Molecular Simulation Studies of Peptide Aggregation in Small Finite Sized Systems and Near Surfaces

Dissertation

zur Erlangung des Grades eines
Doktor der Naturwissenschaften
(Dr. rer. nat.)

eingereicht beim
Fachbereich Chemie der Technische Universität Dortmund

vorgelegt von

MS (Med. Chem.) Gurpreet Singh
New Delhi, India

Dortmund, 2008

First Referee:	Prof. Dr. R. Winter
Second Referee:	Prof. Dr. A. Geiger
Date of Examination:	05–02–2009

Acknowledgement

The work described in this thesis was carried out in the lab of Prof. Dr. Roland Winter, under his direct supervision. I am thankful to him for providing me with an opportunity and resources to work on this project. This thesis wouldn't have taken its present form without the inputs of Dr. Ivan Brovchenko and Dr. Alla Oleinikova, I am greatful to both of them for their guidance. Dr. Dietmar Paschek helped me with the purchase and setup of the Linux cluster, and performing simulations with RPMD and Gromacs packages. Replica exchange simulations were carried out on LiDO cluster maintained by the University. I am greatful to Maximilian Andrews and Dr. Roland Krivanek for helping me with the maintenance of computers, and Dr. Karsten Vogtt, Michael Sulc, Dr. Claus Czeslik and Diana Radovan for enlightening discussions. I thank Dr. Nagarajan Periasamy, Dr. Nadeem Javid and Dr. Mahesh Kulharia for their help and support. I am thankful to Diana Radovan for proof reading the thesis and to Christian Reichhart for translating the summary of this thesis to German.

I am also thankful to all the members of the group for providing a pleasent working environment. Finally, the financial support from International Max Planck Research School in Chemical Biology is duly acknowledged.

Contents

Chapter 1

Introduction

Proteins were first described and named by the Swedish chemist Jöns Jakob Berzelius in 1838. The word protein comes from the Greek word '$\pi\rho\omega\tau\epsilon\iota o\varsigma$', which means 'standing in the front', 'in the lead'.[1] More than a century of research has shown that proteins are central to all living systems due to their involvement in all aspects of cell life, immune protection, cell-cell communication, physical support (hair and skin), and muscle movement. The solubility of proteins has been measured for more than a century, and the concept played a prominent role in the development of protein physical chemistry.[2] The majority of proteins in the human body are biologically active in a narrow range of temperature and pressure, and a 'folded' conformation is necessary for biological activity.

Relatively recently, several diseases have been associated with protein aggregation. Some proteins or protein fragments have been found to undergo a transition from the normally soluble forms to insoluble fibrils or plaques, which accumulate in various organs such as liver, spleen and brain. The mature aggregates often have well-defined fibrillar structures, and are generally known as amyloids, hence the term amyloidosis is used to describe many of the clinical conditions with which they are associated (see Ref. 3 and references therein). Protein aggregation is also a relevant issue in pharmaceutical industry, as it also affects the shelf-life of a protein solution, which is directly related to the economic viability of a typical protein based pharmaceutical product.[4] Hence, the understanding of protein sequence-structure relationships and the phase behavior of proteins has become a major research area in the field of biophysics.

1.1 Proteins

Proteins and peptides are built from smaller, individual amino acid units, with each amino acid containing an α-carbon bonded to an α-amino group, an α-carboxyl group, one hydrogen, and a side chain, R, of varying lengths and functionalities. There are twenty naturally occurring amino acids. Their names are commonly abbreviated with either a three-letter code or a one-letter code. Depending on the chemical structure of the side chain, the amino acids can be broadly divided into two different classes. The first class comprises those with hydrophobic side chains (non-polar) e.g. Ala (A), Val (V), Leu (L), Phe (F), etc. The second class comprises amino acids having hydrophilic side chains (polar) that are either charged (Lys (K), Glu (E), Asp (D) and Arg (R)) or uncharged (Gln (Q), His (H), etc). The amino acid glycine (G), the simplest of all the twenty amino acids, has only a hydrogen atom as the side chain. The growth of a protein occurs by a condensation reaction between the free amine and the free carboxylic group of two distinct amino acids, resulting in the formation of amide bonds. The secondary structure of the α-helix is formed by the hydrogen bonding between the NH and CO groups of the same strand, whereas the β-strand forms hydrogen bonds with other β-strands. These β-strands stabilize each other through intermolecular hydrogen bonding, forming β-sheets. Tertiary structures are built on these secondary structures, with the corresponding formation of higher-order structures such as $\beta\alpha\beta$-units, β-hairpins, and α-helix bundles. If a protein has two or more polypeptide subunits, their spatial arrangement is referred to as quaternary structure. On the basis of higher level structures, proteins can be classified into two major groups: fibrous proteins, having polypeptide chains arranged in long strands or sheets, and globular proteins, having polypeptide chains folded into a spherical or globular shape. The two groups are structurally distinct: fibrous proteins usually consist largely of a single type of secondary structure; globular proteins often contain several types of secondary structures. The two groups also differ functionally; the structures that provide support, shape, and external protection to vertebrates are made of fibrous proteins, whereas most enzymes and regulatory proteins are globular proteins.[5]

1.2 Protein folding

"Protein folding" is a term used to describe the process by which a polypeptide chain adopts its three-dimensional "native" conformation. To carry out their functions, proteins must fold rapidly and reliably. They must satisfy a kinetic requirement (folding must be completed within a reasonable time) and a thermodynamic requirement (the folded conformation must be stable under physiological conditions). Due to the vast amount of configurational space available to a protein chain, predicting protein folding from its primary sequence still remains one of the most important challenges in science.[6] It has been shown experimentally that, for some proteins, the stable native state of the protein survives in a closed range in the pressure-temperature plane.[7] In these studies, it was assumed that there are only two possible states of the protein, the native and the denatured state, and that the transition occurs without any intermediate states. The Gibbs free energy difference (ΔG) between the denatured and the native state was calculated. The transition line was defined by $\Delta G = 0$ and the curve was found to be elliptic-like in shape. In the framework of this theory, heat, pressure and cold denaturation are the three ways of protein unfolding.[8]

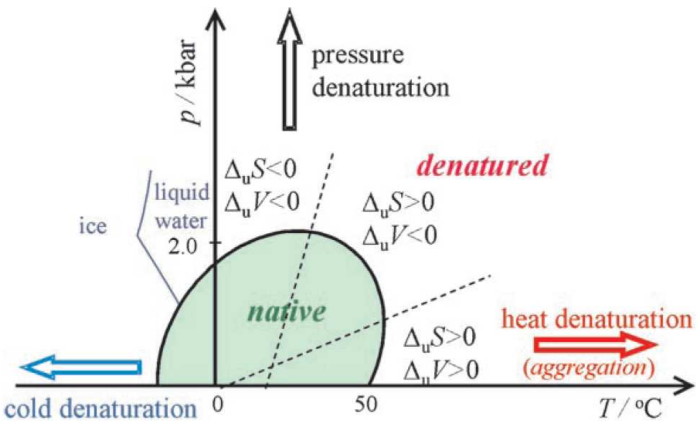


Figure 1.1: Schematic pressure/temperature stability diagram of a typical monomeric protein. The different routes of unfolding of the native protein (green area) as well as corresponding thermodynamic properties are depicted. Heat denaturation is often accompanied by irreversible aggregation. The water/ice phase transition line is also given.[9]

Solvent plays a crucial role in determining the state of a polymer. In polymer physics, it is well known that the compactness of a polymer greatly depends on the type of liquid in which it is placed. If there is a strong attractive interaction between solvent molecules and the polymer chain, the solvent is called a "good solvent". Polymers swell or expand in a good solvent, in a way similar to the swelling caused by repulsion between various segments of the chain itself. On the other hand, if there is a much weaker attractive interaction between solvent molecules and the polymer than between segments of the polymer molecules themselves, the solvent is called a "poor solvent". In a poor solvent, the polymer chain will be forced to adopt a more compact conformation. This is equivalent to an attractive interaction between the segments of the chain.[10,11] Thus, various interactions between constituent amino acid residues of the protein, and with the surrounding solvent, determines the preferable structure of protein in a given thermodynamic state. These interactions include electrostatic and van der Waal interactions between various atoms, formation of hydrogen bonds within the polypeptide chain or with solvent molecules, and hydrophobic interactions. The term "hydrophobic interaction" refers to the total interaction between two or more interacting non-polar solutes in an aqueous solution. The net attractive interaction between hydrophobic particles was initially explained using the "iceberg" model. It was believed that water molecules near hydrophobic particles are "ordered", i.e., in order to avoid the enthalpic penalty of losing hydrogen bonds, water molecules arrange themselves around a hydrophobe in a relatively ordered fashion, causing a loss of entropy, which is to some extent compensated by enhanced hydrogen bonding in the hydration layer, resulting in the formation of semicrystalline hydration shell. Thus, if the two hydrophobes comes together, the "structured" water in the region between them is returned to the bulk, leading to an entropy increase.[12,13] This view has been discredited as many evidences have shown that water does not undergo any structural enhancement near hydrophobic solutes. The favorable enthalpy of hydrophobic hydration can be attributed to van der Waals interactions between water and an apolar solute. The water molecules simply orient themselves near the apolar solutes in such a way that the O–H bonds are tangential to the solute surface. The increase in either the number or the size of hydrophobic particles causes the loss of hydrogen bonds between water molecules, an energetically unfavorable event, and results in aggregation of apolar particles (see References 14 and 15 for detailed reviews).

One of the main ideas in the theory of protein folding is the energy landscape theory.

The protein energy landscape is considered to resemble a rugged funnel, i.e. riddled with traps in which the protein can transiently reside. Native contacts and local conformation energies are considered more stabilizing than non-native contacts; hence there is an overall slope of the energy landscape towards the native structure. In the early stages of protein folding, the funnel guides the protein through many different sequences of traps towards the low-energy folded (native) structure. There are no pathways, but a multiplicity of folding routes, and discrete pathways only appear later in the folding process, when most of the protein has already achieved a correct configuration.[16]

1.3 Protein aggregation

In 1854, Rudolph Virchow introduced the term "amyloid" to denote a macroscopic tissue abnormality that exhibited a positive iodine staining reaction. However, it was later demonstrated that amyloids consist mainly of proteins rather than carbohydrates. The variety of pathological conditions were related to the inability of certain proteins to fold correctly or to remain correctly folded, resulting in their aggregation.[3,17] The insolubility of amyloid fibrils has hindered their experimental evaluation for a long time. The X-ray diffraction examination of amyloid fibrils has led to the general conclusion that fibrils are made up of a cross-β structure in which the polypeptide chains are organized as β-sheets arranged parallel to the fibril axis, having their constituents β-strands perpendicular to the fibril axis. The characteristic cross-β diffraction pattern and the fibrillar morphology are accepted as a diagnostic hallmark of amyloids, suggesting that the fibrils formed from different protein precursors share a degree of similarity.[18]

The process of amyloid fibril formation follows a nucleated growth mechanism, as depicted in Figure 1.3. In a typical nucleated growth mechanism, a threshold in protein concentration, called "critical concentration", is required for aggregation to be initiated. The critical concentration depends on several factors, such as protein structure, presence of co-solvents, pH, and presence of interface. If the protein concentration slightly exceeds the critical concentration, there is a lag time before polymerization occurs. During this lag time, the addition of preformed fibrillar species or "seeds" causes a shortening of the lag phase. The length of the lag time drastically depends on the concentration of protein solution. A protein solution above its critical concentration could remain kinetically

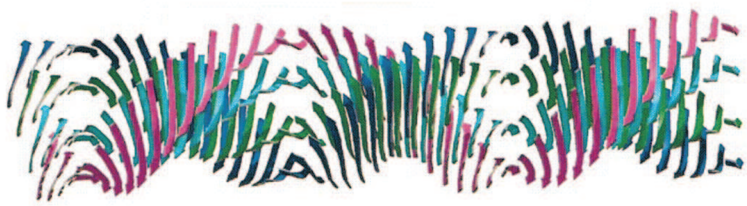


Figure 1.2: Model of the generic amyloid fibril structure depicting four β sheets running parallel to the fibril axis with component β strands perpendicular to the fibril axis.[18]

soluble for a considerable amount of time depending on the level of supersaturation.[19] Once the critical nucleus is formed, the monomer concentration decreases exponentially. After aggregation is complete and the equilibrium is established, the solution contains predominantly monomers and large aggregates. Nucleation dependent protein aggregation is also a characteristic of many processes such as protein crystallization, microtubule assembly, and sickle cell hemoglobin formation.[21]

Experimental studies have shown that the temperature-concentration phase diagram for a variety of globular proteins follows the pattern depicted in Figure 1.4.[22] This phase diagram can be obtained by using a very short range interaction between protein molecules in comparison to their diameter. If the range of attraction is long, the phase diagram resembles the elementary phase diagram of atoms and molecules where van der Waals interactions predominate. As illustrated in Figure 1.4, the binodal and spinodal lines lie completely within the region of fluid-solid phase separation and the critical point is also below the coexistence curve between the low-density fluid and the solid. At high temperatures, the fluid to solid phase transition is not interrupted by the glass transition. As the temperature is lowered below solubility limits, the solution becomes metastable; however, the fluid-fluid coexistence curve persists in the metastable regime and for particular temperatures and concentrations, the protein solution separates into a liquid, dilute in protein, and a liquid, rich in protein.[20,22] It can also be speculated that the phase diagram of amyloid forming proteins that do not have any native structure should be similar to the above mentioned phase diagram, with the exception that the solid phase

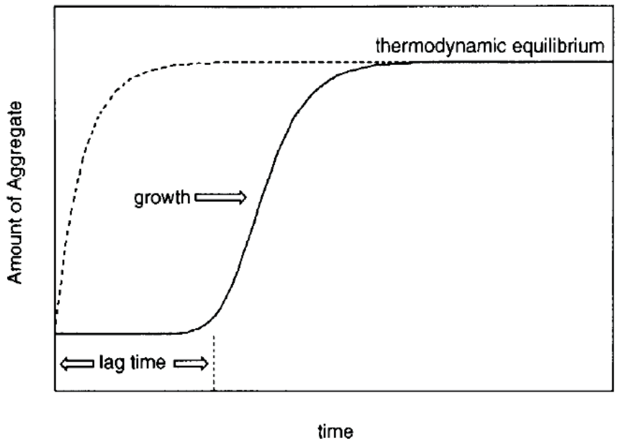


Figure 1.3: Experimentally observable formation of an aggregate for a nucleation dependent process, above its critical concentration. Aggregate formation at low and high concentrations is indicated by a solid and dashed line, respectively. Addition of seed also results in rapid aggregation.[19]

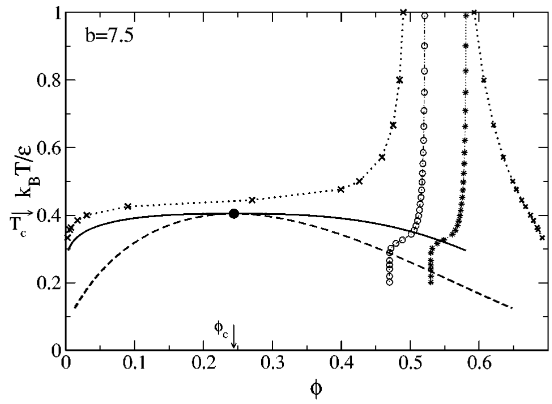


Figure 1.4: Phase diagram for a Yukawa fluid with screening parameter b=7.5. The crosses represent the fluid-solid phase transition, the continuous line is the binodal, and the dashed one is the spinodal. The filled circle is the critical point. The glass line (open circles) as evaluated for mode coupling theory and the glass line shifted to obtain the experimental packing fraction $\phi = 0.58$ (stars).[20]

would not be a crystalline phase, but a solid fibril. In this case, the morphology of the resulting fibrils could depend on the anisotropy of the protein structure and the kinetics of the system.

Aggregation and precipitation of proteins also occur upon heating (see Figure 1.1). The aggregation behavior of proteins in liquid water is similar to the demixing phase transition of binary mixtures, in general, and to aqueous solutions of organic molecules, in particular. Many aqueous solutions of simple organic molecules (pyridines, tetrahydrofuran, etc.) show demixing upon heating[23,24]. Aqueous solutions of large polymeric macromolecules (poly-N-isopropylacrylamide, polyoxyethylene, etc.) also show demixing upon heating, accompanied by drastic changes in polymer conformation[25–27]. The occurrence of a first-order demixing phase transition in aqueous solutions of proteins is therefore not surprising, being a common phenomenon in binary systems. Upon demixing, aqueous solutions of organic molecules separate into a water-rich phase and an organic-rich phase. When the thermodynamic conditions are close to ambient conditions, the water-rich phase is a liquid phase. The state of the organic-rich phase mainly depends on the phase state of the corresponding organic substance at ambient conditions. It may be a vapor phase (for example, in water-methane mixture), a liquid phase (for example, in aqueous solutions of pyridines) or an amorphous (solid-like) phase (aqueous solutions of macromolecules). Dry biomolecular substances are typically in an amorphous state at ambient conditions. Therefore, upon demixing, their aqueous solutions separate into a liquid water-rich phase having the critical concentration of biomolecules, and a solid-like organic-rich phase, which could be amorphous or ordered (crystalline or fibrillar).

1.4 Islet amyloid polypeptide

Human islet amyloid polypeptide (hIAPP), also known as amylin, is a pancreatic hormone secreted by islet β-cells. It is a 37 amino acid residue polypeptide, having a disulfide bond between the two cysteine residues at position 2 and 7 and an amidated C-terminus, which is important for its hormonal functions. As hormone, it modulates carbohydrate metabolism in muscle and liver by stimulating glycogen breakdown and acting as a non-competitive antagonist of insulin; additionally, IAPP is involved in processes associated with feeding and maintenance of bone cells, renal proximal tubular cells, and islet β-cells.[28] Fibrillar deposits containing hIAPP have been found in over 95% of type 2 diabetic subjects postmortem and their abundance correlates with the severity of the disease.[29] Studies indicate that islet amyloid is usually juxtaposed with β cell membranes. Sequence comparison between amylin from different species suggested that amylin (25-29) could serve as a core recognition motif and lack of proline residues at one or more of positions 25, 28, and 29 attributes to fibril formation in hIAPP.[28] A sequence comparison between human and rat amylin is depicted in Figure 3.1. The amino acid sequences of all known amylin molecules are highly conserved at their N- and C-termini, but vary among themselves to a greater extent in the central segment. In contrast to human amylin, rodent amylin does not form fibrils. It was further demonstrated that penta- and hexapeptide sequences, hIAPP(23-27) and hIAPP(22-27), are sufficient for formation of β-fibrils.[30] However, studies have also shown that proline substitutions outside the core domain diminishes fibril formation and single residue substitutions in rat IAPP from corresponding positions of human IAPP render rat IAPP competent for fibril formation, pointing towards the contribution of other regions in fibril formation.[31,32] Several other fragments of hIAPP have also been shown to form fibrils.[29,33,34] Amyloid formation by short peptides derived from region 10-19 of hIAPP are strongly influenced by pH. His-18 is considered to play a crucial role in modulating the pH dependent effect on aggregation.[35]

There are several reports indicating that hIAPP is able to interact with membranes,[36–40] and it seems that the interaction results in the formation of fibrillar amyloid deposits in the extracellular matrix of the β-cells, leading to membrane damage. Details of the nature of this interaction are still not fully understood. Some studies have suggested that the binding of hIAPP to the cell membrane is followed by loss of lipids from the membrane,[41] whereas other reports propose insertion of hIAPP into the membranes[42]. It has also been observed that phospholipids are able to catalyze hIAPP amyloid formation, leaving the lipid bilayer intact.[43] Also, it is still unclear which of the hIAPP species, monomers, oligomers, protofibrils, or mature fibrils are mainly involved in these membrane interactions. Several studies have indicated that hIAPP oligomers, and not the fibrils, may be involved in the interaction with membranes.[43–45] According to Porat et al.[39], oligomers are believed to be intermediate species in the formation of hIAPP amyloid fibrils. Another report pointed to the possible role of hIAPP monomers in membrane interactions.[46] It has also been suggested that the process of hIAPP amyloid formation and not the presence of a particular hIAPP species is related to its cytotoxicity.[42,43,46] These various results indicate that the interaction between hIAPP and lipid membranes is still far from being fully understood.

1.5 Computer simulations of protein aggregation

There have been several lattice studies on protein aggregation depicting the relationship between the stability of the native state and the propensity of aggregation suggesting distinct mechanisms for polymerization and its dependence on protein sequence .[47–49] MD simulation methods have also been used to predict the possible atomic scale amyloid organization for many peptide fragments. In these studies, several models of β-clusters were built and MD simulations were used to investigate the structure and stability of such models.[50,51] Simulations starting from random orientations of peptides have also been carried out to investigate the initial steps in aggregation, either using implicit solvent and discrete molecular dynamics[52] or explicit solvent molecular dynamics simulations.[53] Wu *et al.* have reported the formation and elongation of ordered peptide aggregates using MD simulations with explicit solvent.[54,55] The aggregation of small peptide fragments has also been studied using replica exchange molecular dynamics simulations with implicit solvent,[56,57] and also with explicit solvent.[58]

1.6 Thesis objectives

Protein aggregation and fibril formation seem to be generic in origin as they can be induced in proteins of diverse sequences and structures. The demixing of a protein aqueous solution is similar to the demixing observed in two component systems. Thus, protein aggregation should be studied as a function of concentration. In molecular simulation studies, the system size is an additional factor that can have a drastic effect on the simulation results. Even though studies of peptide aggregation using molecular simulation techniques are reported in the literature, the effect of concentration and system size on aggregation is generally neglected. Even the parameters needed to quantify aggregation in simulation studies have not yet been fully established.

Therefore, the first objective of this thesis is to establish parameters that can be used to quantify the aggregation process in simulation studies of peptide aggregation. Once the parameters have been established, systematic studies on the effect of concentration and system size on aggregation can be studied. These studies will also reveal the extent to which the finite size effect can distort simulation results meant to reproduce the properties of macroscopic systems. The finite size effects have been well understood for Ising magnets and Lennard-Jones fluids, but have never been studied for complex systems like aqueous solutions of peptides in full atomistic details. The second objective of this thesis is to study the effect of temperature on peptide aggregation. The change in aggregation parameters as a function of temperature would provide information about the transition temperature and the nature of the transition involved. As various experimental studies have shown that protein aggregation is generally affected by the presence of interfaces, the third and final aim of this thesis is to study the effect of model surfaces on the peptide aggregation process.

This thesis has been divided into five chapters: this chapter provides a general overview of protein aggregation; the second chapter gives a brief introduction of molecular simulation techniques and associated issues; the selection of appropriate parameters for the quantification of peptide aggregation, effect of concentration and system size on peptide aggregation in the bulk solution are described in the third chapter; the effect of temperature on peptide aggregation is described in chapter four; the last chapter describes the effect of surfaces on peptide aggregation by using two different types of amyloidogenic peptide fragments, in both hydrophobic and hydrophilic pores.

Chapter 2

Molecular Simulations

Molecular simulation is a generic term encompassing both Monte Carlo (MC) and Molecular Dynamics (MD) computing methods. These methods can be used to provide exact results for problems in statistical mechanics which would otherwise only be solved by approximate methods, or are intractable. Molecular simulations can provide a direct route from microscopic details of a system to macroscopic properties of experimental interest. It is also used to obtain information about processes that are difficult to probe experimentally.

The MD simulation technique follows the natural time evolution of the system, which allows the prediction of static and dynamic properties of the system directly from underlying interactions between molecules. However, sampling problems in MD are connected to the difficulty of spanning broad scales of time and this problem is hard to surmount because of the need to adhere to the governing dynamics. On the other hand, the MC method uses random number generators to sample the phase space and clever algorithms can be developed to optimize the efficiency of the sampling.

2.1 Theory

The microscopic state of a system can be specified in terms of the position and momentum of the constituent set of particles. If we make the approximation that a classical description is adequate, the Hamiltonian H of a system of N particles can be written as a sum of the kinetic and potential energy functions of the set of coordinates $\mathbf{q}_i$ and momenta $\mathbf{p}_i$ of each particle i:[59]

$$H_{(\mathbf{q},\mathbf{p})} = K_{(\mathbf{p})} + U_{(\mathbf{q})}, \tag{2.1}$$

$\mathbf{q}$ and $\mathbf{p}$ are generalized coordinates and conjugate momenta, respectively. To obtain thermodynamic averages over an ensemble, which is characterized by the macroscopic variables (N, V, T), where N is the number of particles, V is the volume and T is the temperature of the system, it is necessary to know the probability of finding the system at each and every point in the phase space. This probability distribution is given by the Boltzmann distribution function

$$\rho(\mathbf{q},\mathbf{p}) = \frac{exp\left[\frac{-H(\mathbf{q},\mathbf{p})}{k_B T}\right]}{Z}, \tag{2.2}$$

where the canonical partition function Z is the integral over the phase space:

$$Z = \int d\mathbf{q} \int d\mathbf{p}\ exp\left[\frac{-H(\mathbf{q},\mathbf{p})}{k_B T}\right]. \tag{2.3}$$

Once this distribution function is known, it can be used to calculate the phase space averages of any dynamic variable $A(\mathbf{q},\mathbf{p})$ of interest:

$$\langle A(\mathbf{q},\mathbf{p})\rangle = \int d\mathbf{q} \int d\mathbf{p}\ \rho(\mathbf{q},\mathbf{p}) A(\mathbf{q},\mathbf{p}). \tag{2.4}$$

However, in order to calculate these thermodynamic averages, it is necessary to simultaneously know the Boltzmann probability (Eq. 2.2) for each and every state $\{\mathbf{q},\mathbf{p}\}$, which is not feasible for many particle systems. An alternative strategy for calculating system wide averages is to follow the motion of a single point through the phase space instead of averaging over the whole phase space at once, i.e., the motion of a single point (a single molecular state) through the phase space is followed as a function of time, and the averages are calculated only over those points that were visited during the excursion. Averages calculated in this way are called "dynamic averages". The motion of a single point through the phase space is obtained by integrating the equation of motion of the system.[60] Dynamic averages of any dynamical variable $A(\mathbf{q},\mathbf{p})$ can now be calculated along this trajectory as follows:

$$\langle A(\mathbf{q},\mathbf{p})\rangle = \frac{1}{\tau}\int_0^\tau A\Big(\mathbf{q}(t),\mathbf{p}(t)\Big)dt, \tag{2.5}$$

where τ is the duration of the simulation. It is hoped that the point that is being dynamically followed will eventually cover all the phase space and that the dynamic average will converge to the desired thermodynamic average. A key concept that ties the two averaging strategies together is the *ergodic hypothesis*. This hypothesis states that, for an infinitely long trajectory, the thermodynamic ensemble average and the dynamic average become equivalent to each other.

2.2 Force fields

The core of any force field is the potential energy function used to describe the relationship of the structure R to the energy U of the system of interest. However, a potential energy function alone does not make a force field. It is the combination of the potential energy function with the parameters used in that function, as described below, that yield a force field;

$$\begin{aligned} U(R) = &\sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 \\ &+ \sum_{dihedral} K_\chi\left(1+\cos(n\chi-\delta)\right) + \sum_{impropers} K_{imp}(\vartheta-\vartheta_0)^2 \\ &+ \sum_{nonbonded}\left(\epsilon_{ij}\left[\left(\frac{Rmin_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{Rmin_{ij}}{r_{ij}}\right)^{6}\right]\right) + \frac{q_i q_j}{\epsilon r_{ij}}. \end{aligned} \tag{2.6}$$

The equation 2.6 is an example of a potential energy function. It is comprised of a collection of simple functions to represent a minimal set of forces that can describe molecular structures. Bonds, angles, and out-of-plane distortions (improper dihedral angles) are treated harmonically, while dihedral or torsional rotations are described by a sinusoidal term. Interactions between atoms use a Lennard-Jones (LJ) 6-12 term to describe the atom-atom repulsion and dispersion interactions combined with electrostatics treated via a Coulombic term.[61]

In Eq.2.6, b is the bond length; θ is the valence angle; χ is the dihedral or torsional angle; ϑ is the improper angle; r_{ij} is the distance between atom i and j. Parameters are the terms that represent the actual force field. Included in them are the bond force constant and the equilibrium distance, K_b and b_0, respectively; the valence angle force constant and equilibrium angle, K_θ, and θ_0, respectively; the dihedral force constant, multiplicity and phase angle, K_χ, n, and δ, respectively; the improper force constant and equilibrium improper angle, K_{imp} and ϑ_0, respectively. Collectively, these represent the internal or intra-molecular parameters. Non-bonded parameters between atoms i and j include the partial atomic charges, q_i, q_j and the LJ well-depth, ϵ_{ij}, and minimum interaction radius, $Rmin_{ij}$, used to treat the van der Waals interactions. These terms are also referred to as interaction or external parameters. Typically, ϵ and $Rmin$ are obtained for individual atom types either by fitting to an experimental data or from quantum chemical calculations, which are then combined to yield ϵ_{ij} and $Rmin_{ij}$ for the interacting atoms via combining rules.[62]

2.3 Periodic boundary conditions

The correct treatment of boundaries and boundary effects is crucial to simulation methods. In order to simulate bulk phases it is essential to choose boundary conditions that mimic the presence of an infinite bulk surrounding the N particle model system. This is achieved by employing periodic boundary conditions. The volume containing the N particles is treated as the primitive cell of an infinite periodic lattice of identical cells. If a particle leaves the box during the simulation, then it is replaced by an image particle that enters from the opposite side. The number of particles within the central box thus remains constant. The intermolecular interactions which are of short range are truncated beyond a certain cutoff distance.

2.4 Long-range interactions

Intermolecular Lennard-Jones interactions are of short range and decay ($1/r^6$) rapidly with the distance r. Truncating these interactions to zero beyond a certain cutoff does not induce large errors. However, this is not the case for Coulombic interactions, which vary strongly at small distances and decay slowly at long distances. The Ewald summation is one of the methods by which long range interactions can be handled appropriately. The Ewald summation method splits the slowly converging sum over the Coulomb potential into two exponentially converging sums

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r}. \tag{2.7}$$

The splitting function f is chosen in a way that the first part $\frac{f(r)}{r}$ should be negligible or zero, beyond a certain cutoff distance. The second part $\frac{1-f(r)}{r}$ should be a slowly varying function of r, such that its Fourier transform can be represented by only a few $\mathbf{k}$ vectors. This permits an efficient calculation of this contribution in the reciprocal space. The function f is a complementary error function

$$erfc(r) = \frac{2}{\sqrt{\pi}} \int_r^{\infty} \exp{(-t)^2} dt. \tag{2.8}$$

The potential energy of the primary box is now given by:

$$\begin{aligned}
V &= V^r + V^k + V^s + V^d && (2.9)\\
V^r &= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{n=0}{}' \frac{q_i q_j}{4\pi\epsilon_0} \frac{erfc(\alpha|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|} && (2.10)\\
V^k &= \frac{1}{2} \sum_{k\neq 0} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\epsilon_0} \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2}\right) cos(\mathbf{k}\cdot\mathbf{r}_{ij}) && (2.11)\\
V^s &= -\frac{\alpha}{\sqrt{\pi}} \sum_{k=1}^{N} \frac{q_k^2}{4\pi\epsilon_0} && (2.12)\\
V^d &= \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i}{4\pi\epsilon_0} \mathbf{r_i} \right|^2, && (2.13)
\end{aligned}$$

where V^r is the contribution from real space, V^k the contribution from the reciprocal space, V^s the self energy and V^d the dipole correction. The term α is often referred to as Ewald parameter and it tunes the relative weight of the real space and the reciprocal space contribution. The real space summation involving the error function converges very rapidly, and beyond a certain cutoff distance its value can be considered negligible. The sum of Gaussian functions in real space includes the interaction of each Gaussian with itself, thus it needs to be subtracted. In the reciprocal space summation, the vectors $\mathbf{k}$ are reciprocal vectors and are given by $\mathbf{k} = 2\pi\mathbf{n}/L$. If the surrounding medium has an infinite relative permittivity, then no correction term is required. However, if the surrounding medium is vacuum, then V^d must be added.[63]

2.5 Finite size effects

Molecular simulations deal with a much smaller number of degrees of freedom ($N \leq 10^6$) than a typical experiment ($N \approx 10^{23}$), an aspect that can lead to artifacts in simulations. Finite system size effects are particularly large near phase transitions and percolation thresholds. In an infinite system, a first-order transition is characterized by a delta function singularity in the first derivatives of the thermodynamic potential, whereas for finite systems, delta function singularities are rounded off.

The work of formation ΔG of a cluster of size n, according to classical nucleation theory, is given by:

$$\Delta G = -n\Delta\mu - \gamma A, \tag{2.14}$$

where $\Delta\mu$ is the difference in the chemical potentials of the vapor and the liquid phase, γ is the surface tension and A is the interfacial area. In a canonical ensemble with a finite number of particles far below the thermodynamic limit, the formation of a new phase causes a rapid consumption of the metastable phase. It can cause discrepancies between the results obtained from simulations and those of experiments. MacDowell *et al.*[64] have shown that for a system having a volume V and the total density ρ, the free energy cost ΔA_{hom} required to increase the density above the coexistence vapor density ρ_g^c, and the free energy cost ΔA_{inh} if the excess vapor condenses to form a droplet, can be given by:

$$\Delta A_{hom} = \mu_c(\rho - \rho_g^c)V + \frac{1}{2}\frac{V}{\kappa_g^c}\left(\frac{\rho - \rho_g^c}{\rho_g^c}\right)^2 \tag{2.15}$$

$$\Delta A_{inh} = \mu_c(\rho - \rho_g^c)V + k_g\gamma\left(\frac{\rho - \rho_g^c}{\rho_l^c - \rho_g^c}\right)^{2/3} V^{2/3}, \tag{2.16}$$

where μ_c and κ_g^c are the chemical potential and the vapor compressibility at coexistence,

k_g is the geometric constant that allows the surface to be expressed in terms of $V^{2/3}$, and ρ_c^l is the coexistence liquid density. For small deviations from ρ_g^c, the equilibrium state of the system is a supersaturated homogeneous vapor as $\Delta A_{hom} < \Delta A_{inh}$. It must be emphasized that the system is in equilibrium and not in a metastable state. For larger deviations from ρ_g^c , ΔA_{hom} becomes greater than ΔA_{inh} and condensation takes place. This is also a system size effect and it vanishes in the thermodynamic limit. It has also been shown that the density at which condensation takes place, ρ^t, scales as $\rho^t - \rho_g^c \propto V^{-1/4}$; hence, the transition moves towards coexistence with increasing the system size.

2.6 Methods

2.6.1 Molecular Dynamics simulations

Molecular Dynamics simulations solve Newton's equation of motion of molecules to generate new configurations. For a system of N interacting particles:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \; i = 1 \cdots N, \tag{2.17}$$

where m_i is the mass of the particle i, $\mathbf{F}_i$ is the force and $\mathbf{r}_i$ is the position vector of the particle i. The forces are the negative derivative of given potential functions $U(\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)$:

$$\mathbf{F}_i = -\frac{\partial U}{\partial \mathbf{r}_i}. \tag{2.18}$$

Solving Newton's equation of motion requires a numerical procedure for integrating the differential equation. Finite difference techniques are mainly used for this purpose. These techniques are used to generate molecular dynamics trajectories with continuous potential models, which is assumed to be pairwise additive. The integration is broken down into many small stages, each separated in time by a fixed time δt. The total force on each particle in the configuration at a time t is calculated as the vector sum of its interactions with other particles. From the force, accelerations of particles are determined, which are then combined with the positions and velocities at a time t to calculate the positions and velocities at a time $t + \delta t$. The force is assumed to be constant during the time step, which is generally in the range of 0.5–2.0 femtoseconds. The forces on the particles in their new positions are then determined, leading to new positions and velocities at time $t+2\delta t$, and so on. The *leap frog* modification of the *Verlet* algorithm is probably the most widely used method for integrating the equations of motion. The velocities $\mathbf{v}(t+\frac{1}{2}\delta t)$ are first calculated from the velocities at time $t-\frac{1}{2}\delta t$ and acceleration at time t, as follows:

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t-\frac{1}{2}\delta t) + \delta t\mathbf{a}(t), \tag{2.19}$$

the positions $\mathbf{r}(t+\delta t)$ are then deduced from velocities just calculated together with the positions at time $\mathbf{r}(t)$ using Equation 2.20. Thus, the velocities *leapfrog* over the position (hence the name)[59]

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t+\frac{1}{2}\delta t). \tag{2.20}$$

2.6.2 Monte Carlo simulations

The configurational part of the partition function can also be denoted by Z:

$$Z \equiv \int d\mathbf{r}^N \exp[-\beta U(\mathbf{r}^N)], \tag{2.21}$$

where $\beta = 1/k_B T$ and $\mathbf{r}^N$ stands for coordinates of all N particles. If random points are generated in the configurational space according to the probability density $\varrho(\mathbf{r}^N)$, where

$$\varrho(\mathbf{r}^N) \approx \frac{\exp[-\beta U(\mathbf{r}^N)]}{Z}\ , \tag{2.22}$$

then the average property of a variable A is:

$$\langle A \rangle \approx \frac{1}{L}\sum_{i=1}^{L} n_i A(\mathbf{r}^N), \tag{2.23}$$

where the number of points n_i generated per unit volume around a point $\mathbf{r}^N$ is equal to $L\varrho(\mathbf{r}^N)$ and L is the total number of points generated. The main idea of importance sampling is to generate points in configurational space with a relative probability proportional to the Boltzmann factor. If o and n are two configurations of the system, then the transition probability to go from o to n, $\pi(o \rightarrow n)$ is:

$$\pi(o \rightarrow n) = \alpha(o \rightarrow n) \times acc(o \rightarrow n), \tag{2.24}$$

where $\alpha(o \rightarrow n)$ is the probability to perform a trial move from o to n and $acc(o \rightarrow n)$ is the probability of accepting such a trial move. In equilibrium, the average number of accepted trial moves that results in the system leaving the state o must be equal to the number of accepted trial moves from all other states n to the state o. Generally, the detailed balance condition is applied:

$$\varrho(o)\pi(o \rightarrow n) = \varrho(n)\pi(n \rightarrow o), \tag{2.25}$$

where the average number of accepted moves from o to n is exactly canceled by the number of reverse moves. In the Metropolis scheme, the transition probability of going from o to n is given by:

$$\begin{aligned} \pi(o \rightarrow n) &= \alpha(o \rightarrow n) & \varrho(n) \geq \varrho(o) \quad (2.26) \\ &= \alpha(o \rightarrow n)[\varrho(n)/\varrho(o)] & \varrho(n) < \varrho(o). \end{aligned}$$

If α is chosen to be symmetric, then the equation 2.25 can be rewritten as:

$$\begin{aligned} \varrho(o)acc(o \rightarrow n) &= \varrho(n)acc(n \rightarrow o) \\ \frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} &= \frac{\varrho(n)}{\varrho(o)} \\ &= \exp\{-\beta[U(n) - U(o)]\}. \end{aligned} \tag{2.27}$$

The metropolis scheme is implemented by attempting a trial displacement and calculating the change in energy $\Delta U = [U(n) - U(o)]$. If $\Delta U < 0$ then the move is accepted. Otherwise a random number X is generated and the move is accepted if $\exp(-\beta\Delta U) \geq X$.

2.6.3 Replica Exchange simulations

In order to carry out meaningful simulation studies of biomolecular systems containing thousands of atoms, a sampling of a huge configurational space is required, and therefore such studies are extremely time consuming. This problem is more acute at ambient and low temperatures, as the system could get trapped in local energy minima and might not be able to escape from it within the duration of the simulation run. Similar problems exist in other fields of physics and a parallel tempering technique (or the replica exchange algorithm) was proposed.[65,66] This method was applied to the simulation of biomolecules in 1997,[67] and extended to molecular dynamics simulations.[68]

Replica Exchange Molecular Dynamics (REMD) is an enhanced sampling technique based on the parallel tempering Monte Carlo method where multiple copies (or replicas) of identical systems are simulated in parallel at different temperatures.[68] Periodically state-exchange move are attempted, where two neighboring replicas exchange their thermodynamic states (their temperatures). The acceptance rule for each state-exchange move between two neighboring states i and j is chosen to be

$$P(acc) = min\left\{1, \exp(1/k_BT_i - 1/k_BT_j)(U(r_i^N) - U(r_j^N))\right\}. \tag{2.28}$$

The state-exchange acceptance probability has been shown to obey the detailed balance condition for an extended ensemble of canonical states.[69]

Chapter 3

Simulations of Peptide Aggregation in Water

Molecular simulations can be an efficient tool to characterize aggregation and conformational changes of biomolecules in water at a molecular level. However, any simulation study is unavoidably affected by the finite size of the simulated system. Therefore, the evolution of system properties towards those encountered in the macroscopic limit should be properly understood and taken into account. The possible occurrence of phase transitions in the system studied must also be taken into account by choosing an appropriate ensemble. Nowadays, simulation studies of aggregation of biomolecules can be performed in simple ensembles, such as those with constant volume or constant pressure. Typically, these simulations are performed under conditions of strong oversaturation, i.e., deeply inside the two-phase region, where the system undergoes demixing. When a finite size system is in the two-phase region, the behavior of the minor phase noticeably differs from that encountered in the corresponding macroscopic system.[64,70–74] For example, in a macroscopic one-component fluid, the oversaturated vapor separates into two coexisting phases, the liquid and the vapor phase, such as a liquid droplet surrounded by saturated vapor. However, when the fluid volume is finite, a new stable state of the system emerges at the same density: an evaporated (at least partially) droplet surrounded by oversaturated vapor. The state with the droplet and the state without the droplet are both stable and replace each other with time. The occurrence probability of these states is determined by the level of oversaturation and by the system size. An increase of the system size eventually stabilizes the droplet state, which is the only stable state in the macroscopic limit. A state without droplet is just an artificial state, whose properties have no relation to those expected in the macroscopic limit. A similar behavior may be expected for an aqueous solution of peptides in small volumes at concentrations exceeding solubility limits.

To study the effect of the finite system size on the aggregation of peptides, we chose aqueous solutions of amyloidogenic peptides, which are highly insoluble in water. Aggregation and subsequent amyloid formation is a central phenomenon in a number of diseases, such as Alzheimer's, Parkinson's and type II Diabetes Mellitus, and seems to be a key factor in the development of the symptoms of these diseases.[75] Upon formation of amyloid fibrils, protein molecules adopt ordered, stacked cross-β-sheet structures.[21,75–77] It has been discovered that certain short sequence fragments contained within a respective protein can form amyloids even in isolated forms. Many fragments of Aβ and hIAPP,[30,33,78–80] fragments consisting of residues 10-20 and 105-115 of the protein transthyretin,[81] residues 7-13 of the yeast protein, Sup35, are few such examples.[82] Most of these amyloidogenic peptides have also been shown to be cytotoxic. This makes them particularly useful for studies of amyloid formation and elongation. However, the fibrils formed by short peptide fragments may or may not have the same fibril morphology and the same cytotoxicity as those of the parent protein. Nevertheless, studies of such short fragments may lead to understanding of fundamental processes taking place in amyloid formation in large systems. Short peptide sequences also provide a major technical advantage in MD simulations, as the size of the system containing small fragments will be much smaller than if the whole protein were to be simulated. Aqueous solutions of the amyloidogenic fragment of hIAPP corresponding to residues 15 to 19 were chosen for the present studies. It is one of the shortest peptide fragments capable of self-assembly. Experimental studies have reported that the aqueous solution of the fragment forms broad ribbon-like fibrils, with a short lag time.[33] All the aspects mentioned above make this fragment an ideal candidate for studying protein aggregation by using molecular simulations.

Human IAPP: KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY
Rat IAPP : KCNTATCATQRLANFLVRSSNNLGPVLPPTNVGSNTY

Figure 3.1: Sequences of human and rat IAPP. Conserved mutations are colored in blue and nonconserved mutations in red. The sequence of the simulated fragment is depicted by the yellow box.

3.1 System setup

All atomic MD simulations were performed using Amber99s (C. Simmerlings modifications of ϕ and ψ dihedral parameters of the original Amber99).[83] The N- and C-termini of the peptide fragments were capped with acetyl and N-methylamine groups, respectively. All starting configurations were solvated with the TIP3P water model and the simulations were carried out in an isothermic-isobaric ensemble and under periodic boundary conditions, at 330 K and 1 atm pressure, using GROMACS software package.[84] The Particle Mesh Ewald (PME) method was used to treat long range electrostatic interactions.[85] SHAKE was applied to constrain all the bonds to their equilibrium bond lengths allowing a time step of 2.0 fs. Initial equilibration was carried out using Berendsen's thermostat for 1 ns, followed by the production run of various lengths using Nose-Hoover temperature coupling and Parrinello-Rahman pressure coupling. The number N_p of peptides in the simulation box varied from 1 to 56 and the number N_w of water molecules varied from 591 to 29702. A total number of 12 peptide-water systems were studied (see 3.1). The peptide weight concentration C is effectively equal to zero in the case of a single peptide in a simulation box, representing an infinite dilution, and varied from ~ 2.5 to 42 % in systems with $N_p > 1$. Accordingly, the lateral size L of the cubic simulation box was in the range of 2.7 to 9.9 nm.

Initial configurations of these systems were prepared by random insertion of peptides in a cubic box such that the shortest peptide-peptide distance exceeds a certain minimal value, which depends on concentration. 1 to 10 different initial configurations were used for different systems. Equilibration periods, estimated from the time evolution of various system parameters, varied from 10 to 25 ns, depending on the system size and peptide concentration. Configurations were saved every 2 to 5 ps and equilibration periods were excluded from the analysis.

N_p	$N(sim)$	box(Å)	N_w	$C_w\%$	$Time(ns)$
1	10	26.9	591	5.8	300
3	5	37.4	1590	6.4	250
3	5	29.7	750	12.7	250
6	1	65.0	8600	2.5	100
6	3	48.8	3536	5.8	300
6	3	37.2	1480	12.9	600
6	3	29.2	626	25.9	300
12	3	59.7	6448	6.4	450
12	6	47.5	3079	12.4	750
12	3	36.7	1232	26.2	150
12	1	30.8	586	42.7	160
56	1	99.3	29702	6.4	145

Table 3.1: Table summarizing the simulations carried out depicting the following columns: number of peptides in the simulation box N_p, number of simulations $N(sim)$, averaged box size box, average number of water molecules in the simulations N_w, concentration of protein in weight percent $C_w\%$, total duration of simulation runs in nanoseconds $Time(ns)$.

For the analysis of peptide clustering based on the number of hydrophobic contacts as a criterion for the connectivity between two peptides, a hydrophobic contact was defined as follows: one hydrophobic contact between two peptides exists, when the distance between two sidechain carbon atoms involved (sidechains of F, L and V) does not exceed the sum of their van der Waals radii plus 2.8 Å. The radius of gyration R_g of the largest peptide cluster, as well as its maximal extension L_{max}, measured as the maximal distance between two heavy atoms of peptides in the cluster were also calculated. The secondary structures were determined using the corresponding distributions of ϕ and ψ dihedral angles in the Ramachandran plot. A residue was considered to contribute to α-helices, when -110° $\leq \phi \leq -30°$ and -90° $\leq \psi \leq -10°$; to β-sheets, when -180° $\leq \phi \leq -100°$ and 60° $\leq \psi \leq 180°$; to polyproline II structures, when -100° $\leq \phi \leq -30°$ and 60° $\leq \psi \leq 180°$. The solvent accessible surface area (SASA) of all peptides was obtained with a probe radius of 1.4 Å. Water molecules were considered as belonging to the hydration shell of the peptides, if the shortest distance between a water oxygen atom and at least one of the heavy atoms of the peptides, did not exceed 4.5 Å. The volume of the surface layer of water can be approximated to be $V_h = SASA \cdot D$. The width D of the water layer is considered to be 0.3 nm. Thus, the density of the hydration shell ρ_h is calculated as $\rho_h = N_h \cdot m_{wat}/V_h$. The peptide-peptide and peptide-water hydrogen bonds (H-bonds) were identified using the following criteria: a H-bond exists if the distance between donor and acceptor did not exceed 0.35 nm and if the donor-hydrogen-acceptor angle exceeds 120°.

3.2 Characterization of aggregation

The snapshots of the largest simulated system ($N_p = 56$) are shown in Figure 3.2. Within the first 0.05 ns, the randomly oriented peptides in the initial configuration form well distinguished clusters. These clusters grow and merge with time (see snapshots at $t = 1$ ns and $t = 2.5$ ns). Two large droplets can be seen at $t = 10$ ns, which later merge to form a single droplet at $t = 25$ ns. The simulation was carried out at a concentration manyfold higher than the expected critical concentration. It is expected that all the peptides should belong to the peptide aggregate. However, the system exhibits two stable states, which interchange with time: state A, where all the peptides are in one large cluster (lower-middle snapshot in Figure 3.2) and state B, where few peptides break up from the large cluster (lower-right snapshot in Figure 3.2). The behavior of the system can be considered as analogous to that of finite Lennard-Jones fluids at a constant density in the two phase region or finite Ising magnet at constant magnetization, where only the state A is relevant in the macroscopic systems and state B is an artifact of the finite size of the simulation box and is not seen in macroscopic systems. As the system size increases, the effect of state B on the system properties diminishes and vice versa. The snapshots of the simulation at the same peptide concentration, $C_w \approx 6\%$, but with only 6 peptides, are shown in Figure 3.4. Visual comparison with Figure 3.2 shows that the degree of peptide aggregation is lower in small systems, as a relatively large fraction of the total aggregate repeatedly breaks up. However, the systematic and quantitative study of the effects of concentration and system size on peptide aggregation can only be performed by introducing parameters reflecting the degree of clustering in simulations.

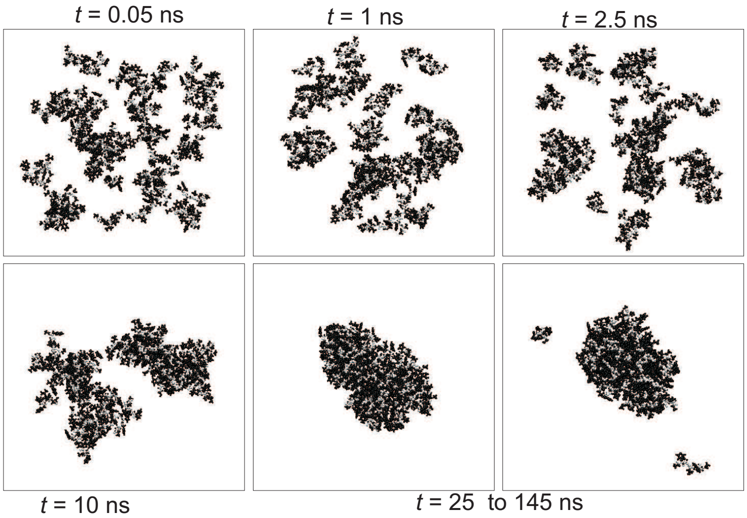


Figure 3.2: Time evolution of peptide clustering in the system with $N_p = 56$ and $C \approx 6\%$. After the nucleation process ($t < 25$ ns), the system exhibits two stable states (lower-middle and lower-right snapshots): all peptides are in one cluster (lower-middle snapshot); a few small clusters are separated from the large peptide cluster (lower-right snapshot). Backbones and side chains are shown in white and black, respectively.

3.2.1 Radius of gyration

The radius of gyration (R_g) of all peptides can be used to characterize aggregation; however, in MD simulations of several peptide fragments with periodic boundary conditions, the radius of gyration is not well defined (see Figure 3.3). The following approach was used to calculate the radius of gyration: one of the peptides was selected as a reference and the box was translated in such a way that the center of mass of the reference peptide coincides with the center of the box. The radius of gyration of all the peptides based on the i^{th} reference peptide can be calculated as:

$$Rg_i = \sqrt{\frac{\sum_{j=1}^{N} m_j \left(\mathbf{Ra}_i - \mathbf{r}_j\right)^2}{\sum_{j=1}^{N} m_j}}, \tag{3.1}$$

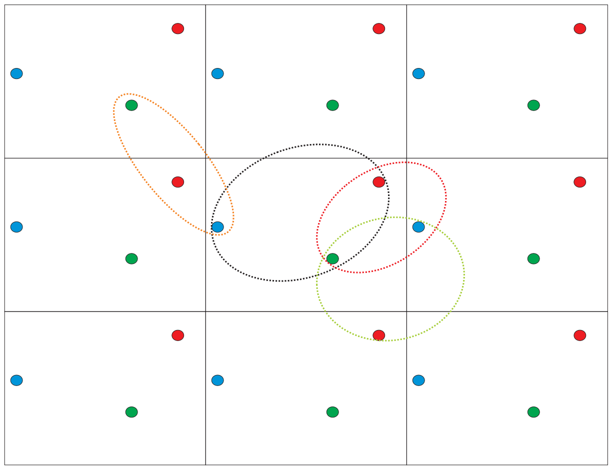


Figure 3.3: Figure depicting three particles in a box surrounded by its images. Depending on the selection of the images of particles, as depicted by colored ellipses, distinct radii of gyration can be obtained.

where N is the total number of atoms, m_j is the mass of atom j, $\mathbf{Ra}_i$ is the center of mass of all peptides based on the i^{th} reference peptide, calculated form the following equation:

$$\mathbf{Ra}_i = \frac{1}{J}\sum_{j=1}^{J}\mathbf{R}_j. \tag{3.2}$$

J is the total number of peptide fragments and $\mathbf{R}_j$ is the center of mass of a peptide fragment j. Note that the center of mass of reference peptide i is also the center of the box. Rg_i is calculated using each peptide fragment as reference and the minimum is reported: $R_g = min(Rg_i)$, where i is iterated sequentially over all the peptides.

Figure 3.4 depicts the time evolution of the radius of gyration for the two systems at the same concentration, calculated by using the above mentioned method for one of the simulations. Two states of the system are clearly visible for the small system, i.e., the "droplet" state where the total R_g of the peptides is low and all the peptides are together forming one large cluster, and the "dissolved" state where the peptides split into smaller clusters with a high total R_g of peptides. The system switches between these two states throughout the course of the simulation, whereas for the large system R_g varies in a relatively narrow range. The probability distribution $P(R_g)$ of R_g for the large system

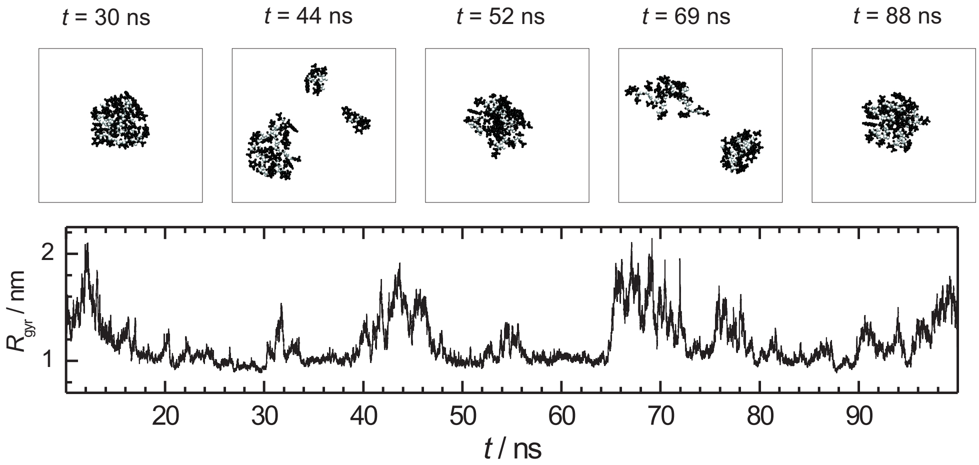


Figure 3.4: Time evolution of peptide clustering in the system with $N_p = 6$ and $C \approx 6\%$. Upper panel: Snapshots of peptides at various time steps, as indicated in the figure. Backbones and side chains are shown in light grey and black, respectively. Lower panel: Time evolution of the radius of gyration R_g of all peptides.

is highly symmetric and can be well fitted to a Gaussian function (dashed area in the lower right panel of Figure 3.5). In contrast, $P(R_g)$ for the smaller system is asymmetric (upper right panel of Figure 3.5), with the narrow peak of ~ 1 nm corresponding to a state where all the peptides form a compact aggregate, whereas the wide tail of $P(R_g)$ extending towards large R_g values corresponds to a state without a compact aggregate. To estimate the existence probabilities of these states, the distribution should be decomposed into two constituents. For the large system where the peptide aggregate dominates and the distribution is symmetric, the area under the Gaussian can be taken as the existence probability R_1 of a state with peptide aggregate. In the smaller system with 6 peptides, R_1 can be approximated by fitting the Gaussian function to the first maxima of R_g.

The characterization of the degree of peptide aggregation by use of the distribution $P(R_g)$ does not require any criterion for the connectivity between peptides. This is the main advantage of the proposed method, as the choice of an adequate connectivity criterion is not trivial. On the other hand, this approach suffers from a rather ambiguous extraction of the Gaussian contribution to $P(R_g)$. In particular, it is not applicable in very small systems (with $N_p = 3$ in our studies), where this contribution to the $P(R_g)$

distribution is indistinguishable.

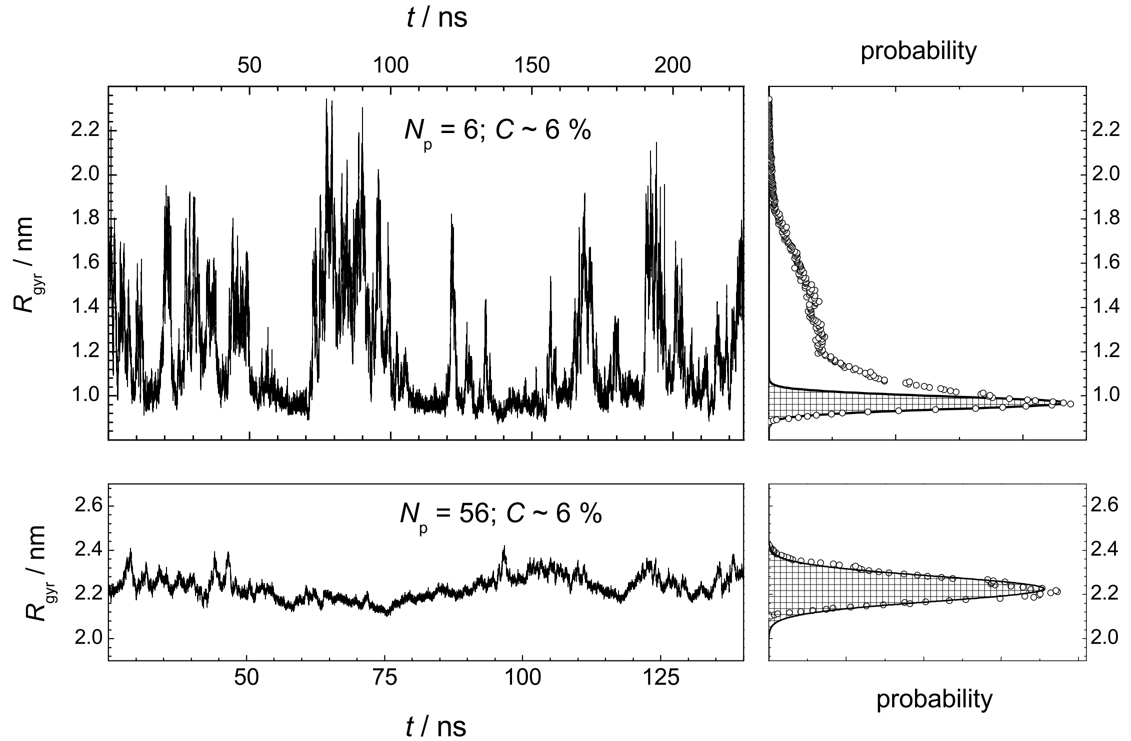


Figure 3.5: Left panels: Time evolution of the radius of gyration R_g of all peptides for two systems with the same peptide concentration, but a different number of peptides. For the system with $N_p = 6$, three time intervals of 75 ns, each from different simulation runs, are placed successively. Right panels: Probability distribution of R_g for the dependences shown in the left panels (solid points). Fit of the symmetrical part of the probability distribution to a Gaussian is shown by the dashed area.

3.2.2 Cluster analysis: search for connectivity criteria

Cluster analysis is the most basic approach for characterizing the arrangement of particles in a system. It is widely used in various fields of statistical physics. A system of particles can be described as an ensemble of clusters. Typically, the definition of clusters is based on a certain criterion for the connectivity between two particles. This criterion can be based on interparticle distance, potential and relative kinetic energies of two particles, the existence of H-bonds, etc. The particular choice of a connectivity criterion depends on the problem considered. In studies of condensation phenomena, two particles are usually considered as belonging to the same cluster, if the distance between them does not exceed a certain critical value r_{crit}, comparable to the distance between two particles in the condensed phase.

When considering aggregation of amyloidogenic peptides, the condensed organic-rich phase appears as fibrillar aggregates. Therefore, the connectivity criterion should be related to the structure of fibrils. There are two characteristic interpeptide distances in the case of fibrils: the interstrand distance between the neighboring peptides in the β-sheets is ~ 0.5 nm and the intersheet distance is ~ 1.0–1.1 nm. All peptides in the fibril belong to one cluster, if r_{crit} is close to the latter distance. Hence, in the cluster analysis of peptide aggregate formation in water, two peptides can be considered as belonging to the same cluster, when the distance between their centers of mass is $\lesssim 1.0$–1.1 nm. Additionally, we can use a qualitatively different connectivity criterion, based on the number n_c of hydrophobic contacts between two peptides. However, typical values of n_c in fibrils are not known. The choice of a particular value of r_{crit} or n_c should satisfy well-established aspects of peptide aggregation; in particular, the degree of peptide aggregation must increase with increasing peptide concentration.

The clustering analysis allows distinguishing the largest peptide cluster, which mimics an organic-rich phase of the studied system. Accordingly, various properties of the largest peptide cluster (shape, fractal dimension, density, etc.) can be studied. Additionally, the fraction f of the peptides in the largest cluster can be used as a parameter reflecting the degree of aggregation.

3.2.2.1 Distance between centers of mass

To choose the value r_{crit} for the distance between the center of mass of two peptides, which adequately characterizes peptide aggregation, we consider the largest system studied with $N_p = 56$. In this system, almost all peptides form one large aggregate permanently (see Figure 3.2), hence, the existence probability R of the aggregate state is equal to 1. The dependence of the existence probability R to find at least fN_p peptides in the largest cluster of this system for various choices of r_{crit} is shown in Figure 3.6. When $r_{crit} <$ 0.9 nm, the fraction f of the peptides in the largest cluster never exceeds ~ 0.2. This contradicts visual observation and should be considered as unreasonable. When $r_{crit} >$

1.2 nm, the fraction f is always close to 1. At $R = 1$, a drastic increase of the fraction f occurs, when r_{crit} changes from 0.9 to 1.1 nm, in agreement with literature values of intersheet distances in fibrils.[86] The presence of the majority of peptides $(0.5 < f < 1.0)$ in the largest cluster can be considered as a signature of the state with peptide aggregate. Even in the largest system studied, f is never equal to 1, since a small fraction of peptides periodically splits from the main aggregate (see Figure 3.2). Therefore, any value of f, which exceeds 0.5 and is not very close to 1, seems to be reasonable for the definition of the existence probability R of a state with peptide aggregate. In Ising magnets, a fraction of 2/3 of the excess magnetization forms a droplet at the transition point.[73,74] Therefore, we use $f = 2/3$ for the definition of R.

The dependence of the existence probability R defined in such a way on r_{crit} is shown in the upper and middle panels of Figure 3.7 for systems with $N_p = 6$ and $N_p = 12$, respectively, and with several peptide concentrations C. As expected, R increases, when

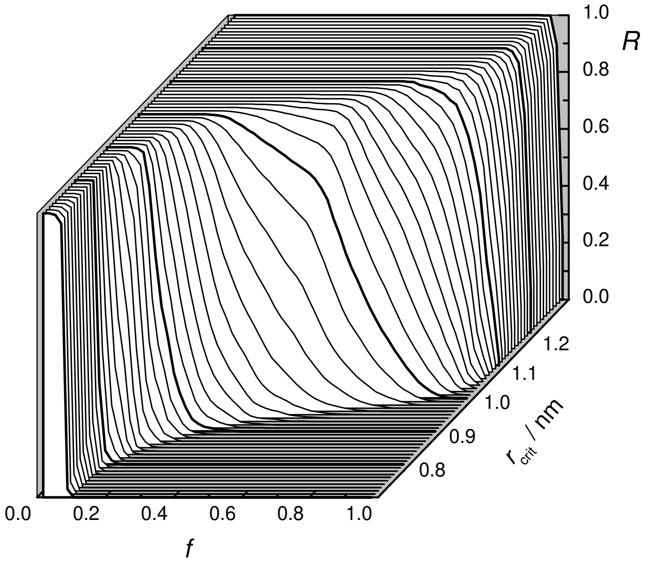


Figure 3.6: Probability R to find $> fN_p$ peptides in the largest peptide cluster at various choices of the distance r_{crit} between the center of mass of two peptides, which is used as a connectivity criterion

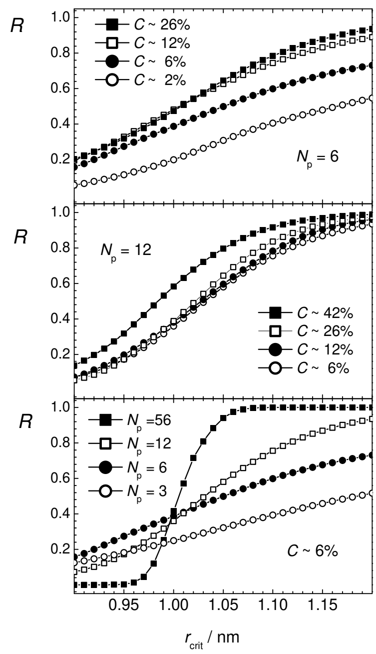


Figure 3.7: Dependence of the aggregation parameter R_2 (with $f = 2/3$) on the connectivity criterion r_{crit}, which is the distance between the center of mass of two peptides.

the connectivity criterion weakens, i.e., r_{crit} increases. A physically reasonable choice of the value r_{crit} should provide an increase of R with increasing peptide concentration at constant number of peptides. This condition is satisfied for $r_{crit} > 1.0$ nm, which can be considered as the lower limit for r_{crit}. The existence of such a lower limit for r_{crit} is not expected for systems of simple isotropic molecules and evidences a multilevel structural organization of peptide aggregates. This can be seen, when systems with approximately the same peptide concentration ($C \approx 6\%$) but with a different number of peptides are compared (lower panel in Figure 3.7). When the connectivity criterion applied is stricter ($r_{crit} < 1.0$ nm), the systems with smaller peptide numbers N_p exhibit increased aggregation in comparison with larger systems. This is due to the fact that in the systems with just a few peptides, the formation of a single β-sheet with interstrand distance of about 0.5 nm is the main form of aggregation. In contrast, in larger systems, more then one β-sheet can be formed, and the use of $r_{crit} < 1.0$ nm artificially breaks the peptide aggregate into separate β-sheets even in the case of an ideal fibril. For the largest system studied with $N_p = 56$, the aggregation parameter R shows a pronounced sigmoid-like dependence on r_{crit} with an inflection point at ~ 1.0 nm. Such sigmoid-like dependence, although less steep, is still seen in the system with $N_p = 12$, but it disappears in the systems $N_p = 6$ and 3 (see lower panel in Figure 3.7).

The choice of the connectivity criterion in studies of peptide aggregation should be meaningful also in the limit $N_p \to \infty$, since simulation studies are typically aimed to reproduce the properties of macroscopic systems. Therefore, a connectivity criterion of $r_{crit} > 1.0$ should be used, keeping in mind that their use overestimates peptide aggregation in small systems. In the largest system studied with $N_p = 56$, the state with peptide aggregate exists almost permanently with a probability of $R = 1$, when r_{crit} exceeds 1.1 nm (see lower panel in Figure 3.7). Below, we use the connectivity criterion $r_{crit} = 1.1$ nm for the distance between the centers of mass of two peptides and the minimal fraction $f = 2/3$ of peptides in the largest cluster, to estimate the existence probability of the state with aggregate which is denoted as the aggregation parameter R_2.

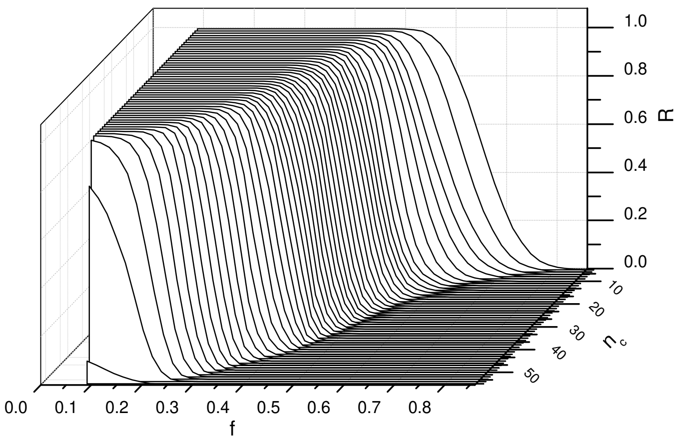
3.2.2.2 Number of hydrophobic contacts

The connectivity criterion based on the distance between the centers of mass does not take into account specific interpeptide interactions, such as interpeptide H-bonds or direct contact of the atomic groups of two peptides. For the description of the formation of ordered peptide aggregates, the connectivity criteria which deal with specific interpeptide interactions might be important. As a first step in this direction, we have also analyzed peptide aggregation using the number n_c of hydrophobic contacts between two peptides as connectivity criterion.

Thus, if we consider the largest system with $N_p = 56$, the dependence of the existence probability R to find at least fN_p peptides in the largest cluster of this system for various choices of number of contacts n_c is shown in Figure 3.8. The existence probability rapidly approaches 1 as n_c is lowered from 30 to 20 for the f between 0.4–0.7. The dependence of the existence probability R (with $f = 2/3$) on peptide concentration is shown in the upper and middle panels of Figure 3.9 for systems with $N_p = 6$ and $N_p = 12$, respectively. With decreasing the number of contacts n_c, the probability R approaches 1, as expected. The physically justified enhancement of aggregation with increasing the peptide concentration is observed, when $n_c \leq 10$. On the other hand, in the system with $N_p = 56$, the requirement $R = 1$ is satisfied for $n_c \leq 15$. We use $\geq$10 hydrophobic contacts between two peptides as a connectivity criterion and the corresponding existence probability of the state with aggregate is denoted as aggregation parameter R_3.

3.2.3 Peptide aggregation: concentration and system size

The dependences of the aggregation parameters R_1, R_2 and R_3 on peptide concentration C are shown in Figure 3.10. All three aggregation parameters used show a qualitatively similar behavior: the aggregation is fostered with increasing C, when the number of peptides is fixed. Although such a dependence was, in fact, imposed by the choice of the connectivity criteria for parameters R_2 and R_3, this is not the case for the aggregation parameter R_1, which characterizes the degree of aggregation without imposing any connectivity criterion. The enhancement of aggregation with increasing peptide concentration is physically obvious and we are not aware of any mechanism, which can decrease the aggregation propensity when the number of peptide is fixed but the amount of solvent decreases. This aspect of peptide aggregation did not get proper attention in simulation studies of aggregation of peptides or other particles in liquid water so far. The effect of concentration on peptide aggregation is illustrated by the dependence of the size S_{max} of the largest peptide cluster on C for the systems with $N_p = 6$ (Figure



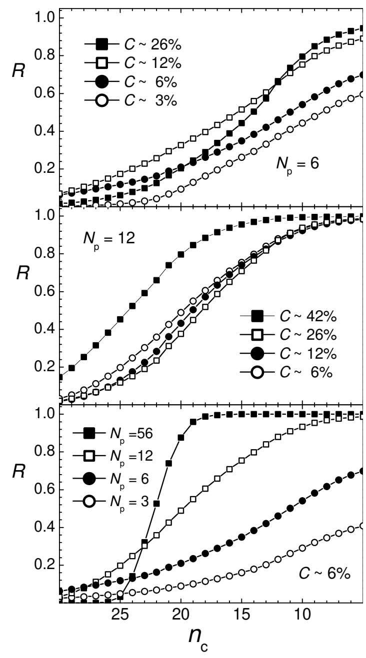


Figure 3.8: Probability R to find $> fN_p$ peptides in the largest peptide cluster at various choices of the number of hydrophobic contacts n_c between the two peptides, used as a connectivity criterion.

Figure 3.9: Dependence of the aggregation parameter R on the number of hydrophobic contacts between two peptides, used as a connectivity criterion.

3.12). Obviously, $S_{max} = 1$ upon infinite dilution ($C = 0\%$) and $S_{max} = 6$ in the absence of solvent ($C = 100\%$). The steepness of the dependence $S_{max}(C)$ is determined by the degree of solubility of the peptides considered. For more soluble peptides, we may expect a more gradual increase of S_{max} with C. The data shown in Figure 3.10 also demonstrate that the degree of aggregation increases strongly when the concentration is fixed but the number of peptides in the system increases. This is shown explicitly in Figure 3.11, where the aggregation parameters R_1, R_2 and R_3 are given as a function of the box size L. All three different aggregation parameters depend on the system size in a drastic way. In fact, the effect of system size on aggregation (Figure 3.11) is as strong as the effect of peptide concentration (Figure 3.10). Decreasing the system size has the same effect as decreasing the peptide concentration: the peptides become apparently more "dissolved" in water. The physical origin of this phenomenon is the effect of the finite system size on the minority phase in the two-phase region. The same behavior is seen in simple fluid and magnet

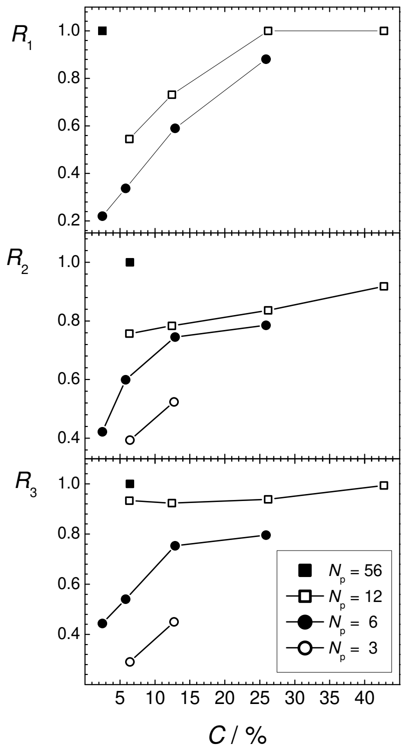


Figure 3.10: Dependence of the aggregation parameters R_1, R_2 and R_3 (upper, middle and lower panel, respectively) on the peptide concentration C at fixed numbers of peptides (indicated in the figure).

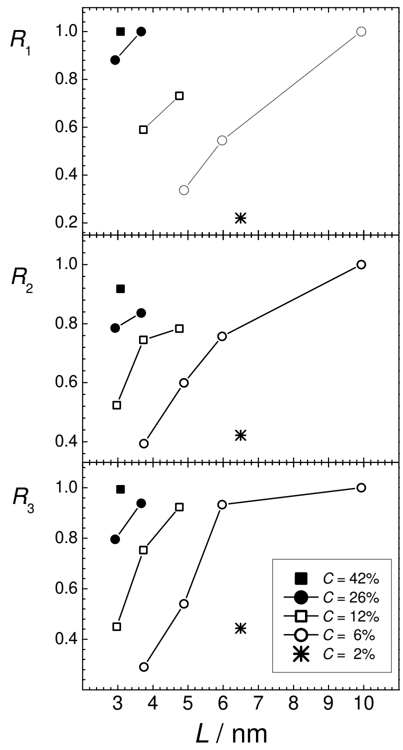


Figure 3.11: Dependence of the aggregation parameters R_1, R_2 and R_3 (upper, middle and lower panel, respectively) on the system size L at fixed peptide concentrations (indicated in the figure).

systems, when they are at constant density or magnetization, respectively.[64,70–74] Since the concentration and the system size strongly affect the degree of peptide aggregation as measured by different parameters, various other properties of the peptide-water system should also be strongly sensitive to these two factors. The dependence of the average number of interpeptide hydrogen bonds per peptide, n_H^{pp}, on the system size is shown in the lower panel of Figure 3.13 for various peptide concentrations. These dependences are qualitatively similar to those of the aggregation parameters shown in Figure 3.12, i.e., n_H^{pp} increases with increasing concentration C and with increasing system size L. The average number of peptide-water H-bonds, n_H^{pw}, is also strongly sensitive to C and L, although

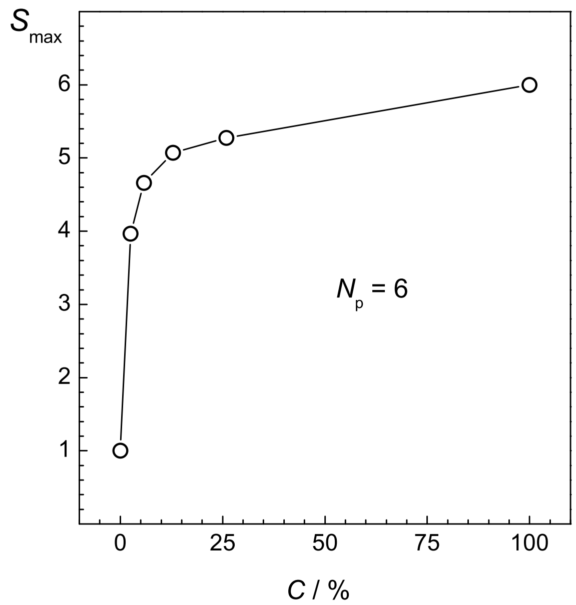


Figure 3.12: Dependence of the size S_{max} of the largest peptide cluster on the peptide concentration C at fixed peptide number N_p.

this dependence is opposite to that of n_H^{pp}. The clear correlation between the aggregation parameters and other properties of the peptide-water system evidences the reasonable choice of the aggregation parameters.

The secondary structure content of peptides is an important property, which is used in the characterization of peptide aggregation both in simulations and experiments. We have found that the β-sheet and α-helical contents are strongly sensitive to the system size and concentration, and vary from 0.28 to 0.42 and from 0.11 to 0.33, respectively. Nevertheless, there is a clear anti-correlation between the α-helical and β-sheet contents of the peptide-water system (see upper panel in Figure 3.14). In turn, the β-sheet content is proportional to the average number of interpeptide H-bonds per peptide n_H^{pp} (see lower panel in Figure 3.14).

Important information about the driving forces of peptide aggregation can be obtained from the analysis of the properties of hydration water. The water-mediated attraction between hydrophobic groups of peptides as well as Coulombic and H-bonding interactions between peptides are the main driving forces of peptide aggregation. An analysis of the effective hydrophobicity/hydrophilicity of the surface of peptide aggregates may give

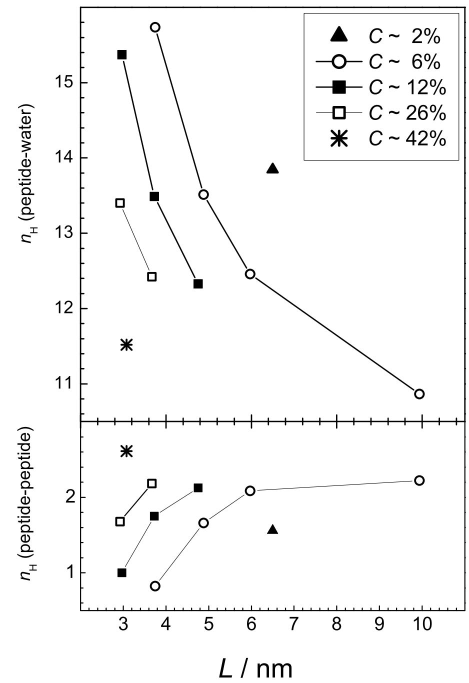


Figure 3.13: Dependence of the average number n_H^{pw} of peptide-water (upper panel) and the average number n_H^{pp} of peptide-peptide (lower panel) H-bonds on the system size L at several peptide concentrations.

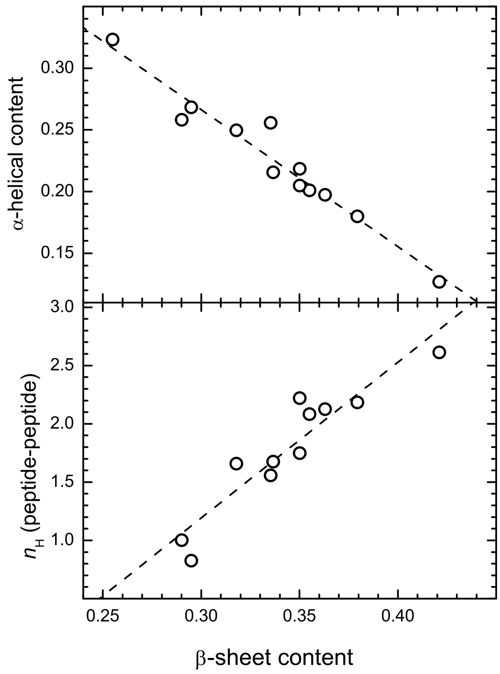


Figure 3.14: Correlation between the β-sheet content and the average number of peptide-peptide H-bonds n_H^{pp} (lower panel) and correlation between the β-sheet and α-helical contents (upper panel) in all systems studied.

insight into the mechanism of aggregation. If the peptide surface exposed to water becomes more hydrophilic upon aggregation, the attraction between hydrophobic groups may be considered as the main driving force in aggregation. Otherwise, aggregation should be attributed mainly to the interpeptide Coulombic and H-bonding interactions. The strength of water-peptide interaction can be characterized by the number of water-peptide H-bonds n_H^{pw} per unit area of solvent accessible surface. The dependence of this number on the system size L at peptide concentration $C \approx 6$ % is shown in the lower panel of Figure 3.15, where the value for a single peptide is shown by a horizontal line. With increasing system size (and, accordingly with increasing aggregation), the surface of peptides exposed to water becomes more hydrophilic. We can conclude that the hydrophilicity of the peptide surface exposed to water is enhanced upon aggregation and that the hydrophobic attraction between hydrophobic groups of the peptides studied is the main driving force of their aggregation. This could be attributed to the presence of hydrophobic amino acid residues and hydrophobic caps in the peptide.

The density ρ_h of water in the hydration shell of peptides can also be used as a measure of the effective strength of peptide-water interaction: more hydrophobic surfaces cause a decrease in the density of hydration water. The density ρ_h of hydration water can be simply estimated, if the solvent accessible peptide surface as well as the number of water molecules in the shell of a certain width near the surface are known. The dependence of the density ρ_h of hydration water on the system size L at a peptide concentration $C \approx 6\%$ is shown in the upper panel of Figure 3.15. The trend towards more dense hydration water upon aggregation is clearly seen, corroborating our conclusion about the leading role of hydrophobic attraction in the aggregation of the studied fragments, as derived from the analysis of water-peptide H-bonding (see lower panel in Figure 3.15).

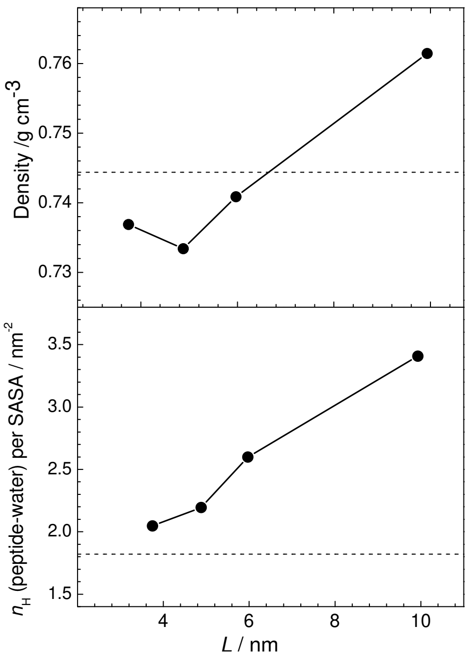


Figure 3.15: Dependence of the average number of peptide-water H-bonds, n_H^{pw}, normalized by the peptide SASA, on the system size L (lower panel) and dependence of the average density ρ_h of hydration water (upper panel) on the system size L at peptide concentration $C \approx 6$ %. The dashed lines represent corresponding values observed in the simulations of single peptides.

3.2.4 Properties of the largest peptide cluster

A clustering analysis provides the possibility to explore various properties of the largest peptide cluster, which can be considered as an embryo of the organic-rich fibrillar phase. The time evolution of the size S_{max} of the largest peptide cluster (lower panel in Figure 3.16) as well as the time evolution of its radius of gyration R_g (upper panel in Figure 3.16) can be used to characterize the equilibration process. Similar to the radius of gyration R_g of all peptides, the properties of the largest peptide cluster achieve saturation in about 25 ns. Note that the R_g of all peptides and of the peptides in the largest cluster are very close due to the strong degree of aggregation in the largest system studied (upper panel in Figure 3.16).

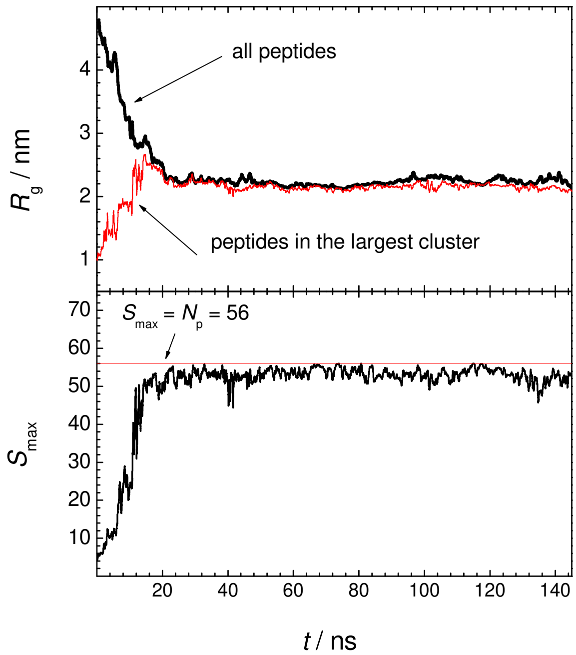


Figure 3.16: Upper panel: Time evolution of the radius of gyration, R_g, of all peptides (black line) and of the peptides in the largest cluster (red line). Lower panel: Time evolution of the size S_{max} of the largest peptide cluster.

The structural properties of the largest peptide cluster may be characterized by the relation between its average mass M and volume V. In a first approximation, V can be estimated from the radius of gyration R_g of the largest peptide cluster, assuming that it is a spherical body: $V = (4/3)\pi\ (R_g\sqrt{5/3})^3$. The obtained dependence $M(V)$ in the various systems studied is shown in Figure 3.17. If the peptide cluster is a three-dimensional object, its density is equal to the slope of the dependence $M(V)$, which is $\rho = 0.66$ g/cm^3. Since a spherical shape of the peptide cluster was imposed, the estimation given above is a lower limit for the aggregate density, as the sphere provides the maximal volume at a fixed radius of gyration. Information about the shape of the largest cluster can be obtained from the analysis of the maximal extension L_{max} of the largest cluster. The dependence of the average value of L_{max} on the average value of R_g in the various systems is shown in the upper panel of Figure 3.18.For comparison, the dependence expected for a spherical object, whose $L_{max} = 2R_g\sqrt{3/5}$, is also shown. The dependence $L_{max}(R_g)$ for the peptide cluster deviates from the one expected for spherical objects, indicating an elongated shape of the peptide clusters.

To gain further insight into the mass distribution within the largest cluster, we plotted the dependence of M on the radius of gyration, R_g, in a double logarithmic scale (lower panel in Figure 3.18). The slope of this dependence is equal to the fractal dimensionality of the object. The fit of the dependence $M(R_g)$, shown in the lower panel in Figure 3.18, to a power law yields the fractal dimension of the largest peptide cluster equal to 2.8 (solid line), which notably differs from 3 (dashed line), corresponding to compact three-dimensional objects. A fractal-like structure of the peptide aggregate is also supported by the dependence $L_{max}(R_g)$, which is not linear, but $L_{max} \sim R_g^{1.1}$, indicating that the fractal dimension of the peptide clusters is < 3. In fact, the fractal dimensionality of proteins is always < 3 and, therefore, a low fractal dimension of peptide clusters is not surprising.[87]

3.2.5 Time evolution of secondary structure

In order to evaluate the time evolution of the secondary structure of the peptides, each residue is assigned to a particular secondary structure using SEGNO,[88] and a peptide is assigned to a particular secondary structure, if three consecutive residues have the same secondary structure and no other consecutive secondary structure is present, else, the peptide is assigned to a coil. Two peptides are considered to be forming β-sheets

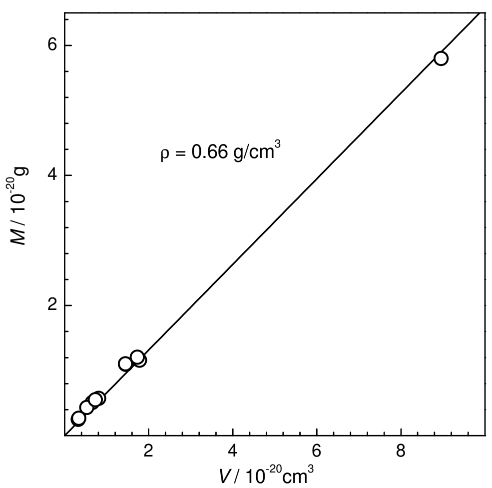


Figure 3.17: Dependence of the average mass M of the largest peptide cluster in various systems on its volume, estimated as the volume of a sphere with the corresponding radius of gyration (circles). The lower limit for the density of the largest peptide cluster is estimated from the slope of the linear fit (solid line).

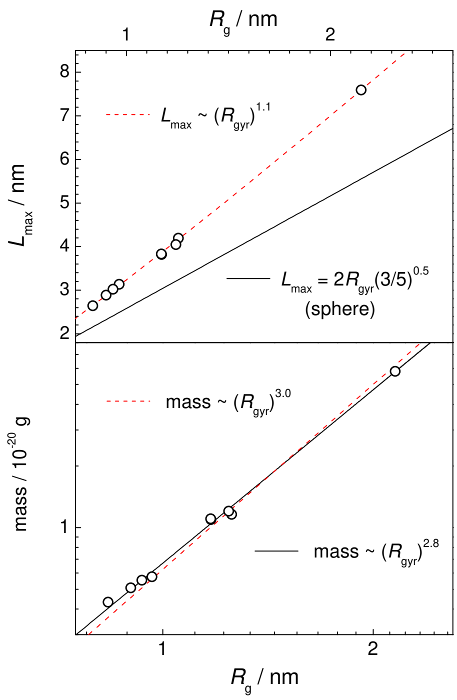


Figure 3.18: Dependence of the maximal extension L_{max} of the largest peptide cluster on its radius of gyration, R_g, in various systems. Lower panel: Dependence of the mass M of the largest peptide cluster on its radius of gyration, R_g, in the double logarithmic scale.

if both peptides are assigned to β-strands by SEGNO and at least two polar backbone hydrogen bonds are present between them. Figure 3.19 shows the time evolution of two representative simulations, Run1 and Run2, at $C = 12\%$. In the initial configuration of the system, all the peptides were in an isolated strand conformation. A rapid increase in random coil conformation, accompanied by a corresponding decrease in isolated strands, is observed within the first 5 ns in both simulations. In Run1, β-sheets increase steadily up to 100 ns, which correlates well with the decrease in random coils. There seems to be no other secondary structure surge during this transformation, indicating the direct transition from coil to β-sheets. Run2 has a very low fraction of β-sheets and most of the

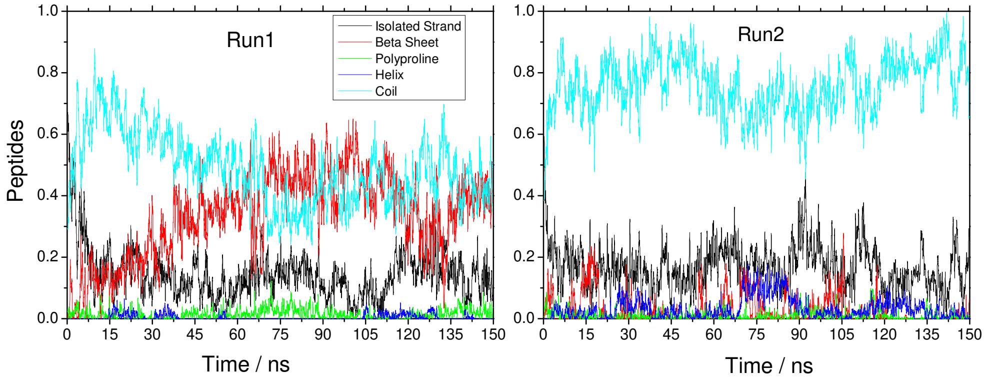


Figure 3.19: Plot of the fraction of peptides in a particular secondary structure for Run1 and Run2 as a function of time. Data was smoothed using Savitzky-Golay filtering with a fourth degree polynomial and 25 points on each side.

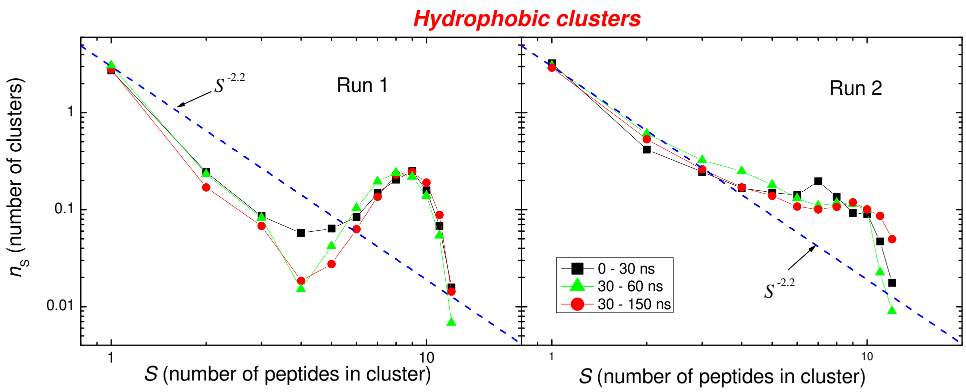


Figure 3.20: Log-log plot of n_S, the number of peptide molecules in a cluster of size S. Clusters were defined using a hydrophobic contact cutoff of ≥ 20 between two peptides.

peptides have no distinct secondary structure; however, the relative amount of peptides in the helix conformation is higher. The cluster size distribution for both simulations at various time intervals is shown in Figure 3.20. The power law $n_S \approx S^{-2.2}$ represents the behavior of n_S at the percolation threshold.[89] A hump of n_S at a large S and its drop-off at intermediate S values indicate the presence of a large aggregate in the case of Run1, which is missing in Run2. The increase in the amount of β-sheets with time is only seen in simulations where the droplet survives for multiple nanoseconds.

3.3 Discussion

The results of the simulation studies presented have two main implications. The first implication is related to the necessity to take into account the finite size of the simulated system, when one intends to reproduce properties of a real macroscopic system. The second implication is related to natural systems, where the finite system size is an intrinsic property.

3.3.1 Finite system size *in silico*: simulation studies of binary systems

Generally, the simulation studies are aimed at reproducing the properties of macroscopic systems. There are various factors, which complicate the realization of this goal, and it is important to know to what extent these factors make the properties of the simulated system different from those of its macroscopic analog. For example, the ability of available force fields to reproduce various system properties is approximate even for simple fluids and their mixtures. Another complication arises from the necessity to reproduce correctly the phase behavior of a system of interacting particles that requires application of sophisticated simulation methods (such as Monte Carlo simulations in the grand canonical ensemble). Besides, there is another general problem, which accompanies *a priori* any simulation studies aimed to reproduce the properties of macroscopic systems; this is the finite size of the simulated systems, which even in the distant future will not noticeably approach the size of a macroscopic system.

The effect of the finite system size on its properties is well known in statistical and computational physics. Finite size scaling allows approaching the properties of the macroscopic system. This can be achieved by simulations of several systems of different sizes (measured, for example, by their linear extensions L) with subsequent extrapolation of the results to the macroscopic limit $L \to \infty$. The effect of the finite size depends on the thermodynamic state of the system. For example, it is especially strong in the vicinity of the critical point due to the suppression of fluctuations by the finite system size. On the other hand, when the system is in a thermodynamic state that is distant from the phase transition, system properties are not strongly affected by its finite size.

Typically, simulation studies of aqueous solutions are performed in constant-volume or constant-pressure ensembles. As an aqueous solution is a mixture, the concentration of solutes is a key parameter of such a system, which is also kept constant. Simulations of a single solute in water can be used for reproducing the properties of a macroscopic system only at infinite dilution. In this case, the finite size does not affect the system properties noticeably when the box size essentially exceeds the size of the solute. The addition of just another solute molecule to the system changes the status of simulations in a drastic way. Now the clustering (aggregation) of solutes is possible and strongly depends on the concentration of the solutes and the thermodynamic state of the system.[89–91] If this state is distant from the demixing phase transition, the probability to find a cluster containing S molecules drops in a drastic way with increasing S and the vast majority of solute molecules exists as monomers or belongs to small clusters. Of course, the finite system with only a few solutes fails to reproduce the cluster size distribution of a macroscopic system. However, this affects mainly the large clusters whose population in solution is low, whereas the clustering of the majority of solute molecules is not strongly perturbed.

The distorting effect of the finite size on the system properties becomes enormous as the system enters the two-phase region.[64,70–74] This situation is typically encountered in the simulation studies of strongly aggregating solutes in liquid water, which are usually performed in one simulation box with the solute concentration deeply inside the two-phase region. However, the effect of a finite system size is generally neglected not only in simulations of biomolecules, but also in simulation studies of simple small hydrophobic solutes in liquid water (see, for example, Refs. 92–94). Moreover, the effect of concentration on the aggregation of solute molecules was considered in only few studies.[95,96] In large enough systems, the solution may separate into two coexisting phases with an explicit interface between them. However, this is not the case in the majority of simulation studies of solute aggregation in water, where the number of solutes is usually small. The finite system size suppresses the minority (organic-rich) phase and produces an artificial stable state of the system in which the organic-rich phase is dissolved. The existence probability of this state increases with decreasing the system size and with decreasing solute concentration (within the two-phase region). To make the simulation studies of such systems relevant to real systems, it is necessary to consider the type of phase transition, the character of the coexisting phases, and to perform a correct extrapolation of the simulation results to the macroscopic limit.

3.3.2 Finite system size *in silico*: simulation studies of peptide aggregation

In the case of amyloidogenic peptides in water, the organic-rich phase appears as a solid-like ordered aggregate. Typically, the organic-rich phase is a minority phase and its properties are strongly affected by the finite system size. When the peptide concentration exceeds the critical one, the peptides should form highly ordered fibrils in the macroscopic limit. However, the number of peptides is relatively small in simulations. Just a few peptides in a simulation box can only to some extent exhibit properties of the peptide aggregate expected in the macroscopic limit. Therefore, simulation studies of aqueous solutions of peptides must include the analysis of the effect of the system size on the degree of aggregation and on all other system properties, and the artificial system properties, produced solely by the finite size effect, should be excluded from the consideration. In a finite system of an aqueous solution of peptides whose concentration is within the two-phase region, there are two stable (equilibrium) states. One state is a state with the peptide aggregate, which represents an organic-rich phase. Of course, this phase, reproduced by only a few peptides, differs strongly from the ordered peptide aggregate seen in macroscopic systems. However, this state will evolve towards its analog in the macroscopic limit upon increasing the system size (the number of peptides in the simulation box at fixed concentration). In the other state, the organic-rich phase is dissolved, either partially or completely. This state is a pure artifact of the finite system size and has no analog in the macroscopic limit, as it disappears with increasing the system size. The time intervals of the molecular dynamics trajectory, where the system exists in such an artificial state, can be determined using, for example, the time evolution of the total radius of gyration of all peptides (see Figure 3.4). The exclusion of these time intervals from the analysis should help approaching the properties of the macroscopic system. Note that an equilibrium between the previously mentioned states was also observed in simulation studies of three amyloidogenic peptides in implicit water,[57] although the origin of such a behavior was not discussed.

The most important finding of our studies is the drastic effect of the system size on peptide aggregation. This effect seems to be responsible for the instability of aggregates consisting from just a few peptides, as seen in simulations.[97–99] This effect of the finite size of a system which is in the two-phase state, is known and well understood for Ising magnets and LJ fluids, has not been studied so far in more complex systems. To the best of our knowledge, this is the first study showing the manifestation of this effect in binary mixtures of complex peptide-water systems with low solubility of the peptide. This system was chosen due to the importance of simulation studies of the aggregation of biomolecules in water. Our simulation studies of peptide aggregation with various peptide concentrations clearly show that the concentration affects all system properties in a drastic way. This effect is well understood and seems to be obvious. The degree of peptide aggregation (Figure 3.10), H-bonding (Figure 3.13), secondary structure content and other system properties strongly depend on peptide concentration. The dependence of the size of the largest peptide cluster on concentration at a fixed number of peptide (Figure 3.11) clearly illustrates the necessity to account for the concentration effect in the simulation studies of aggregation phenomena. This effect is unavoidable, when the number of aggregating particles exceeds 1. In particular, any property of a system with just two aggregating particles are concentration dependent.

When studying the effect of concentration and system size on aggregation, it is important to introduce an adequate parameter characterizing the degree of aggregation. Although the use of the radius of gyration of all peptides for this purpose does not require a criterion for connectivity between peptides, it does not allow obtaining of the cluster size distribution and an analysis of the properties of the largest peptide cluster. A more detailed analysis of aggregation requires the introduction of a connectivity criterion between two peptides. Such a choice is not unambiguous, even for simple molecules, and it becomes even more difficult for biomolecules. In a first approximation, we have used the distance between the center of mass of two peptides as a measure of connectivity. The disadvantage of this measure is the inability to distinguish the ordered character of the peptide aggregate. Clearly, the search for more adequate connectivity criteria in studies of formation of ordered peptide aggregates is necessary. These criteria should be derived from the structure of the macroscopic ordered peptide aggregates (fibrils) and include the interatomic distances between two peptides and H-bonds in particular. It is to be noted that the effect of the finite system size on the ordered character of a peptide aggregate is so far unknown. We may assume that a more ordered aggregate will be more affected by this effect. However, this assumption should be tested in simulations.

3.3.3 Finite system size *in vivo*: aggregation of proteins in cells

The effect of the system size on peptide aggregation complicates the attempts to reproduce the properties of macroscopic systems by simulations of finite systems. However, this is an intrinsic property of the finite system. Therefore, if the real system of interest is not macroscopic and contains a relatively small number of peptides, their aggregation will be suppressed by the finite system size as well. This situation may be relevant in the case of peptides in small volumes, such as biological cells or their compartments. *In vivo*, amyloidogenic peptides can be found both in intracellular and extracellular fluids and these two pools of peptides often seem to be mutually related.[100,101] The extracellular fluid may be regarded as an essentially macroscopic system and formation of the peptide-rich phase via the phase transition should largely follow the regularities normally encountered in *in vitro* experiments with bulk aqueous solutions of the amyloidogenic peptides. The picomolar concentrations of islet amyloid polypeptide in the plasma[102] and of $A\beta$ in the cerebrospinal fluid[103,104] should be considered as critical for their aggregation, because the extracellular fluid is in direct contact with amyloid plaques. The production of amyloidogenic peptides in cells and their accumulation with time[100,105,106] evidence that the intracellular concentration noticeably exceeds the critical one. The absence (or at least the slowing down) of extensive fibrillation in cells might be due to the effect of the relatively small volume of cells or their compartments. Due to the high insolubility of amyloidogenic peptides, the presence of dozens or hundreds of peptides in a cell may already provide conditions of strong oversaturation. Hence, a small cell volume should suppress peptide aggregation in general, and formation of ordered aggregates, in particular. The escape of these peptides into the extracellular fluid (for example, due to the destruction of the cell membrane upon its death), makes fibrillation unavoidable at the same peptide concentration. If oversaturation is not too strong, the lag time of fibril formation may still take many years.

The suppression of aggregation by a small system size has a general physical origin and, therefore, is unavoidable. However, it is rather difficult to estimate the magnitude of this effect, which depends on the system and detailed solution conditions considered.

Additionally, other factors (first of all, surface effects) may be equally or even more important. We may expect that peptide adsorption on some surfaces is favorable for the ordered character of their aggregation, whereas confinement in a small volume should suppress the formation of ordered aggregates. Further studies are necessary to identify those cases where the finite size of biological cells noticeably affect intracellular peptide aggregation.

Chapter 4

Effect of Temperature on Peptide Aggregation

The phase state of a multicomponent solution at a given temperature is the result of the interplay of various interactions between the constituent particles. As the temperature is increased, the kinetic energy should become dominant over the potential energy, and at a certain temperature, the solution should become macroscopically homogenous.

Amyloidogenic peptides are highly insoluble in liquid water and their critical concentration is generally expected to be in the micromolar to nanomolar range.[21] Under ambient conditions, their aqueous solutions are separated into a water-rich phase, containing the critical peptide concentration, and an organic-rich phase, which appears as amyloid fibrils. It can be expected that amyloid fibrils dissolve in liquid water upon heating at a particular temperature and pressure, depending on the peptide concentration. In this chapter, the temperature-induced demixing phase transition between the aggregated and the disaggregated states of the peptide system was examined with the REMD simulation method, using an explicit water model.

4.1 System setup

The simulated system consists of 12 peptide fragments having the amino acid sequence FLVHS, corresponding to residues 15-19 of hIAPP, and 3079 water molecules, resulting in a peptide concentration of 12.4 w/w% . The N- and C-termini of the peptide fragments were capped with acetyl and methylamide groups, respectively. A modified AMBER force field[83] was used for the peptides along with the TIP3P water model.[107] The Replica exchange molecular dynamics (REMD) simulation was performed using the RPMDRUN package.[108] The temperature spacing was calculated by performing initial uncoupled simulations for 0.5 ns at 36 different temperatures ranging from 275 to 991 K. The average energies were fitted with a polynomial and Eq. 2.28 was solved iteratively for the temperature distribution for which $P(acc) = 0.20$. Sixty replicas distributed over a temperature range of 286.7 K to 645.5 K were used with a state exchange probability of 0.1 leading to a time of about 3 ps for each replica between two state exchanges. The SHAKE algorithm was applied to constrain all bonds in order to allow a time step of 2.0 fs. The electrostatic interactions were treated by Particle Mesh Ewald (PME) summation and the simulation was performed for 49 ns, corresponding to a total simulation time of 2.94 µs.

The intrinsic volume V_p of the peptides was calculated as:

$$V_p = L_{box}^3 - V_h - (N_0 - N_w)/\rho_b, \tag{4.1}$$

where V_h is the volume of the hydration layer, calculated using the slab approximation (see section 3.1 of Chapter 3 for details), N_0 and N_w are the numbers of water molecules in the cubic simulation box (with the edge L_{box}) and in the hydration shell, respectively, and ρ_b is the numerical density of bulk water determined at the same thermodynamic states in constant-volume simulations of TIP3P water. The radius of gyration, hydrophobic contacts, hydrogen bonds, and secondary structure were calculated using the methods described in section 3.1 of Chapter 3.

4.2 Results

In order to study the equilibrium thermodynamic states of the system, the part of the MD trajectory where system properties vary monotonically with time should be excluded. The time dependence of various system properties was analyzed by dividing the simulation run into five non-overlapping segments and calculating the average properties of each segment. Of all the measured properties, the inter-peptide hydrophobic contacts and hydrogen bonds needed the longest equilibration time ($\approx$ 25 ns). Other parameters, such as the SASA and the radius of gyration, equilibrated much faster. Figure 4.1 shows the average number of inter-peptide hydrophobic contacts and SASA of all peptides for each segment of the simulation run. Based on the time evolution of all parameters studied,

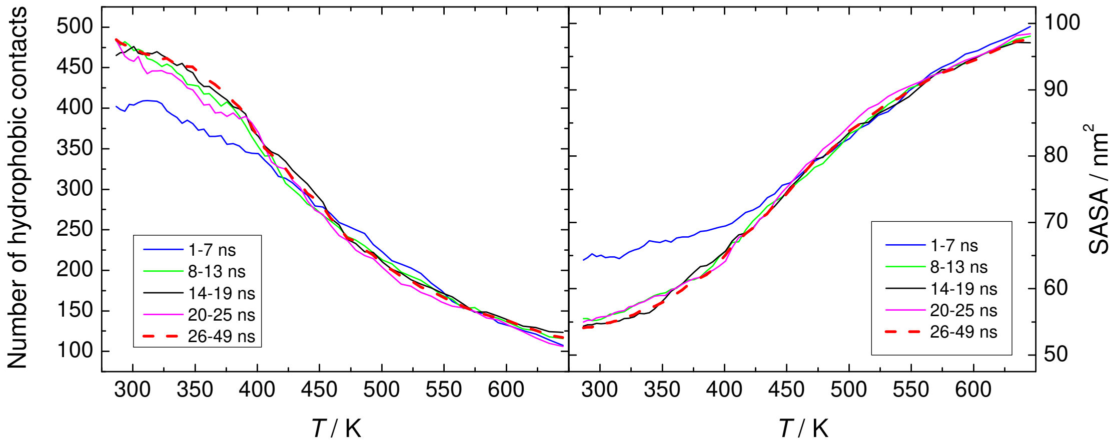


Figure 4.1: Average number of inter-peptide hydrophobic contacts and SASA as a function of temperature in four non-equilibrium segments of the total MD simulation run and in the equilibrium run between 26 to 49 ns.

the equilibration period was estimated to be 25 ns. Only the equilibrium results obtained from the last 24 ns of the REMD simulation run are further discussed.

The radius of gyration R_g of all peptides and its mean-square fluctuations ($\sigma = \langle R_g^2 \rangle$ - $\langle R_g \rangle$ ^{2}) were shown to be useful parameters for characterizing the phase transition in aqueous solution of peptides.[56,57] The R_g shows a sigmoid-like increase with increasing temperature, indicating a transition from the state with compact aggregate with $R_g \approx$ 1.25 nm to a disaggregated state with $R_g \approx 2$ nm (Figure 4.2). The fit of the $R_g(T)$ to a sigmoidal function (dashed line) indicates a midpoint of the transition at $T \approx 450$ K (vertical dashed line) and a smearing of the transition within a range $\pm$ 55 K. A transition temperature can be estimated using the mean-square fluctuations σ, which pass through a maximum at $T = 486$ K (vertical solid line). The discrepancy between the two estimations is $\sim$ 35 K, which could be due to notable deviations of $R_g(T)$ from a sigmoidal shape. The behavior of the radius of gyration as a function of the reversed temperature $1/T$ was also analyzed (Figure 4.3, solid squares). The $R_g(1/T)$ dependence can be almost perfectly described by a sigmoidal function, with a better goodness of fit as compared to that of the $R_g(T)$ dependence. The inflection point was found to be at $T \approx 465$ K (Figure 4.3, solid vertical line). Similarly, the dependence of SASA on the reversed temperature

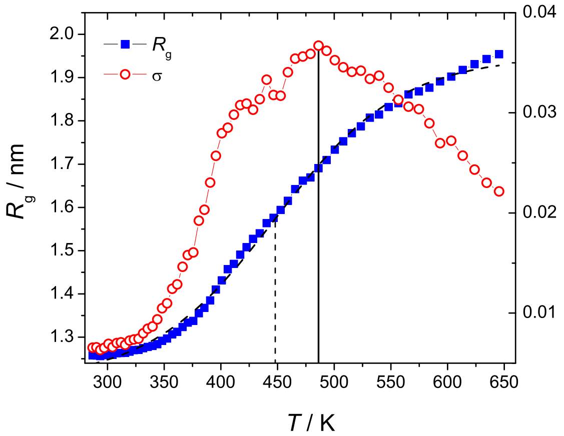


Figure 4.2: Temperature dependence of the radius of gyration R_g of all peptides (squares) and its mean-square fluctuations σ (circles). The temperature where σ passes through a maximum is indicated by a vertical solid line. The sigmoid fit of $R_g(T)$ and location of its inflection point are shown by dashed lines.

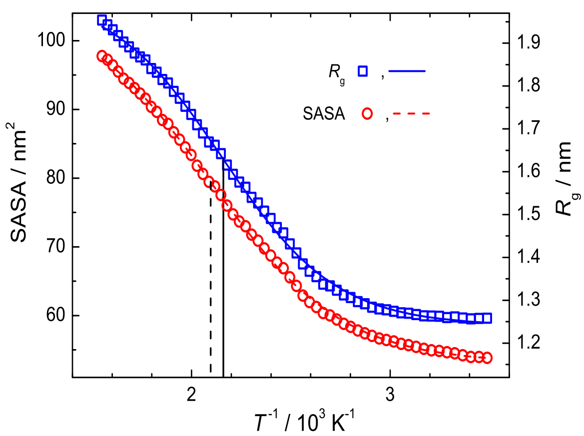


Figure 4.3: Radius of gyration R_g and SASA of all peptides as functions of reversed temperature $1/T$ (symbols). The fits to a sigmoid function and locations of the inflection points are shown by solid and dashed lines, respectively

$1/T$ is much closer to a sigmoid (Figure 4.3) and the goodness of fit is about three times better as compared to the SASA(T) dependence. The inflection of the sigmoidal fit of SASA($1/T$) is located at $T \cong 480$ K (Figure 4.3, vertical dashed line).

Peptide aggregation can be quantitatively characterized utilizing cluster analysis. The probability distributions $P(S_{max})$ to find the largest peptide cluster containing S_{max} peptides, obtained with a connectivity criterion based on the distance between the centers of mass of the peptides, are shown in Figure 4.4 for selected temperatures. At the lowest temperature studied ($T = 287$ K), the largest cluster consists of all 12 peptides in $\sim 50\%$ of the observed configurations and only rarely contains less than 5 peptides. Conversely, at the highest temperature studied ($T = 646$ K), largest clusters consisting of just 2 or 3 peptides dominate. At $T = 440$ K, the largest peptide cluster strongly fluctuates, consisting of 4 to 10 peptides with comparable probabilities. The temperature dependence of the probabilities to have an aggregate including more than 2/3 of the peptides, based on the connectivity criteria of either the distance between the centers of mass of two peptides ($r_{crit} \leq 1.1$ nm) or the number of hydrophobic contacts between the two peptides ($n_c \geq 10$), are shown in Figure 4.5. Fits to a sigmoidal function give inflection points at 415 and 435 K, respectively.

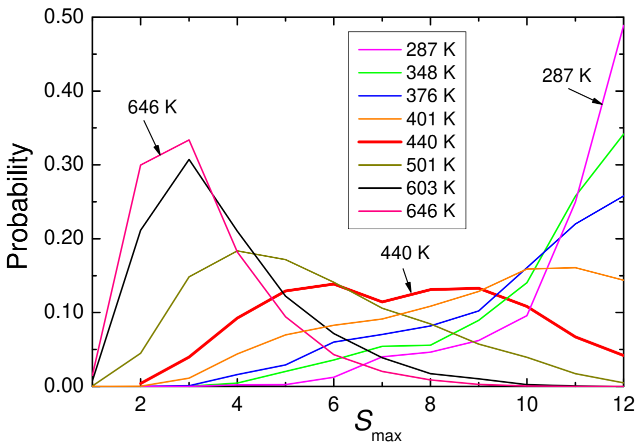


Figure 4.4: Probability that the largest cluster consists of S_{max} peptides, calculated at T = 287, 348, 376, 401, 440, 501, 603 and 646 K.

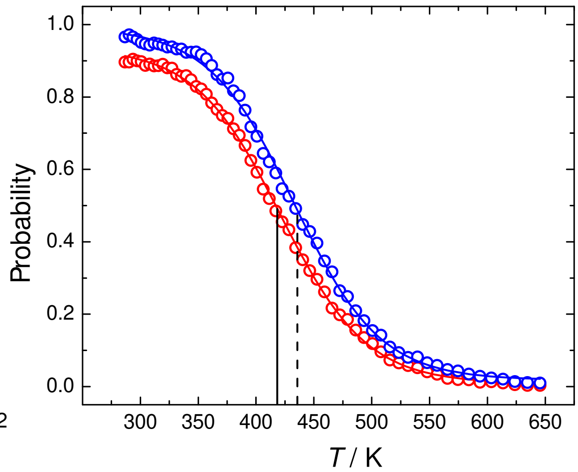


Figure 4.5: Temperature dependence of the probabilitiy that the largest peptide cluster includes $> 2/3$ of all molecules (more than 8 peptides). The criterion for the connectivity between two peptides was based on the distance between the centers of mass of the peptides (red circles) and on the number of hydrophobic contacts (blue circles). Fits to the sigmoidal function are shown by solid lines. The inflection points are indicated by vertical lines.

The cluster size distribution n_S, which is the probability to find a cluster containing S peptides, can also serve for characterizing aggregation. Examples of the n_S distributions are shown in Figure 4.6 using a double logarithmic scale. Such distributions are very sensitive to a phase transition. In the one-phase region, where the aqueous solution is homogeneous at the macroscopic scale, the distribution of n_S deviates downwards from the power law $n_S \sim S^{-2.2}$.[109] Such a behavior is seen from $T = 646$ K up to 566 K. Phase separation becomes evident with the appearance of a peptide droplet, and is indicated by a sharp increase of n_S at large S values, e.g., the distributions from $T = 287$ K to 340 K. Near the phase transition, n_S shows a power-law behavior $\sim S^{-2.2}$ in the widest range of

S. However, such a scenario is only true for dilute macroscopic systems. Due to the finite size of the simulation box, the probability distribution of large clusters gets distorted and a hump appears on the n_S distribution at large S values. This hump does not indicate the presence of a droplet, however, and can also be observed at temperatures much higher than that of the phase transition. These humps can be seen for temperatures $T < 557$ K, therefore, 557 K should be considered as an absolute upper temperature limit for the presence of an aggregate.

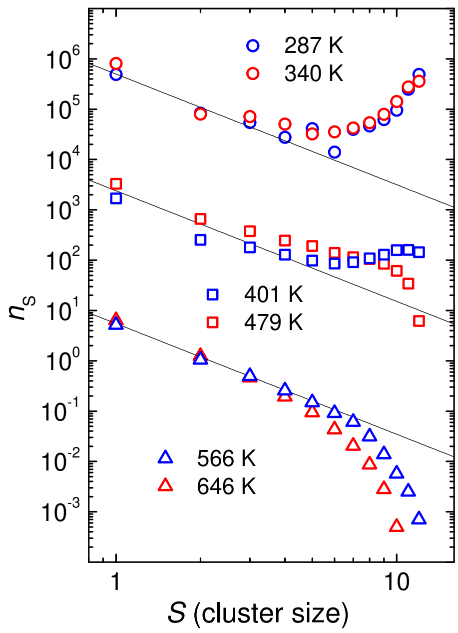


Figure 4.6: Size distribution n_S of the peptide clusters for selected temperatures (symbols) in a double-logarithmic plot. The power law $n_S \sim S^{-2.2}$ is shown by lines.

The average number of peptide-water hydrogen bonds per peptide, n_H^{wp}, closely follows the temperature dependence of SASA in the temperature range 287 K $< T <$ 430 K (Figure 4.7). In particular, one additional water-peptide H-bond corresponds to an increase of SASA of about 1.1 nm^2. The temperature induced increase of SASA above 430 K is not accompanied by the creation of new water-peptide H-bonds. This means that the peptide is not able to form H-bonds with more than about 14 water molecules. The effective hydrophilicity of the surface can be characterized by the number of H-bonds with water per unit SASA. The increase of the ratio n_H^{wp}/SASA upon cooling indicates an increasing hydrophilicity of the exposed surface upon aggregation (Figure 4.7, lower panel). These results are in agreement with previous studies of aggregation of the same peptide at $T = 330$ K, proving that the aggregation of the peptide studied occurs mainly via contacts between hydrophobic groups. The average number of intermolecular peptide-peptide H-bonds per peptide, n_H^{pp}, is ~ 2.5 at low temperatures (Figure 4.8, upper panel), continuously decreases and finally vanishes upon heating, indicating that most of the peptides are in monomeric form at high temperatures. The average number of intra-peptide H-bonds, n_H^{intra}, increases slightly up to 450 K, beyond which it remains almost unchanged. The population of various elements of secondary structure is shown as a function of temperature in the lower panel of Figure 4.8. In the aggregated state, which dominates at low temperatures, β-strands are the dominating structural elements. The fraction of the PP_{II} structure is about 25% at low temperatures and decreases by a few percents upon heating. The population of random coil conformations increases from 18 to 26 % upon heating, whereas the α-helix content shows a non-monotonic behavior, i.e., it increases from $\sim 19\%$ at $T = 287$ K, passes through a maximum ($\sim 23\%$) at $T \approx 475$ K, and drops to $\sim 21\%$ at the highest temperature studied. The similarity of the

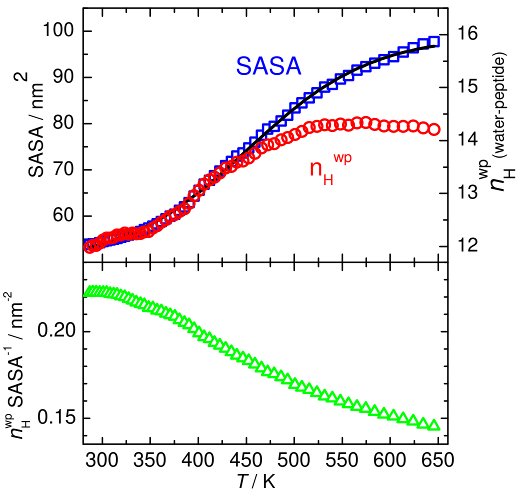


Figure 4.7: Temperature dependences of the SASA of all peptides (open squares), the number of peptide-water H-bonds per peptide n_H^{wp} (open circles), and their ratio (open triangles).

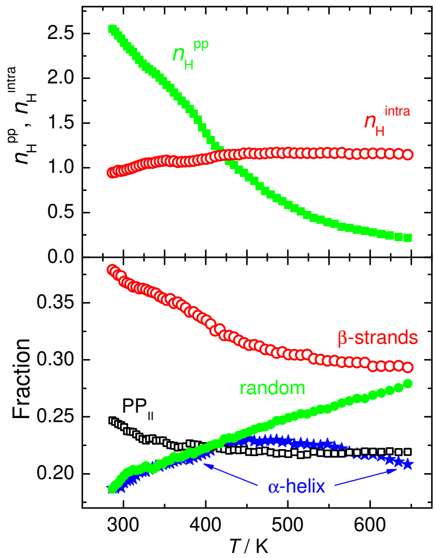


Figure 4.8: Number of inter- (n_H^{pp}) and intra-peptide H-bonds (n_H^{intra}), and populations of various elements of secondary structure, as a function of temperature are shown in upper and lower panels, respectively.

temperature dependences of the β-strand content and the number n_H^{pp} of inter-peptide H-bonds (Figure 4.8) indicates their coupling, which is expected as the β-sheets are formed due to the creation of inter-peptide H-bonds. The degree of such coupling is clearly seen, when the β-strand content is shown as a function of inter-peptide H-bonds per peptide, n_H^{pp} (Figure 4.9, lower panel). The dependence is linear with a slope of ~ 0.036 over the entire temperature range studied. As the presence of intra-molecular H-bonds is an important condition for the existence of α-helical structures, the α-helical content as a function of n_H^{intra} is shown in the upper panel of Figure 4.9. Three temperature intervals with an almost linear relation between the α-helical content and n_H^{intra} (dashed line) can be distinguished. Up to about 450 K, both properties increase with temperature due to the dissolution of the peptide aggregate, and a qualitative change of the correlation between the α-helical content and n_H^{intra} is observed at ~ 375 K. Upon heating beyond 475 K, the α-helical content starts to decrease, whereas n_H^{intra} remains almost unchanged.

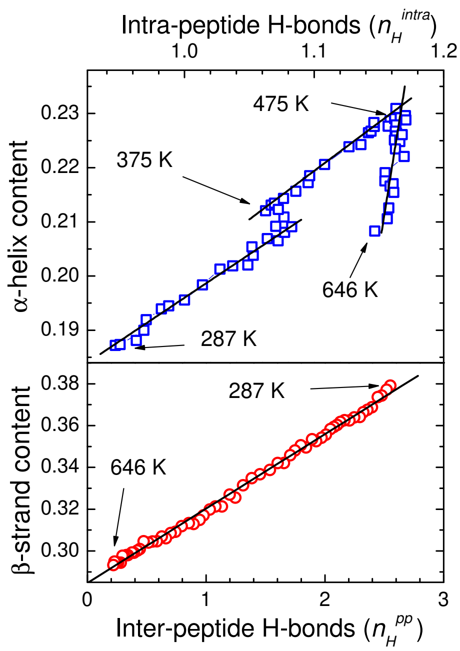


Figure 4.9: Correlations between α-helix and β-strand contents and the numbers of intra- (n_H^{intra}) and inter-peptide H-bonds (n_H^{pp}), respectively.

Additional information about the peptide surface and the character of peptide aggregation can be obtained from the behavior of the volumetric properties of the hydration water and the intrinsic volumetric properties of the peptides.[110] The density of water in the hydration shells of the peptides, $\rho_h = N_w/V_h \cdot m_w$, where m_w is the mass of a water molecule, is shown in Figure 4.10 as a function of temperature. ρ_h was found to be

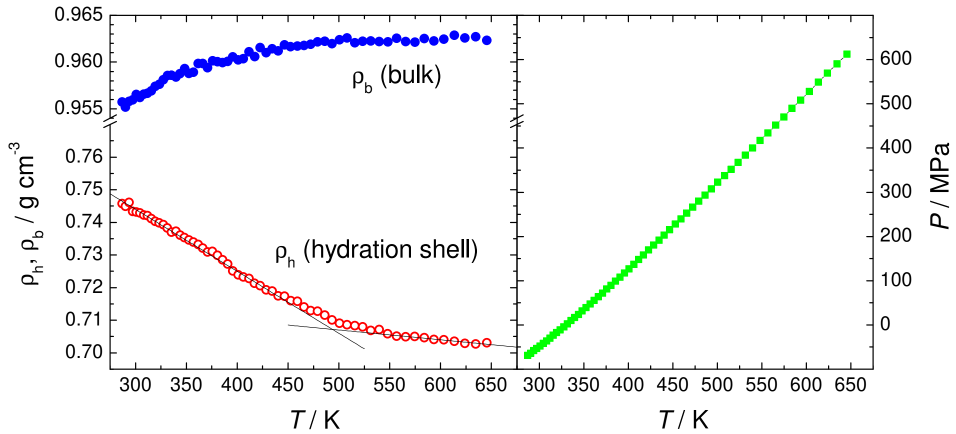


Figure 4.10: Dependence of the water density in the bulk ρ_b (solid circles) and in the hydration shell ρ_h (circles) on temperature along the simulated thermodynamic path. Pressure-temperature diagram of REMD simulations (right panel).

notably lower ($\sim$ 25-30 %) than the bulk density, ρ_b, in the same thermodynamic states, in agreement with the pronounced hydrophobic nature of the peptide surface. The densities of both the bulk and the hydration water strongly depend on the thermodynamic state of the system, which is characterized by temperature and pressure in the present constant-volume REMD simulations. The density of bulk water increases with temperature, due to a strong increase of pressure with temperature (Figure 4.10). In contrast, the density of hydration water decreases with temperature, and the thermal expansion coefficient of the hydration water, α_h, is estimated to be -2$\cdot$10^{-4} K^{-1} up to $T \approx 450$ K, and about one order of magnitude smaller at higher temperatures. The thermal expansivity of the hydration water observed in the present REMD simulations is notably smaller than that obtained near other peptides using constant temperature molecular dynamics simulations at $p =$ 0.1 MPa,[110,111] which is probably due to the very high pressures at high temperatures in the REMD simulations. However, the qualitative change in the temperature dependence of ρ_h that takes place between 450 and 550 K cannot be attributed to the specific change of the thermodynamic state of system, as the pressure p increases almost linearly with temperature.

The intrinsic volume V_p of the peptides varies non-monotonically along the thermodynamic path studied: it slightly decreases upon heating up to $T \approx 350$ K and decreases by about 30% upon further heating. The temperature dependence of $\ln V_p$ is shown in Figure 4.11. The slope is equal to the peptide's intrinsic thermal expansion coefficient α_p. The temperature dependence of $\ln V_p$ is almost linear between 375 and 575 K, with α_p = -2.3·10^{-3} K^{-1} (solid line in Figure 4.11). The temperature dependence of the intrinsic density ρ_p of the peptides is shown in the lower panel of Figure 4.11. The density of the peptides is about 1.75 g/cm^3 and increases with temperature and pressure upon the dissolution of the aggregate.

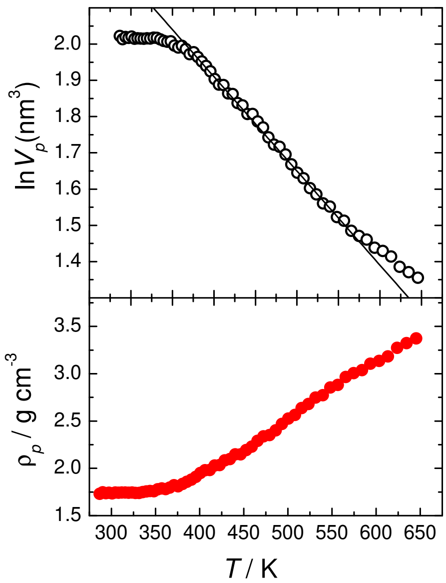


Figure 4.11: Temperature dependencies of lnV_p (upper panel) and the average density ρ_p of the peptides (lower panel).

4.3 Discussion

The simulation studies of an aqueous solution of the amyloidogenic peptides show that upon cooling, the solution undergoes a transition from a one-phase state, with peptides completely dissolved in water, to a two-phase state, where a solid-like peptide aggregate is in equilibrium with a saturated solution of the peptide. Various properties of the peptide system show a behavior similar to a conventional first order phase transition of fluids, which is strongly smeared out due to the finite size of the simulation box. The rounding of the transition is caused by the destabilization of the minor phase (peptide aggregate) in any simulated finite system due to thermodynamic reasons.[71] This is an intrinsic problem of any simulation study and it can be minimized by using larger systems. Additional rounding of the transition may originate from the exchange of configurations between the one-phase and the two-phase thermodynamic states in REMD simulation methods.

Various parameters were used to characterize the peptide aggregation and to locate the temperature of the demixing transition which separates these two distinct states of the system. One type of parameters reflects the properties of the whole system of peptides, e.g. the total radius of gyration or SASA, the total number of hydrophobic contacts or H-bonds, etc. Such an approach is valid, when the critical peptide concentration in the water-rich phase (the solubility limit) is so small that it is effectively reproduced by pure water, while all the peptides in the simulation box almost permanently belong to the organic-rich phase. This situation corresponds to the aqueous solution of amyloidogenic peptides at low temperatures. However, upon heating, the critical peptide concentration increases and the essential part of the peptides at this point belongs to the water-rich phase. Hence, the ability of parameters that utilize the properties of all the peptides to describe the formation of a peptide aggregate degenerates as they become less sensitive to the formation of the peptide aggregate in the organic-rich phase. This results in the smearing out of the temperature dependences of these parameters in a wide temperature range, and a shift of the midpoint of the transition towards higher temperatures (see dependences in Figures 4.2 and 4.3). Therefore, the inflection points of these dependences (450 to 485 K) give overestimated temperatures for the corresponding phase transition. In a similar analysis of the temperature-induced transition of amyloidogenic peptides in implicit water, the transition was found to be smeared out in an essentially narrower temperature interval, despite the use of a smaller number of peptides,[56,57] which could be due to the lower peptide concentrations used in these studies.

The appearance of the aggregated phase can be seen in the size distribution of the peptide clusters (Figure 4.6). The distortion of these distributions due to the small system size containing a few peptides only prevents an accurate location of the transition temperature. The clustering analysis can be used to recognize the largest peptide aggregate that represents an organic-rich phase and offers the possibility to analyze its properties. The probability distribution of the size of the largest peptide cluster (Figure 4.4) shows largest fluctuations at ~ 440 K. The temperature dependence of the probability to have an aggregate including more than 2/3 of peptides, based on the connectivity criterion of the distance between the centers of mass of two peptides ($r_{crit} \leq 1.1$ nm), gives an inflection point at $T = 415$ K. As this is one of the most robust parameters for characterizing aggregation of this particular peptide, the midpoint of the demixing phase transition can be expected to be located at ~ 415 K.

Similar to the previously reported studies of the same peptide at $T = 330$ K, the hydrophilicity of the total exposed surface increases upon aggregation (Figure 4.7). This conclusion is also supported by the increase in the density of hydration water upon aggregation (left panel in Figure 4.10). It is to be noted that this is not related to a decrease in temperature, as a corresponding decrease in pressure leads to an even more drastic decrease in the density of bulk water.

The thermal expansivity of the peptides in an aggregated state is close to zero (Figure 4.11), indicating a solid-like structure of the aggregate. As the aggregate dissolves with increasing temperaure, the thermal expansivity α_p of the peptides becomes negative. This is expected for a random chain, and was also observed for the amyloidogenic peptide $A\beta_{42}$ and an elastin-like peptide in water.[110] In the latter case, a random chain behavior was confirmed by analyzing the structural properties of the peptide.[112] Therefore, a negative value of α_p obtained for the peptide system studied indicates a random chain conformational behavior of single peptides in the disaggregated state, which dominates at high temperatures.

The density of the peptide aggregate (1.75 g/cm^3) largely exceeds the density (0.66 g/cm^3) estimated for the same peptide at 330 K in the previous Chapter. This could be due to the fact that the latter value was obtained by estimating the V_p from the radius of gyration R_g of the largest peptide cluster and assuming its shape to be spherical. Since a sphere provides the maximal volume at a fixed radius of gyration, the value estimated was the lower limit for the aggregate density. However, in this study, V_p was calculated from Eq. 4.1, where the volume of the hydration shell was calculated using a slab approximation with a cutoff 0.35 nm, which could lead to systematic errors in absolute values of V_p. The dissolution of the peptide aggregate upon heating leads to the disappearance of the loosely packed structure with numerous voids, inaccessible to water molecules. Accordingly, the observed intrinsic density of the peptides, ρ_p, increases upon heating (Figure 4.11, lower panel).

β-sheet conformations are predominant in the peptide aggregate and the dissolution of the aggregate upon heating leads to the formation of α-helical structures (upper panel in Figure 4.8). Some qualitative changes in this process, presumably related to the demixing transition, occur between 350 and 375 K. A drastic change of the correlation between the α-helical content and the number of intra-peptide H-bonds occurs at about 475 K and 2.5 kbar. This indicates a change in the conformational behavior of individual peptides in liquid water, which occurs essentially above the demixing transition temperature. Approximately at the same state point, the dependence of the density of hydration water also changes qualitatively (see the left panel in Figure 4.10), whereas this is not the case for bulk liquid water. These changes point towards a possible relation between the state of the hydration water and the conformational behavior of the peptides.

Chapter 5

Simulations of Peptide Aggregation Near Surfaces

Spontaneous adsorption of protein molecules at interfaces is generally observed as an aqueous solution of the protein comes into contact with an interface.[113] Protein side chains vary widely in their hydrophobicity, which renders them amphiphilic. Thus, there is a general tendency of proteins to reside at the interface between two phases of different polarity. As the folded conformation of the proteins is only marginally stable against unfolding, an interface-induced conformational change of protein can lead to an optimized contact between the interface and the protein.[114] The presence of surfaces can induce a concentration gradient in a protein solution, and in turn affect its aggregation. The two dimensional nature of the surfaces can also induce orientational ordering of peptides, and thus lead to the formation of two dimensional aggregates.

Since the effects of surfaces on peptide aggregation are multifacet, it would be useful to analyze them separately. The direct interaction between peptides and surfaces includes contributions from Coulombic interactions, dispersion interactions, hydrogen bonding, etc. The presence of the solvent causes the appearance of solvent-induced interactions between the peptides and the surface. Hence, the *effective* peptide-surface interaction is determined by relative strengths and ranges of direct peptide-surface, solvent-surface and peptide-solvent interactions.

In order to study the effects of solvent induced peptide-surface interactions on peptide aggregation, two types of peptide fragments were used, i.e., fragment A, with the amino acid sequence NFGAIL, having a molecular weight of 634 Dalton, and fragment B, with the amino acid sequence GNNQQNY and a molecular weight of 837 Dalton. Fragment

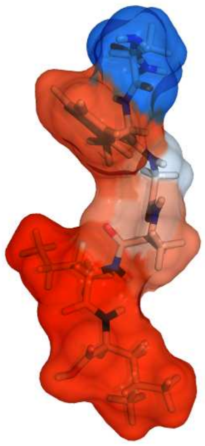


Figure 5.1: Fragment A, the hydrophobic fragment.

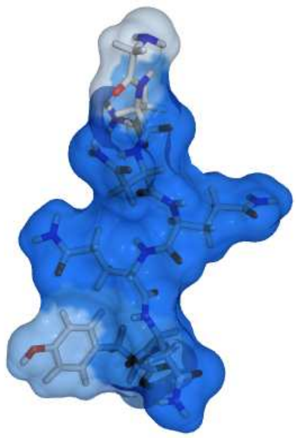


Figure 5.2: Fragment B, the hydrophilic fragment.

A is part of hIAPP (residue number 25-29) and fragment B is part of the Sup35 yeast protein (residue number 7-13). Sup35 is a 685 residue yeast protein. It is a subunit of the translation termination factor and its normal cellular role is to terminate translation. Yeast harboring the $[PSI^{+}]$ determinant exhibits increased translation read-through of stop codons due to aggregation and subsequent loss of function of Sup35. The first 123 residues constitute the prion determining domain and are sufficient for fibril formation.[115] The sequence GNNQQNY has been shown to display the same amyloid properties as Sup35.[81] Even though both fragments are capable of forming fibrils, they differ greatly in their amino acid composition. Fragment A mainly consists of hydrophobic amino acids, whereas fragment B is made up of polar amino acids. Figures 5.1 and 5.2 show the molecular surfaces of both the fragments, colored with values according to the hydrophobicity scale of Kyte and Doolittle,[116] with colors ranging from dodger blue for the most hydrophilic amino acid residue (-4.5) to white at 0.0, and towards orange-red (4.5) for the most hydrophobic residue . In the following text, fragments A and B are referred to as the "hydrophobic" and the "hydrophilic" fragment, respectively.

5.1 System setup

All simulations were performed using the Gromacs software, OPLS force field[117,118] and SPCE water molecules.[119] The short range cutoff of 1.2 nm was used for LJ and Coloumbic interactions. LINCS algorithm was used to constrain all the bonds to their equilibrium bond lengths, allowing an integration time step of 2 fs. The neighbor list was updated every 20 fs and the coordinates were recorded every 2 ps. The following systems were simulated:

1. Water in slit-like pores.
2. The liquid-vapor interface of water.
3. Peptides in bulk water.
4. Peptides in slit-like pores and at the liquid-vapor interface.

5.1.1 Water in slit-like pores

The interaction of water molecules with the surfaces was represented using the 9-3 potential

$$\begin{aligned} U(r) &= \frac{4\pi\epsilon n_{wall}\sigma^3}{6}\left[\frac{2}{15}\left(\frac{\sigma}{r}\right)^9 - \left(\frac{\sigma}{r}\right)^3\right] \\ &= \pi n_{wall}\left(\frac{A_{ij}}{45r^9} - \frac{B_{ij}}{6r^3}\right), \end{aligned} \tag{5.1}$$

where $A_{ij} = \sqrt{A_i A_j}$, $A_i = 4\pi\epsilon_i\sigma_i^{12}$, $B_{ij} = \sqrt{B_i B_j}$, $B_i = 4\pi\epsilon_i\sigma_i^6$. The "hydrophilic wall" was represented by σ and ϵ values of 0.373 nm and 0.030800 kJ/mol, respectively, whereas the "hydrophobic wall" interacts with σ and ϵ values of 0.373 nm and 5.2 kJ/mol, respectively (see Figure 5.3). The number density of the wall n_{wall} was set to 100 nm^{-3}. Simulations were carried out on SPCE water molecules in a cubic box of length 6 nm with the model walls at 0 and 6 nm in the z dimension. PME with 3dc geometry was used to calculate long range Coloumbic interactions. In 3dc geometry, the reciprocal sum is still performed in 3d, but a force and potential correction are applied in the z dimension to produce a pseudo-2d summation.[120] For Ewald summation, a scaling factor of 3 was

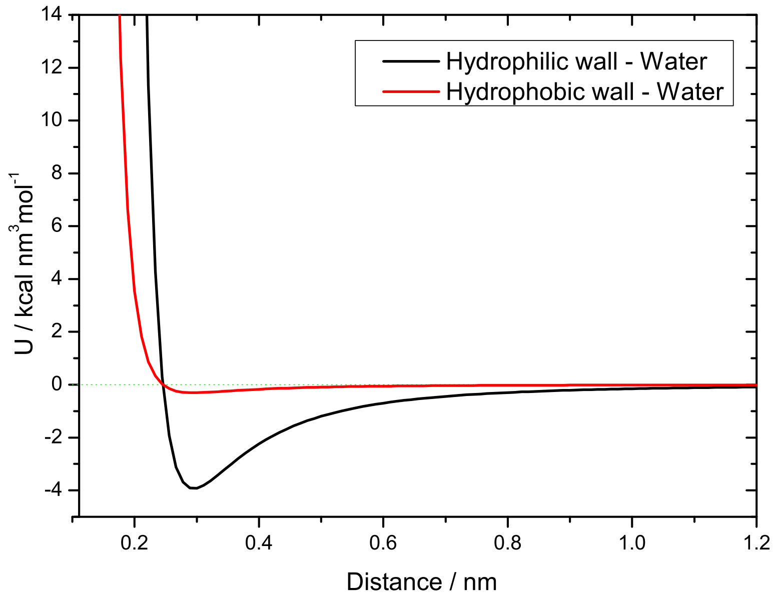


Figure 5.3: Potential energy as a function of distance, between an oxygen atom of water and the hydrophilic or the hydrophobic wall are representated by black and red lines, respectively.

used to scale down the interaction in the z direction. The system was simulated for 20 ps at 600 K and subsequently for 300 ps at 330 K for initial equilibration. Water molecules were added or removed from the system until the density at the center of the box matched that of the bulk density of SPCE water molecules (see Figure 5.4).

5.1.2 Liquid-vapor interface

The cubic box of length 6 nm was filled with 6400 SPCE water molecules. The box was extended in the z direction in such a way that an empty space of 6 nm exists between the two periodic images in z direction. The box was equilibrated for the duration of 10 ns at 330 K in an NVT ensemble using periodic boundary conditions in all three directions.

5.1.3 Peptides in bulk water

Six copies of one fragment were added randomly into a cubic box of 6 nm length in such a way that the fragments were at least 0.8 nm away from each other. The boxes were solvated with water molecules. Five simulations were started at different initial velocities for each of the systems, i.e., five simulations for the hydrophobic fragment and five simulations for the hydrophilic fragment. Each system was initially equilibrated for 1 ns with position restrains on the peptide fragments. After 1 ns, the position restrains were removed, and the system was further simulated for the duration of 70 ns. The number of water molecules for the systems with hydrophobic and hydrophilic fragments were 6439 and 6394, respectively. All the simulations were carried out in an NVT ensemble. An equilibration period of 30 ns was estimated from the time evolution of various system parameters and was excluded from the analysis of system properties.

5.1.4 Peptides in slit-like pores and at liquid-vapor interface

The interaction between wall and peptide atoms was represented by $A_{ij} = 0.0$ kJ mol^{-1} nm^{12} and $B_{ij} = 0.0$ kJ mol^{-1} nm^{6}, which lead to no interaction between the peptide atoms and the pore walls. Six copies of the peptide fragment were added randomly into the box in such a way that each fragment is at least 0.8 nm away from other fragments and at least 1.5 nm away from the walls. The water molecules overlapping with the peptides were removed. The system was equilibrated for 1 ns using Berendsen's temperature coupling[121] with a coupling constant of 0.5 ps. In the case of peptides with a liquid-vapor interface, position restrains were used on peptides during the initial 100 ps. After 1 ns of equilibration using Berendsen's temperature coupling, the system was switched to Nose-Hoover temperature coupling[122,123] with a coupling constant of 1.5 ps and thereupon further simulated for 70 ns. Five NVT simulations, each starting with a different distribution of initial velocities at 330 K and 1 atm pressure, were carried out for each of the following systems:

1. Fragment A with hydrophilic walls: 6689 water molecules and 6 peptide fragments.
2. Fragment A with hydrophobic walls: 6189 water molecules and 6 peptide fragments.
3. Fragment A with a liquid-vapor interface: 6184 water molecules and 6 peptide fragments.

4. Fragment B with hydrophilic walls: 6643 water molecules and 6 peptide fragments.
5. Fragment B with hydrophobic walls: 6145 water molecules and 6 peptide fragments.

5.2 Results

The density profiles of liquid water in the two pores and the density profile in the slab of liquid water are shown in Figure 5.4. There is a pronounced density depletion of liquid water near the hydrophobic surface caused by the domination of the effect of missing neighbors over the weak water-surface attractive potential. The depletion of water density is more gradual near the liquid-vapor interface. Conversely, there are two pronounced and highly ordered water layers near the hydrophilic surface. The high value of the local density at the maximum of the first peak does not correspond to an enhancement of water density in the first layer, but rather to a narrow localization of water molecules in a plane parallel to the surface.

The radius of gyration R_g of the peptides, calculated using the method described in Chapter 3, decreases rapidly within 30 ns in all the simulations, while the number of inter-peptide hydrogen bonds increases during the same time period, indicating that the peptides aggregate in all the simulation runs. The equilibration period of 30 ns was estimated from the time evolution of various system parameters. The time dependence of the probability distribution of the center of mass of the peptides in pores and in the slab of liquid water is shown in Figure 5.5. After the random insertion of hydrophobic peptides into the liquid water in the hydrophobic pore or in the slab of liquid water, system equilibration requires up to 30 ns. After equilibration, all peptides are adsorbed at the interfaces. Usually the peptides are adsorbed on the two opposing interfaces, but sometimes all the peptides are adsorbed at only one of them. The situation is quite different in hydrophilic pores, where both hydrophilic and hydrophobic peptides quickly become localized in the center of the pore. A similar trend is also observed for hydrophilic peptides in the hydrophobic pore.

The density profiles of peptide in pores, calculated after the equilibration period of 30 ns, are shown in Figures 5.6 and 5.7. The degree of peptide localization in the pore center is quite similar for both the hydrophilic and hydrophobic peptides in the hydrophilic pore. Hydrophilic peptides show weaker localization in the center of the hydrophobic pore. An

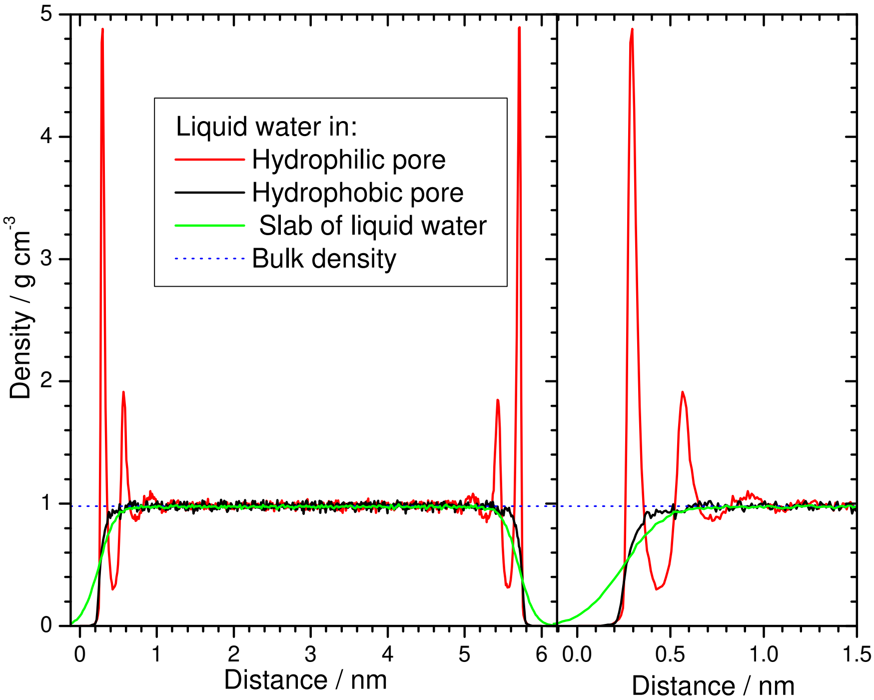


Figure 5.4: Density profile of liquid water in a hydrophobic pore, hydrophilic pore, and in a slab of liquid water. Pore walls are located at 0 and 6 nm, respectively. The density profile of the liquid slab is shifted by 3.5 nm. The same graph is shown at an enlarged scale in the right panel.

opposite behavior is observed for the hydrophobic peptides in the hydrophobic pore or in the slab of liquid water: in all simulation runs, peptides are strongly adsorbed at the interfaces (Figure 5.7). The degree of peptide localization near interfaces essentially exceeds that in the pore center, as can be seen from the comparison of the width of the density distributions shown in Figures 5.6 and 5.7, respectively. The probability distribution of the angle α between the pore surface and the vector connecting two most distant peptide heavy atoms are shown in Figure 5.8. When peptides are repelled from the pore walls and localized in its center, the orientations of their longest axes are highly isotropic (left panel in Figure 5.8) and only a slight preferential orientation of these axes parallel to the wall can be noticed. The situation is quite different in the case of a strong adsorption of peptides at the interfaces (right panel in Figure 5.8). The localization near the interfaces essentially enhances the orientational ordering of peptides and makes their longest axes align parallel to the interfaces.

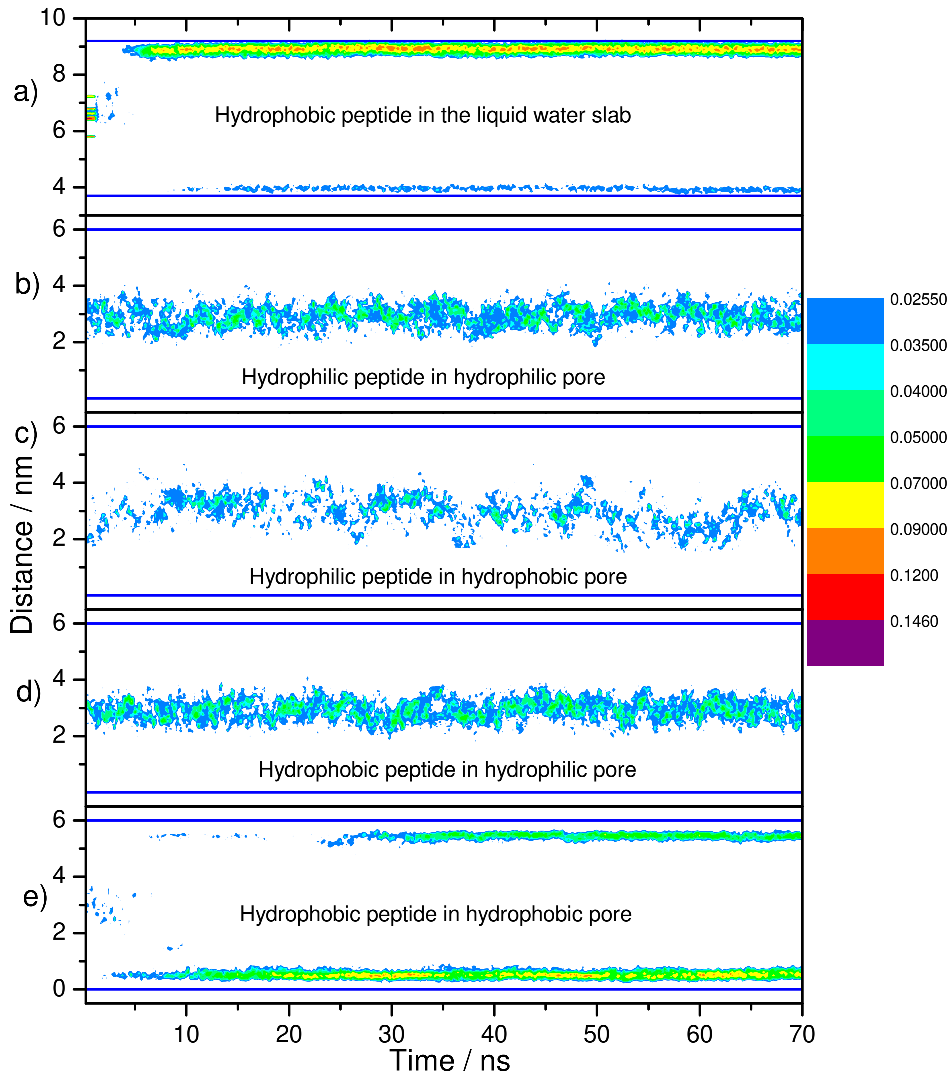


Figure 5.5: Probability distribution of the center of mass (COM) of the peptides as a function of time. Blue lines depict the walls in the case of slit-like pores, and the midpoint of the liquid-vapor interface in the case of the liquid slab. The probabilites of COM are shown using the color range depicted above.

The droplet probability R to have more than four peptides in a cluster based on the distances between the centers of mass of two pepties as a function of the distance r_{crit} used as cutoff is shown in Figure 5.9. The degree of aggregation of the hydrophobic and hydrophilic peptides in bulk solutions is quite similar (see upper panel in Figure 5.9). It

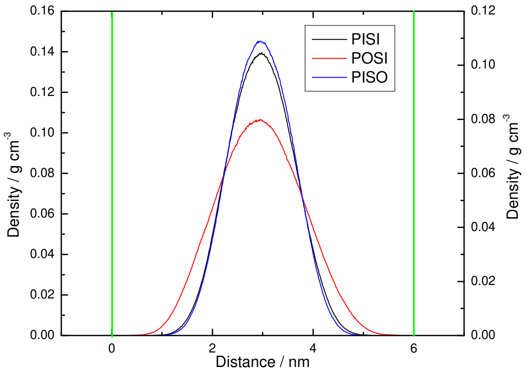


Figure 5.6: Density profiles of peptides in aqueous solution in various pores: hydrophilic peptides in hydrophobic (PISO) and hydrophilic pores (PISI and left vertical axis); hydrophobic peptides in the hydrophilic pore (POSI and right vertical axis). The scales of the left and right are proportional to the molecular weights of the hydrophilic and hydrophobic peptides. The green vertical lines represent the position of the pore walls.

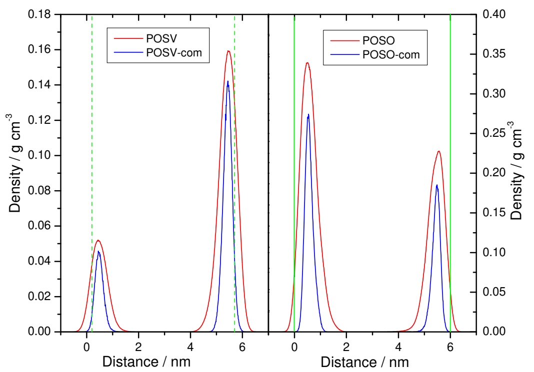


Figure 5.7: Density profiles of hydrophobic peptides in the slab of aqueous solution (left panel, POSV) and in aqueous solution in the hydrophobic pore (right panel, POSO). The density profiles of the centers of mass of the peptides are shown by the blue lines (POSV-com and POSO-com). The midpoints of the liquid-vapor interface and the position of the pore walls are shown by the green dashed and solid lines, respectively.

can be compared with that of the FLVHS peptides studied in Chapter 3. The degree of aggregation is noticeably weaker in the latter case in a wide range of r_{crit}. This could be due to a weaker propensity of these peptides to aggregate, the presence of methyl caps in the FLVHS peptides, or due to the different force fields. The aggregation of hydrophilic peptides becomes weaker in both hydrophilic and hydrophobic pores (see middle panel in Figure 5.9), as it can be inferred from a shift of the dependence $R(r_{crit})$ to higher values of r_{crit}. For hydrophobic peptides, the situation is opposite (lower panel in Figure 5.9). The aggregation of hydrophobic peptides becomes slightly stronger upon confinement in hydrophilic pores, and it is enormously enhanced by confinement in the hydrophobic pore.

The average peptide-water and inter-peptide hydrogen bonds, peptide-water contacts and the density of the hydration layer of the peptides for the various systems investigated are shown in Figure 5.10. The average number of peptide-water and inter-peptide hydrogen bonds is higher in the simulation runs with hydrophilic fragments, due to a higher number of polar atoms. Hydrophilic peptides also have fewer peptide-water hydrogen bonds upon bulk phase aggregation compared to the aggregation in pores.

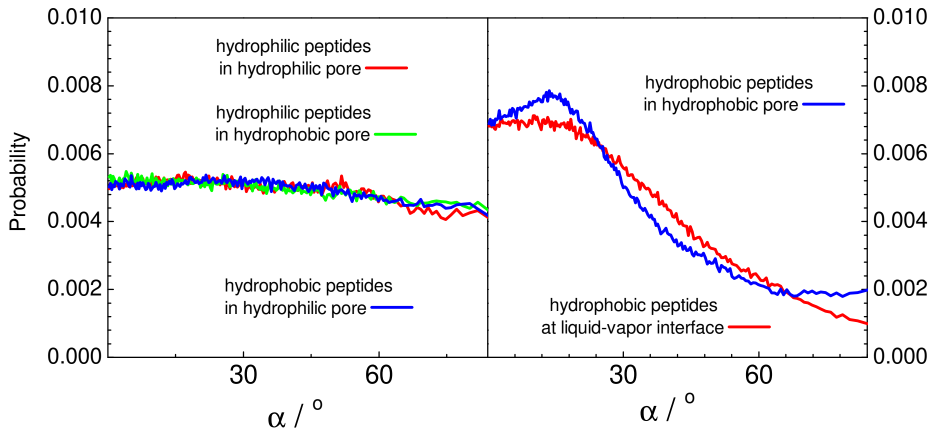


Figure 5.8: Probability distribution of the angle α between the pore wall or the liquid-vapor interface and the vector connecting two most distant peptide heavy atoms.

The number of water molecules in contact with the aggregate follows a similar trend: we observed a 12% increase in the number of water molecules around the aggregate in pores as compared to the aggregate in the bulk. In the case of hydrophobic fragments, the number of peptide-water hydrogen bonds and water molecules in contact with the peptides decreases upon adsorption to the hydrophobic surface. As the peptides are aligned parallel to the surface, the formation of β-sheets is facilitated, thus compensating for the decreased peptide-water hydrogen bonding by formation of inter-peptide hydrogen bonds. As a result, an increase in the number of peptides adopting a β-sheet conformation is observed upon adsorption.

The density of the hydration layer of the peptides for the various simulation runs (except for the hydrophobic peptides in hydrophobic pores) is shown in Figure 5.10 (bottom left panel). In the case of hydrophilic peptides, the higher shell densities found for aggregates in pores as compared to aggregates in the bulk can be attributed to the suppression of aggregation of hydrophilic peptides in pores. The trend towards a more dense hydration shell upon increased aggregation has also been observed in the case of the hydrophobic fragment studied in Chapter 3 (see Figure 3.15). A similar trend is observed for the hydrophobic fragment in the hydrophilic pore, where aggregation is enhanced, as

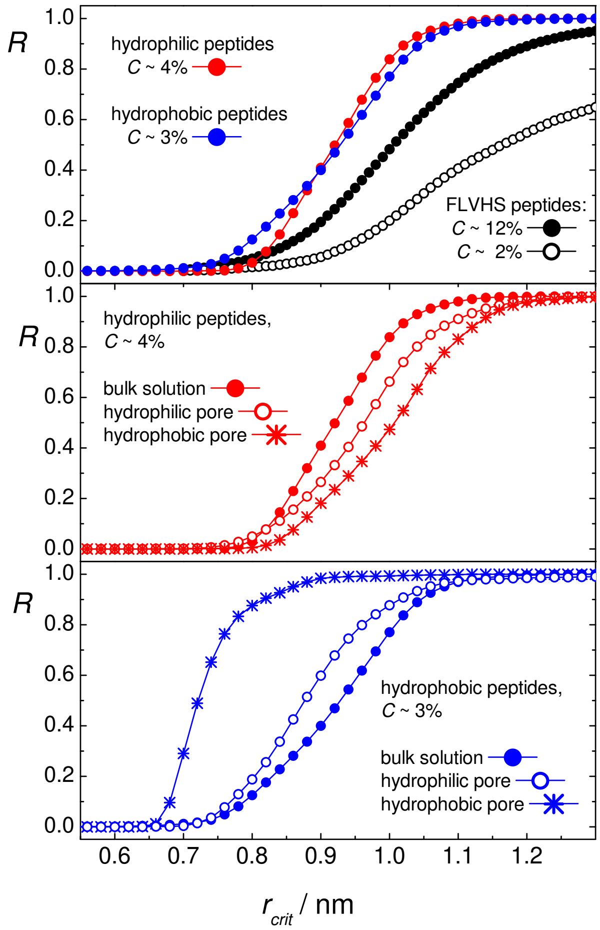


Figure 5.9: Dependence of the aggregation parameter R on the distance r_{crit} used as a criterion for inter-peptide connectivity. Upper panel: bulk aqueous solutions of peptides. Middle panel: aqueous solution of hydrophilic peptides in hydrophilic and hydrophobic pores. Lower panel: aqueous solution of hydrophobic peptides in pores.

compared to the bulk simulations.

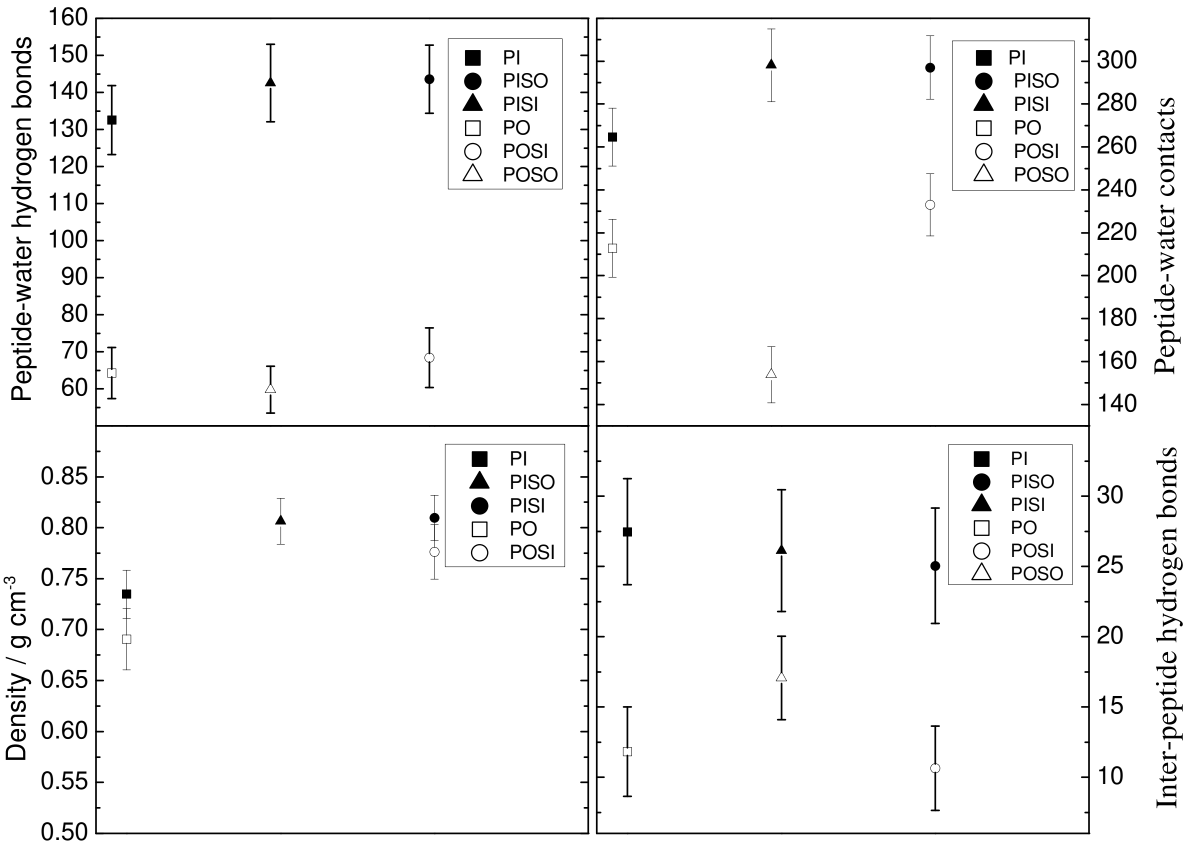


Figure 5.10: Average number of peptide-water hydrogen bonds (top left panel), water molecules in contact with peptides (top right panel), shell density (bottom left), and inter-peptide hydrogen bonds (bottom right) for each type of run labelled as follows: PI (hydrophilic fragment in bulk water), PISI (hydrophilic fragment in the hydrophilic pore), PISO (hydrophilic fragment in the hydrophobic pore), PO (hydrophobic fragment in bulk water), POSO (hydrophobic fragment in the hydrophobic pore), POSI (hydrophobic fragment in the hydrophilic pore). Error bars represent RMSD from average values.

5.3 Discussion

The effect of water mediated peptide-surface interactions is determined by the balance of peptide-surface, peptide-water and water-surface interactions. For simplicity, the direct peptide-surface interactions are switched off in these studies. The hydrophilic peptide is composed of residues having multiple polar atoms in its side chains, which makes water-peptide interactions much more favorable as compared to the case of the hydrophobic peptide. This prevents the adsorption of hydrophilic peptides at both hydrophilic and hydrophobic surfaces, as the adsorption would decrease the number of favorable peptide-water hydrogen bonds. On the contrary, the hydrophobic peptide has many apolar atoms, which makes direct peptide-water interactions much weaker as compared to water-water interactions. Hence, adsorption of hydrophobic peptides decreases the overall amount of hydrophobic surface exposed to water, and the formation of inter-peptide β-sheets leads to further stabilization. However, this trend can be overcome if the surface is hydrophilic enough. Near a hydrophilic silica-like surface, the tendency of the surface to be hydrated overcomes the tendency of the hydrophobic peptides to be dehydrated, which leads to their desorption from the surface.

Depending on the nature of the peptide and the surface, either an enhancement or a retardation of peptide aggregation can be observed. The adsorption of hydrophobic peptides on hydrophobic surfaces foster the aggregation drastically by facilitating the formation of inter-peptide β-sheets, which is a key structural element of amyloid fibril formation. A small increase in the aggregation propensity is also seen by confining hydrophobic peptides in hydrophilic pores, whereas confinement of hydrophilic peptides in either hydrophobic or hydrophilic pores decreases their aggregation propensity.

Summary

Protein aggregation is involved in various biological processes and is particularly associated with many diseases such as Alzheimer's, Parkinson's and type II Diabetes Mellitus. The aggregation of amyloidogenic proteins resulting in the formation of amyloid fibrils has been shown to follow a nucleated growth mechanism. The separation of an aqueous solution of a protein into solid fibrils and a saturated protein solution, in equilibrium with each other, is quite similar to the demixing phase transition of binary mixtures. As protein aggregation in cells is affected by various factors such as surfaces, the finite size of the cells and the presence of cosolutes, it is important to understand their effects on the aggregation process. Protein aggregation can be studied using computer simulations of model peptide fragments, which behave in a similar way as the parent proteins themselves. These studies can be very helpful in providing details at the molecular level. In this thesis, the effects of finite size, temperature, and the influence of surfaces on protein aggregation were studied.

In any computer simulation studies involving phase transitions, finite size effects must be properly considered, before any attempt is made to extrapolate the results of the simulations to macroscopic systems. Therefore, the effect of finite size on peptide aggregation was studied by molecular dynamics simulations of aqueous solutions of peptide fragments, at various concentrations and system sizes. Three different parameters were used to quantify the aggregation process and calculate the degree of aggregation. The aggregation was enhanced with increasing the peptide concentration at a fixed system size, and with increasing the system size at a fixed peptide concentration. The observed decrease in the degree of peptide aggregation with decreasing system size was attributed to the appearance of an artificial state in which the minor phase, i.e. the peptide aggregate, was "dissolved". As the system size approaches macroscopic dimensions, this artificial

state eventually disappears. Thus, at a concentration where an aqueous solution of peptides will form an aggregate in a macroscopic system, the aggregation of the peptides will be suppressed in small (microscopic) systems. There are two major implications of these studies: Firstly, the finite system size has a drastic effect on peptide aggregation and is responsible for the instability of aggregates consisting of only a few peptides, hence it must be considered in any simulation study designed to reproduce the aggregation in macroscopic systems. Secondly, if the real system of interest is not macroscopic and contains a relatively small number of peptides, their aggregation will be suppressed by the finite system size as well. This situation may be relevant in the case of peptides in small volumes, such as biological cells or their compartments, where it might be playing a significant role in the suppression of protein aggregation.

The effect of temperature on peptide aggregation was studied using replica exchange molecular dynamics simulations. Upon cooling, the aqueous solution of peptides was found to undergo a transition from a one-phase state, where the peptides were completely dissolved in water, to a two-phase state, where the peptide aggregates are in equilibrium with the saturated solution of the peptides. Various properties of the peptide-water system showed a behavior similar to conventional first order phase transitions of fluids. The transition was smeared out due to the finite system size. The midpoint of the transition was located at ~ 415 K.

The effect of hydrophobic and hydrophilic surfaces on peptide aggregation was studied by simulating aqueous solutions of peptides in slit-like pores. Two types of peptide fragments having an amino acid composition that renders them either hydrophobic or hydrophilic, were used in these studies. The adsorption of hydrophobic peptides onto hydrophobic surfaces was accompanied by enhanced aggregation and β-sheet formation. The confinement of hydrophobic fragments in hydrophilic pores slightly increased their aggregation propensity, whereas confinement of hydrophilic fragments retarded their aggregation, irrespective of the nature of the surface.

The studies presented in this thesis demonstrate the pronounced effects of finite size and surfaces on peptide aggregation. Based on the results obtained, qualitative predictions can be made for changes in peptide aggregation, both in small volumes and near surfaces, which could be helpful in understanding the process of peptide aggregation in biosystems and finding innovative ways to extend the shelf-life of protein-based pharmaceutical products.

Zusammenfassung

Die Aggregation von Proteinen spielt in verschiedenen biologischen Prozessen eine Rolle und wird insbesondere mit vielen Krankheiten wie Alzheimer, Parkinson oder Type II Diabetes Mellitus in Verbindung gebracht. Es wurde gezeigt, dass die Aggregation amyloidogener Proteine im Falle der Ausbildung von amyloiden Fibrillen einem keimbasierten Wachstumsmechanismus folgt. Die Aufspaltung einer wässrigen Proteinlösung in eine gesättigte Proteinlösung und feste Fibrillen, die sich miteinander im Gleichgewicht befinden, ist vergleichbar mit der Entmischung in binären Systemen. Weil die Proteinaggregation in Zellen zusätzlichen Einflüssen wie der Anwesenheit verschiedener Oberflächen und Cosolventien und einer begrenzten Zellgröße unterliegt, ist es wichtig, den Effekt dieser Einflüsse zu verstehen. Einen Beitrag zur Aufklärung kann auch durch Computersimulation von Modellpeptidfragmenten, die sich ähnlich verhalten wie das Protein, aus dem sie entnommen wurden, geleistet werden. Solche Studien haben den Vorteil, dass sie Informationen auf molekularer Ebene liefern können. In dieser Arbeit wurden die Auswirkungen einer endlichen Systemgröße, der Temperatur und der Wechselwirkung mit Oberflächen auf die Aggregation untersucht.

In jeder Computersimulation, die sich mit Phasenübergängen beschäftigt, müssen Oberflächeneffekte in die Überlegungen auf geeignete Weise einbezogen werden, bevor die Ergebnisse auf makroskopische Systeme übertragen werden können. Aus diesem Grunde wurde im Rahmen dieser Arbeit der Einfluss einer begrenzten Systemgröße auf die Aggregation von Peptiden mit Hilfe der molekulardynamischen Simulation wässriger Peptidlösungen unter Verwendung verschiedener Konzentrationen und Systemgrößen untersucht. In diesem Zusammenhang wurden drei verschiedene Parameter verwendet, um den Aggregationsprozess zu quantifizieren und den Aggregrationsgrad zu berechnen. Zunehmende Aggregationsgrade ließen sich durch Erhöhung der Peptidkonzentration bei gleich bleibender Systemgröße oder durch Erhöhung der Systemgröße bei gleich bleibender Peptidkonzentration erreichen. Die Abnahme des Aggregationsgrades mit der Abnahme der Systemgröße wurde auf das Auftreten eines künstlichen Zustandes zurückgeführt, in dem die weniger stark vertretene Phase, in diesem Fall das Peptidaggregat, beeinflusst wird. Sobald sich die Systemgröße an die Größe makroskopischer Systeme annähert, verschwindet dieser künstliche Zustand schließlich. Folglich wird eine Aggregation, die bei einer bestimmten Konzentration in einem makroskopischen System auftritt, bei der gleichen Konzentration in kleinen (mikroskopischen) Systemen unterdrückt. Aus den durchgeführten Studien ergeben sich hauptsächlich zwei Schlussfolgerungen: Zum einen hat die endliche Systemgröße einen drastischen Einfluss auf die Aggregation von Peptiden und ist verantwortlich für die Instabilität von Aggregaten, die aus nur wenigen Peptiden bestehen. Wie bereits oben angesprochen, muss dies bei der Übertragung auf makroskopische Systeme berücksichtigt werden. Zum anderen ist auch klar, dass sobald ein reales nicht makroskopisches System von Interesse ist, die Aggregation ebenfalls durch die endliche Systemgröße unterdrückt werden kann. Dies mag im Falle von Peptiden in kleinen Volumina, wie zum Beispiel in biologischen Zellen oder deren Kompartimenten, von Bedeutung sein. Auf diese Weise könnte der begrenzten Systemgröße im Falle biologischer Zellen eine Schlüsselrolle bei der Verhinderung der Ausbildung von Proteinaggregaten zukommen.

Der Einfluss der Temperatur auf die Aggregation von Peptiden wurde mittels replica exchange-molekulardynamischer Simulation betrachtet. Während des Abkühlens durchlief die wässrige Proteinlösung einen Übergang von einem einphasigen Zustand, in dem die Peptide vollständig in Wasser gelöst sind, zu einem Zustand, in dem sich eine feste Phase (Peptidaggregat) mit einer flüssigen Phase (wasserreiche Peptidlösung) im Gleichgewicht befindet. Das Peptid-Wasser-System zeigte ein mit einem Phasenübergang erster Ordnung in Flüssigkeiten vergleichbares Verhalten. Allerdings war der Phasenübergang durch den Effekt einer endlichen Systemgröße verschmiert. Der Mittelpunkt des Übergangs war bei ~ 415 K zu finden.

Die Auswirkung hydrophober und hydrophiler Oberflächen auf die Aggregation von Peptiden wurde untersucht, indem eine molekulardynamische Simulation der Peptidlösung zwischen zwei unendlich ausgedehnten Platten durchgeführt wurde. Hierzu wurden zwei Peptidfragmente verwendet, die sich so in ihrer Aminosäuresequenz unterscheiden, dass sie jeweils entweder eher hydrophile oder eher hydrophobe Eigenschaften aufwiesen. Im

Falle des hydrophoben Peptids führte die Adsorption an einer hydrophoben Oberfläche zur erhöhten Aggregation und Ausbildung von β-Faltblattstrukturen. Beim Einschluss hydrophober Peptide zwischen hydrophilen Oberflächen wurde hingegen nur ein leichter Anstieg der Aggregationstendenz verzeichnet. Für hydrophile Proteine konnte unabhängig von der Art der Oberfläche festgestellt werden, dass die Aggregationstendenz leicht zurückging.

Zusammengefasst demonstrieren die Ergebnisse dieser Arbeit den deutlichen Einfluss von endlicher Systemgröße und Oberflächen auf die Aggregation von Peptiden. Auf Basis dieser Ergebnisse lassen sich qualitative Vorhersagen über das Aggregationsverhalten in kleinen Volumina oder in der Nähe von Oberflächen machen. Dies könnte schließlich dabei helfen, den Prozess der Aggregation in biologischen Systemen besser zu verstehen und neue, innovative Wege zu finden, die Lebensdauer von proteinbasierten pharmazeutischen Produkten zu verlängern.

Bibliography

[1] Tanford, C., and J. A. Reynolds, 2001. Nature's Robots: A History of Proteins. Oxford University Press, USA.

[2] Arakawa, T., and S. N. Timasheff, 1985. Theory of protein solubility. *Methods Enzymol.* 114:49–77.

[3] Sipe, J. D., and A. S. Cohen, 2000. Review: history of the amyloid fibril. *J. Struct. Biol.* 130:88–98.

[4] Chi, E. Y., S. Krishnan, T. W. Randolph, and J. F. Carpenter, 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.* 20:1325–1336.

[5] Stryer, L., 1996. Biochemistry. W. H. Freeman and Company, 4th edition.

[6] 2005. So much more to know. *Science* 309:78–102.

[7] Hawley, S. A., 1971. Reversible pressure–temperature denaturation of chymotrypsinogen. *Biochemistry* 10:2436–2442.

[8] Smeller, L., 2002. Pressure-temperature phase diagrams of biomolecules. *Biochim. Biophys. Acta* 1595:11–29.

[9] Mishra, R., and R. Winter, 2008. Cold- and pressure-induced dissociation of protein aggregates and amyloid fibrils. *Angew. Chem. Int. Ed. Engl.* 47:6518–6521.

[10] Doi, M., 1996. Introduction to polymer physics. Oxford science publications.

[11] Bower, D. I., 2002. An introduction to polymer physics. Cambridge university press.

[12] Frank, H. S., and M. W. Evans, 1945. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *J. Chem. Phys.* 13:507–532.

[13] Kauzmann, W., 1959. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* 1:1 – 63.

[14] Blokzijl, J. B. F. N. E. W., 1993. Hydrophobic Effects. Opinions and Facts. *Angew. Chem. Int. Ed. Engl.* 32:1545–1579.

[15] Ball, P., 2008. Water as an Active Constituent in Cell Biology. *Chemical Reviews* 108:74–108.

[16] Onuchic, J. N., Z. Luthey-Schulten, and P. G. Wolynes, 1997. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* 48:545–600.

[17] Chiti, F., and C. M. Dobson, 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–366.

[18] Sunde, M., L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, and C. C. Blake, 1997. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* 273:729–739.

[19] Jarrett, J. T., and P. T. Lansbury, 1993. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* 73:1055–1058.

[20] Foffi, G., G. D. McCullagh, A. Lawlor, E. Zaccarelli, K. A. Dawson, F. Sciortino, P. Tartaglia, D. Pini, and G. Stell, 2002. Phase equilibria and glass transition in colloidal systems with short-ranged attractive interactions: application to protein crystallization. *Phys. Rev. E* 65:031407.

[21] Harper, J. D., and P. T. Lansbury, 1997. Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* 66:385–407.

[22] Dumetz, A. C., A. M. Chockla, E. W. Kaler, and A. M. Lenhoff, 2008. Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. *Biophys. J.* 94:570–583.

[23] Narayanan, T., and A. Kumar, 1994. Reentrant phase transitions in multicomponent liquid mixtures. *Phys. Rep.* 249:135–218.

[24] Brovchenko, I. V., and A. V. Oleinikova, 1997. Structural changes of the molecular complexes of pyridines with water and demixing phenomena in aqueous solutions. *J. Chem. Phys.* 106:7756–7765.

[25] Winnik, F. M., 1990. Fluorescence studies of aqueous solutions of poly(N-isopropylacrylamide) below and above their LCST. *Macromolecules* 23:233–242.

[26] Luna-Barcenas, G., J. C. Meredith, I. C. Sanchez, K. P. Johnston, D. G. Gromov, and J. J. de Pablo, 1997. Relationship between polymer chain conformation and phase boundaries in a supercritical fluid. *J. Chem. Phys.* 107:10782–10792.

[27] Maeda, Y., T. Nakamura, and I. Ikeda, 2001. Changes in the hydration states of Poly(N-alkylacrylamide)s during their phase transitions in water observed by FTIR spectroscopy. *Macromolecules* 34:1391–1399.

[28] Cooper, G. J., 1994. Amylin compared with calcitonin gene-related peptide: structure, biology, and relevance to metabolic disease. *Endocr. Rev.* 15:163–201.

[29] Jaikaran, E. T., C. E. Higham, L. C. Serpell, J. Zurdo, M. Gross, A. Clark, and P. E. Fraser, 2001. Identification of a novel human islet amyloid polypeptide beta-sheet domain and factors influencing fibrillogenesis. *J. Mol. Biol.* 308:515–525.

[30] Tenidis, K., M. Waldner, J. Bernhagen, W. Fischle, M. Bergmann, M. Weber, M. L. Merkle, W. Voelter, H. Brunner, and A. Kapurniotu, 2000. Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. *J. Mol. Biol.* 295:1055–1071.

[31] Abedini, A., and D. P. Raleigh, 2006. Destabilization of human IAPP amyloid fibrils by proline mutations outside of the putative amyloidogenic domain: is there a critical amyloidogenic domain in human IAPP? *J. Mol. Biol.* 355:274–281.

[32] Green, J., C. Goldsbury, T. Mini, S. Sunderji, P. Frey, J. Kistler, G. Cooper, and U. Aebi, 2003. Full-length rat amylin forms fibrils following substitution of single residues from human amylin. *J. Mol. Biol.* 326:1147–1156.

[33] Mazor, Y., S. Gilead, I. Benhar, and E. Gazit, 2002. Identification and characterization of a novel molecular-recognition and self-assembly domain within the islet amyloid polypeptide. *J. Mol. Biol.* 322:1013–1024.

[34] Nilsson, M. R., and D. P. Raleigh, 1999. Analysis of amylin cleavage products provides new insights into the amyloidogenic region of human amylin. *J. Mol. Biol.* 294:1375–1385.

[35] Abedini, A., and D. P. Raleigh, 2005. The role of His-18 in amyloid formation by human islet amyloid polypeptide. *Biochemistry* 44:16284–16291.

[36] Engel, M. F. M., H. Yigittop, R. C. Elgersma, D. T. S. Rijkers, R. M. J. Liskamp, B. de Kruijff, J. W. M. Hppener, and J. A. Killian, 2006. Islet amyloid polypeptide inserts into phospholipid monolayers as monomer. *J. Mol. Biol.* 356:783–789.

[37] Quist, A., I. Doudevski, H. Lin, R. Azimova, D. Ng, B. Frangione, B. Kagan, J. Ghiso, and R. Lal, 2005. Amyloid ion channels: a common structural link for protein-misfolding disease. *Proc. Natl. Acad. Sci. U. S. A.* 102:10427–10432.

[38] Green, J. D., C. Goldsbury, J. Kistler, G. J. S. Cooper, and U. Aebi, 2004. Human amylin oligomer growth and fibril elongation define two distinct phases in amyloid formation. *J. Biol. Chem.* 279:12206–12212.

[39] Porat, Y., S. Kolusheva, R. Jelinek, and E. Gazit, 2003. The human islet amyloid polypeptide forms transient membrane-active prefibrillar assemblies. *Biochemistry* 42:10971–10977.

[40] Lopes, D. H. J., A. Meister, A. Gohlke, A. Hauser, A. Blume, and R. Winter, 2007. Mechanism of islet amyloid polypeptide fibrillation at lipid interfaces studied by infrared reflection absorption spectroscopy. *Biophys. J.* 93:3132–3141.

[41] Sparr, E., M. F. M. Engel, D. V. Sakharov, M. Sprong, J. Jacobs, B. de Kruijff, J. W. M. Hppener, and J. A. Killian, 2004. Islet amyloid polypeptide-induced

membrane leakage involves uptake of lipids by forming amyloid fibers. *FEBS Lett.* 577:117–120.

[42] Harroun, T. A., J. P. Bradshaw, and R. H. Ashley, 2001. Inhibitors can arrest the membrane activity of human islet amyloid polypeptide independently of amyloid formation. *FEBS Lett.* 507:200–204.

[43] Knight, J. D., and A. D. Miranker, 2004. Phospholipid catalysis of diabetic amyloid assembly. *J. Mol. Biol.* 341:1175–1187.

[44] Mirzabekov, T. A., M. C. Lin, and B. L. Kagan, 1996. Pore formation by the cytotoxic islet amyloid peptide amylin. *J. Biol. Chem.* 271:1988–1992.

[45] Kayed, R., E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman, and C. G. Glabe, 2003. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* 300:486–489.

[46] Green, J. D., L. Kreplak, C. Goldsbury, X. L. Blatter, M. Stolz, G. S. Cooper, A. Seelig, J. Kistler, and U. Aebi, 2004. Atomic force microscopy reveals defects within mica supported lipid bilayers induced by the amyloidogenic human amylin peptide. *J. Mol. Biol.* 342:877–887.

[47] Broglia, R. A., G. Tiana, S. Pasquali, H. E. Roman, and E. Vigezzi, 1998. Folding and aggregation of designed proteins. *Proc. Natl. Acad. Sci. U. S. A.* 95:12930–12933.

[48] Harrison, P. M., H. S. Chan, S. B. Prusiner, and F. E. Cohen, 1999. Thermodynamics of model prions and its implications for the problem of prion protein folding. *J. Mol. Biol.* 286:593–606.

[49] Dima, R. I., and D. Thirumalai, 2002. Exploring protein aggregation and self-propagation using lattice models: phase diagram and kinetics. *Protein Sci.* 11:1036–1049.

[50] Zanuy, D., B. Ma, and R. Nussinov, 2003. Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL. *Biophys. J.* 84:1884–1894.

[51] Zanuy, D., and R. Nussinov, 2003. The sequence dependence of fiber organization. A comparative molecular dynamics study of the islet amyloid polypeptide segments 22-27 and 22-29. *J. Mol. Biol.* 329:565–584.

[52] Nguyen, H. D., and C. K. Hall, 2004. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U. S. A.* 101:16180–16185.

[53] Colombo, G., I. Daidone, E. Gazit, A. Amadei, and A. D. Nola, 2005. Molecular dynamics simulation of the aggregation of the core-recognition motif of the islet amyloid polypeptide in explicit water. *Proteins* 59:519–527.

[54] Wu, C., H. Lei, and Y. Duan, 2004. Formation of partially ordered oligomers of amyloidogenic hexapeptide (NFGAIL) in aqueous solution observed in molecular dynamics simulations. *Biophys. J.* 87:3000–3009.

[55] Wu, C., H. Lei, and Y. Duan, 2005. Elongation of ordered peptide aggregate of an amyloidogenic hexapeptide NFGAIL observed in molecular dynamics simulations with explicit solvent. *J. Am. Chem. Soc.* 127:13530–13537.

[56] Baumketner, A., and J.-E. Shea, 2005. The influence of different treatments of electrostatic interactions on the thermodynamics of folding of peptides. *J. Phys. Chem. B* 109:21322–21328.

[57] Cecchini, M., F. Rao, M. Seeber, and A. Caflisch, 2004. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* 121:10748–10756.

[58] Tsai, H.-H. G., M. Reches, C.-J. Tsai, K. Gunasekaran, E. Gazit, and R. Nussinov, 2005. Energy landscape of amyloidogenic peptide oligomerization by parallel-tempering molecular dynamics simulation: significant role of Asn ladder. *Proc. Natl. Acad. Sci. U. S. A.* 102:8174–8179.

[59] Allen, M. P., and D. J. Tildesley, 1989. Computer Simulation of Liquids. Oxford University Press, USA.

[60] Oren M. Becker, M. W., Benoit Roux, editor, 2001. Computational Biochemistry and Biophysics. CRC.

[61] Mackerell, A. D., 2004. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25:1584–1604.

[62] D. van der Spoel, B. H., E. Lindahl, 2005. Gromacs User Manual version 3.3. `www.gromacs.org`.

[63] Leach, A. R., 2001. Molecular Modelling principles and applications. Prentice Hall.

[64] MacDowell, L. G., P. Virnau, M. Muller, and K. Binder, 2004. The evaporation/condensation transition of liquid droplets. *J. Chem. Phys.* 120:5293–5308.

[65] Swendsen, R., and J. Wang, 1986. Replica Monte Carlo of spin-glasses. *Phys. Rev. Lett.* 57:2607–2609.

[66] Marinari, E., and G. Parisi, 1992. Simulated tempering - a new Monte Carlo scheme. *Europhys. Lett.* 19:451–458.

[67] Hansmann, U. H. E., 1997. Parallel temperaing algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281:140–150.

[68] Sugita, Y., Y. Okamoto, 1999. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314:141–151.

[69] Frenkel, D., and B. Smit, 2001. Understanding Molecular Simulation. Academic Press, 2nd edition.

[70] Furukawa, H., and K. Binder, 1982. Two-phase equilibria and nucleation barriers near a critical point. *Phys. Rev. A* 26:556–566.

[71] Binder, K., 2003. Theory of the evaporation/condensation transition of equilibrium droplets in finite volumes. *Phys. Acta* 319:99–114.

[72] MacDowell, L. G., V. K. Shen, and J. R. Errington, 2006. Nucleation and cavitation of spherical, cylindrical, and slablike droplets and bubbles in small systems. *J. Chem. Phys.* 125:034705.

[73] Nußbaumer, A., E. Bittner, T. Neuhaus, and W. Janke, 2006. Monte Carlo study of the evaporation/condensation transition of Ising droplets. *Europhys. Lett.* 75:716–722.

[74] Nußbaumer, A., E. Bittner, and W. Janke, 2008. Monte Carlo study of the droplet formation-dissolution transition on different two-dimensional lattices. *Phys. Rev. E* 77:041109.

[75] Chiti, F., and C. M. Dobson, 2006. Protein misfolding, functional amyloid, and human disease. *Ann. Rev. Biochem.* 75:333–366.

[76] Jansen, R., W. Dzwolak, and R. Winter, 2005. Amyloidogenic Self-Assembly of Insulin Aggregates Probed by High Resolution Atomic Force Microscopy. *Biophys. J.* 88:1344–1353.

[77] Grudzielanek, S., V. Smirnovas, and R. Winter, 2006. Solvation-assisted Pressure Tuning of Insulin Fibrillation: From Novel Aggregation Pathways to Biotechnological Applications. *J. Mol. Biol.* 356:497–509.

[78] Balbach, J. J., Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, and R. Tycko, 2000. Amyloid fibril formation by A beta 16-22, a seven-residue fragment of the Alzheimer's beta-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry* 39:13748–13759.

[79] Westermark, P., U. Engstrm, K. H. Johnson, G. T. Westermark, and C. Betsholtz, 1990. Islet amyloid polypeptide: pinpointing amino acid residues linked to amyloid fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* 87:5036–5040.

[80] Azriel, R., and E. Gazit, 2001. Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. *J. Biol. Chem.* 276:34156–34161.

[81] Gustavsson, A., U. Engstrm, and P. Westermark, 1991. Normal transthyretin and synthetic transthyretin fragments form amyloid-like fibrils in vitro. *Biochem. Biophys. Res. Commun.* 175:1159–1164.

[82] Balbirnie, M., R. Grothe, and D. S. Eisenberg, 2001. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. U. S. A.* 98:2375–2380.

[83] Simmerling, C., B. Strockbine, and A. E. Roitberg, 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124:11258–11259.

[84] Lindahl, E., B. Hess, and D. van der Spoel, 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* 7:306 – 317.

[85] Essmann, U., L. Perera, M. L. B. andTom Darden, H. Lee, and L. G. Pedersen, 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577 – 8593.

[86] Serpell, L. C., 2000. Alzheimer's amyloid fibrils: structure and assembly. *Biochim Biophys Acta* 1502:16–30.

[87] Enright, M. B., and D. M. Leitner, 2005. Mass fractal dimension and the compactness of proteins. *Phys. Rev. E* 71:011912.

[88] Cubellis, M. V., F. Cailliez, and S. C. Lovell, 2005. Secondary structure assignment that accurately reflects physical and evolutionary characteristics. *BMC Bioinformatics* 6 Suppl 4:S8.

[89] Brovchenko, I., A. Geiger, and A. Oleinikova, 2004. Clustering of water molecules in aqueous solutions: Effect of water-solute interaction. *Phys. Chem. Chem. Phys.* 6:1982–1987.

[90] Fisher, M. E., 1967. The theory of equilibrium critical phenomena. *Rep. Prog. Phys.* 30:615–730.

[91] Oleinikova, A., I. Brovchenko, A. Geiger, and B. Guillot, 2002. Percolation of water in aqueous solution and liquid–liquid immiscibility. *J. Chem. Phys.* 117:3296–3304.

[92] Pangali, C., M. Rao, and B. J. Berne, 1979. A Monte Carlo simulation of the hydrophobic interaction. *J. Chem. Phys.* 71:2975–2981.

[93] Smith, D. E., and A. D. J. Haymet, 1993. Free energy, entropy, and internal energy of hydrophobic interactions: Computer simulations. *J. Chem. Phys.* 98:6445–6454.

[94] Li, J.-L., R. Car, C. Tang, and N. S. Wingreen, 2007. Hydrophobic interaction and hydrogen-bond network for a methane pair in liquid water. *Proc. Natl. Acad. Sci. USA* 104:2626–2630.

[95] Wallqvist, A., 1991. Molecular dynamics study of a hydrophobic aggregate in an aqueous solution of methane. *J. Phys. Chem.* 95:8921–8927.

[96] Raschke, T. M., J. Tsai, and M. Levitt, 2001. Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutes in water. *Proc. Natl. Acad. Sci. USA* 98:5965–5969.

[97] Zheng, J., B. Ma, C.-J. Tsai, and R. Nussinov, 2006. Structural Stability and Dynamics of an Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup-35. *Biophys. J.* 91:824–833.

[98] Rohrig, U. F., A. Laio, N. Tantalo, M. Parrinello, and R. Petronzio, 2006. Stability and Structure of Oligomers of the Alzheimer Peptide Abeta16-22: From the Dimer to the 32-Mer. *Biophys. J.* 91:3217–3229.

[99] Nguyen, P. H., M. S. Li, G. Stock, J. E. Straub, and D. Thirumalai, 2007. Monomer adds to preformed structured oligomers of Abeta-peptides by a two-stage dock-lock mechanism. *Proc. Natl. Acad. Sci. USA* 104:111–116.

[100] Wilson, C. A., R. Doms, and V. M.-Y. Lee, 1999. Intracellular APP processing and Abeta production in Alzheimer disease. *J. Neuropathol. Exp. Neurol.* 58:787–794.

[101] Oddo, S., A. Caccamo, A. F. Smith, K. N. Green, and F. M. LaFerla, 2006. A Dynamic Relationship between Intracellular and Extracellular Pools of Abeta. *Am. J. Pathol.* 168:184–194.

[102] Hanabusa, T., K. Kubo, C. Oki, Y. Nakano, K. Okai, T. Sanke, and N. K., 1992. Islet amyloid polypeptide (IAPP) secretion from islet cells and its plasma concentration in patients with non-insulin-dependent diabetes mellitus. *Diabetes Res. Clin. Pract.* 15:89–96.

[103] Ida, N., T. Hartmann, J. Pantel, J. Schroder, R. Zerfass, H. Forstl, R. Sandbrink, C. Masters, and K. Beyreuther, 1996. Analysis of heterogeneous betaA4 peptides in human cerebrospinal fluid and blood by a newly developed sensitive Western blot assay. *J. Biol. Chem.* 271:22908–22914.

[104] Lue, L.-F., Y.-M. Kuo, A. E. Roher, L. Brachova, Y. Shen, L. Sue, T. Beach, J. H. Kurth, R. E. Rydel, and J. Rogers, 1999. Soluble Amyloid beta Peptide Concentration as a Predictor of Synaptic Change in Alzheimer's Disease. *Am. J. Pathol.* 155:853–862.

[105] Tienari, P. J., N. Ida, E. Ikonen, M. Simons, A. Weidemann, G. Multhaup, C. L. Masters, C. G. Dotti, and K. Beyreuther, 1997. Intracellular and secreted Alzheimer beta-amyloid species are generated by distinct mechanisms in cultured hippocampal neurons. *Proc. Natl. Acad. Sci. USA* 94:4125–4130.

[106] Echeverria, V., and A. C. Cuello, 2002. Intracellular A-beta amyloid, a sign for worse things to come? *Mol. Neurobiol.* 26:299–316.

[107] Jorgensen, W. L., J. Chandreskhar, J. D. Madura, R. W. Impey, and M. L. Klein, 1983. *J. Chem. Phys.* 79:926.

[108] Paschek, D. Volume,Temperature-Replica Exchange Molecular Dynamics with the GROMACS 3.2.1 Simulation Package. RPMDRUN - A mini-HOWTO. `http://ganter.chemie.uni-dortmund.de/pas`.

[109] Fisher, M., 1967. The theory of condensation and critical point. *Physics* 3:255–283.

[110] Brovchenko, I., R. R. Burri, A. Krukau, A. Oleinikova, and R. Winter, 2008. Intrinsic thermal expansivity and hydrational properties of amyloid peptide Aβ42 in liquid water. *J. Chem. Phys.* 129:(in press).

[111] Mitra, L., A. Oleinikova, and R. Winter. Intrinsic volumetric properties of trialanine isomers in aqueous solution. *ChemPhysChem* (in press).

[112] Krukau, A., I. Brovchenko, and A. Geiger, 2007. Temperature-induced conformational transition of a model elastin-like peptide GVG(VPGVG)$_3$ in water. *Biomacromolecules* 8:2196–2202.

[113] Malmsten, M., editor, 2003. Biopolymers at interfaces. Marcel Dekker, Inc.

[114] Czeslik, C., 2004. Factors ruling protein adsorption. *Z. Phys. Chem.* 218:771–801.

[115] Glover, J. R., A. S. Kowal, E. C. Schirmer, M. M. Patino, J. J. Liu, and S. Lindquist, 1997. Self-seeded fibers formed by Sup35, the protein determinant of [PSI+], a heritable prion-like factor of S. cerevisiae. *Cell* 89:811–819.

[116] Kyte, J., and R. F. Doolittle, 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132.

[117] Jorgensen, W. L., and J. Tirado-Rives, 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.

[118] Kaminski, G., R. Friesner, J. Tirado-Rives, and W. Jorgensen, 2001. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* 105:6474–6487.

[119] Berendsen, H. J. C., J. R. Grigera, and T. P. Straatsma, 1987. The missing term in effective pair potentials. *J. Phys. Chem.* 91:6269–6271.

[120] ln Chul Yeh, and M. L. Berkowitz, 1999. Ewald summation for systems with slab geometry. *J. Chem. Phys.* 111:3155.

[121] Berendsen, H. J. C., J. Postma, A. DiNola, and J. Haak, 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

[122] Nose, S., 1984. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 52:255–268.

[123] Hoover, W., 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* 31:1695–1697.