CHAPTER 1

Introduction
1.1 Proteins.

Proteins are large organic polymers made up of a linear chain of amino acids joined together by peptide bonds between the carboxyl and amino groups of the adjacent residues. Proteins are synthesized on the ribosome based on sequence information (primary structure) encoded within genes. They are the most abundant component within cells comprising more than half of the dry weight. Proteins form an essential part of all living organisms, and participate in a wide variety of biological processes. Their functions range from the catalysis of biochemical reactions to formation of the scaffolding that maintains cellular shape [1, 2]. The functions that proteins performed are mostly determined by their structures.

The overall three dimensional structure of a protein is usually described in a hierarchical manner with local structural elements such as α-helix and the β-sheet being referred to as secondary structure. The arrangement of these secondary structural elements gives rise to tertiary structure and the assembly of individual protein units is referred to as quaternary structure [1].

The folding and unfolding of proteins is critical to a wide array of biological processes, including protein trafficking and cellular regulation [3]. In addition, diseases such as Alzheimer’s disease and Parkinson’s disease are directly related to the presence of unfolded or misfolded proteins [3, 4]. Thus elucidating the relationship between the sequence of a protein and its structure is of fundamental importance in biology, as is understanding the mechanism by which proteins fold to their native conformations.

1.2 The structure of peptides in solution.

The formation of short stretches of secondary structure in a protein may act as nucleation sites for the formation of native structure, which is believed to be the first phase of protein folding. For this reason, short chains of amino acids, also called peptides, that can form elements of secondary structure (such as α-helix and the β-sheets [5]) are widely used as models to study the initial stages of protein folding and even to probe protein-protein interactions. Peptides are also widely used in clinical
applications and the pharmaceutical formulation of peptides and proteins is an important part of modern drug development. As a consequence there is much interest in understanding and predicting the structural properties of peptides in solution.

1.3 Experimental techniques to determine peptide structure and their uncertainties.

Peptides can be purified according to their size, solubility, charge and binding affinity. Their mass can be precisely determined by using electrospray ionization to generate aerosols of ionized proteins or protein fragments, which are then analyzed using mass spectrometry to measure their mass-to-charge ratios. Their amino acid sequence can be determined directly by methods such as Edman degradation, which sequentially removes one residue at a time from the amide end of the peptide, allowing the nature of the cleaved amino acid to be determined. However, the three dimensional structure of an individual peptide or a protein in solution cannot be determined directly. While a range of methods can be used to obtain structural information, ultimately this information must be interpreted in regard to a specific structural model.

• Information on the secondary structure.

Information on the presence of helices and β-sheet can be obtained using a range of spectroscopic techniques, such as infrared, visible and ultraviolet absorption spectra, or circular dichroism, by correlating the absorption intensity in specific regions of the electromagnetic spectra with presence or absence of specific elements of secondary structure. However, each method is best suited to a particular application; thus, any one method cannot be used to fully quantify these secondary structural transitions simultaneously.

For example, infrared spectroscopy (IR) is sensitive to changes in the electric dipole moment of a molecule associated with intramolecular vibrations such as when a bond lengthens and contracts. Molecular bonds are not static, fixed constraints, but undergo a variety of harmonic and anharmonic motions. When electromagnetic radiation of the corresponding wavelength is projected on to the bond, energy is absorbed. By
observing the frequencies at which the absorption occurs, quantitative information concerning the environment of atoms associated with a particular bond type can be inferred. IR primarily gives information about collective motions of small numbers of atoms, through bonds, angles etc. Water, which is the principle solvent for biological systems such as proteins, absorbs strongly in the infrared and presents a major challenge in IR spectroscopy of proteins. Recently, a polarized infrared radiation method was developed which can reduce the effects associated with the very strong absorption by water. Using polarized IR, peaks corresponding to $\alpha$-helical hydrogen-bonded $–\text{C}=\text{O}$ groups can be observed more clearly. However, the speed of the measurement, 32 times per second, is too slow for the observation of fast peptide folding. Therefore, IR results mainly refer to the average of an ensemble at the specific time period.

Similarly, proteins and peptides with aromatic amino acids (Phe, Tyr and Trp) are intrinsically fluorescent when excited with UV light. If the compound is more concentrated, or if the number of aromatic residues in the peptide is increased, more light will be absorbed by the sample. For example, if a protein containing a single tryptophan is denatured with increasing temperature, the tryptophan will be exposed to an aqueous environment as opposed to a hydrophobic protein interior. Thus a red-shift emission spectrum will appear. In contrast, a blue-shift occurs when protein is embedded in the hydrophobic vesicle or micelle [6]. As a consequence, the nature of microenvironment of these residues in a protein can be estimated.

Polarized light can be divided into the left circularly polarized light (which follows a left-hand helix), and right circularly polarized light (which corresponds to a right-handed helicity), according to their rotations in the direction in which they propagate. Left and right circularly polarized light interact differently with chiral molecules. When equal amounts of left and right circularly polarized light of a selected wavelength are alternately radiated into a chiral sample, one of the two polarizations is absorbed more than the other one. This differential absorption of left- and right-handed circularly polarized light is Circular Dichroism (CD). CD spectra (in the UV region) can be used to estimate the fraction of a molecule that is in the $\alpha$-helix conformation, the $\beta$-sheet conformation, the beta-turn conformation, or some other (e.g. random coil) conformation and used to study the dynamics such as of how the
secondary structure of a molecule changes as a function of temperature or of the concentration of denaturing agents [7]. However, CD cannot give detailed information such as the location of the secondary structure helices within the molecule, or how many helices in the protein, and due to the reason that solutions containing membrane structures are often strongly scattering in CD, it is not suitable for analysis of membrane embedded protein.

In general, each of these spectral methods is limited to measuring the sample at a time averaged level. Therefore, it is difficult to quantify structural transitions as they occur. This leaves a fundamental gap in our understanding of the underlying physics of secondary structure transitions that must be addressed by alternate methods.

- **Structural information at an atomic level.**

Structural information at the atomic level can be obtained using either X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

X-ray crystallography is a method of determining the arrangement of atoms within a crystal. When a beam of X-rays of either a single wavelength or a range of wavelengths passes through a protein crystal, it scatters in many different directions, giving a particular diffraction pattern depending on the spacing of the clouds of electrons surrounding the atoms. When the orientation of the crystal is changed systematically, the diffraction patterns arising from the angles of refraction and intensities of the scattered beams are also changed. This allows a three-dimensional image of the electron clouds of the molecule (electron density map) to be constructed. Using a series of refinement techniques, a set of Cartesian coordinates for every non-hydrogen atom in the protein is derived by fitting a molecular model to this electron density map. Although X-ray crystallography is a powerful method to elucidate the structure of a protein, the model fitting process biases the resulting structure and, in addition, it is often difficult to obtain a high-quality crystal of the protein of interest. As crystallization requires the formation of a (meta) stable lattice of protein structures; proteins that possess unfolded tertiary structures with many degrees of freedom in solution cannot be crystallized. Even when a crystal can be obtained, the conformational changes associated with crystallization can be pronounced for small
peptides. This problem is reflected by the near absence of high resolution X-ray structures of small peptides (around 20 a.a.) in the Protein Data Bank.

• **Information of structures in solution.**

Nuclear magnetic resonance (NMR) is by far the most successful technique used to elucidate the structures of small peptides in solution. When nuclei with a nonzero spin (for example, hydrogen ($^1\text{H}$) in proteins) are placed in a magnetic field they align with the field and electromagnetic pulses can be used to probe the local chemical environment. Specifically, the frequency required to achieve resonance is determined by the shielding effect of the surrounding electrons and the coupling to the external magnetic field (chemical shift). Geometric information can be obtained from spin-spin coupling which arises from the interaction of different spin states of the nuclei through the chemical bonds and results in the splitting of NMR signals ($J$-coupling). Furthermore, the relaxation time of a given spin-state depends on motions within the molecule and can provide information on the dynamics of a protein. Finally, the transfer of magnetization from one nuclear spin to another through dipole-dipole interactions (nuclear overhauser effect) provides spatial information such as an estimate of the distance between two neighboring nuclei. By analyzing the set of distances between specific pairs of hydrogen atoms, an ensemble of models can be produced that fits the NMR data, and an average model can be developed by determining the mean position of each atom across all models and adjusting the data to obey normal bond distances and angles. In comparison to X-ray crystallography, NMR contains a wealth of information about the dynamics of a protein or peptide in solution. For very small peptides, their tendency to self-assemble and fast dynamics makes it difficult to deduce high-resolution structures from data collected using what is a polydispersed sample [8]. Thus NMR is mainly suitable for determining the structures of larger water soluble peptides and smaller proteins.

All of the above, experimental techniques have the limitation that the information gathered must be interpreted in regard to a specific structural model. As a consequence, the theoretical study of peptide/protein structure and dynamics is a necessary and important companion to experimental investigations.
1.4 Current understanding of the mechanism of protein/peptide folding.

- **Background.**

The ability of peptides to fold spontaneously has been known for more than eighty years [9]. Despite this, the mechanism by which proteins fold remains unknown. As early as the 1920’s Wu [10] proposed that native proteins consist of regular structural elements, and that protein denaturation, such as induced by urea or high temperature, involves a process of unfolding and refolding. This led to the “thermodynamic hypothesis” [11] in which the native structure is assumed to be thermodynamically stable, i.e. it is the global free energy minimum. This suggests that one might be able to predict the native state [10] of a protein by calculating the global free energy minimum. Most proteins have, however, been shown to be only marginally stable: the native state of a protein under physiological conditions is only 20 to 40 kJ/mol more stable than the unfolded state and the average stabilization per residue is just 0.4 kJ/mol. This is less than the background thermal energy (0.5 kT per degree of freedom [2]) making the prediction of the folded conformation highly challenging.

- **Forces involved in protein folding.**

The specific conformation into which a particular amino acid sequence will fold, if any, is dependent on a fine balance between many competing intermolecular interactions. Dill [9] and Kauzmann [12] have argued that the dominant driving force is hydrophobicity. The evidence to support this view includes a large decrease in the heat capacity of the system associated with protein folding, which corresponds to the transfer of nonpolar sidechains from an aqueous to a nonpolar environment [13, 14]. In general the majority of nonpolar residues in proteins of known crystal structure are sequestered into the core of the protein, where they are largely protected from contact with water [15, 16]. Besides hydrophobicity, other forces such as electrostatic interactions also affect stability. Electrostatic interactions affect the protein stability in diverse ways. For example, electrostatic repulsions that arise when proteins are at non-physiological pH may destabilize a folded protein [12]. Alternately, ion-ion interactions such as salt bridge formation, can stabilize the folded protein [17]. The electrostatic interactions involved in hydrogen bonding together with van der Waals
interactions contribute to the formation of secondary structures and help maintain the conformational state of a protein [18]. Protein folding involves the collapse of the protein from a large volume to extremely compact and ordered native state [19], thus steric constraints and repulsive interactions in the folded state will cause a dramatic loss of entropy in the protein [9]. Therefore, while protein folding can be considered to be driven by hydrophobic forces and steric constraints [9], in reality the native structure is determined by a fine balance of many subtle factors.

• *The kinetics of protein folding.*

The processes of folding can be understood as finding one of multiple paths through a rugged energy landscape, with the protein encountering a variety of kinetic traps and energy barriers on its way to the native state [20]. In the case of a protein that folds without encountering any significant free energy barrier under a particular set of conditions, such as low temperature, the process may be described as a downhill folding [21]. If a protein folds in the absence of a free energy barrier under all conditions, the process is described as one-state folding (or global downhill folding). In one-state folding, there is a unimodal population distribution of the protein at all temperatures and denaturant concentrations, suggesting one conformational ensemble with a continuous unfolding transition [22]. Some ultrafast folding proteins may fold downhill [23]. In contrast, when a single free energy barrier exists, the folding process may be described as a two-state folding process. Here it is assumed only two protein ensembles (folded and unfolded) exist and a sharp unfolding transition is a component of the folding process [24]. For very small single domain proteins (less than 100 a.a.), their folding kinetics can often be described as a two state transition on timescales that can be probed experimentally [24]. This type of protein folding can be very fast, especially in the case of the formation of the basic elements of secondary structure: short α helices, β-hairpins and loops ($\tau_{\text{folding}} < 100 \mu s$) [23, 25], with helical proteins folding faster than β-hairpins or αβ proteins [26-30]. As the number and height of energy barriers increases, the folding process becomes complicated, muti-ensembles (intermediates) and muti-route unfolding transitions occur. The timescale of this type of protein folding is commonly on the order of several minutes or hours. This is because the protein has to pass through a number of intermediate states before reaching its native state. It is thought that large proteins, when outside the cell, may
fold through a series of meta-stable intermediate states [31].

Currently, due to the marginal stability and complexity of protein structure, protein/peptide folding remains a challenging problem for both theoretical and experimental investigations.

1.5 Methods to model peptide folding.

A wide variety of computational models have been used in an attempt to understand how proteins fold. These range from atomistic models, through to coarse-grained models, and simple lattice based models [32, 33].

• Atomistic models.

The highest level of accuracy is obtained when using an atomistic representation of the protein and solvent. The parameters used in these models are fitted to a combination of high level quantum chemistry calculations and a wide range of experimental data on small molecular systems [33, 34]. The high level of detail means that one must consider a large number of particles and propagate the system using a small time step (1-2 fs). This makes the direct simulation of folding processes involving proteins that fold on time scales of microseconds or longer very difficult. Therefore, the use of full atomistic models has been restricted to the study of small systems, often in combination with techniques to enhance sampling, such as replica exchange [35], local elevation [36], or meta dynamics [37]. The advantage of a fully atomistic simulation is that the results from the simulation can be used to compare the experiments directly.

• Coarse-grained models.

An atomistic representation of protein residues and solvent can be replaced with a lower-resolution coarse-grained model. In such models one averages over groups of atoms, effectively smoothing the potential energy functions and accelerates the speed of the simulation. Coarse-grained models have been developed to investigate longer time- and length-scales than are possible with atomic models but which are critical to
many biological processes, such as modelling interactions between lipid membrane proteins. Because the potential is smoothed, a larger time step (10-50 fs) is commonly used in the simulation [38]. However, due to its low-resolution in reproducing the physical properties of the protein and the solvent, transitions of protein folding may be poorly represented.

- **Simple lattice model**

Representation of protein residues can be further simplified as a self-avoiding chain on a lattice with beads representing different types of residues, such as hydrophilic and hydrophobic residues. Even simple lattice model can capture some of the essential features of protein folding [39-41]. These models allow for the efficient sampling of conformational space. They often have a well-defined global energy minimum which can be found by enumerating all energy states and calculating the corresponding free energies using Monte Carlo simulations. While the complex kinetics of certain large proteins can be studied using such approaches, these models do not contain sufficient information to describe individual proteins and therefore can not be directly related to experiment.

### 1.6 Molecular dynamics simulation techniques.

Molecular dynamics (MD) simulation techniques are widely used to study the equilibrium and non-equilibrium structural, dynamical and thermodynamic properties of molecular systems in both atomistic and coarse-grained simulations. In a MD simulation, Newton’s equations of motion (equation 1.2) are integrated numerically in order to generate the time dependent behavior of a system of $N$ interacting atoms.

$$ F_i = m_i \frac{d^2 r_i}{dt^2} $$ (1.2)

In equation 1.2, $r_i$ represents the Cartesian coordinates of a particle $i$ with mass $m_i$ at a specific time $t$. $F_i$ is the force acting on the particle which is given by the negative gradient of the potential energy function $V(r_1, r_2, ... r_N)$ of the system of $N$ particles.

$$ F_i = -\frac{\partial V(r_1, r_2, ... r_N)}{\partial r_i} $$ (1.3)

A simple but efficient algorithm to integrate Newton’s equations of motions in MD
simulations can be derived by taking a Taylor expansion of the velocity \( v_i ( t_n - \Delta t/2 ) \) at the point \( t = t_n \) and subtracting this from a Taylor expansion of \( v_i ( t_n + \Delta t/2 ) \) at \( t = t_n \). Neglecting terms of third and higher order in the time step \( \Delta t \) it can be shown that:

\[
v_i ( t_n + \Delta t/2 ) = v_i ( t_n - \Delta t/2 ) + m_i^{-1} f_i ( t_n ) \Delta t \tag{1.4}
\]

Using a similar approach, an equivalent expression for the position can be obtained:

\[
r_i ( t_n + \Delta t ) = r_i ( t_n ) + v_i ( t_n + \Delta t/2 ) \Delta t \tag{1.5}
\]

Together, equations 1.4 and 1.5 form the leap-frog scheme, where the velocity and position are calculated at different jumps, like in a game of leapfrog.

The reliability of a MD simulation is dependent on two factors. The accuracy of the interaction function or force field used to describe the potential energy of the system \( V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \), and the time scale that can be simulated. In general, the size and complexity of biomolecular systems requires the use of semi-empirical functions that have been parameterized to reproduce the thermodynamic properties of the systems of interest.

Non-bonded interactions are generally modeled using a Lennard-Jones potential \( V_{LJ} \)

\[
V_{LJ}(r_{ij}) = C_{ij}^{(12)} \frac{r_{ij}^{12}}{r_{ij}^{6}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \tag{1.6}
\]

to describe the van der Waals interactions and a Coulombic potential

\[
V_{coulomb} = \sum_{i \neq j} \frac{1}{4\pi \varepsilon_0} \frac{q_i q_j}{r_{ij}} \tag{1.7}
\]

to describe the electrostatic interactions. The Lennard-Jones parameters \( C_{ij}^{(12)}, C_{ij}^{(6)} \) are dependent on pairs of atom types. \( r_{ij} \) is the distance between particles. \( q_i, q_j \) are the charges of the particles. \( \varepsilon_0 \) is the permittivity in free space (vacuum).

Bonded interactions are represented by a combination of covalent bond-stretching, angle-bending, proper dihedrals, and improper dihedrals. As an example, the potential for the GROMOS force field can be written as:
Where $V_{\text{bond}}$ is the bond stretching potential. $V_{\text{angle}}$ is the bond angle bending potential. $V_{\text{dihedral}}$ is the proper dihedral angle torsion potential; $V_{\text{improper}}$ is the harmonic improper dihedral bending potential. $i$ is a particle in the system of $N$ particles. $b_i$ is the bond length, $\theta_i$ is the bond angle, $\varphi_i$ is the dihedral angle and $\xi_i$ is the improper dihedral angle. $b^0_i$ is the ideal bond length, $\theta^0_i$ is the ideal bond angle, and $\xi^0_i$ is the ideal improper dihedral angle. Moreover, $k_b$, $k^\theta$, $k^\varphi$ are the force constant for bonds and angles and $k^\xi$, $k^\xi$ are the dihedral force constants [42].

The ability of the force field to reproduce the potential energy surface is dependent on two factors. One is the functional form used to describe the potential energy of the system as a function of the atomic coordinates. The second is the parameters that are used in conjunction with this set of functions. Force fields can be classified based on the nature of the particles included in the model: "All-atom" force fields provide parameters for every atom in a system, including hydrogens, while "united-atom" force fields treat the hydrogen and carbon atoms in methyl and methylene groups as single interaction centers. "Coarse-grained" force fields provide a more abstract representation where groups of 3 or 4 atoms are represented by a single interaction site. Coarse-grained models trade atomistic detail and accuracy for computational efficiency and as noted previously are increasingly used to model large scale systems.

The most widely used biomolecular force fields include the CHARMM force field (Chemistry at HARvard Macromolecular Mechanics) developed by the group of Martin Karplus at Harvard University [43]; AMBER force field (Assisted Model Building and Energy Refinement) developed by the group of Peter Kollman at the
University of California, San Francisco [44, 45]; the OPLS (Optimized Potential for Liquid Simulations) force field developed by the group of William Jorgensen at Yale University [46]; and the GROMOS force field (GROingen Molecular Simulation library) developed by the group of Wilfred van Gunsteren at the University of Groningen and ETH Zurich. In this thesis, all the simulations were performed using the GROMOS force field [47] in association with the molecular dynamics simulation package GROMACS (GROningen MAchine for Chemical Simulations) version 3.2.1 [48, 49].

1.7 Simulation of peptide and protein folding

The first molecular dynamics simulations of a protein were performed by Andrew McCammon et al. in 1977 who simulated a 58 residue protein, bovine pancreatic trypsin inhibitor, in vacuum for 8.8 ps [50]. Due to the limitations of CPU power and the theoretical models at that time, the simulation took more than a month to complete. The same simulation run on current computers will be more than $10^5$ times faster, taking less than a minute. By 1983, theoretical models had been further developed by van Gunsteren et al. to a level whereby a 20 ps MD simulation of bovine pancreatic trypsin with 560 water molecules had become possible [51], and in 1997, computational advances had progressed to a point where a 100 ns MD simulation was performed on a helical-heptapeptide to investigate its folding process [52]. This marked a turning point in MD simulations, introducing them as a powerful technique to investigate the peptide and protein structure properties and folding kinetics.

Nowadays, with constant improvements in algorithm design, force fields and computational speed, microsecond simulations of fully hydrated proteins involving tens of thousands of atoms are now possible. The size of the systems that can be simulated have also increased allowing the characterization of more complex biological system such as the spontaneous aggregation of phospholipids into bilayers [53], peptide induced pore formation [54, 55], and spontaneous curvature of membranes [56] on a nanosecond timescale.
To evaluate the accuracy of current MD investigations of protein folding, Snow et al. has analyzed how well simulations predict protein folding kinetics and thermodynamics compared to experiment [57]. He concluded that though there are many encouraging cases where the simulated folding time was in agreement with the experimental estimations, such as the folding time of peptide BBA5 of 7.5(±4.2) µs by Pande and coworkers, which is in excellent agreement with the experimental folding time of 7.5(±3.5) µs [58]; prediction of the absolute rate is still computationally demanding and requires the advent of new methods and further increases in computer power [57]. In addition, more experimental data are needed to further validate simulations.

The dramatic improvements in protein simulations over the last 30 years has validated molecular dynamics as an integral tool for investigating the protein folding process, although the current simulation times are still too short for a complete search of the conformational space of a large protein throughout its folding process. However, at present, ab initio prediction techniques are possible for small peptides and proteins (about 100 a.a.). With the other computational techniques like homology modeling or replica exchange developing, MD simulations can provide a complimentary aspect to the experimental method used to investigate small protein folding.

### 1.8. Environmental effects on peptide folding

From the landmark study demonstrating that ribonuclease could be refolded after denaturation while still preserving enzyme activity, Anfinsen et al. [11] showed that, for many proteins, all the information required to direct the folding of a protein into the native state is present in the primary sequence of the protein. Nevertheless, the precise structure adopted by a protein in vitro or in vivo will strongly depend on the local environment. The primary focus of this thesis is to investigate the role the local environment plays in facilitating the folding of peptides and proteins.

Before starting an investigation of the various environment effects on peptide folding, the choice of force field for MD simulations, and its effect on peptide folding
simulations have been examined. The results are discussed in Chapters 2 and 3.

**Chapter 2** The effect of force field on the structure and dynamic properties of small peptides.

Using a test set of 5 peptides and 36 proteins, the ability of the GROMOS 43A1, 45A3 and 53A6 parameter sets and a modified version of the 53A6 parameter set proposed by Dr. Hao Fan and Dr. Bojan Žagrovčić to reproduce the conformational preferences of a series of small peptides has been examined. The factors that determine the nature of the phi and psi distributions have also been analyzed and a new set of dihedral potentials that give rise to better agreement between the simulations and the available X-ray data are suggested.

**Chapter 3. Estimating the folding rates of small peptides: Comparing simulation to experiment.**

In Chapter 3, five peptides, which show distinct conformation preferences experimentally, were examined using long time scale (µs) MD simulations to probe how fast these peptides obtain their native state from an extended conformation.

**Chapter 4. Alternative approaches to promote protein folding.**

Peptides fold differently in different solvent environments. Many experiments have shown that it is possible to stabilize folded proteins by placing them in nonpolar solvents such as benzene, hexane, and carbon tetrachloride [9]. This is because intermolecular hydrogen bonding and the formation of secondary structure are promoted as the polarity of the solvent environment reduced. However, it is well known that under some conditions, proteins will not fold into their biochemically functional forms but obtain alternative stable conformations [59]. These conditions include extremes of pH, high solvent temperature, the high solute concentrations, and the presence of chemical denaturants. Molecular chaperons are a set of proteins that interact with specific target proteins during the process of folding. Chaperons are believed to act by inhibiting nonproductive interactions and thus, allowing the protein to fold more efficiently to its native structure. Here we investigate a recently reported approach involving oscillating the solvent environment, to mimic the chaperon effects
in the MD simulations, as a way of accelerating folding and to make the folding of larger proteins accessible to simulation timescales.

Chapter 5. The effect of polyethylene glycol (PEG) spacers on the conformational properties of small peptides.

Peptides fold differently in various spatial environments. Factors such as charged or blocked termini play an important role in small peptide folding. At high concentrations, unfolded and partially folded peptides tend to aggregate; this increases the intermolecular interactions and prevents native peptide folding. Moreover, coupling a linker to the end of the sequence in order to change the properties of a peptide is also commonly used in the peptide synthesis industry. This may also cause changes to their folded conformation. Polyethylene Glycol (PEG) is used as an inert spacer in a growing number of biotechnological applications such as for the display of peptide epitopes in micro array techniques. Using molecular dynamics (MD) simulation techniques, we have investigated the influence of the PEG spacer on the conformation properties of the peptides to which it is attached. A series of five peptides with differing physical-chemical properties were examined.

Chapter 6. The interfacial structure of self-assembled switchable peptide surfactants.

Interfaces can promote or inhibit peptide folding, depending on the amino acid sequence of a peptide and the molecular structure of the interface. For example, the hydrophobic residues of a peptide are preferentially solvated by hexane, lipids or air, and the orientation of these interfacial hydrophobic residues may induce a significant effect on peptide folding. This phenomenon is evident in the formation of amyloids in Alzheimer’s disease, which is due to the misfolding and assembly of the Aβ peptide, which is accelerated by membrane interfaces [60, 61]. Recently a range of surface active peptides have been designed which form interfacial networks that can be switched between a high- and low elasticity state, giving the possibility to stabilize or destabilize emulsions and foams in a controlled manner. In this work, we report molecular dynamics (MD) simulations and neutron reflection (NR) studies of the interfacial structure of two stimuli-responsive peptides, AM1 and Lac21E [62]. The simulations were used to investigate the structural and dynamic properties of peptides
both in solution and at an air/water interface. The elastic properties of the film were estimated at different pH values from the fluctuations in the lateral pressure and compared to experimental measurements of the Young’s modulus.

**Chapter 7. Conclusions and perspectives.**

In conclusion, the primary sequence of a protein provides sufficient information for it to fold into its native structure. Protein folding is not an isolated phenomenon: there are many possible ways of folding and the local environment plays an important role in the adjustment of the structural preferences and the kinetics of the proteins folding. These factors will be discussed here.
References


59. Shortle, D., The denatured state (the other half of the folding equation) and its role in protein stability. FASEB Journal, 1996, 10, 27-34.


