sum_{NonBonded} \left\{ \varepsilon_{ij} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right\} \end{aligned} \tag{10}$$

where the internal coordinates are: b the bond length, θ the valence angle; ϕ the dihedral angle, φ the improper angle, ρ the UB distance and r_{ij} the distance between atoms i and j. The parameters K_b and b_o are the bond force constant and equilibrium bond distance; K_θ and θ_o valence angle force constant and equilibrium angle; the dihedral angle barrier height K_ϕ, the multiplicity n and phase angle δ; K_φ and φ_o the improper force constant and the equilibrium improper angle; K_{UB} and ρ_o UB force constant and equilibrium UB distance. All these parameters describe the intramolecular force field. The external parameters are the partial atomic charges q_i of atom i, the LJ well-depth ε_{ij} and the minimum interaction radius R^{min}_{ij} of the vdW interaction term.

The specific form of each of the contributions to the total potential energy function can be seen in the Figure 4.

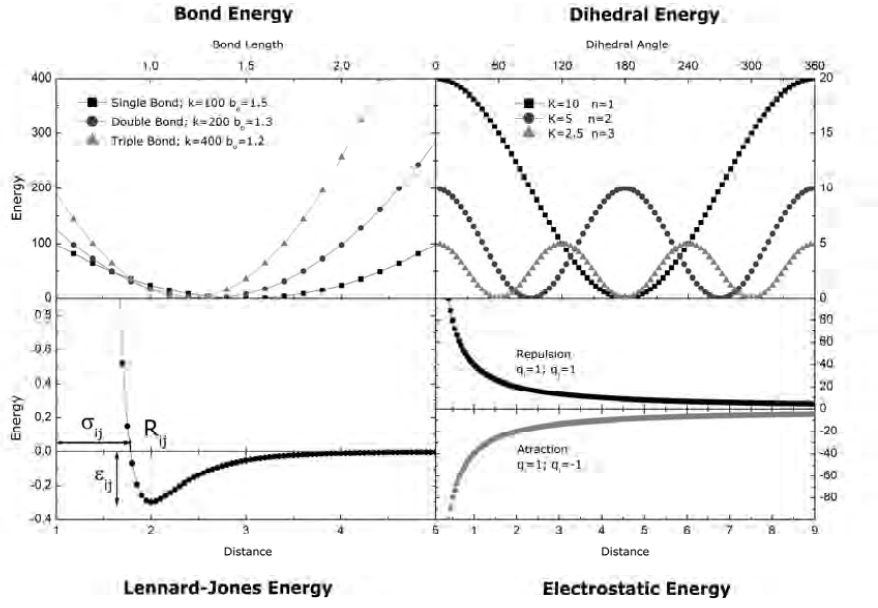


Figure 4. Specific form of the energy terms of the potential energy.

This form of the potential energy function is adopted by the most currently used biomolecular simulation packages as CHARMM(Brooks et al., 1983), AMBER(Cornell et al., 1995), GROMOS(van Gunsteren, 1987) and OPLS(Jorgensen and Tirado-Rives, 1988).

Although these force fields do not include explicit hydrogen bond terms, the coulombic and LJ treatment of the nonbonded interactions seems to appropriately reproduce hydrogen bond features. In general these force fields are all-atom force fields; however, most of them have also united atom versions. In this case the only the polar hydrogens are explicitly described, saving computational time.

Biomolecular force fields have been designed to deal mainly with the condensed phase, for this reason solvation is an important issue. Solvation can be considered explicitly or implicitly. The latter is mainly used for cases where a broad conformational sampling is important, as for instance for protein folding. The former option will give a more accurate description of the biological system. The water model chosen for the simulations is also important, due to the need of consistence between the water model and the force field. The force fields have been developed with a specific model of water – e.g. CHARMM, AMBER and OPLS with the TIP3P water model or TIP4P(Jorgensen et al., 1983) also for OPLS; similarly GROMOS and ENCAD with the SPC/E(Berendsen et al., 1987) and F3C(Levitt et al., 1997) water models respectively – and one has to be aware of it when setting up the simulation.

The nonbonded interactions are very important for the properties of biomolecular systems. However, the calculation of these is the most expensive part of the calculations and normally truncation schemes are used to reduce the number of interacting pairs. Smoothing functions that brings the energies and forces gradually to zero are commonly used(Steinbach and Brooks, 1994; Norberg and Nilsson, 2000). Also there is another way for treating the long range electrostatic interactions, this mehod is the particle mesh Ewald (PME) method, when periodic boundary conditions are used (PBC).

Determination and optimization of the external parameters is thus of crucial importance since they ultimately will dictate the thermodynamical properties as well as the structure the biomolecular systems will adopt. A common procedure is to obtain these parameters from *ab initio* calculations. However, with respect to the partial charges caution has to be taken since the concept of partial charge is not well-defined in QM. In fact, charges are highly dependent on the level of theory chosen as well as on the model of charge chosen. Particularly important is to remember that QM calculations are performed in vacuum and that the force fields try to mimic the condensed phase. Also for this reason, in general the QM calculations underestimate the molecular dipole moment. (See also chapter about Quantum Mechanics Calculations: charges)

Instead of performing *ab initio* calculations in the isolated molecule of interest it is better to use the so called supramolecular approach. In the supramolecular approach QM calculations are performed on model compounds, normally a small compound together with water or a dimer. The charges are then fitted to reproduce the interaction energies and geometries of the compound-water system or the dimer. In addition, by this method electronic polarization is implicitly included in charges. CHARMM and OPLS use this approach and the small compounds usually are functional groups with total charge equal to zero. Again, the model of water used for the study of the compound-water system will influence the results, and this has to be kept in mind.

Once the charges are calculated, the LJ parameters are derived to reproduce experimental data such as heat of vaporization, density, isocompressibility or heat capacity. However, again these parameters are not universal and are highly correlated to

the atomic partial charges, which indicates that parameters from different force fields should not be combined.

When charges and LJ parameters of the functional groups are satisfactorily optimized they can be directly transferred to other novel molecules that include these functional groups.

The intramolecular parameters for proteins are obtained from fitting to experimental and QM data for small compounds. For example the force constants should reproduce vibrational spectroscopy data.

Continuous improvement in parameter optimization is needed in order to obtain simulations closer to experiments. For instance the CHARMM22 force field with the CMAP correction has been demonstrated to substantially improve the sampling of the backbone dihedrals, ϕ and ψ. This 2D dihedral energy grid correction in particular has diminished the sampling of π-helical conformation quite typically sampled in simulations without this correction. Also, inclusion of polarizability can improve the quality of the simulation.

Energy Minimization

The *potential energy surface*, which is the just the hypersurface of the potential energy of the system as a function of the coordinates of the system, is a very important concept for molecular mechanics. The stationary points on this energy surface, in particular the global minimum, will determine the native conformation of the protein. Moreover, the way the system wanders over the surface will dictate the dynamics of the system, and thermodynamical properties can be extracted from the analysis of these trajectories.

If the system is in a high energy conformation we may want to find a conformation with a lower energy, to relax strain in the system. Several minimization algorithms of varying complexity exist. For molecular mechanics the *steepest descent* (SD) method and *conjugate gradients* methods, as the *adopted based Newton-Raphson* (ABNR) method, are often used (for details see (Leach, 2001)).

Molecular Dynamics Simulations

The aim of MD simulations is to observe the time evolution of the system. That is the calculation of the set of N atomic coordinates $\left\{\vec{r}_i(t)\right\}_{i=1\ldots N}$.

The trajectory is obtained by numerical integration of atomic accelerations given by Newton's second law, in which the force on an atom is obtained from the gradient of the potential energy. The *Verlet algorithm* is a widely used method for this numerical integration. It is based on a Taylor expansion of the position. The velocities are not computed until the new positions have been calculated.

$$\vec{r}(t+\delta t) = \vec{r}(t) - \vec{\mathrm{v}}(t)\delta t + \frac{1}{2}\ddot{\vec{r}}(t)\delta t^2 + \ldots \qquad \textbf{(11)}$$

$$\vec{r}(t+\delta t) = 2\vec{r}(t) - \vec{r}(t-\delta t) + \delta t^2 \ddot{\vec{r}} \qquad \textbf{(12)}$$

and

$$\vec{\mathrm{v}}(t) = \frac{\vec{r}(t+\delta t) - \vec{r}(t-\delta t)}{2\delta t} \qquad \textbf{(13)}$$

A variation of method is the *leap-frog algorithm*:

$$\vec{r}(t+\delta t) = \vec{r}(t) + \vec{v}(t+\frac{1}{2}\delta t)\,\delta t$$
$$\vec{v}(t+\frac{1}{2}\delta t) = \vec{v}(t-\frac{1}{2}\delta t) + \ddot{\vec{r}}(t)\,\delta t \qquad \textbf{(14, 15)}$$

With this method the new velocities $\vec{v}(t+\frac{1}{2}\delta)$ are first calculated and then the new positions $\vec{r}(t+\delta t)$. This method includes the velocities explicitly and there is no need of computing subtractions. The disadvantage is that the positions and velocities are not synchronized, and thus the kinetic energy can not be determined simultaneously with the potential energy (determined form the positions).

The size of the time step chosen is important in the integration algorithm. Too short time steps will slow down the simulation, however if too long can lead to instabilities in the algorithm since rapid events can be missing due to the big size of the step. The choice of the time step size should be correlated with the fastest motions occurring in the system and in general should be an order of magnitude shorter than these motions. However, in biomolecules, there are extremely fast motions of high-frequency bond vibrations. To avoid this problem it is possible to constrain the bonds vibrations and to consider the bond only at equilibrium, thus larger time steps can be considered.

The SHAKE constraint is the most commonly used constraint. This is a holonomic constraint that keeps fix the bond length between two atoms or the angles. In particular, it is applied to the bond lengths of all the bonds between a heavy atom and hydrogen atoms. This will not have any impact on the simulation since this kind of vibrations is not correlated with the rest of the motions in the protein. It is also common to apply SHAKE to water molecules in such a way that the molecule is kept rigid. This will reduce the degrees of freedom of the system. A feasible time step for flexible biomolecules is 1 fs, but if SHAKE is employed the time step can be increased by a factor of 2.

To simulate a realistic system huge amount of solvent molecules should be considered. However this is computationally very expensive, since most of the time of computation is gone in calculating the new positions of the solvent and all the solvent-solvent interactions. A common approach to mimic solution bulk properties in the simulations is to use periodic boundary conditions (PBC) (Figure 5). The approach consists in having a system which is replicated in all directions to completely fill space. There are only a finite number of possible geometric volumes that can completely tessellate (fill) the space in this way. These are: the cube, the parallelepiped, the hexagonal prism, the truncated octahedron, the rhombic dodecahedron and the elongated dodecahedron. By this procedure every time that a particle (solvent molecule) leaves our principal unit (box) to an adjacent unit, another from the box at the opposite side enters in our system, insuring the number of particles in our system to be constant. It is undesirable that a particle could see its own image when calculating the pair-wise interactions. It is recommended to use a symmetric form, particularly the rhombic dodecahedron despite is computationally more expensive. The reason is that due to rotations that the polypeptide can undergo during the simulation it may end up sticking out of the unit cell and make contacts with its own images if an asymmetric shape is used for the unit cell. Care has to be taken with the size of the boxes.

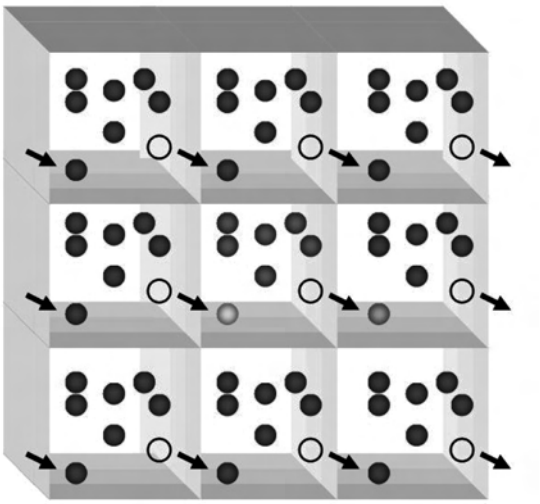


Figure 5. Periodic boundary conditions (PBC).

As previously mentioned truncation or PME methods are applied to deal with the calculation of the coulombic electrostatic contribution. The PME method aim to calculate all the pair-wise interactions between each particle with the particles of the simulation system and with all the images of the periodic infinite array imposed by the PBC. To give a general outline the series $\sum_{n}^{N=no\ cells} \sum\sum {}^{q_i q_j}\!/_{r}$ of the coulombic term converges very slowly, thus the series is rearranged in a new way to converge much faster:

$\sum \frac{1}{r} = \sum \frac{f(r)}{r} + \sum \frac{1-f(r)}{r}$ where f(r) is a Gaussian function ($\exp(-\alpha x^2)$).

What is done is that every charge of the system is supposed to be surrounded by a Gaussian charge distribution of the same magnitude but opposite sign. (A gaussian charge distribution function f(r) is added in the space real). But then, this distribution has to be subtracted (See Figure 6). The summation of the $\frac{1-f(r)}{r}$ term is done in the reciprocal space.

The real space term converges rapidly with large gaussian widths; however the reciprocal space term does the opposite. Recommended gaussian width is about 5/L with L the length of the box (in Å) in the case of cubic symmetry. This method can introduce artifacts due to the periodicity of the system. Water layers around the molecule should be at least 7Å to avoid this problem(Mackerell, 2004).

This methodology and the use of spherical cutoffs with force-shifting truncation show no differences if the cutoff radii are larger than 12 Å(Norberg and Nilsson, 2000).

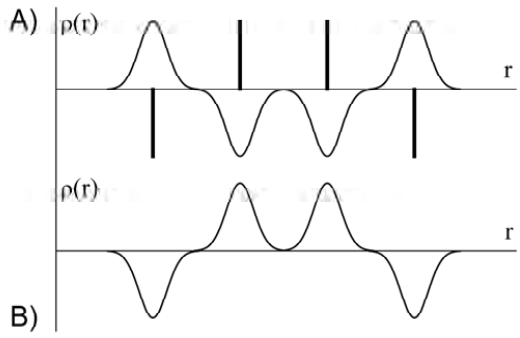


Figure 6. Charge distribution of the Ewald summation method. A) Original charges plus the Gaussian charge distribution of opposite sign. The summation is performed in the real space. B) Cancelling Gaussian charge distribution. The summation is performed in the reciprocal space.

Thermodynamic information can be obtained from the analysis of the trajectories of the simulations. However, it is necessary to connect the microscopic-macroscopic thermodynamic quantities and this is not straightforward. Ensemble average of thermodynamic quantities can be obtained from our simulations assuming equality between ensemble average <X> and time average $\overline{X}$ via the *Ergodic hypothesis*.

The *Ergodic hypothesis* states that the time average $\overline{X}$ is equal to the ensemble average <X> or expectation value:

$$\overline{X} = \lim \frac{1}{\tau} \int_0^{\tau} X(\vec{\mathrm{p}}^{\mathrm{N}}(t), \vec{\mathrm{r}}^{\mathrm{N}}(t))\, dt \qquad \textbf{(16)}$$

$$\langle X \rangle = \iint \mathrm{d}\vec{\mathrm{p}}^{\mathrm{N}} \mathrm{d}\vec{\mathrm{r}}^{\mathrm{N}} X(\vec{\mathrm{p}}^{\mathrm{N}}, \vec{\mathrm{r}}^{\mathrm{N}})\, \rho(\vec{\mathrm{p}}^{\mathrm{N}}, \vec{\mathrm{r}}^{\mathrm{N}}) \qquad \textbf{(17)}$$

where in our case X is the thermodynamic property of the system containing N atoms, $\vec{\mathrm{p}}$ and $\vec{\mathrm{r}}$ are the momenta and positions respectively of every atom and $\rho(\vec{\mathrm{p}}^{\mathrm{N}}, \vec{\mathrm{r}}^{\mathrm{N}})$ the probability density of the ensemble.

Different ensembles are used when performing simulations. In the *microcanonical ensemble* (NVE) the number of particles N, the volume of the system V, and the energy of the system E are maintained constant along the simulations. In this case the most significant property is the *entropy S*, in particular the maximum entropy principle; for the *canonical ensemble* (NVT) N, V and temperature T are constant. The state function of the canonical ensemble is the *Helmholtz free energy A*. However for physiological conditions the term PV is not a big contribution to the free energy and free energy G and Helmholtz A can be assumed equivalent. In turn, the *isothermal-isobaric ensembles* (NPT) have constant N, pressure P, and T. The *Gibbs free energy G* is the state function in this case. This last kind of simulation is also useful when the density of a system at the beginning of a simulation is not well adjusted.

The comparison of the thermodynamic properties between experiments and the simulations can indicate the accuracy of our simulation. But also, thermodynamic quantities that have not or can not be measured experimentally can be calculated from the simulations and enlighten biophysical phenomena. In any case, we should remember that we get thermodynamic qualitative information from the simulations, if the exact quantity is needed it is better to obtain it experimentally.

Sampling and Convergence

How long should be the simulation to obtain representative data to reproduce the characteristics of the system? The answer is not unique. It depends on which kind of phenomena it is intended to be reproduced, on the complexity of the energy landscape, on the existence of multiple minima and ultimately on the size of the energy barriers. The rule of thumb is that the simulation time should be an order of magnitude higher than the time scale of interest.

3.2 Quantum Mechanics Calculations

Ab initio calculations are very appropriate to study chemical reactions, geometries and partial atomic charges of small molecules. These can be used later as first step for parametrization of new molecules in the empirical force fields.

The *ab initio* calculations are very accurate however their disadvantage is that as pointed out before *ab initio* calculations are limited to relatively small systems due to their expensive computational cost.

In quantum mechanics the electrons of the system are explicitly included in the calculations. Therefore, it is possible to study properties dependent on electronic distribution. For this reason these calculations are very useful on investigation of chemical reactions.

The basis of quantum mechanics is the *Schrödinger equation*:

$$H|\Psi\rangle = \mathrm{E}|\Psi\rangle \tag{18}$$

Where H is the Hamiltonian operator:

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V \tag{19}$$

To solve this equation it is necessary to find the wave function of the particle, or eigenfunction, $|\Psi>$, and the energy, or eigenvalue, E, that satisfy the equation.

Nevertheless, this equation can only be solved analytically for a system with only one electron, therefore has no exact solution for any molecular system.

The *Born-Oppenheimer approximation* is the key to allow the calculation of the eigenvalues and eigenfunctions of molecular systems. This approximation assumes that nuclear and electronic motion is uncoupled. This assumption is based in the fact that the nucleus is much heavier than the electron and for this reason the electron wave function depends only on the nuclear positions but not on their momenta. The ultimate implication is that on this approximation the nuclei are assumed fixed and it is only necessary to solve the Schrödinger equation for the electronic part, immersed in the nuclear electrostatic field.

$$\Psi_{Tot}^{nuclei,e^-} = \Psi^{e^-}\Psi^{nuclei}$$

And the total energy of the system is:

$$E_{Tot} = E_{e^-} + E_{nuclei}$$

In turn, the electronic energy is the kinetic and potential energy of the electrons in the electrostatic field generated by the nucleus and the electron-electron interaction.

Nonetheless, to treat systems with more than one electron further approximations such as the Hartree-Fock (HF) approximation are necessary. This method is also known as the Molecular orbital (MO) approximation. In this approximation the many electron problem is replaced by one electron problem, where every electron interacts with the rest via the average field produced by the rest of electrons. The system of eigenvalues and eigenfunctions is solved iteratively, what is known as the Self Consistent Field (SCF).

However this approximation disregards an important phenomenon, the instantaneous repulsion between electrons. The energy change caused by this instantaneous repulsion is called *correlation energy*. There are many methods that calculate correlation energy. The most known are *Perturbation Theory* and *Density Functional Theory* (DFT). The Perturbation theory method considers the electron correlation as a perturbation to the HF energy; MP2 is the lowest Møller-Plesset non-zero order correction to the HF energy of the Perturbation Theory. The DFT method in contrast, computes the electron correlation energy and in some cases the exchange energy as an empirical functional of the electronic density. DFT has emerged as a promising alternative to the conventional *ab initio* methods because it is computationally simpler, especially when electron correlation energy is treated and at high levels of accuracy.

In our *ab initio* calculations for the urea molecule we have used the hybrid B3LYP(Stephens et al., 1994) DFT method that consists in the Becke's exchange(Becke, 1988) 3 parameter functionals(Becke, 1993) and the non-local correlation provided by the LYP(Lee et al., 1988) correlation functional.

Basis sets

The HF equations are differently solved for atoms than for molecules. For atoms there is an analytical approximation that uses Slater type orbitals (STO) for the radial part of the wave function and spherical harmonics for the angular part. On the other hand, for molecules every spin orbital is assumed to be a linear combination of atomic STO orbitals (LCAO). This minimal amount of wave functions that describes any other wave function is known as basis set. Since the STO are still computational expensive, these are replaced by Gaussian functions (GTO). Most quantum mechanics packages use contracted basis sets, that is, during the calculation the coefficients of the LCAO and the width of the Gaussians are not allowed to vary. For these calculations a minimal basis set is used. A minimal basis set just contains the strictly necessary amount of the functions to accommodate all the filled orbitals in each atom. This can cause problems since: (a) atoms as Li will be described by the same amount of basis functions as O, even though this last atom contain many more electrons; (b) the orbitals can not expand or contract in response to the environment and (c) non-spherical properties of the electronic distribution can not be described.

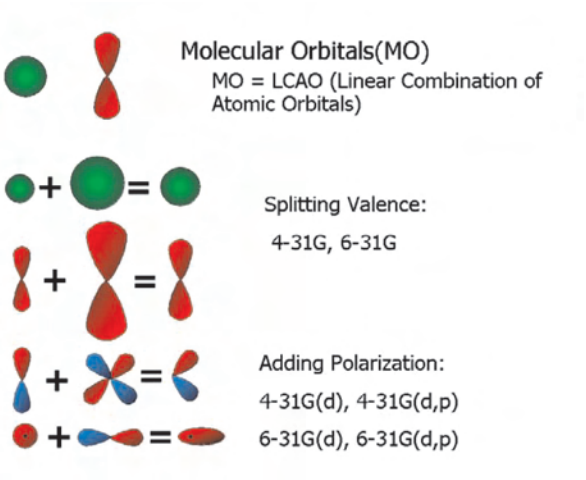


Figure 7. Basis sets.

One possible solution is to use more functions to describe the valence electrons and to keep the single function for inner shell. In our simulations we have used 6-31G, which means that. Polarization is also included by adding functions with higher angular moment. We have used the 6-31G(d) or 6-31G* basis set, which means that six Gaussians are used for the core shells and for the valance electrons the contracted part by two Gaussians and the diffusive part by one Gaussian, in addition polarization is used for the non-hydrogen atoms.

Charges

In classical force-field based calculations electrostatic interactions are calculated using Coulomb's law and point charges placed at the positions of the nuclei. The charge of an atom in a molecule is however not an observable quantity. There are different ways to obtain atomic charges form quantum mechanics calculations. But to which extent are these derived charges accurate? Wiberg and Rablen (Wiberg and Rablen, 1993) have suggested that relatively accurate charges should not be strongly dependent on the used basis set, that the electronegative trends should be accomplished, that the dipole of the molecule should be approximately reproduced and that the molecular potential energy created by that charge representation outside the van der Waals radii reproduce approximately the real potential energy map of the molecule.

The oldest method for deriving charges is the Mulliken population analysis (Mulliken, 1962), which is strongly dependent on the basis set. An alternate approach is to use the charge density distribution, which is an observable quantity, but still a definition of the atom is required to convert the charge density into point charges. There are also procedures that derive the charges by fitting the electrostatic potential, which is also an observable quantity, CHELP(Chirlian and Francl, 1987) and CHELPG(Breneman and Wiberg, 1990) are of this kind. The CHELP charges were observed to be dependent on the molecule orientation. For the calculations in this thesis we have used the CHELPG population analysis scheme since we want a point charge centered at the nucleus position for further simulations in the empirical force field.

4 ENVIRONMENT

The environment surrounding proteins is an important factor for protein structure and stability. Principally, water is the most common biological environment of proteins and they normally achieve their native state in this medium. However, proteins are also found in more or less partially unfolded states, which in fact can be seriously harmful due to aggregation and fibril formation. But still, these partially unfolded states can be necessary for biological process such as signal transduction, translocation across membranes, transcriptional activation and cell cycle regulation(Vendruscolo and Dobson, 2005). Also intrinsically disordered proteins, commonly found in eukaryotes which may change their structure upon binding, are very important for molecular recognition (Dyson and Wright, 2002). Aqueous-urea solutions induce unfolding, and the process of protein unfolding as well as the structure of unfolded proteins can be studied immersing them, experimentally or in silico, in this milieu. Nevertheless, not all proteins are water-soluble. Since the explosion of the genomic era it is known that membrane spanning domains are encoded by ~30% of most genomes. The principal property of membrane bilayers is the high hydrophobic character conferred by the lipid chains as well as its low dielectric constant. The preferred conformation adopted by proteins inside this medium is helical, although β-barrels are also formed.

Alcohols such as hexafluoroisopropanol (HFIP) are known to stabilize secondary structure, in particular helices. Experimental studies have demonstrated how in some extent alcohol-water solutions mimic biological membranes(Perham et al., 2006) since the protein native structure changes in such a environments are similar to those expected if the protein will be inserted into a lipid bilayer.

In this thesis we have performed MD simulations of different peptides in water, urea-water solutions and HFIP-water solution as well as in a low dielectric constant medium that mimics the interior of lipid-bilayers.

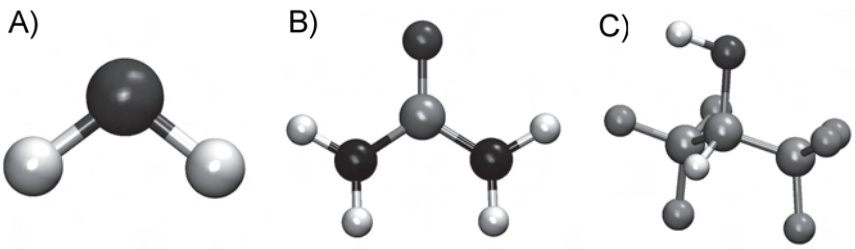


Figure 8. A) Water molecule. B) Urea molecule. C) HFIP molecule.

4.1 Chemical Denaturant Agents: Urea

The study of unfolded and partially folded states has become an important issue in the field of protein folding. The residual preferential interactions remaining in the unfolded states can reveal the initial events necessary to fulfill a correct folding or that can lead, although rarely to misfolding. Furthermore, the characterization of these states and their relation with the native state provides insights that are important for kinetics of folding, stability and behavior(Gsponer and Caflisch, 2001; Vendruscolo and Dobson, 2005). Simulation of unfolding processes is a common method used to get atomistic information about this process. The two procedures most widely used are thermal

denaturation and chemical denaturation (Caflisch and Karplus, 1995; Tirado-Rives et al., 1997; Zhang et al., 2001; Bennion and Daggett, 2003; 2004; Daggett, 2006).

Although urea and guanidinium chloride are widely used as chemical denaturant agents the molecular mechanism by which they induce denaturation is not completely understood.

General Aspects

Urea is a highly polar molecule. Indeed, the experimentally obtained urea dipolar moment are 3.83 D in the gas phase(Brown et al., 1975) and 4.2 D in solution(Gilkerson and Srivastava, 1960), values obviously higher than water (water dipole moment in gas phase is 1.8 D and between 1.9-3.1 D in liquid phase). Urea also has a high hydrogen bonding potential: as hydrogen bond donor through its 4 amide hydrogens and as hydrogen bond acceptor trough the oxygen atom. In addition, urea mixes very well with water. Furthermore, the free energy of transferring hydrocarbons from water to denaturant aqueous solution is approximately linear with the denaturant concentration with a constant of proportionality that strongly depends on the accessible surface area. In addition, calorimetric studies demonstrated that also the enthalpy and entropy of transfer are proportional to the urea concentration. However, high concentrations of denaturant, typically 6-8 M, are required to observe denaturation.

Molecular Action Mechanism

The understanding of the unfolding mechanism induced by urea at molecular level has always been dichotomized. Two main concepts have guided these studies: the so called the *Indirect* and *Direct mechanisms*. In this first one, urea acts indirectly altering the structure of the solvent, making the water more hydrophilic, therefore facilitating the solvation of hydrocarbons. Whereas in the second, the improved hydrocarbon solubility, is due to direct action of urea(Roseman and Jencks, 1975): directly making hydrogen bonds to the peptide, removing water form the protein surface, it has even been suggested that it induces repulsion between residues (Wallqvist et al., 1998; Mountain and Thirumalai, 2003). Recently several studies suggest that both mechanisms play a role in the unfolding process directed by urea. Thermodynamic data can be explained in terms of both theories. However, neither theoretically or experimentally it has been demonstrated which of these models explain more accurately the effect of urea in the chemical denaturation process(Idrissi, 2005).

The indirect mechanism was originally proposed by Frank and Franks (Frank and Franks, 1968). According to them urea is a "water structure breaker", a chaotropic agent. The tetrahedral hydrogen bonding character of water can not be maintained in the presence of urea. This theory was questioned by the SKSS model (Schellman, 1955; Kreschek and Scheraga, 1965; Stokes, 1967), in which the properties of urea-water solutions arise from the formation of urea dimers or even higher order aggregates.

The idea of urea as a water structure breaker was extensively studied and several groups have dismissed this theory (Kuharski and Rossky, 1984a; b; Tsai et al., 1996; Wallqvist et al., 1998; Zou et al., 1998; Grdadolnik and Marechal, 2002; Kallies, 2002; Zou et al., 2002; Batchelor et al., 2004), while other groups still keep this theory alive. In fact, stiffness of water dynamics (Idrissi et al., 2000) weakening of the water structure (Bennion and Daggett, 2003), more bent hydrogen bonds that water makes in the first solvation shell than in bulk water (Vanzi et al., 1998) and reduction of order in the structures formed by water(Salvi et al., 2005) have been reported.

On the other hand, only few studies (Astrand et al., 1994; Åstrand et al., 1994) had focused on the SKSS approach, although recently, the urea aggregation issue has been recaptured (Weerasinghe and Smith, 2003b; a; Smith et al., 2004).

The direct mechanism comprises two aspects: the "interaction" of urea with apolar solutes and second, the interaction of urea with polar solutes.

The *urea-apolar interaction* has an entropic character. The system regains entropy(Roseman and Jencks, 1975; Kuharski and Rossky, 1984a; b; Zou et al., 1998; Zou et al., 2002) due to the displacement of approximately 4 water molecules by the larger urea molecule from the apolar solvation shell. The water had stronger rotational restraints being constrained to be on the apolar region of the protein surface. In addition, solute size is an important factor in the free energy of transfer and free energy of cavity formation (Ikeguchi et al., 2001; Shimizu and Chan, 2002; Batchelor et al., 2004). Larger cosolvent sizes and lower hydrogen bond donor/receptor density will also favour cavity formation (Roseman and Jencks, 1975; Kuharski and Rossky, 1984a; b). The reason is intimately related with the balance between the loss of orientational freedom suffered by the solvent and the attempt to keep its hydrogen bonding ability when a solute or cosolvent molecule is introduced in the system.

In contrast *urea-polar interaction* is mainly of enthalpic character (Zou et al., 1998; Zou et al., 2002; Bennion and Daggett, 2003). These interactions enhance solvation due to the more favorable hydrogen bonds that the peptide amide units make with urea than with water (Tobi et al., 2003). The reason has been attributed to the urea being itself a soluble amide (Zou et al., 1998; Tobi et al., 2003).

Moreover, the urea accumulation in excess around the peptide has been observed by different MD studies especially around the polar groups (Tirado-Rives et al., 1997; Caflisch and Karplus, 1999; Bennion and Daggett, 2003; Salvi et al., 2005).

4.2 Fluorinated Cosolvents: HFIP

In contrast to urea-water solutions, the alcohol-water solutions are characterized by low polarizabilities. The relative dielectric constant of trifluoroethanol-water and HFIP-water solutions decreases linearly with concentration. For 20% v/v solutions the relative dielectic constant is around 75 or 66 for TFE-water solutions and HFIP-water solutions respectively(Hong et al., 1999), lower than the relative dielectric constant of water (80 for pure water).

Fluorinated alcohols such as TFE and HFIP stabilize and promote formation of α helices but they also denature the native conformation of proteins(Buck, 1998). Recently, experimental and MD studies have attributed these properties to the strong clustering in water of these compounds (Hong et al., 1999; Roccatano et al., 2005) and to preferential solvation(Fioroni et al., 2002; Roccatano et al., 2005; Perham et al., 2006).

The mechanism suggested by the MD simulations for the α-helical stabilization is the decreased accessibility of water to the helical backbone due to the coating of the protein by HFIP molecules, which is not completely homogeneous. In particular, the HFIP molecules are found in higher concentrations close to the regions with higher α-helical tendency and therefore it is site-specific. Moreover the peptide-peptide interactions that enhance helical formation are not disturbed by the HFIP, maybe due to the big size of this cosolvent. In summary, the effect of HFIP on α-helical stabilization has also an entropic and enthalpic character. The first from the removal of water from protein surface, the second because the higher stabilization of intra-molecular hydrogen bonds.

Comments on urea vs. HFIP molecular action mechanisms

Surprisingly there is not much difference between the molecular action mechanisms of both solutions. However, the effect they induce on the protein structure is clearly different. We may speculate and just attribute the different behavior to the difference in dielectric constant between both solutions. For proteins immersed in high dielectric medium such as the urea-water solution Coulombic interactions are attenuated therefore the intra-molecular hydrogen bonds will be weakened. Consequently in the case of HFIP-water solution the scenario is the opposite, the charges will 'see' much better the other charges and the hydrogen bonds then will be reinforced. Undoubtedly, the big size of these molecules is a factor to the apparent similarity between both cosolvent-solutions action mechanisms.

This notion is supported by recent work of Schellman(Schellman, 2003) which demonstrates how the stability of the changes induced by mixed solvents (reagents or osmolytes in aqueous solution) depends on the balance between the changes in the excluded volume and the contact interaction. For denaturation the increase in excluded volume is always positive, however the interaction of a reagent such as urea with the protein is strong enough to compensate the change in volume, leading to a destabilizing effect.

4.3 Membranes

All cells, eukaryotic and prokaryotic, are enclosed in membranes. The membranes consist in arrangements of lipids and proteins. These proteins make membranes permeable to specific compounds, and transport inside-outside the membrane as well as transmittance of information is possible because of them. Therefore they are important targets for pharmaceutical companies.

The lipid bilayers consist of two leaflets of tightly packed lipid molecules, where the hydrophobic lipid tails point towards the inside of the bilayers, whereas the hydrophilic headgroups stay at both external surfaces of the bilayers. The thickness of the hydrophobic interior part is 25-30 Å(Popot and Engelman, 2000), while the polar region thickness is about 10-15 Å. Obviously, this restrains the length of the transmembrane (TM) proteins which are located inside the membrane normal to the bilayer plane (TM helices form a tilt angle on average of 24º). TM proteins consist of series of 20-24 residues (Popot and Engelman, 2000) highly hydrophobic, flanked by polar and highly charged regions. The lipid bilayers are asymmetric and whereas the inner face of the bilayers in contact cytoplasm are negatively charged, the outer face has a total neutral charge(Singer and Nicolson, 1972). Accordingly, the positively charged regions of membrane proteins will normally be forming loops in the inner cytoplasmatic side of the membrane. The low dielectric medium inside the lipid bilayers promote α-helical formation, since it is the easiest way to fulfill the hydrogen bond formation of the main chain. The hydrogen bonds formed in such a low dielectric environment are so energetically favorable that the entropic penalty of restraining a polypeptide with low helical propensity in a helical conformation is substantially reduced(Popot and Engelman, 2000). In turn, according to the two-stage model of peptide insertion into membranes proposed by Popot and Engelman(Popot and Engelman, 1990), the α helices are easily inserted in the membrane due to the hydrophobic effect and later they assemble forming homo- or heteromers such as α-helical bundles, which will confer a large stability to the complexes due to the very favorable interhelical hydrogen bond formation in the low dielectric medium. In absence of interactions with water other interactions than the hydrogen bonds, such as dipole-dipole interactions between helices,

lipid/protein interactions and especially the van der Waals interactions have to overcome the unfavorable loss in entropy of keeping the TM together. In particular, van der Waals interactions are supposed to be responsible for the detailed close packing of TM α helices(Popot and Engelman, 1990). However, alternative architectural motifs such as closed β-barrels are also formed by transmembrane proteins.

The two distinct zones of the bilayers (hydrophobic-hydrophilic) will affect the preference of the amino acids as we have already mentioned. The hydrophobic residues Leu, Ile, Val and Ala together with Gly and Phe (Sternberg and Gullick, 1990; Liu and Deber, 1998; Eilers et al., 2000)are the most abundant amino acids in the transmembrane regions of TM proteins, but also some polar residues as Ser, Thr and Cys can be found in the interior. Also Pro is abundant in transmembrane proteins. There is an apparent contradiction between the concept of Gly and Pro residues as "helix-structure breakers" for soluble proteins and forming frequently α helices in the membranes interior. In fact Gly and Pro residues act as "notches" in TM proteins facilitating the close contact and helix-helix packing(Javadpour et al., 1999). Moreover, Pro can promote helical kinks and Gly confer flexibility to the TM helices.

Hydrogen bond networks between helices of the type C_α-H_α···O have been pointed out as important factor for helical oligomerization in lipid bilayers(Fleming and Engelman, 2001; Senes et al., 2001). This ensemble of interactions can contribute significantly to the stability of helical assemblies because in a low dielectric medium this type of hydrogen bond is around a half of the energy of a regular amide hydrogen bond(Scheiner et al., 2001). Consequently small residues as Gly, Ala and Ser can be found frequently forming these networks.

Burial of charged residues inside the membrane will have such an energetic penalty that they are not normally found inside the membrane, but on the interface making contacts with water and the charged headgroups.

5 NMR RELAXATION AND MOTIONAL MODELS

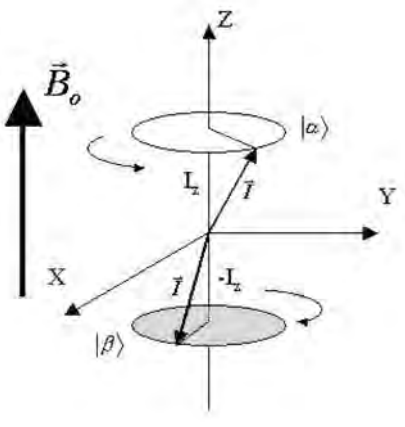
An NMR relaxation process is the process by which the net magnetization of the spins $\mathbf{M}_z$ of a system immersed in an external constant magnetic field $\mathbf{B}_o$ returns to its equilibrium position after having been perturbed by the application of a radio frequency pulse at a particular radio frequency ω_o called the Larmor frequency and applied transversally to $\mathbf{B}_o$. Since the relaxation rate is intimately bound to the physical environment of the nuclei as well as to their motions NMR spectroscopy can provide information about the physical structure and the dynamics of the system,.

The nuclei involved in NMR spectroscopy experiments are those with an odd number of protons or neutrons. In particular the half-spin nuclei ^{1}H, ^{13}C and ^{15}N are the atoms used for the study of proteins.

Microscopically what happens is that when an external magnetic field $\mathbf{B}_o$ is applied the nuclear magnetic momenta, or spins, will interact with the external magnetic field and for spin-half nuclei two possible energy levels will be populated with energies $\pm\frac{1}{2}\hbar\gamma B_o$ respectively. The proportionality constant γ is the gyromagnetic ratio, and is a property of the nucleus.

If the system would not be in equilibrium transitions between these two energy levels will occur in order to recover the equilibrium. The energy emitted or absorbed in this transition is $\hbar\gamma B_o$. Thus a line at frequency $\gamma B_o / 2\pi$ Hz will be observed in the NMR spectrum. This is indeed a radio frequency. The product $-\gamma B_o$ is the Larmor frequency ω_o.

NMR spectroscopy analyzes the behavior of these magnetic moments in the applied magnetic field by measuring the absorption/emission of these radio frequency signals associated with the transitions between the energy levels of the spins. In the system there are sources of local random magnetic fields that fluctuate at frequencies close to the Larmor frequency.



These local fields constitute a thermal bath were excitation can be released allowing the system to return to the equilibrium. This is the basis of the *longitudinal relaxation* also called *spin-lattice relaxation*. Similarly, the decay to zero of the transversal components is called *transverse relaxation* or *spin-spin relaxation*. The transverse relaxation mechanism, however, is more complex and has two contributions, a secular contribution similar to the longitudinal relaxation, and a non-secular. The random motion decreases the rate of the secular contribution(Keeler, 2005).

Figure 9. Splitting of energy levels in presence of an external magnetic field.

The rates at which this relaxation process occur are denoted by $R_z=1/T_1$ for the longitudinal relaxation and $R_{xy}=1/T_2$ for the transverse relaxation

5.1 Scalar coupling or J coupling

The asymmetric electronic cloud around a nucleus produces a shielding effect on the nucleus. The perturbation will be transmitted to another bound atom since the electron density distributed along the bond is also perturbed. In summary, one spin can propagate one excited state to another by this mechanism. This is called J-coupling and the result it is a symmetric splitting of the resonance line. It is called scalar because this type of interaction is independent of the orientation with respect to the external magnetic field. Also scalar couplings are independent of the distance between atoms, since excitation is transmitted through the network of electrons. J-couplings are large for directly bound atoms but also there are important J-couplings between atoms coupled 3 bonds apart. They are called 3-bond J-couplings, ^{3}J, and depend of the dihedral angle between the two atoms. The expression of the ^{3}J-coupling is given by the Karplus equation:

$$J(\theta) = A\cos^2(\theta) + B\cos(\theta) + C \qquad \textbf{(20)}$$

where A, B and C are empirically derived parameters.

J-couplings are important because they give an estimation of the bonding network and the molecular conformation.

5.2 Dipole-Dipole and Chemical Shift Anisotropy mechanisms

For half-spin nuclei the most common relaxation mechanism are the *dipolar* or *dipole-dipole (DD) mechanism* and the *chemical shift anisotropy (CSA) mechanism*.

The dipole-dipole mechanism is due to the interaction between two spins (or magnetic moments). Since it is a two dipole interaction, the interaction energy is proportional to $1/r^3$, with r the distance between spins. The distance dependence restricts this type of mechanism to spins closer than about 5 Å. Furthermore, this interaction will also depend on the orientation of the vector **r** with respect to the initial $\mathbf{B}_o$ as well as on both spins' gyromagnetic ratios. This kind of relaxation phenomenon causes cross relaxation that leads to the nuclear Overhauser effect (NOE).

The electronic charge distribution around the nucleus in the presence of a strong magnetic field $\mathbf{B}_o$ will generate a local field that will enhance or weaken the magnetic field felt by the nucleus. This will produce a shift in the line of the spectra. This is called CSA. The change in local field while the molecule tumbles will depend on the anisotropy of the chemical shift. It is strongly dependent on the chemical environment surrounding the nuclei.

5.3 Correlation Time and Spectral Density Function

Rotational diffusion consists in an ensemble of particles in a liquid moving randomly, that is undergoing Brownian motion. This kind of motion induce the local magnetic fields to fluctuate at frequencies of the order of the Larmor frequencies, therefore relaxation can be achieved. The average time needed for a particle to rotate a radian is called the *correlation time* τ_c. The *correlation function* defines the random diffusive motion and is characterized by this correlation time. For isotropic rotational diffusion the correlation function is:

$$C(t) = \exp(-\tau / \tau_c) \quad \textbf{(21)}$$

The *Spectral density function* J(ω)is the Fourier Transform of the correlation function C(τ) and is a measure of the amount of motion occurring at frequency ω. For isotropic rotational diffusion J(ω) is:

$$J(\omega) = \frac{2\tau_c}{1 + \omega^2 \tau_c^2} \quad \textbf{(22)}$$

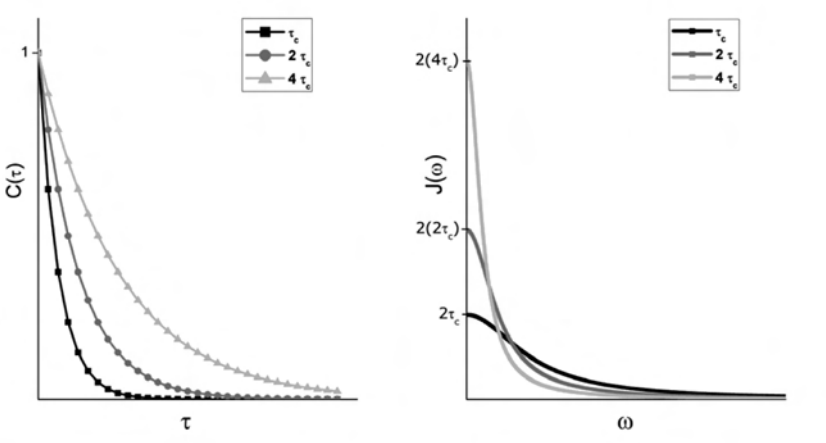


Figure 10. Correlation function and Spectral distribution function at different correlation times

For a given frequency ω_o the Spectral density function has a maximum at $\tau_c\omega_o=1$ when τ_c is varied. For processes with $\tau_c=1/\omega_o$ the relaxation rate is maximum. To undergo an optimal relaxation process the motion has to be not too fast ($\tau_c<<1/\omega_o$) neither too slow ($\tau_c>>1/\omega_o$). In addition, two motional regimes can be distinguished with respect to this quantity: the *fast motion* or *extreme narrowing limit* where $\tau\omega_o<<1$ and the *slow motion* or *spin diffusion limit* with $\tau\omega_o>>1$. For the fast motion limit $J(\omega) \approx 2\tau_c$ or $J(\omega)=J(0)$, which implies that the spectral density function is independent of the Larmor frequency. For the slow-motion limit $J(\omega) \approx 2/\omega^2\tau_c = J(0)/\omega^2\tau_c^2$, thus $J(\omega)<<J(0)$. Small proteins are in the slow motion limit.

5.4 Relaxation rates

For a system of two interacting half spins, I and S, the magnetization evolution has two main contributions: the *self-relaxation* of the spins, $R_z^{S,I}$ and the *cross relaxation*, σ_{12}. The relaxation-rates due to the dipolar-dipolar interactions can be written in terms of spectral density functions:

$$R_z^{(I)} = \frac{b^2}{20}\left[3J(\omega_{o,I}) + J(\omega_{o,I} - \omega_{o,S}) + 6J(\omega_{o,I} + \omega_{o,S})\right]$$

$$R_z^{(S)} = \frac{b^2}{20}\left[3J(\omega_{o,S}) + J(\omega_{o,I} - \omega_{o,S}) + 6J(\omega_{o,I} + \omega_{o,S})\right] \quad \textbf{(23, 24, 25)}$$

$$\sigma_{12} = \frac{b^2}{20}\left[6J(\omega_{o,I} + \omega_{o,S}) - J(\omega_{o,I} - \omega_{o,S}))\right]$$

Here b is the size factor and $b^2 = \frac{\mu_o \gamma_I \gamma_S h}{8\pi^2 r^3}$, where γ_I and γ_S are the gyromagnetic ratios of the nuclei, r the distance between them and *h* Planck's constant.

The cross term, $\sigma_{12;}$ can be positive or negative. The function takes negative values for long correlation times, therefore in the slow-motion limit; whereas it is positive in the fast-motion limit.

The transversal rate constant for two spins that relax through dipolar relaxation are:

$$R_{xy}^{(I)} = \frac{b^2}{20}\frac{1}{2}\left[\underbrace{4J(0)+6J(\omega_{o,S})}_{Secular} + \underbrace{3J(\omega_{o,I})+J(\omega_{o,I}-\omega_{o,S})+6J(\omega_{o,I}+\omega_{o,S})}_{Non-secular=\frac{1}{2}R_x^{(I)}}\right]$$

$$R_{xy}^{(S)} = \frac{b^2}{20}\frac{1}{2}\left[\underbrace{4J(0)+6J(\omega_{o,I})}_{Secular} + \underbrace{3J(\omega_{o,S})+J(\omega_{o,I}-\omega_{o,S})+6J(\omega_{o,I}+\omega_{o,S})}_{Non-secular=\frac{1}{2}R_x^{(S)}}\right] \quad \textbf{(26, 27)}$$

Or

$$R_{xy}^{(I)} = \frac{b^2}{20}\frac{1}{2}\left[4J(0)+6J(\omega_{o,S})+\frac{1}{2}R_x^{(I)}\right]$$

$$R_{xy}^{(S)} = \frac{b^2}{20}\frac{1}{2}\left[4J(0)+6J(\omega_{o,I})+\frac{1}{2}R_x^{(S)}\right] \quad \textbf{(28, 29)}$$

In the slow motion limit longitudinal and transverse relaxation rates differ substantially, but in the fast motion limit they coincide.

5.5 NOE enhancement

The goal of a heteronuclear steady-state NOE experiment is to enhance the signals that come from heteronuclei such as ^{13}C. The experiment consists in the irradiation of a spin I, commonly a proton, until it saturates. If the sample is irradiated long enough the second spin S, will reach a steady state. Two spectra are recorded in this experiment: the reference spectra, and the irradiated, with the target spin I saturated. The enhancement η is defined as:

$$\eta = \frac{peak\ height\ in\ irradiated\ Sp - peak\ height\ in\ reference\ Sp}{peak\ height\ in\ reference\ Sp}$$

The NOE enhancement in the fast limit is:

$$\eta_{SS} = \frac{\gamma_I}{2\gamma_S} \quad \textbf{(30)}$$

For a ^{13}C-^{1}H couple η_{SS} is ~2 and for ^{15}N-^{1}H η_{SS} is ~-4. The NOE enhancement is reduced once the correlation time increases.

This NMR section has been based on a recent published book about NMR spectroscopy (Keeler, 2005). We recommend it for more detail.

5.6 Motional Models

In order to obtain an analytical description of the spectral density function to describe the dynamics of the system, it is necessary to have a model of the motion.

If the amplitude of internal movement is restricted as it occurs for the helical parts of a protein, the approximation of isotropic rotation can be used, and the expression of the spectral density function is given by Eq. 22

Model-Free approach

Motion in complex and flexible systems can not be described in general by an isotropic rotation. In the Model-Free approach the global motion is described in terms of two independent movements. One is the overall tumbling and the other is the internal motion. The total correlation function is thought to be the product of two uncorrelated correlation functions, one describing the isotropic overall tumbling, characterized by τ_c, the second is the internal motion also considered as Brownian motion, characterized by τ_e. A new parameter, the *generalized order parameter S*, is also introduced which is model independent. This parameter gives an estimation of the vector flexibility and takes values between 0 and 1. If the vector is almost fixed, as inside a helix, then all the movement arises form the molecule overall tumbling and S is equal to 1, if the vector is very flexible then S is approximately equal to zero, and is like the only significant movement comes from the internal mobility of this vector.

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S^2)\tau}{1+\omega^2\tau^2} \right) \quad \textbf{(31)}$$

where

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e} \quad \textbf{(32)}$$

If the internal movement is very fast and belongs to the fast motion limit ($\omega\tau_e << 1$) then $\tau \approx \tau_e$ and

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1+\omega^2\tau_c^2} + (1-S^2)\tau_e \right) \quad \textbf{(33)}$$

More and more complex descriptions of motion based on the Model-Free have been developed called extended model-free. One of them decomposes the internal motion in slow and fast components, characterized by the correlation times,τ_s andτ_f, and by the structure factors, S and S_f. The spectral density function formula is similar to the model free.

5.7 MD simulations

In MD simulations all the motions are computed. All the atomic coordinates are recorded as a function of time and the correlation functions, C(t), and order parameters, S, of a vector as well as their Fourier transform, J(ω) can be directly computed. In contrast to dynamics NMR studies, MD simulations do not need any approximation or motional model to compute these functions. We can then try to identify the relevant motions responsible for the features in the correlation functions. The problem MD

simulations have to face is related with the conformational sampling. It is for this reason that motional features that occurs on a very long time scales can be poorly sampled and will not be reflected in the correlation function.

Another issue is related with the inability of accurately separating the overall tumbling motion from the internal motion when they occur on similar time scales, as may be the case for small and flexible peptides. This may be a source of discrepancy when comparing correlation times from the MD simulations and NMR experiments, which in turn have been calculated using a motional model.

6 FUTURE PERSPECTIVES

From the first 9.2 ps of simulation of bovine pancreatic inhibitor (BPTI) in vacuum 30 years ago the length of biomacromolecular simulations is now close to 100 ns with explicit solvent and up to μs with implicit solvation models (Karplus and McCammon, 2002), although also there exist simulations of the order of 100 μs with explicit solvent(Sorin and Pande, 2005) with a global distributed computing network simulated in ~20000 PCs. This increase of the simulation time, in some cases one order of magnitude larger than the experimental folding time, can really give full information about the kinetics and dynamics of the systems describing the biological phenomena as they occur in reality. Therefore the increasing power of computers can help in this context to obtain more physically meaningful simulations. In particular the larger computer power could be applied to the study of large protein multisubunit complexes (molecular machines) which are starting to be known to be necessary to achieve function in cells rather than individual proteins or other big systems such as the ribosome. Further MD application evolution should be directed to simulations at cellular level. Different groups have already started to give insights about the formation of phospholipid bilayers (Marrink et al., 2000; Bogusz et al., 2001). Also further knowledge about the structure and dynamics of channels, enzymes and other biological elements important in the cells is expected to be reached.

Furthermore developments of new algorithms that speed up the simulations without too many simplifications would help in this direction. On the other hand, also new algorithms that take into account more detail about the non covalent interactions as polarization or about high vibrational modes (Janezic et al., 2005) or further force field developments will give a more accurate description of the process under study as well as larger conformational sampling.

Recent simulation studies as well as protein engineering methods point out that the denatured state is far from being a complete random coil, and that non-native contacts but even native-like contacts are still present in this state. Also network analysis studies have shown the heterogeneous character of the denatured state which may lead to different folding pathways. The characterization of the denatured state would be necessary for the understanding of the folding process as well as to determine the reasons and possible solutions against diseases arising from fibril and amiloyd formation such as cystic fibrosis or Alzheimer. Another interesting point is to relate folding and unfolding, what are the similarities and the dissimilarities, if unfolding is the reverse in time of folding as suggested by the principle of irreversibility claimed by Daggett(Daggett, 2006).

With respect to the role of solvation it can be expected that also other solvents that could contain ions and cosolvents will be more commonly used in simulations in the future. Algorithms for constant pH simulations are also being developed; more realistic cellular processes could be reproduced such as the low pH environment observed at cellular viral entrance in shorter time. In particular, could be important the development of implicit models of solvation that could reproduce cosolvent-water solutions properties, such as urea-water solution. In this case, the solvation term would be easily computable, and it would be applicable to study the previous mentioned dilemma about folding/unfolding. Also, unfolding experiments by chemical reactants could be successfully performed in a short time range and a large understanding about the folding/unfolding process could be acquired. It would also be of interest to explore hybrid methods in which the first solvation shells of the protein are explicit solvent molecules while apart the solvent further away is considered implicitly.

Also the new used technique of network analysis and graph analysis would have important implications for the clarification of the protein folding process(Vendruscolo et al., 2002; Rao and Caflisch, 2004; Caflisch, 2006). This technique has the advantage that does not need the projection of the energy landscape on adequate reaction coordinates, which is difficult. Until now most of studies have focused on peptides, further applications to protein would also shed light on the similarities and differences between peptide and protein folding process.

In summary I believe that the field of computational simulations is just stating to mature and soon simulations on much larger systems and particularly complexes of biomolecules involved at different cellular processes together with experiments will be performed giving us a atomistic detail about how this process occur. Also the knowledge of transmembrane protein structure the mechanism of permeation of transmembrane pores will really contribute to the development of drugs that can enter into the cells. Moreover, the simulations would be larger, close to real time scale giving more insight to problems where large conformational sampling is needed such as the problem of misfolding and fibril formation and therefore it would help to the design of new strategies and drugs to eradicate the diseases related with this. Ultimately the combination of different computational techniques and experiments would lead someday to the prediction of the tridimensional structure as well as the function of a protein from its primary sequence. This will definitely contribute to the human wellness.

7 RESULTS AND CONCLUSIONS

7.1 Paper I: Urea Parametrization for Molecular Dynamics Simulations

Protein chemical denaturation is achieved by introducing proteins in a mixed solvent with high concentration of guanidinium chloride or urea. Despite the extended use of this procedure, the molecular mechanism by which these denaturants and in particular urea, induces denaturation is not yet completely understood. Although MD simulation can be a useful tool to elucidate the principal characteristics of this process at atomic level and different studies have tried to shed light on this subject, the question remains open.

The formation of urea dimers or higher polymerization complexes have been pointed out as an important factor for the properties of aqueous-urea solutions. The parametrization should play an essential role for a correct description of this, in particular the consistency between the parameters used for urea and the force field, as well as the balance in the strengths of the urea-urea, urea-water and urea-protein interactions.

An accurate description of the urea dimer formation is fundamental for the characterization of urea aqueous solutions. Despite the importance of formation of urea dimers or higher polymerization complexes, none of the models used for urea in the MD simulations has taken into account this fact.

The aim of this work is the design of a urea molecule model that reproduces substantial dimer formation while being consistent with the CHARMM22 force field and the TIP3 water model to use in MD simulations of protein unfolding processes. These parameters are compared with two other urea potentials obtained from density functional theory (DFT). Using the supramolecular approach a set of atomic charges (referred as the MD set of charges) is obtained to reproduce the urea dimer and urea-water complex interaction energies reported by Åstrand and coworkers(Astrand et al., 1994). In this case the most energetically favorable urea dimer is a cyclic dimer whereas the head to tail dimer is more unfavorable. The other two sets of charges were obtained from *ab initio* calculations, in particular from the planar conformation (C_{2v}) and non-planar conformation (C_2) of the urea molecule immersed in a self consistent reaction field that mimics water.

MD simulations of a box of 2M urea-aqueous solution with the three parameters sets were performed. Only the simulation with the MD parameter set displayed significant urea dimer formation. The simulations also designate the cyclic dimer as the most favorable although the dimer is slightly bent which allows better interaction than the purely planar dimer. However, in the simulations a relatively large content of tail-to-head dimers was found, although the probability for this kind of complex formation was lower than cyclic formation. In summary, this set of charges gives a good balance between urea-urea and urea-water interactions, with the TIP3 water model. Urea water solution displays quasi-ideal behavior since urea solvation and self-solvation are practically concentration independent. The model of water used for the simulations is also important for water and urea diffusion. The MD simulations with the ab initio set of charges had very poor and unstable dimer formation. In addition, when dimers form tail-to-head formation was preferred over cyclic dimer formation. Consequently these sets of

charges are inadequate for MD simulations of urea-aqueous solutions. On the other hand our *ab initio* calculations indicate that the urea molecule in water solution may display a planar conformation as seen in crystal.

7.2 Paper II: Effect of Urea on Peptide Conformation in Water: Molecular Dynamics and Experimental Characterization

Protein misfolding and aggregation is the source of serious diseases such as cystic fibrosis and Alzheimer, emphasizing the importance of the need of a detailed characterization of the nature of the denatured state. Also detailed knowledge about the ensemble average of the denatured state is needed for the understanding and the description of refolding simulations.

Proteins are normally denatured at high concentrations of denaturant or at high temperatures. The protein solubility increases with denaturant concentration. But how does urea denature proteins? Two main theories have guided the investigation of the molecular action mechanism of urea during the last four decades. One is the so called *indirect mechanism* were urea is supposed to act indirectly altering the structure of water, which in turn reduces the hydrophobic effect, while the second theory known as the *direct mechanism* proposes a model were urea interacts directly with the protein solvating together with water the protein. In turn this second approach has two different characters one mainly enthalpic involving hydrogen bond formation between urea-polar groups of the protein, and a second of entropic character arising from the interactions between urea and apolar groups. Every urea will remove almost four waters form the surface of an apolar residue, therefore these water molecules will regain entropy weakening the hydrophobic effect. Also other factors as urea and solute size will be important for the energy needed for cavity formation.

We have performed MD simulations of two peptides in water at 277K and 300K and in urea aqueous solution in order to shed light on the process by which urea induces denaturation. The peptides under study were a C-peptide analogue and a sequence variant of the former which has the same composition but different sequence. The C-peptide is the N-terminal fragment (residues 1-13) of the bovine pancreatic ribonuclease A (RNaseA); this peptide when isolated in aqueous solution at low temperature and pH shows high α-helical content. It has been largely studied to elucidate the importance of structural preference of amino acids or other factors such as specific side chain-side chain interactions, helical capping and helix dipole interactions among other factors for secondary structure formation. Different studies have pointed out the interaction between the residues Glu 2 - Arg 10 and between the aromatic rings of Phe 8 and His12, as the source of the high stability of this peptide. These interactions are pH dependent. The second peptide used in the simulations was observed experimentally not to contain α-helical structure.

In our simulations interdependent electrostatic side chain-side chain interactions were responsible of the high stability of the C-peptide analogue, confirming that side chain interactions are the factors that confer specificity to the native state (Dobson et al., 1998). Simulations suggested that a possible mechanism for the C-peptide denaturation would be initiated by the breakage of the Glu 2 - Arg 10 interaction: the Lys 7 residue would compete with Arg 10 for the formation of a hydrogen bond with Glu 2. Nevertheless, partial helical unwinding was observed for the second peptide that propagates cooperatively from the C-terminal towards the N-terminal. The conformational sampling was observed to be sequence and solvent dependent.

The water ability to form hydrogen bonds to the peptide is enhanced in urea-aqueous solution. Furthermore, the number of peptide-water hydrogen bonds increased substantially as well as the duration of particularly the peptide amide-water hydrogen

bonds. This can be important for the denaturation process since stable α-helical amides have a tendency to be highly protected from the solvent. Also the mobility of the water molecule is restricted close to the peptide surface when urea is present as cosolvent. Moreover, urea accumulated in excess around the peptide forming long-lived hydrogen bonds to it. The mechanisms by which urea and high temperature induce unfolding are of different nature. In summary, from our results we conclude that the urea induced denaturation process is a combination of both indirect and direct mechanisms.

7.3 Paper III: Molecular Dynamics Simulations of the E1/E2 Transmembrane Domain of the Semliki Forest Virus

In order to shed light on the kind of interhelical interactions achieved in the transmembrane (TM) spanning domain complex of glycoproteins E1 and E2 of Semliki Forest Virus (SFV) that are important for maintaining the SFV infectivity, MD simulations of the putative α-helical TM domain of this complex, the wild type as well as mutants, were performed in a dielectric medium that mimicked the hydrophobic lipid bilayer (ε=1). The mutations consisted mainly of Gly to Leu mutations for which experimental data is available. These Glycines are highly conserved among the alphavirus family. The two E1 variants consisted in four mutations ($E1^{4L}$) and in eleven mutations ($E1^{11L}$) close to the N terminal. Experimentally it was known that the viral budding of the virus with the $E1^{11L}$ peptide was severely inhibited whereas there was only a slight reduction of budding with the $E1^{4L}$ peptide. First, we analyzed the isolated TM spanning domains of each peptide. The simulations reveal that all the isolated transmembrane spanning segments formed helices, with the E1 wild-type being the most flexible. Thereafter, from multiple E1/E2 complex simulations we calculated the probabilities of contact between the residues at E1 and E2 peptides. The results clearly indicated that the possible residues located at the helical interfaces displayed the typical motif of the left-handed coiled-coils *abcdefg*. The helices packed together in a left handed fashion with an interhelical angle close to -20° for the wild type complex. Further investigation of the contacts between helices indicated that the $E1^{4L}/E2^{wt}$ complex was not as tightly packed as the wild type complex and that the $E1^{11L}/E2^{wt}$ complex displayed very poor contacts. This was in agreement with posterior experimental results that have found that in fact the $E1^{4L}/E2^{wt}$ complex and not only the $E1^{11L}/E2^{wt}$ complex is substantially less stable than the wild type.

In summary, we suggest that the TM region of the E1/E2 wild type complex is formed by two helices that intimately pack together their N-terminal regions in a parallel left-handed two-stranded rope. The close packing between helices was achieved by a special arrangement of small and medium size residues (Gly, Ala, Ser and Leu) in a complementary way at the significant positions of the coiled-coil heptad (***a****bc****de****f****g***). This spatial arrangement allows the formation of a network of C_α-H···O hydrogen bonds between the E1 and E2 peptides that may have an important role in the stabilization of the complex. The complementary between residues at the E1 and E2 interhelical interfaces is an indication of a compromise between E1 and E2 that is the key for insuring heterodimerization and consequent infectivity against possible mutations of the transmembrane spanning domains of the viral glycoproteins E1 and E2.

To test this compromise we constructed a new mutant ($E2^{3L}$) and we performed simulations on the $E1^{wt}/E2^{3L}$ and $E1^{4L}/E2^{3L}$ complexes. Our hypothesis is that the $E1^{wt}/E2^{3L}$ complex will exhibit similar features to the $E1^{4L}/E2^{wt}$ complex, that is the two helices pack their N-terminals in a two-stranded rope fashion although with weaker interactions than the wild type complex, while the $E1^{4L}/E2^{wt}$ complex should at most form scarcely and weak contacts analogously to the $E1^{11L}/E2^{wt}$ complex. Our results confirm our hypothesis.

Our findings for the packing of the TM spanning domains of E1 and E2 were extrapolated to sketch TM helical packing in the other related alphaviruses. Although the E2 sequence it is not very well conserved among these viruses, the kind of interactions that the E1/E2 complexes could achieve if following the same pattern as the SFV are of the same kind as those found in the SFV packing regions. Moreover, when some of the

related virus E1 proteins displayed a bulky side-chain at the putative packing region, E2 compensates this effect by introducing smaller residues in the interhelical face to which the former would pack against.

As final remarks, just to outline that our results correctly predicted defects in heterodimerization of the mutated E1/E2 complexes that were manifested by the decrease in stability of the virus (Sjoberg and Garoff, 2003), and explain the non-infectivity of other of the mutants. We believe that our simulation strategy can correctly discriminate between correct and improper packing of the TM of SFV and we suggest that our method can be used as a first step for the design of new mutants as well as to roughly sketch the TM spanning domains of the other related alphaviruses.

7.4 Paper IV: Dynamics of the Flexible Peptide Motilin by NMR, Fluorescence Anisotropy and Molecular Dynamics Analysis

Oligopeptides and proteins dynamics can be characterized experimentally at atomic level by NMR and fluorescence anisotropy decay (FAD). However, the description of the dynamics of oligopeptides is not straightforward since they are more flexible, and exhibit a greater conformational variation than large proteins. In addition, the correlation times obtained by FAD are in general shorter than from NMR. Recently both approaches have been reconciled (Damberg et al., 2002). We performed MD simulations using the motilin peptide to study at atomic level the dynamics of the peptide and to verify than in fact the motions described by both experimental approaches agree. We also discuss the differences when using different motion models, as well as using different parts of the peptide.

Our results indicated that the N-terminal of the peptide in the 30% HFIP water solution is highly flexible, while the fragment from residue 9-20 forms a rather stable α helix.. This is in agreement with NMR experimental results. However, analysis of the backbone RMSD and the backbone dihedrals distribution along the whole trajectory (~45 ns) indicated the existence of at least three different conformations sampled by the peptide. The different configurations differed mainly in the N-terminal structure. The dynamics described by two vectors of the Tyr 7 ring used to mimic the NMR and FAD experiments agreed, although one has to be careful with the vector selection since there are vectors in the ring that are insensitive to of the χ_1 angle. Therefore the choice of vector of the ring is very important to get agreement between the correlation times. NMR relaxation parameters of all the backbone N-H vectors were also obtained and clearly indicated the dissimilar dynamical behavior of the N-terminal and the helix. Correlation times were also obtained for the C^{α}-H^{α} vector of Leu 10, which is located in a more stable region of the peptide than Tyr 7, giving significantly different results to those of the Leu 10 backbone N-H vector and those from the Tyr 7, as seen experimentally. This underlines the difficulty of obtaining reliable and robust dynamic parameters in the case of flexible peptides.

MD simulations of the peptide were performed in water too. The structural analysis shows a much more flexible peptide with less helical formation than in HFIP, mainly the peptide looses helicity close to the C-terminal. However, comparison with the experimental J-couplings although in qualitative agreement, indicated an over estimation of the helical character of the peptide. The motion of the peptide in this solvent has some degree of anisotropy. The correlation times of the overall tumbling were much shorter than in the HFIP-water solution. However, the data was difficult to interpret. Similarly the choice of vector in the Tyr ring or at Leu 10 gives different correlation times.

Our calculations reflect the difficulty of obtaining accurate or physically significant correlation times. The main problem arises from the decoupling between overall and internal motion necessary for the calculation of the order parameter. Also comparison with experiment indicates insufficient conformational sampling of our simulations which is due to their limited duration. The big advantage of the calculation of relaxation and dynamic parameters by MD is that in our case the spectral distribution function is directly the Fourier transform of the correlation function and it does not need any motion-model to interpret the data as it happens experimentally.

8 ACKNOWLEDGEMENTS

Now at the end of this thesis, I look backwards and I see all the difficulties and the effort that I have putted on it. It is difficult to think that I would be able to do it without the big and always invaluable contributions of you, especially: my family, my friends and my supervisor. I want to use this occasion to thank all you:

In first place, I would like to thank my supervisor, Prof. Lennart Nilsson, for giving me the opportunity to work in his molecular modelling group, for having placed his trust in me, for his guidance, unquestionable support, unlimited kindness and fantastic sense of humor. Thank you.

I want to thank my closest CHARMMing friends: Jan Norberg for everything, all the help, reading my papers, answering my questions, listening to me, etc, etc, etc. But above all for your unconditional friendship. Also for all the candies I robbed you! And Ansuman Lahiri, for teaching me about CHARMM, computers, research, bengali-English and your frienship. Thank you for making my life more enjoyable when I first arrived here.

Also to my sweet CHARMMing Joanna Sarzynska, for every thing we shared together along the way. I was lucky having you close to me.

I also wish to thank:

Prof. Inger Porsh-Hällström, for believing so strongly in me and all encouragement and understanding.

Angel Gutierrez for help with computers, sharing secrets and friendship. Máquina!

Kurt Berndt, especially for sharing with me his enthusiasm for teaching, I really enjoyed your lectures and I learned a lot being your assistant. Also for the C-peptide project.

Colleagues and former members of the Molecular Modelling group: Srikanta, Pekka, Johan, Nicolas, Boel, Katarina, Sofia, Olof, Milan. In particular, to Jianxin and Cristian for interesting discussions and nice company. And especially to Vineet for help ☺

All the current and former CSB colleages and the administration staff, especially Erik Lundgren, Prof. Rudolf Ladenstein, Maria Flocco, Andrey Karshikoff, Linda Arnfors, Mikael Karlström, Katerina Morgunova, Xiaofeng Zhang, Tobias Elgan, Prof. Hans Hebert.

All my Spanish friends for friendship, parties, afternoons in the park and "buenas risas". Y por mantener el espíritu español vivo. To Karin, Catti, Frida, Anna, Linda and all those friends who have received me so kindly and whose friendship has contributed to make my life more agreeable. To my Salamanca friends for all these years of friendship in spite of the distance.

Jordi Benach for help, support in the normal life and science and for always having a smile to give.

My porcelain Princess Therese Pham for everything, endless help and friendship. I wish that a big shiny sun will always enlighten your life.

A mi queridísima Cecilia. Por tu inagotable amistad y apoyo a lo largo de "media vida".

My Spanish and Swedish family, because you are great.

A mi hermana Marta por ser la mejor hermana del mundo. Por tantas y tantas cosas, que sería difícil de escribir en un par de líneas. ¡Es fantástico que siempre estés ahí! También a Javier por haber entrado triunfante en la family.

Quiero agradecer muy especialmente a mi padre todo el amor que nos diste. Pero también: Gracias papá porque tu me enseñaste a aprender.

Y a mi madre. ¿Cómo darte GRACIAS? Gracias por todo tu amor y entrega. Por tu ayuda sin precio. Por ser increíble.

And finally, thank you, my beloved little family with all my corazón y mi alma. To my dear Björn for all these years, your unconditional support, for always being by my side (JGD) and for such incredible, fantastic and wonderful two treasures: Olivia and Claudia. ¡¡¡ Os quiero mucho, mucho, mucho!!!

¡ Va por vosotros! ¡ Por las ilusiones!

9 REFERENCES

Albert R., H. Jeong, and A. L. Barabasi. 1999. Internet - Diameter of the World-Wide Web. *Nature* 401:130-131.

Aqvist J., and V. Luzhkov. 2000. Ion permeation mechanism of the potassium channel. *Nature* 404:881-884.

Astrand P. O., A. Wallqvist, and G. Karlstrom. 1994. Nonempirical Intermolecular Potentials for Urea-Water Systems. *Journal of Chemical Physics* 100:1262-1273.

Batchelor J. D., A. Olteanu, A. Tripathy, and G. J. Pielak. 2004. Impact of protein denaturants and stabilizers on water structure. *J Am Chem Soc* 126:1958-1961.

Becke A. D. 1988. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Physical Review A* 38:3098-3100.

Becke A. D. 1993. Density-Functional Thermochemistry .3. The Role of Exact Exchange. *Journal of Chemical Physics* 98:5648-5652.

Bennion B. J., and V. Daggett. 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A* 100:5142-5147.

Bennion B. J., and V. Daggett. 2004. Counteraction of urea-induced protein denaturation by trimethylamine N-oxide: A chemical chaperone at atomic resolution. *Proc Natl Acad Sci U S A* 101:6433-6438.

Berendsen H. J. C., J. R. Grigera, and T. P. Straatsma. 1987. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* 91:6269-6271.

Bogusz S., R. M. Venable, and R. W. Pastor. 2001. Molecular dynamics simulations of octyl glucoside micelles: Dynamic properties. *Journal of Physical Chemistry B* 105:8312-8321.

Breneman C. M., and K. B. Wiberg. 1990. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. *Journal of Computational Chemistry* 11:361-373.

Brooks B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* 4:187-217.

Brooks C. L. 2002. Protein and peptide folding explored with molecular simulations. *Accounts of Chemical Research* 35:447-454.

Brown R. D., P. D. Godfrey, and J. Storey. 1975. Microwave-Spectrum of Urea. *J. Mol. Spectrosc.* 58:445-450.

Buck M. 1998. Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Quarterly Reviews of Biophysics* 31:297-355.

Caflisch A. 2006. Network and graph analyses of folding free energy surfaces. *Current Opinion in Structural Biology* 16:71-78.

Caflisch A., and M. Karplus. 1995. Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J Mol Biol* 252:672-708.

Caflisch A., and M. Karplus. 1999. Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure Fold Des* 7:477-488.

Cavalli A., P. Ferrara, and A. Caflisch. 2002. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. *Proteins-Structure Function and Genetics* 47:305-314.

Chan H. S., and K. A. Dill. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins-Structure Function and Genetics* 30:2-33.

Chirlian L. E., and M. M. Francl. 1987. Atomic Charges Derived from Electrostatic Potentials - a Detailed Study. *Journal of Computational Chemistry* 8:894-905.

Cornell W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* 117:5179-5197.

Daggett V. 2006. Protein folding-simulation. *Chemical Reviews* 106:1898-1916.

Daggett V., and A. R. Fersht. 2003. Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences* 28:18-25.

Damberg P., J. Jarvet, P. Allard, U. Mets, R. Rigler, and A. Graslund. 2002. C-13-H-1 NMR relaxation and fluorescence anisotropy decay study of tyrosine dynamics in motilin. *Biophysical Journal* 83:2812-2825.

Daura X. 2006. Molecular dynamics simulation of peptide folding. *Theoretical Chemistry Accounts* 116:297-306.

de Groot B. L., and H. Grubmuller. 2001. Water permeation across biological membranes: Mechanism and dynamics of aquaporin-1 and GlpF. *Science* 294:2353-2357.

Dinner A. R., A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends in Biochemical Sciences* 25:331-339.

Dobson C. M., A. Sali, and M. Karplus. 1998. Protein folding: A perspective from theory and experiment. *Angewandte Chemie-International Edition* 37:868-893.

Dyson H. J., and P. E. Wright. 2002. Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* 12:54-60.

Eilers M., S. C. Shekar, T. Shieh, S. O. Smith, and P. J. Fleming. 2000. Internal packing of helical membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* 97:5796-5801.

Fioroni M., M. D. Diaz, K. Burger, and S. Berger. 2002. Solvation phenomena of a tetrapeptide in water/trifluoroethanol and water/ethanol mixtures: A diffusion NMR, intermolecular NOE, and molecular dynamics study. *Journal of the American Chemical Society* 124:7737-7744.

Fleming K. G., and D. M. Engelman. 2001. Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proceedings of the National Academy of Sciences of the United States of America* 98:14340-14344.

Frank H. S., and F. J. Franks. 1968. Structural approach to the solvating power of water for carbohydrates - urea as a structure breaker. *J. Chem. Phys.* 48:4746.

Garcia A. E., and K. Y. Sanbonmatsu. 2001. Exploring the energy landscape of peptides in explicit solvent. *Abstr. Pap. Am. Chem. Soc.* 221:U395-U395.

Gilkerson W. R., and K. K. Srivastava. 1960. The Dipole Moment of Urea. *Journal of Physical Chemistry* 64:1485-1487.

Grdadolnik J., and Y. Marechal. 2002. Urea and urea-water solutions-an infrared study. *Journal of Molecular Structure* 615:177-189.

Gsponer J., and A. Caflisch. 2001. Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *Journal of Molecular Biology* 309:285-298.

Hong D. P., M. Hoshino, R. Kuboi, and Y. Goto. 1999. Clustering of fluorine-substituted alcohols as a factor responsible for their marked effects on proteins and peptides. *Journal of the American Chemical Society* 121:8427-8433.

Idrissi A. 2005. Molecular structure and dynamics of liquids: aqueous urea solutions. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* 61:1-17.

Idrissi A., F. Sokolic, and A. Perera. 2000. A molecular dynamics study of the urea/water mixture. *Journal of Chemical Physics* 112:9479-9488.

Ikeguchi M., S. Nakamura, and K. Shimizu. 2001. Molecular dynamics study on hydrophobic effects in aqueous urea solutions. *J Am Chem Soc* 123:677-682.

Islam S. A., M. Karplus, and D. L. Weaver. 2004. The role of sequence and structure in protein folding kinetics: The diffusion-collision model applied to proteins L and G. *Structure* 12:1833-1845.

Janezic D., M. Praprotnik, and F. Merzel. 2005. Molecular dynamics integration and molecular vibrational theory. I. New symplectic integrators. *Journal of Chemical Physics* 122.

Javadpour M. M., M. Eilers, M. Groesbeek, and S. O. Smith. 1999. Helix packing in polytopic membrane proteins: Role of glycine in transmembrane helix association. *Biophysical Journal* 77:1609-1618.

Jeong H., B. Tombor, R. Albert, Z. N. Oltval, and A. L. Barabasi. 2000. The large-scale organization of metabolic networks. *Nature* 407:651-654.

Jorgensen W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* 79:926-935.

Jorgensen W. L., and J. Tirado-Rives. 1988. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *Journal of the American Chemical Society* 110:1657-1666.

Kallies B. 2002. Coupling of solvent and solute dynamics - molecular dynamics simulations of aqueous urea solutions with different intramolecular potentials. *Phys. Chem. Chem. Phys.* 4:86-95.

Karplus M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* 9:646-652.

Karplus, M. and D. L. Weaver. 1994. Protein-Folding Dynamics - the Diffusion-Collision Model and Experimental-Data. *Protein Science* 3(4): 650-668.

Keeler J. 2005. Understanding NMR Spectroscopy. Chichester, West Sussex: John Wiley & Sons Ltd.

Kreschek G. C., and H. A. Scheraga. 1965. *J. Phys. Chem.* 69:1704.

Kritzer J. A., J. Tirado-Rives, S. A. Hart, J. D. Lear, W. L. Jorgensen, and A. Schepartz. 2005. Relationship between side chain structure and 14-helix stability of beta(3)-peptides in water. *Journal of the American Chemical Society* 127:167-178.

Kuharski R. A., and P. J. Rossky. 1984a. Molecular dynamics study of solvation in urea water solution. *J. American Chemical Society* 106:5786-5793.

Kuharski R. A., and P. J. Rossky. 1984b. Solvation of Hydrophobic Species in Aqueous Urea Solution: A Molecular Dynamics Study. *J. American Chemical Society* 106:5794-5800.

Lazaridis T., G. Archontis, and M. Karplus. 1995. Enthalpic contribution to protein stability: Insights from atom-based calculations and statistical mechanics. Advances in Protein Chemistry, Vol 47. San Diego: ACADEMIC PRESS INC. p 231-306.

Lazaridis T., and M. Karplus. 2003. Thermodynamics of protein folding: a microscopic view. *Biophys. Chem.* 100:367-395.

Leach A. R. 2001. Molecular Modeling Principles and Applications. Hall P, editor. Harlow, England: Pearson Education Limited.

Lee C. T., W. T. Yang, and R. G. Parr. 1988. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Physical Review B* 37:785-789.

Levitt M., M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett. 1997. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *Journal of Physical Chemistry B* 101:5051-5061.

Liu L. P., and C. M. Deber. 1998. Uncoupling hydrophobicity and helicity in transmembrane segments - alpha-helical propensities of the amino acids in non-polar environments. *Journal of Biological Chemistry* 273:23645-23648.

Ma J. P., P. B. Sigler, Z. H. Xu, and M. Karplus. 2000. A dynamic model for the allosteric mechanism of GroEL. *Journal of Molecular Biology* 302:303-313.

Mackerell A. D. 2004. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* 25:1584-1604.

Makhatadze G. I., and P. L. Privalov. 1995. Energetics of protein structure. Advances in Protein Chemistry, Vol 47. San Diego: ACADEMIC PRESS INC. p 307-425.

Marrink S. J., D. P. Tieleman, and A. E. Mark. 2000. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *Journal of Physical Chemistry B* 104:12165-12173.

McCammon J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of Folded Proteins. *Nature* 267:585-590.

Mountain R. D., and D. Thirumalai. 2003. Molecular dynamics simulations of end-to-end contact formation in hydrocarbon chains in water and aqueous urea solution. *J Am Chem Soc* 125:1950-1957.

Mulliken R. S. 1962. Criteria for Construction of Good Self-Consistent-Field Molecular Orbital Wave Functions, and Significance of Lcao-Mo Population Analysis. *Journal of Chemical Physics* 36:3428-&.

Norberg J., and L. Nilsson. 2000. On the truncation of long-range electrostatic interactions in DNA. *Biophysical Journal* 79:1537-1553.

Onuchic J. N., and P. G. Wolynes. 2004. Theory of protein folding. *Current Opinion in Structural Biology* 14:70-75.

Perham M., J. Liao, and P. Wittung-Stafshede. 2006. Differential effects of alcohols on conformational switchovers in alpha-helical and beta-sheet protein models. *Biochemistry* 45:7740-7749.

Popot J. L., and D. M. Engelman. 1990. Membrane-Protein Folding and Oligomerization - the 2-Stage Model. *Biochemistry* 29:4031-4037.

Popot J. L., and D. M. Engelman. 2000. Helical membrane protein folding, stability, and evolution. *Annual Review of Biochemistry* 69:881-922.

Rao F., and A. Caflisch. 2004. The protein folding network. *Journal of Molecular Biology* 342:299-306.

Ravasz E., and A. L. Barabasi. 2003. Hierarchical organization in complex networks. *Phys. Rev. E* 67.

Religa T. L., J. S. Markson, U. Mayor, S. M. V. Freund, and A. R. Fersht. 2005. Solution structure of a protein denatured state and folding intermediate. *Nature* 437:1053-1056.

Roccatano D., M. Fioroni, M. Zacharias, and G. Colombo. 2005. Effect of hexafluoroisopropanol alcohol on the structure of melittin: A molecular dynamics simulation study. *Protein Science* 14:2582-2589.

Roseman M., and W. P. Jencks. 1975. Interactions of Urea and Other Polar Compounds in Water. *J. Am. Chem. Soc* 97:631-640.

Salvi G., P. De los Rios, and M. Vendruscolo. 2005. Effective interactions between chaotropic agents and proteins. *Proteins-Structure Function and Bioinformatics* 61:492-499.

Scheiner S., T. Kar, and Y. L. Gu. 2001. Strength of the (CH)-H-alpha center dot center dot O hydrogen bond of amino acid residues. *Journal of Biological Chemistry* 276:9832-9837.

Schellman J. A. 1955. *Comp Rend. Trav. Lab. Carlsberg Ser. Chim.* 29:223.

Schellman J. A. 2003. Protein stability in mixed solvents: A balance of contact interaction and excluded volume. *Biophysical Journal* 85:108-125.

Senes A., I. Ubarretxena-Belandia, and D. M. Engelman. 2001. The C alpha-H center dot center dot center dot O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. *Proceedings of the National Academy of Sciences of the United States of America* 98:9056-9061.

Shimizu S., and H. S. Chan. 2002. Origins of protein denatured state compactness and hydrophobic clustering in aqueous urea: inferences from nonpolar potentials of mean force. *Proteins* 49:560-566.

Singer S. J., and G. L. Nicolson. 1972. Fluid Mosaic Model of Structure of Cell-Membranes. *Science* 175:720-&.

Sjoberg M., and H. Garoff. 2003. Interactions between the transmembrane segments of the alphavirus E1 and E2 proteins play a role in virus budding and fusion. *J. Virol.* 77:3441-3450.

Smith L. J., H. J. C. Berendsen, and W. F. van Gunsteren. 2004. Computer simulation of urea-water mixtures: A test of force field parameters for use in biomolecular simulation. *Journal of Physical Chemistry B* 108:1065-1071.

Sorin E. J., and V. S. Pande. 2005. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophysical Journal* 88:2472-2493.

Steinbach P. J., and B. R. Brooks. 1994. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *Journal of Computational Chemistry* 15:667-683.

Stephens P. J., F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. 1994. Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *Journal of Physical Chemistry* 98:11623-11627.

Sternberg M. J. E., and W. J. Gullick. 1990. A Sequence Motif in the Transmembrane Region of Growth-Factor Receptors with Tyrosine Kinase-Activity Mediates Dimerization. *Protein Eng.* 3:245-248.

Stokes R. H. 1967. *Aust. J. Chem* 20:2087.

Tirado-Rives J., M. Orozco, and W. L. Jorgensen. 1997. Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry* 36:7313-7329.

Tobi D., R. Elber, and D. Thirumalai. 2003. The dominant interaction between peptide and urea is electrostatic in nature: a molecular dynamics simulation study. *Biopolymers* 68:359-369.

Tsai J., M. Gerstein, and M. Levitt. 1996. Keeping the shape but changing the charges: A simulation study of urea and its iso-steric analogs. *Journal of Chemical Physics* 104:9417-9430.

Wallqvist A., D. G. Covell, and D. Thirumalai. 1998. Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. *Journal of the American Chemical Society* 120:427-428.

van Gunsteren W. F. 1987. GROMOS. Groningen Molecular Simulation Program Package. *University of Groningen: Groningen, 1987*.

Vanzi F., B. Madan, and K. Sharp. 1998. Effect of the protein denaturants urea and guanidinium on water structure: A structural and thermodynamic study. *Journal of the American Chemical Society* 120:10748-10753.

Weerasinghe S., and P. E. Smith. 2003a. Cavity formation and preferential interactions in urea solutions: Dependence on urea aggregation. *Journal of Chemical Physics* 118:5901-5910.

Weerasinghe S., and P. E. Smith. 2003b. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* 107:3891-3898.

Vendruscolo M., and C. M. Dobson. 2005. Towards complete descriptions of the free-energy landscapes of proteins. *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 363:433-450.

Vendruscolo M., N. V. Dokholyan, E. Paci, and M. Karplus. 2002. Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E* 65.

Wiberg K. B., and P. R. Rablen. 1993. Comparison of Atomic Charges Derived Via Different Procedures. *Journal of Computational Chemistry* 14:1504-1518.

Zhang Z., Y. Zhu, and Y. Shi. 2001. Molecular dynamics simulations of urea and thermal-induced denaturation of S-peptide analogue. *Biophys Chem* 89:145-162.

Zou Q., B. J. Bennion, V. Daggett, and K. P. Murphy. 2002. The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. *J Am Chem Soc* 124:1192-1202.

Zou Q., S. M. Habermann-Rottinghaus, and K. P. Murphy. 1998. Urea effects on protein stability: hydrogen bonding and the hydrophobic effect. *Proteins* 31:107-115.

Åstrand P. O., A. Wallqvist, and G. Karlström. 1994. Molecular-Dynamics Simulations of 2-M Aqueous Urea Solutions. *Journal of Physical Chemistry* 98:8224-8233.