Introduction

1.1 Fold — un-fold

Proteins are compounds of tremendous significance for all life on this planet and are found in all the living organisms and viruses known to us. Many proteins function as enzymes or as the building blocks of biological matter. Often the function depends on the specific three-dimensional structure of the protein. The structure of a protein is based on a linear sequence of amino acids in a polymeric chain. The specific sequence and the spatial arrangement — topology, or fold — of a peptide chain determine its three-dimensional structure. Due to the many degrees of freedom in this system, the relationships between the sequence, structure [Levinthal 1968] and function are complex and hard to predict. Although numerous protein structures have been experimentally solved, understanding how sequence leads to structure and thus to function remains one of the most fundamental challenges in biology.

The folding of a protein chain is a dynamic process that takes place in all living cells under so-called native conditions. Under non-native conditions, alternative folding pathways can be accessed that sometimes result in non-native structures with new functional properties, though more often the result is unfolding, i.e. the total loss of structure. Such changes on the molecular level can lead to systemic malfunction and disease in the organism, prominent examples are the pathological prion and amyloid proteins. In the case of pathogens, knowledge of the function of the pathogenic enzymes can open ways to interfere with infection, or to cure the disease. In the quest to explore and understand the functions and properties of proteins, alternative structures have been discovered that are based on non-native folds of known sequences, e.g. the molten globule. Modified enzymes are specifically engineered for applications in the molecular biology lab, the food, pharmaceutical and other industries.

Some physico-chemical conditions that are known to affect folding are temperature, pH, solvent or salt concentration, and complex environments and interfaces like membranes, li-
1.1 Fold — un-fold

Posomes or chaperones. Solvents that are typically used for biochemical unfolding in the laboratory are alcohols, acidic, basic or alkaline solvents, and their mixtures. These have been widely studied in chemical experiments, and by physical methods such as spectroscopy, e.g. X-ray diffraction, circular dichroism (CD), fluorescence, UV-visual (UV-VIS), Fourier transform infra-red (FTIR), or nuclear magnetic resonance (NMR) spectroscopy. Computational modelling is a further method of investigating the macroscopic properties of a compound, e.g. in Monte Carlo or molecular dynamics (MD) simulations.

This thesis studies the onset of structural changes in two native protein structures in MD simulation, addressing the following questions:

QI. Is urea a suitable denaturant for the partial unfolding of bacterial cytochrome c-550?

QII. Does the protonation of histidine residues trigger the restructuring of a pH-dependent viral fusion protein?

**Partial unfolding of a bacterial peroxidase by urea**

Cytochrome c is an integral component of the electron transport chain in mitochondrial and bacterial membrane systems. It is a redox-active protein carrying a heme group, the catalytic centre of the protein. In addition, c-type cytochromes show peroxidase activity in the presence of hydrogen peroxide H₂O₂ when the heme-iron is in the ferric state Fe³⁺ [George 1953; Harbury & Loach 1960; Ubbink et al. 1992];

In contrast to essential peroxidases, C-type cytochromes are very stable and remain highly soluble even under conditions of extreme heat, acidity and basicity. At the same time, the peroxidase activity of C-type cytochromes is approximately 1000-fold lower than the activity of, for instance, microperoxidases. This low activity was related to the limited accessibility of the heme pocket for peroxide substrates [Diederix et al. 2001, 2002a,b].

In experiment, the partial unfolding of bacterial cytochrome c lead to an approximately 1000-fold increase in activity, suggesting an expansion of the heme pocket that allowed the peroxide substrate to access the active site more easily [Diederix et al. 2004, 2002a; Worrall et al. 2005a]. If the conformation of this peroxidase could be easily and reversibly manipulated, the activity could be enhanced and regenerated multiple times. In this thesis MD simulations were used to examine the effect of the denaturant urea on the structure of a bacterial cytochrome c-550, in order to understand the relationship between partial unfolding and peroxidase activity.
Low-pH dependent activation of a viral fusion protein

Another physico-chemical condition that is known to influence protein folding and structure is pH. Some proteins require acidic pH to obtain a specific functional fold. Such proteins are found in a number of viruses that require low pH to infect the host cell, among them the Influenza viruses and members of the Alphavirus and Flavivirus genera. Dengue virus is a flavivirus and the most common mosquito-borne virus that infects humans. It causes dengue fever and dengue haemorrhagic fever, that can lead to death. During the last two decades dengue has emerged globally as a major public health concern [CDC 2008; WHO 2008]. No drugs for the treatment of the dengue diseases or vaccines to prevent infection by the dengue virus have yet been found.

The flaviviral envelope is covered with proteins that are thought to mediate the fusion of the viral and the cell membrane that allows the viral RNA to access and infect the cell. The host cell takes up the virus through endocytosis, enclosing the virus in an endosome. There, increasing acidification triggers the fusion of the virus with the endosomal membrane. Low pH usually denatures and potentially deactivates proteins, but in the case of the flaviviral envelope protein low pH is believed to trigger a large-scale conformational change that leads to its activation for fusion [Schibli & Weissenhorn 2004; Zimmerberg et al. 1993]. A number of models have been proposed [Gibbons et al. 2003; Harrison 2008; Helenius 1995; Poubourios et al. 1999], but the exact mechanism of viral protein-mediated membrane fusion are still unknown.

However, several essential aspects of viral fusion are known that could be of particular interest in the design of antiviral drugs. For instance i) the structural change on exposure to low pH is irreversible, and ii) activation is quickly followed by inactivation. In other words, it is possible to prevent infection by prematurely activating the viral fusion protein before the virus comes into contact with the host cell [Heinz 2003; Hurrelbrink & McMinn 2001; Mandl 2005; Mandl et al. 2001; McMinn 1997].

A particularly interesting aspect of low-pH-dependent viral fusion is that the pH of fusion is similar to the pK_a of protonation of histidine in water (pK_a = 6.0). This motivated the hypothesis that the protonation of histidine residues might trigger the activation of the viral fusion protein [Kampmann et al. 2006]. This principle of the so-called histidine-switch might present the link between the pH-dependency of the fusion process and a specific structural mechanism of the viral envelope protein.

MD simulations of a flaviviral envelope protein were performed to elucidate the effect of histidine protonation on the conformation of a low-pH-dependent viral fusion protein. The aim was to determine i) what effect histidine protonation has on the structure of the protein; and ii) whether these effects are related to the conformational change of the envelope protein.
observed in experiment after acidification.

## 1.2 Molecular dynamics simulation

Molecular dynamics (MD) simulation is a computational method that is used to obtain a detailed view of the dynamic behaviour of molecules. This method has proven especially useful in the modelling of complex biomolecular systems such as proteins, membranes or nucleic acids. In an MD simulation a deterministic time series of configurations is generated of a compound or a system of interest, in a temporal and spatial resolution that are usually unattainable in real macroscopic experiments.

MD simulation is based on physical interactions on the molecular level, which defines the accuracy of the method. Therefore the frequency of molecular vibrations sets the upper limit for the iterative time step of computation, which lies in the order of femto-seconds. The time-scale of a process and the size of a system that can be simulated are limited by the computer time available, i.e. the computational resources in real time, and by the speed with which the calculations are processed. For the simulation of a protein in atomistic detail, as performed in the studies presented in this thesis, several 100 nano-seconds of data could be generated, which compares to the time-scale of the motions of a loop, or the relative motions between two domains.

In an MD simulation the classical trajectory of a molecule is modelled by solving Newton’s equations of motion [Newton 1678]

\[
F = ma \\
= m\ddot{v}
\]

for the force \( F \), acceleration \( a \), mass \( m \) and velocity \( v \). Using the leap-frog algorithm, the velocity \( v \) and position \( r \) of a particle are updated at time step intervals of \( \Delta t \) according to

\[
v(t + \frac{\Delta t}{2}) = v(t - \frac{\Delta t}{2}) + \frac{F}{m}\Delta t
\]

\[
r(t + \Delta t) = r(t) + v(t + \frac{\Delta t}{2})\Delta t
\]

which is based on the Verlet algorithm [Verlet 1967]. The particles have physical-chemical properties, e.g. mass and charge, and their interactions are determined by a force field \( V(r) \):

\[
F = -\frac{\partial V}{\partial r}
\]
The particles are modelled as Lennard-Jones particles, i.e. impenetrable spheres that interact via an attractive van der Waals potential. Chemical bonds are described by harmonic potentials. Interactions between charged particles are modelled via a classical coulombic electrostatic potential acting between point charges. In order to lower the number of computations, non-bonded pair-interactions beyond a distance cut-off are neglected, and the long range electrostatic interactions in a dielectric are approximated by a reaction field. The degrees of freedom between hydrogens that are bound to heavy atoms are considered insignificant for the dynamics of the protein and are integrated in the so-called united-atom scheme to further save on the number of computations. The system is set in a periodic box to eliminate boundary artefacts, and the box dimensions are chosen such that the solute does not interact with its periodic image. The number of particles $N$, the pressure $p$ and the temperature $T$ are kept constant.

This level of description is used to generate the classical trajectory of an isothermic-isobaric (NPT) ensemble of chemically inert molecules. It is considered suitable for the modelling of thermodynamic equilibrium processes like conformational changes and molecular motions. Although the results of this iterative method are in principle deterministic, the interactions in a large molecule are too complex for an analytic prediction of the states. Therefore MD simulation presents an extremely helpful tool for the study of molecules.

While the underlying models are empirical, MD simulations have successfully reproduced experimental results [Daura et al. 1999; Güntert et al. 1997; Laasonen et al. 1993; MacKerell et al. 1998; Marrink et al. 2004; Marszalek et al. 1999; Oostenbrink et al. 2004; Rappe & Goddard III 1991; Srinivasan et al. 1998; Verlet 1967]. Often the MD trajectory of one simulated molecular specimen is integrated to infer the thermodynamic properties of a statistical ensemble; then the trajectory is assumed to be ergodic [von Neumann 1932]. In the evaluation of a simulation it is assumed that ergodicity is approximated through more extensive sampling, e.g. in longer trajectories. Due to the limited accuracy of floating-point operations, the time-series obtained is usually chaotic. In practice this turns into an advantage with respect to the sampling of the thermodynamic states, as a chaotic trajectory will not be caught in a periodic loop but continue to explore more phase-space with time.

While ergodicity is relevant for the thermodynamic characterisation of a system in equilibrium, many questions addressed by MD simulations do not aim at a complete thermodynamic description and may therefore neglect the ergodicity issue. For instance the simulations undertaken for this thesis investigate systems that are not in equilibrium. Here the question is whether a specific state is reached while the system equilibrates, and a positive answer presents a proof of concept. The above question QI is answered when the desired state is obtained, i.e. partially unfolded cytochrome c. QII is answered when any significant deviation from the initial state occurs, i.e. an altered conformation of the sE protein.
1.2 Molecular dynamics simulation

Apart from providing results on the atomistic, single molecule level, an important advantage of simulations in general is that conditions can be applied which may be dangerous or difficult to control in experiment. In addition, simulations can help find hypotheses, and thereby reduce the number of animal or human experiments, thus addressing important ethical issues in bio-medical research and drug development. Therefore simulations are a useful and indispensable alternative to handling real compounds or processes that can be hazardous to the experimentalist and the environment.