Understanding amyloidogenesis through computer simulations
Tackling the sampling problem by using replica exchange umbrella sampling and hydrogen bond biasing potentials

Maarten Wolf
The cover drawing is made by Tjarko van der Pol.

The research reported in this thesis was carried out at DelftChemTech, University of Technology Delft, Julianalaan 136, Delft, The Netherlands.

This work was supported for funding by The Netherlands Organization for Scientific Research as part of the computational life science program (grantnumber 635.100.012).
Understanding amyloidogenesis through computer simulations
Tackling the sampling problem by using replica exchange umbrella sampling and hydrogen bond biasing potentials

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de
Rector Magnificus prof. dr. ir. J.T. Fokkema,
voorzitter van het college voor promoties,
in het openbaar te verdedigen op
vrijdag 20 maart 2009
om 10.00 uur

door

Maarten Gijsbrecht WOLF
doctorandus in de scheikunde

geboren te Heemskerk
Dit proefschrift is goedgekeurd door de promotoren:
Prof. dr. S.W. de Leeuw
Prof. dr. J.D. Laman

copromotor:
Dr. J.A. Jongejan

Samenstelling promotiecommissie:
Rector Magnificus, voorzitter
Prof. dr. S.W. de Leeuw, Technische Universiteit Delft, promotor
Prof. dr. J.D. Laman, Erasmus Universiteit, promotor
Dr. J.A. Jongejan, Technische Universiteit Delft, copromotor
Prof. D. van der Spoel, Uppsala University
Prof. P.G. Bolhuis, Universiteit van Amsterdam
Prof. S.J. Picken, Technische Universiteit Delft
Dr. F. Meersman, Katholieke Universiteit Leuven
Preface

"Amyloidogenesis: A computer simulation study", was the title of the project I started more then four years ago as a PhD. In the project description the objective was formulated as follows: "... to determine the mechanisms underlying the aggregation of Aβ peptides into fibrils at atomistic and mesoscopic levels. ...". An ambitious goal for a project of four years. Nevertheless we aimed for the assessment of the mechanisms underlying amyloidogenesis at the atomistic level.

During my PhD research we used two radically different approaches to understand the basis of amyloidogenesis. Therefore this thesis is divided into Parts, with a separate Part for each approach. The resulting two Parts are preceded by an introductory Part and followed by a Part containing a general discussion.

As both approaches discussed in this thesis are based on standard computational methods these will first be discussed. This is followed by a review of amyloidogenesis, where the relevance and the current knowledge is discussed. How we aim to apply or improve the standard computational methods in order to address some of the still unanswered questions regarding amyloidogenesis is discussed in the last section of the introductory Part.

In our first approach we combined two standard computational methods to increase the efficiency of the calculation. Although we would have liked to use the Aβ peptides for our investigations with this combined method, it already quickly became clear that this was infeasible, as the Aβ peptides are too large for efficient atomistic studies. Therefore we switched to small model peptides that were shown to form fibrils with characteristics similar to those of Aβ. With these small peptides we determined thermodynamic properties relevant in fibril growth, which is discussed in Part II of this thesis and published in two papers.

As it is impossible to study mechanistic or kinetic properties at the atomistic level, even with these small peptides, in our second approach we aimed to develop new computational protocols to study the energy landscape of protein folding and aggregation. This is the subject of Part III. Although this already led to a published paper, a submitted paper and an application for a patent, I could not find the time to use these novel methods to determine the mechanisms underlying the aggregation of model peptides into fibrils at the atomistic level. That can be an ambitious goal in another PhD project.

Maarten Wolf, October 2008
Contents

Preface 5

Contents 6

PART I Introduction

1 Computational methods 11
   Interactions between atoms: Force field .......................... 12
   Molecular dynamics ................................................. 14
   Free energy calculation ........................................... 15
   Replica exchange molecular dynamics .............................. 18

2 Amyloidogenesis 21
   Relevance ............................................................ 21
   Amyloid fibril structure ........................................... 26
   Amyloid fibril formation ........................................... 32
   Goals of this thesis ................................................. 38

PART II Predicting amyloid fibril growth properties

3 Rapid free energy calculation by REMD umbrella sampling 43
   Introduction .......................................................... 43
   Method ................................................................. 45
   Results and discussion .............................................. 49
   Concluding remarks ................................................ 55

4 Predicting fibril growth properties of short peptides 57
   Introduction .......................................................... 58
   Results and discussion .............................................. 60
   Method ................................................................. 62
PART I

INTRODUCTION
CHAPTER 1

Computational methods

Abstract In this chapter the computer simulation methods and principles are discussed, that are the basis for the development of the advanced simulation methods in part II and III of this thesis. The subjects that are reviewed include molecular dynamics, expression of the interactions between atoms into equations, free energy calculation using umbrella sampling and replica exchange molecular dynamics.

The first non-military computer became available in the 1950s, allowing the first computer simulation studies [1]. Pioneering work was performed by Metropolis in his effort to study as many different problems as possible, in order to test the logical structure and demonstrate the capability of computer simulations [1]. The work of Metropolis lead to a famous publication, together with Rosenbluth, Rosenbluth, Teller and Teller, describing the Metropolis Monte Carlo (MC) method [2]. Shortly after this paper, Fermi, Pasta and Ulam described a radically different approach in their paper on the investigation of the dynamics of an anharmonic, one-dimensional crystal [3]. Almost simultaneously, Alder and Wainwright reported the first proper molecular dynamics (MD) simulation on a cluster of hard spheres [4]. The common but erroneous assumption that continuous potentials would result in a more complicated and demanding simulation around that time [5, 6] delayed the examination of a real liquid to the mid 60s. Then Rahman performed an MD simulation of liquid argon [7], although a real crystal system was already simulated in the late 50s by Vineyard, on radiation damage of copper crystals [8].

The success of computer simulations originates from the capability to evaluate mathematical equations to any desired accuracy at great speed. As the basic laws of nature are generally expressed in equations that, apart from a few very simple cases, cannot be solved analytically, systems governed by these laws are amenable to evalu-
Chapter 1. Computational methods

ation with computers. A purely analytical theoretical evaluation of the properties and behavior of such systems based on these laws requires simplifications, thereby making the predictions unreliable. However the computer can evaluate these equations numerically to provide an answer in any desired accuracy, given that the interparticle interactions are known.

The studies mentioned in the first paragraph of this chapter showed that through the introduction of the computer, theoretical problems that required heavy calculation could be solved. For example, Vineyard’s MD simulations on the radiation damage in copper crystals were stimulated by the failed attempts of others to assess this effect analytically or perform the calculations by hand [1, 9]. Although the low computer power back then limited these numerical studies to very small systems, the doubling of computer power every two years (Moore’s Law [10]) eventually lead to studies of more complicated systems, such as biomolecular systems.

Interactions between atoms: Force field

In biomolecular systems, such as proteins in solution (as is the case for amyloidogenesis), the behavior of the system is mainly determined by the physico-chemical interactions between all the atoms. A viable approach is to consider the atoms in a biomolecular system within the framework of classical mechanics, i.e. the movement of the atoms are defined by the classical laws of physics, as opposed to the quantum laws. In order to evaluate this movement by computer simulation the interactions between the atoms need to be expressed in equations that can be solved numerically. Therefore all the atomic interactions are described by knowledge based potential energy functions, also known as force field.

Two different types of interactions between the atoms can be distinguished, viz. bonded and non-bonded. The bonded interactions are specific to molecules and describe the interactions between atoms linked through one or more covalent bonds in one molecule. As the electron density is not explicitly considered in classical simulations at the atomistic level, bond length, bond angle and dihedral angle potentials ($V_{\text{bond}}(r_{ij})$, $V_{\text{angle}}(\theta_{ijk})$ and $V_{\text{dihedral}}(\zeta_{ijkl})$, respectively (Figure 1.1)) are required to describe the vibrations and rotations of covalent bonds. These potentials maintain the correct configuration of a set of atoms connected through covalent bonds, or in other words ensure that the molecule retains a physically relevant structure. Another consequence of neglecting the electron density is that bonds cannot be broken or formed in the course of a simulation, rather they are assigned to the atoms at the start of a simulation.

As hydrogen bonds are also a consequence of the electron density, namely the high density on the acceptor and the low density on the donor, to model the specific interaction during hydrogen bonding at atomistic detail requires special attention. In contrast to covalent bonds, the number of hydrogen bonds as well as the atoms that are involved in hydrogen bonds is highly variable in time. Therefore it is not an option to assign the hydrogen bonds at the beginning of the simulation. As hydrogen bonds are
Interactions between atoms: Force field

![Diagram of interactions](image)

**Figure 1.1 — Interactions in a force field and the relevant equations.** On the left the bonded interactions (bond length, angle and dihedral) are depicted. Here \( r_{ij} \) is the distance between atom \( i \) and \( j \), \( \theta_{ijk} \) the angle spanned by atoms \( i, j \) and \( k \), \( \zeta_{ijkl} \) the dihedral angle of atoms \( i, j, k \) and \( l \), the subscript \( r \) refers to the reference value of these variables, \( \zeta_s \) to the multiplicity of the dihedral angle and \( f_c \) is the force constant associated with the specific potential. On the right the non-bonded interactions (Lennard-Jones and Coulomb) are given, with \( C_{12} \) and \( C_6 \) constants of the Lennard-Jones potential, \( q_i \) the charge of atom \( i \) and \( \epsilon \) the dielectric constant.

also an electrostatic effect, the hydrogen bond behavior is achieved by placing partial charges on the atoms involved, e.g. N–H is charged \( \delta^-–\delta^+ \). This results in an effective dipole, which drives the hydrogen bonding.

The non-bonded interactions describe the interactions between all atoms not involved in bonded interactions. These non-bonded interactions result from the laws of physics and consist of a dispersion, a repulsive and a charge interaction. The combination of the dispersion and the repulsive interaction is generally known as the Vander-Waals interaction and is expressed as for example the Lennard-Jones potential \( V_{LJ}(r_{ij}) \) (Figure 1.1)). The charge interaction is expressed as the Coulomb term \( V_{Coul}(r_{ij}) \) (Figure 1.1)).

The total potential energy function \( V(r) \), with \( r \) denoting all particle positions, is then obtained by summing all possible bonded and non-bonded interactions in the system.

\[
V(r) = \sum_{<i,j>} V_{bond}(r_{ij}) + \sum_{<i,j,k>} V_{angle}(\theta_{ijk}) + \sum_{<i,j,k,l>} V_{dihedral}(\zeta_{ijkl}) + \sum_{i,j} V_{LJ}(r_{ij}) + \sum_{i,j} V_{Coul}(r_{ij})
\]  

(1.1)

Here, the sum over \( V_{bond}(r_{ij}) \) is over all neighboring pairs \( i, j \), i.e. all covalently bonded atoms \( i \) and \( j \). Similarly the sums over \( V_{angle}(\theta_{ijk}) \) and \( V_{dihedral}(\zeta_{ijkl}) \) are over all neighboring atoms \( i, j, k \) (two consecutive covalent bonds) and \( i, j, k, l \) (three consec-
utive covalent bonds), respectively. The sum over $V_{LJ}(r_{ij})$ and $V_{Coul}(r_{ij})$ is over all non-neighboring pairs $i, j$, i.e. all pairs of atom $i$ and $j$ that are separated by more than three covalent bonds.

All these interactions contributing to the potential energy lead to forces acting on the atoms. These forces are obtained by differentiating the potential energy with respect to the position and the force in the $x$-direction on atom $i$, $F_{xi}$, is then the partial derivative of the potential energy to the $x$-coordinate of atom $i$, $x_i$:

$$F_{xi} = \frac{\partial V(r)}{\partial x_i} \tag{1.2}$$

Similar equations have to be solved for the force in the $y$- and $z$-direction. Depending on the method used to explore the properties and behavior of a system the potential energy (for Monte Carlo) or the forces (for molecular dynamics) are required.

**Molecular dynamics**

Molecular dynamics (MD) is a method to evaluate the movement resulting from the atomic interactions by integrating equations of motion. In general MD can be used to evaluate the motion of any particle system. A convenient approach is to integrate the equation of motion derived by Newton in his second law (eq 1.3).

$$F_{xi}(t) = m_i \cdot \frac{d^2 x_i(t)}{dt^2} \tag{1.3}$$

Here $x_i(t)$ and $F_{xi}(t)$ denotes the position and the force acting on atom $i$ in the $x$-direction at time $t$ and $m_i$ is the mass of atom $i$.

In the special case that the velocity and the forces (and hence also the acceleration) are constant in time, e.g. one particle in a homogeneous field, the development of the atom positions in time can be solved analytically, which results in

$$x_i(t) = x_i(0) + v_{xi} t + \frac{1}{2} a_{xi} t^2 \tag{1.4}$$

with $x_i(0)$ the position in the $x$-direction of atom $i$ at time $t = 0$, $v_{xi}$ the velocity of atom $i$ and $a_{xi}$ the acceleration of atom $i$ in the $x$-direction.

However, generally in a many body system, i.e. a system composed of many particle, such as a protein solution the velocity and forces are not constant in time. For instance a collision between two atoms or constraints as a consequence of covalent bonds will give rise to a change in the velocities and the forces of these atoms. It is reasonable to assume that for a very small time step $\Delta t$ these values do not change dramatically and consequently equation 1.4 can be used to calculate the new atom positions after $\Delta t$. Subsequently the forces and velocities must be updated for the new situation before equation 1.4 can be used again for another small time step. Mathematically, this rationale is followed by omitting the terms of order $O(\Delta t^4)$ in the Taylor expansion of equation 1.3, which will result in a comparable equation of motion that can be solved successively (Verlet algorithm) [1].
Free energy calculation

In GROMACS [11], the MD program used for the work discussed in this thesis, an alternative algorithm to perform this operation is used, namely the leap-frog algorithm [12].

\[
v_{xi}(t + \frac{1}{2} \Delta t) = v_{xi}(t - \frac{1}{2} \Delta t) + \frac{F_{xi}(t)}{m_i} \Delta t \tag{1.5}
\]

\[
x_i(t + \Delta t) = x_i(t) + v_{xi}(t + \frac{1}{2} \Delta t) \Delta t \tag{1.6}
\]

Equations 1.6 and 1.5 are solved successively to simulate the motion of the atoms (trajectory). From this trajectory the properties of the system over a certain time interval can be measured, similar to a real experiment.

Although the early simulations were all performed with a constant total energy, something not easily done in experiments, nowadays the properties of a simulation can be varied to match the experimental conditions. For example, the computations can simulate a system at room temperature or at ambient pressure. For the majority of the work described in this thesis the simulations are performed with a constant number of particles (N) in a box of constant volume (V) at room temperature (T), representing a NVT ensemble.

It is straightforward to extract the dynamical properties of a system from a MD simulation, like the diffusion constant through the Einstein and Stokes-Einstein relations [1]. Time independent properties can also be obtained from time dependent trajectories, if the trajectories are sufficiently long and the snapshots used for the measurements are uncorrelated, i.e. separated by sufficient time to evolve. This statement is known as the ergodic hypothesis and allows the extraction of thermodynamic properties like the heat capacity and free-energy differences from MD simulations [1]. For example the internal energy of a system can be calculated by taking the ensemble average of the total energy of the system (eq 1.7). Within a simulation this simply results in taking the average over a sufficient number of uncorrelated snapshots (eq 1.8).

\[
E_{int} = \frac{1}{t_{tot}} \int_{0}^{t_{tot}} E(t) dt \tag{1.7}
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} E(t_i) \tag{1.8}
\]

Here $E_{int}$ is the internal energy of the system, $E$ the total energy, $t_{tot}$ the total sampling time, $N$ the number of independent snapshots and $t_i$ the time of snapshot $i$.

Thus, with molecular dynamics the natural evolution in time of a system is obtained by integrating equations of motion. Various experimental conditions can be simulated and both dynamical and time-dependent properties can be extracted from the resulting trajectory.

Free energy calculation

Energy landscape and potential of mean force

The properties of biomolecular structures are conveniently discussed in terms of the energy landscape, which depicts changes in the energy as transitions take place. In
other words, the energy is considered as a function of the relevant "reaction" coordinates describing the transition. The complexity of biomolecular systems is reflected in the free-energy landscape: it often takes the form of a rugged landscape with many mountain passes and valleys. Stable and metastable states of a biomolecular system, frequently characterized by distinct structures, can be identified as valleys in this landscape. The equilibrium population of these valleys is obtained directly from their relative depths.

Since the complete energy landscape contains the barriers between different states this landscape provides information on the rate of transition between these states as well. The transition from one state to another, e.g. unfolded to folded, can go through any path, however the path of least resistance will be followed most frequently (Figure 1.2). For example, analogous to transitions in a biomolecular landscape, when going from Gerlos to Mayerhofen the path through the valleys passing Zellberg (path 1) is preferred over the path in a straight line (path 2), crossing the mountains. In general the width and the depth of the valley determines the average population, with a wider and deeper valley having a larger population. In Figure 1.2 this is clear from the lower left corner, which is the widest and deepest valley and also the most populated.

The average free-energy along such a reaction path connecting two states is known as the potential of mean force (PMF). In other words, the potential of mean force shows the free-energy barriers on a reaction path and a system moving along this path will have to climb or descend these barriers. The derivative of the potential of mean force on a point of this path will give the average force the system experiences at this point.

**Umbrella sampling**

An important property of any biomolecular system is the relative stability of different structures (states), which is determined by the free-energy difference. In many cases only the free-energy difference between two states is of interest and it suffices to sample only the relevant states adequately, without sampling the rest of phase space.

Computationally the free-energy difference between state A and B can be obtained
Free energy calculation

Figure 1.3 — Probability distributions of sampling state A and B. a The probability distribution of a simulation started in state A, which does not sample state B. b Probability distributions of different umbrella windows, where the biased simulations (dashed line) ensure accurate sampling of the path connecting states A and B. c The unbiased probability distribution obtained by combining the probability distributions in b. From this probability distribution the potential of mean force over the reaction coordinate $\xi$ (equation 1.12) as well as the free-energy difference between state A and B can be calculated (equation 1.9).

from the probability distribution along the relevant coordinates linking these states. This is because the free-energy difference $\Delta G$ between two states A and B determines the population of these two states $p_A$ and $p_B$ in thermodynamic equilibrium. In equilibrium the free-energy difference is then given by

$$\Delta G = k_B T \ln \frac{p_A}{p_B}$$

(1.9)

with $k_B$ Boltzmann’s constant and $T$ the temperature.

However, due to barriers in the free-energy landscape, state B is generally not well sampled by the probability distribution around state A, and vice versa (Figure 1.3a). Hence, the ratio $p_A/p_B$ is either intractable ($p_B = 0$) or subject to a large statistical error ($p_B << p_A$). Equal sampling of state A and B in one simulation can be attained by introducing a biasing potential $V_{bias}$ (or umbrella potential) to the system. After correcting for the biasing potential the free-energy difference can be calculated with sufficient accuracy [13].

A single simulation in which both state A and B are sampled, due to the addition of the umbrella potential, is not very efficient. Therefore a general umbrella sampling involves a path that connects two states, which is explored exhaustively. To insure efficient sampling of the path a number of copies of the system are simulated. Each copy only differs by a different umbrella potential that enhances sampling around a specific reaction coordinate, $\xi$, on the path (Figure 1.3b). Since the free-energy difference is a state function the path can be chosen at random, but a wise choice can shorten the calculations.

A given biased simulation results in a small sampling window that is sufficiently accurate to calculate the potential of mean force (Appendix 1) along that part of the
path. To generate a potential of mean force of the entire path the results of the various windows need to be combined (unbiased). This can be done efficiently by solving equations of the weighted histogram analysis method (WHAM) [14, 15].

\[
\rho^u(\xi) = \sum_{i=1}^{N_{rep}} \frac{\rho^b_i(\xi)}{\sum_{j=1}^{N_{rep}} e^{-\beta(V_{bias,j}(\xi) - f_i)}}
\]

(1.10)

\[
e^{-\beta f_i} = \int dr e^{-\beta V_{bias,i}(\xi)} \rho^u(\xi)
\]

(1.11)

In these equations \(\xi\) is the reaction coordinate connecting the two states A and B, \(N_{rep}\) the number of umbrella windows and \(f_i\) a constant that represents the free energy associated with the introduction of the biasing potential of replica \(i\). With equations 1.10 and 1.11 the biased probability density, \(\rho^b_i(\xi)\), from the different umbrella windows are combined into one unbiased probability density, \(\rho^u(\xi)\), over the complete path (Figure 1.3c). The unbiased probability density is connected to the potential of mean force, \(W(\xi)\) through

\[
W(\xi) = -k_B T \ln \rho^u(\xi)
\]

(1.12)

The relation of the potential of mean force to the free-energy difference strongly depends on the system and will be discussed explicitly in Chapter 3 and 4.

**Replica exchange molecular dynamics**

In general, the transition from one valley in the free-energy landscape to another is the result of thermal motion. However, as the barriers separating these valleys are generally higher than the thermal energy \(k_B T\), transitions are a rare event and only one free-energy minimum is sampled in a limited amount of time. Replica exchange molecular dynamics (REMD) or parallel tempering is a method devised to increase the number of transitions between the valleys in the free-energy landscape during molecular dynamics simulations [1, 16–18], while maintaining the elaborate sampling of the free-energy minima characteristic to MD.

A REMD simulation consists of a number of copies of the same system, where every copy is subjected to a small change in one property, like an environmental condition (e.g. temperature) or an interaction function (e.g. VanderWaals interactions). At least one of these copies is simulated under the conditions of interest, while in the other copies the change in this property is chosen such that the rate of barrier crossing from one free-energy minimum to another is enhanced.

The copies of the system are allowed to evolve for a short time period under these different conditions (in the order of a picosecond), after which attempts are made to swap the conditions of two copies in a Monte Carlo scheme.

\[
p_{acc}(i, j) = \begin{cases} 
1 & \Delta(i, j) \leq 0 \\
e^{-\Delta(i, j)} & \Delta(i, j) > 0
\end{cases}
\]

(1.13)

\[
\Delta(i, j) = \beta_j(V_j(r_i) - V_j(r_j)) + \beta_i(V_i(r_j) - V_i(r_i))
\]

(1.14)
Replica exchange molecular dynamics

Figure 1.4 — Temperature REMD scheme of a protein system. a The three copies of the system at different temperatures are simulated for a short period of time (commonly 1 ps). b $N \cdot (N - 1)$ attempts to exchange the temperature between the three copies are made, with $N$ the number of replicas and the arrow indicating an all-pair exchange. In this example the exchange between $T=400$ and $T=350$ is accepted. c After the exchange the systems are simulated with the new temperature. At $T=400$ barriers are crossed easily and the protein is unfolded. At $T=350$ a new folded structure is found. At $T=300$ the valley associated with the initial folded structure is explored. d Another exchange attempt, where the exchange between $T=300$ and $T=350$ is accepted. e The systems are simulated at the new temperatures. Now the valley of the second folded conformation is explored at $T=300$.

Here $p_{acc}$ is the acceptance probability and $\beta = 1/k_B T$, where $k_B$ is Boltzmann's constant and $T$ the temperature. This exchange depends on the change in total energy upon swapping the difference in properties between system $i$ and $j$, i.e. when the particle positions currently subjected to the conditions of $i$ and $j$ are subsequently subjected to the conditions of $j$ and $i$, respectively.

For example in temperature REMD (T-REMD), which is the first and a very popular REMD method, the copies of the system differ in temperature. By letting the replicas exchange temperatures, according to a Metropolis criterion, a system can diffuse through temperature space and overcome barriers at high temperature, while sampling the interesting stable states at the temperature of interest (as illustrated in Fig. 1.4).

The free energy can be obtained by constructing a histogram along any relevant reaction coordinate, i.e. the probability distribution can be calculated along a reaction path of interest. Nevertheless, by migrating through different replicas the information on the dynamics is lost, but the thermodynamic properties, such as free-energy differences, can thus be extracted from a REMD simulation.
Amyloidogenesis

Abstract Amyloidogenesis is the process of amyloid formation. Amyloid structures occur in many living organisms, sometimes linked to pathological conditions and sometimes crucial for survival. Amyloid structures are composed of many units of the same peptide. The macroscopic pictures of these structures are rigid fibrils, with a specific width and of indefinite length. The ordered nature of these structures is a result of the cross-β-sheet interaction in the direction of the fibril axes and lateral interactions perpendicular to the fibril axes mediated by the peptide side chains. This chapter will start with examples of amyloidogenesis to illustrate the relevance. Then the structure and formation of amyloid structures will be reviewed, with a focus on the Alzheimer Aβ-peptide. Finally the aims of the research described in this thesis will be discussed.

Relevance

Amyloid fibrils are associated with many pathological conditions in humans (see Table 2.1, although some living organisms depend on these fibrils for their survival. Recently, interest in amyloid fibrils has been raised from the field of nano-technology and materials science.

Pathological conditions: Alzheimer disease as a case study

A person suffering from Alzheimer disease slowly returns to the intellectual level of a newborn child. In this process three stages can be distinguished in which patients spend from several months to several years. The first stage is characterized by a mild memory deficiency and change of personality. In the second stage a person starts to experience physical difficulties, such as the inability to perform complex motorical tasks.
In the third stage the intellectual functions of the person completely disappear and in the end persons living this long will adopt the fetal position. [20, 21]

In 1906 Alois Alzheimer, a German psychiatrist and neuropathologist studying the conditions and diseases of the brain, gave his famous lecture on this affliction. He described a woman in her fifties suffering from an unusual disease of the cerebral cortex, causing memory loss, disorientation, hallucinations and ultimately her death aged only 55. The pathological-anatomical investigation of the brain showed various abnormalities, such as a thinner cerebral cortex (Figure 2.1) and accumulation of senile plaques and neurofibrillary tangles. Using the latest staining techniques Alzheimer was the first to show that these so far unidentified senile plaques were amyloid structures [22].

The presence of these amyloid structures was suggested to be the causative agent of the disease. This postulate was further supported by various experiments showing that these amyloid structures are toxic to cultured neuronal cells [23–26] and that injection of amyloid structures into the cerebral cortex of aged rhesus monkeys leads to neuronal loss [27]. More recently, the possibility of another causative agent is being explored. Experiments show compelling evidence that not the mature amyloid structures, but precursors to the amyloid fibril, such as low-molecular-weight oligomers and/or structured protofilaments are the pathological agent [28–31].

The relation of amyloid fibrils and its precursors to pathological conditions, like...
Alzheimer disease, has resulted in widespread attention for this class of structures [29, 32]. More than 20 different diseases are known to be associated with amyloid structures (see Table 2.1). However, currently the available therapeutic agents focus on alleviating the symptoms, rather than targeting the cause of the disorder [33]. In order to develop new therapeutic agents that do attack the molecular mechanism underlying amyloid diseases, a clear understanding of this process is essential.
Chapter 2. Amyloidogenesis

Nano-technology

The small size and highly ordered nature of amyloid structures are favorable attributes for application in nano-technology, where the term nano refers to materials with a size smaller than 100 nm. One nanometer is one-billionth ($10^{-9} = 0.000,000,001$) meter. For example, a sheet of paper is 100,000 nanometers thick and a nanometer is the length a man's beard grows in the time it takes him to raise the razor to his face [35].

Nano-materials have a wide range of applications. To name a few: There are many applications within medicine, as the size of nano-materials is similar to most biological molecules and structures. This makes them very suitable to deliver drugs or as diagnostic devices. Also in electronics nano-materials are becoming increasingly important. For years the speed of computers was doubled every two years, by cramming more processing units (transistors) on a chip. Consequently, these processing units had to be reduced in size, which currently extends into the nano-materials size range [36].

In general manufacturing nano-materials presents a formidable challenge. An exception to this rule are self-assembling systems, such as amyloid fibrils. Nevertheless, as the primary effort in nano-technology is directed at electronic devices, and as amyloid fibrils have poor electronic properties, applications of amyloid fibrils as a nano-material are limited.

Alternatively, amyloid fibrils can be useful tools when manufacturing nano-materials, as has been demonstrated under laboratory conditions with promising results. The amyloid fibrils are then used as a scaffold for the deposition of another substance. This substance can be selected for a specific required quality, to construct a nano-material with a desired property. An example of such a process is the adhesion of gold and silver particles to a cysteine labeled amyloid fibril, resulting in gold nano-wires with a diameter in the order of 100 nm (Figure 2.2) [34]. These gold nano-wires have excellent electronic properties and for instance can be used in nano-electronic applications.
Relevance

Gland Silk type
Tubuliform Egg-sac silk
Ampullate (major) Thick drag line silk for web frame and safety line
Ampullate (minor) Thin drag line silk for web reinforcement
Flagelliform Prey-capture thread
Aggregate Strong-adhesive coating the prey-capture thread
Aciniform Swathing silk for wrapping and immobilizing prey
Pyriform bind material or cement threads to a substrate

Figure 2.3 — Glands and silk types found in spiders. The picture shows the silk producing glands and their products in the orb-weaver Araneus diadematus. In the table all the available silk glands in spiders and the type of silk they produce are listed. Picture taken from [39].

circuits [34,37,38]. Although these results are very promising, the nano-wire still has to be connected to the appropriate devices of the electronic circuits and the manufacturing process has to be adapted to industrial conditions, before any practical application can be expected. Therefore an excellent understanding of the amyloid fibril structure and formation is vital.

Spider silk, an archetype for material scientists

All spiders descend from one common ancestor, with the ability to spin silk [40]. This ability has proved so successful that today there are many different spider species, with
Chapter 2. Amyloidogenesis

40,000 species characterized to date [41]. As new species developed also new uses evolved for the silk and today there are seven different types of glands and spider silk (Figure 2.3) [39, 42, 43]. In addition a spider has the remarkable capability to alter the properties of the silk fibril as needed. For instance a spider adjusts the strength of its drag line to its own weight, in order to make sure it will not break, but at the same time reducing the resource cost of manufacturing the fiber [40].

Spider silk is an amyloid fibril that is secreted from special glands in a spider. All spiders share this ability and although some may use this ability sparsely, the majority depends on it to build nests, set traps or build cocoons. There are seven different spider silks known, each with specific properties optimized for its intended use (see Figure 2.3) [39, 40, 42–44]. For example the orb-weaver Araneus diadematus (the European garden spider) uses drag line silk for the web’s spokes and outer rim and capture-spiral silk for the wheels of the web. The drag line silk is as strong as steel (strength of 1.3 GPa [45]) and provides strength and rigidity to the web, while the capture-spiral silk is sticky, stretchy and tough to capture the prey.

The specific biochemical properties of spider silk, such as strength and elongation before breaking, makes these fibers interesting materials for the chemical engineer. Also appealing is the apparent ease with which spiders produce silk, while operating at room temperature, low pressure and using water as a solvent, which are attractive conditions for manufacturing industrial fibers. In addition the ability to modulate the properties of a fibril will prove valuable in optimizing and improving fiber materials. Finally, as spiders use renewable materials, these amyloid fibrils seem promising for the future.

Despite the many advantages of spider silk, its practical exploitation to date is fairly limited, mainly because spiders cannot be domesticated and it is extremely difficult to produce spider silk industrially. Recently, DNA-recombination techniques have been developed to generate spider silk in high yields, providing a promising method for the future utilization of spider silk [46]. Clearly, understanding the detailed structure and assembling mechanism of amyloid fibrils is a prerequisite for further improvement of their properties and for the development of new applications.

Amyloid fibril structure

In 1854 a German physician, Rudolph Virchow, performed tests on cerebral corpora amylacea that had an abnormal macroscopic appearance [47]. The diagnostic tests on these structures involved iodine staining and the tissue stained pale blue. As this color was usually associated with starch or cellulose (amylum in Latin and amylon in Greek) these structures where named amyloid. As a consequence of the limited specificity of iodine staining, there has been a long discussion on the nature of the building blocks of these amyloid structures. Popular theories involved high fatty acid or carbohydrate content. Finally, the observation of high nitrogen content of these structures lead to the conclusion that the primary building blocks of amyloid structures are peptides [47, 48].

For many years a high-resolution molecular level description of amyloid fibrils was
not available. The only structural information was provided at the macroscopic level by low-resolution imaging techniques, such as electron microscopy, atomic force microscopy and X-ray fiber diffraction. High-resolution imaging techniques could not be employed as amyloid fibrils are not crystalline, excluding X-ray crystallography, and too large for solution state NMR and computer simulations. This situation has changed significantly in the last five years. A major contribution was the development of solid state NMR [49] and the increase in computer power allowing computer simulations of larger systems [50, 51]. Also hydrogen-deuterium exchange [52] as well as proline [53] and cysteine [54] mutation techniques have provided valuable information on the structure and mechanisms of amyloid fibril assembly.

Macroscopic picture

The macroscopic picture of amyloid fibrils displays a large degree of uniformity, irrespective of the specific primary amino acid sequence of the peptide building block. The images of amyloid fibrils from electron microscopy (EM) show elongated unbranched fibrils, with a width of 7-12 nm (Figure 2.4 a). The length of an amyloid fibril is undefined and depends on solution and preparation conditions. Usually amyloid fibrils consist of a number (typically 2-6) of fibrillar substructures, protofilaments, that intertwine to form rope-like fibrils (Figure 2.4 c) [29, 32, 47, 55–57].

The two major reflections observed in the X-ray fiber diffraction pattern of amyloid fibril structures (Figure 2.4 b) indicate a strong \( \beta \)-sheet character. The reflections at 4.7 Å correspond to the hydrogen bond distance in a \( \beta \)-sheet and that at \( \sim 10 \) Å to the spacing between different \( \beta \)-sheets. The value of the latter reflection varies, as the spacing between the \( \beta \)-sheets depends on the side-chain composition flanking the \( \beta \)-sheet. The strong \( \beta \)-sheet character of amyloid fibrils is also supported by circular dichroism (CD) and Fourier transform infrared (FTIR) spectroscopy. The hydrogen bonds involved in the \( \beta \)-sheet structure are directed along the length of the fibril and act between different peptides, as shown by cryo electron microscopy (cryo-EM). This results in the characteristic cross-\( \beta \)-sheet, i.e. a \( \beta \)-sheet that runs over multiple peptides [29,47,56–58].

The physico-chemical properties of the peptide backbone, or more precisely the hydrogen bonding quality of the backbone, are causing the formation of amyloid fibrils. Nevertheless, there are significant differences in the structural details of fibrils of various peptides, attributable to the amino acid side chains [60]. Already a single point mutation can lead to distinct fibril morphology, for example observed for the genetic variants of the \( A\beta \)-peptide (Table 2.2) [61]. The side chains influence for example the length of the \( \beta \)-strand, the \( \beta \)-strand arrangement (i.e. parallel or anti-parallel), the spacing between the \( \beta \)-sheets and the position of loops, turns and other secondary structure elements that are not included in the core of the fibril [29,62].

Also the preparation conditions can have a strong influence on the details of the amyloid fibril. For example differences in temperature [63], preparation protocol (e.g. agitation or stationary conditions) [64] or solvent conditions [65,66] can lead to structurally distinct amyloid fibrils. The different fibril morphologies probably are a conse-
Chapter 2. Amyloidogenesis

Figure 2.4 — Macroscopic properties of amyloid fibrils. a An electron micrograph of amyloid fibrils formed by the islet amyloid peptide (IAPP), showing long, unbranching fibrils, with a width of $\sim 10$ nm. b X-ray fiber diffraction pattern from IAPP amyloid fibrils, showing the 4.7 Å and $\sim 10$ Å reflections. c An impression of an amyloid fibril composed of four intertwined protofilaments [57, 59].

Molecular detail

As a case study on the molecular and atomistic details of an amyloid fibril the A$\beta$-peptide, associated with Alzheimer disease, will be discussed. The structural properties of fibrils composed of (part of) this peptide have been extensively studied with both experimental and computational methods. In Table 2.2 the amino acid sequences of the wild-type A$\beta$-peptides, A$\beta$42 and A$\beta$40 [68], are given together with a list of genetic variants and useful derivatives. The genetic variants share the relation to Alzheimer disease, but show marked differences in fibril morphology, assembly rate, assembly mechanism and toxicity. Therefore, the genetic variants are often examined to find the interactions that are determining specific fibrillogenic behavior. Furthermore, fractions of the A$\beta$ peptide (and other derived small fibrillogenic peptides) help to identify the minimum requirements for fibril formation. As a result of the small size of these fractions, they are also ideal for computer simulation studies, leading to a detailed understanding of the fibril structure and assembly mechanism.
Amyloid fibril structure

<table>
<thead>
<tr>
<th>Genetic variants</th>
<th>Wild-type</th>
<th>Aβ</th>
<th>Aβ42 is more toxic and forms fibrils faster than Aβ40 (&gt;100 fold) [61]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>E22Q*</td>
<td>Aβ</td>
<td>Increased rate of fibril assembly and toxicity with respect to the wild-type (∼3 and ∼15 fold, respectively†) [61, 69–71]</td>
</tr>
<tr>
<td>Iowa</td>
<td>D23N</td>
<td>Aβ</td>
<td>Ditto (1-3 and 5-10 fold, respectively‡) [61, 70, 72, 73]</td>
</tr>
<tr>
<td>Italian</td>
<td>E22K</td>
<td>Aβ</td>
<td>Ditto (1-3 and 5-10 fold, respectively‡) [61, 73]</td>
</tr>
<tr>
<td>Flemish</td>
<td>A21G</td>
<td>Aβ</td>
<td>2.3 fold increase in Aβ secretion [70, 74]</td>
</tr>
<tr>
<td>Arctic</td>
<td>E22G</td>
<td>Aβ</td>
<td>Increased rate of fibril assembly, toxicity (1-2 and 7.5 fold, respectively†) [61, 75] and intracellular accumulation of the Aβ-peptide due to reduced degradation by alpha-secretase [76]</td>
</tr>
<tr>
<td>Derived</td>
<td>Aβ10−35 ‡</td>
<td>Aβ</td>
<td>Smallest Aβ-peptide that still forms parallel cross-β-sheets [77]</td>
</tr>
<tr>
<td></td>
<td>Aβ16−22</td>
<td>Aβ</td>
<td>Smallest Aβ-peptide that still forms amyloid fibrils [78]</td>
</tr>
<tr>
<td></td>
<td>KFFE</td>
<td>Aβ</td>
<td>Smallest peptide known that can form amyloid fibrils [79]</td>
</tr>
<tr>
<td></td>
<td>Aβ21−30</td>
<td>Aβ</td>
<td>Does not form fibrils [80]</td>
</tr>
</tbody>
</table>

Table 2.2 — Alzheimer related peptides. Primary amino acid sequence of Alzheimer Aβ42 and Aβ40 [68], genetic variants and interesting derivatives.

* E22Q denotes a mutation of the glutamic acid at position 22 into a glutamine.
† These values depend strongly on the experimental conditions and can vary substantially.
‡ Aβm−n is short for a peptide comprising residues m to n of Aβ42.

In vivo the Aβ-peptide predominantly occurs as Aβ40 or Aβ42, composed of 40 or 42 amino acids respectively (the primary structure is given in Table 2.2). The structure of the peptide inside the fibril has been largely resolved by solid-state NMR [49, 52, 81] combined with computational studies [82–84]. Both the Aβ40 and Aβ42 show a β-strand—loop—β-strand motif, preceded by an unstructured N-terminus. These structured regions are confirmed by proline scanning experiments [53].

The β-strands interact with neighboring peptides through hydrogen bonds directed along the fibril axes to form parallel β-sheets, resulting in the characteristic cross-β-sheet structure. The cross-β-sheet composed of β-strand2 (residues 28 to 40) has two hydrophobic faces, while that of β-strand1 (residues 10 to 22) has one hydrophobic and one mixed hydrophobic hydrophilic face. The loop allows the two cross-β-sheets to fold on top of each other in a lateral association, where the side chains of the hydrophobic faces engage in a hydrophobic collapse (Figure 2.5). A chain of peptides formed solely through cross-β-sheet interactions and with a width of one peptide is
Figure 2.5 — Cartoon representation of a mature amyloid fibril. The amyloid fibril of Aβ40, as resolved with solid state NMR is shown [81]. The β-strand1–loop–β-strand2 motif is clearly visible with the red and blue arrow representing β-strand1 and β-strand2, respectively, and the loop (green tube) connecting them. Two protofilaments interact laterally through one of the hydrophobic faces of β-strand2.

called a protofilament.

In addition protofilaments can associate further through a lateral interaction between the remaining hydrophobic face of β-strand2 (Figure 2.5). The location of the side chains with respect to each other or to the solvent as examined with cysteine mutagenesis experiment support this structure [54].

The data on the loop structure and the N-terminus region in the solid state NMR experiments does not allow for a unique and precise determination of the structure. For example the loop region is unconstrained, except for the requirement of a salt-bridge between a lysine and an aspartic acid and internal quaternary structure [81]. Therefore the structural constraints imposed by these experiments can be satisfied by a variety of amyloid fibril structures.

In Figure 2.6a two examples of amyloid fibril structures are shown that fulfill the experimental constraints. These models assume a solvent-free interior of the protofilament, however molecular dynamics simulations have shown that the lysine-aspartic acid salt bridge attracts water and that these structures are possibly water-filled nanotubes, as illustrated in Figure 2.6b [84].

**Very small amyloidogenic peptides**

In addition to full-length Aβ-peptides, also some smaller fragments of this peptide (as small as 7 residues [78]) are capable of forming amyloid fibrils [77, 78, 85, 86]. When considering all amyloigenic peptides, there are many more very small peptides (< 10 amino acids) that possess the properties necessary to form fibrils at physiological conditions. These very small amyloidogenic peptides are frequently used as a model system to explore the structure and assembling mechanism of amyloid fibrils, as relatively small peptides composed of 30-40 residues already display a complex and vast structural space, prohibiting straightforward characterization of the structure and assembly mechanism [87, 88].
There are some marked differences in the structural details of fibrils composed of full-length Aβ-peptide and those composed of Aβ-peptide fragments. Where the larger peptides (> 25 residues) form parallel in-register cross-β-sheets often with a turn in the middle (Figure 2.6 a) [49,52,77,82], these smaller peptides generally adopt anti-parallel cross-β-sheets with various registries (Figure 2.6 c) [78,82,85,86]. Since in all cases the width of an amyloid fibril is 7-10 nm [87], it is reasonable to assume that the number of peptides involved in lateral interactions is increased when the size of the peptide is decreased. Nevertheless the characteristic interactions of amyloid fibrils, viz. cross-β-sheet and lateral, occur in all amyloid fibrils.
Chapter 2. Amyloidogenesis

Amyloid fibril formation

To elucidate the aggregation mechanism of peptides into amyloid fibrils, a wide range of experimental approaches is available, including spectroscopy, NMR and light scattering [89]. This has lead to an improved understanding of amyloid fibril formation at the macroscopic level. However, the interactions at the molecular or atomistic level that drive amyloid fibril formation are poorly understood. Elucidating these atomistic interactions requires high resolution techniques, such as NMR and computer simulations, which have only become available recently.

Fibril formation is a nucleation-growth mechanism

The evidence from low-resolution experiments indicates that amyloid fibril formation occurs through a nucleation-growth mechanism. During nucleation a transition takes place in the system, where the proteins in solution form fibrillar structures. These initial fibrillar structures then grow into mature fibrils. Only when the monomer concentration exceeds a critical concentration (in the range of 4-15 µM for full-length Aβ-peptide [90]), fibrils will be formed (Figure 2.7) [90–92], analogous to micelle formation [93]. Generally, the appearance of fibrils is preceded by a lag-phase, a common feature of the nucleation step, corresponding to the time required to form a stable nucleus (Figure 2.7). During the nucleation phase no significant amount of fibrils is present, rather a variety of oligomers are formed, including β-sheet-rich species that provide nuclei for the growth of mature fibrils [29]. After the nucleation step fibril formation proceeds rapidly by merging of protofilaments and addition of monomer to the fibril tips [94]; the growth phase.

By adding seeds to a solution at a saturated monomer concentration, i.e. by adding preformed fibrils or protofilaments to this solution (Figure 2.7), the nucleation step and thus the lag-phase can be surpassed [95]. As the monomers will now grow onto the seed, the structural details of the resulting fibril will be determined by the seed [64]. Additionally, some mutations or a change of the experimental conditions can reduce or eliminate the lag-phase [96–98]. These changes can slow down the growth or accelerate the nucleation process, such that the nucleation step is no longer rate limiting, thereby effectively eliminating the lag-phase [29].

Although amyloid fibril formation happens through a nucleated growth mechanism, the details of the mechanism display a multitude of states and pathways. Generally, the mechanism of a specific fibril formation can be classified into three models (Figure 2.8) [56]. The first is the refolding model, which describes for instance the formation of amyloid fibrils from insulin [59]. The peptide in solution is usually found in a well-defined native state, however via an intermediate partially unfolded structure, fibrils can be formed. The second type is the natively unfolded model, found for example in the case of Alzheimer Aβ-peptide. It is similar to the refolding model, with the exception that the peptide in solution does not possess a native state. The third model is the gain of interaction model, which explains the fibril formation of peptides such as the human superoxide dismutase mutant S134N [99]. The peptide has a well-defined
Figure 2.7 — Different stages of fibril formation. Both unfolded and globular peptides are in a concentration-independent equilibrium with partially folded structures (red arrows). The peptides can agglomerate to form large unstructured clusters, called amorphous aggregates. When a structure has adopted sufficient fibril character that it can grow rapidly into a fibril it is called a nucleus (this can be either one peptide or an aggregate). A nucleus that only grows in the axial direction can form one long cross-β-sheet, a protofilament. Partially unfolded proteins, nuclei and protofilaments can all assemble through lateral and axial interactions to form fibrils. This figure is based on [32].

* Some authors use the term protofilament to indicate two or three intertwined cross-β-sheets.

† The lag-phase can be avoided by adding a seed, which can be anything on the right hand side of the barrier. In practice only fibrils can be isolated, broken into smaller pieces and used as a seed.
Chapter 2. Amyloidogenesis

Figure 2.8 — Classifications of fibril formation. Fibril formation can be categorized into three classes, viz. natively folded, natively unfolded and gain of interaction [56]. The red and black arrows indicate concentration-independent and concentration-dependent equilibriums, respectively. This figure is based on [56].

native state, which is mostly retained in the amyloid fibril. Usually only a small portion of the peptide is subject to conformational changes, exposing small regions that can now interact with other peptides to form amyloid fibrils.

Energy landscape of fibril formation

The assembly of peptides into amyloid fibrils usually involves rigorous changes in the conformation of individual peptides, either as a monomer or in the initial aggregates. Therefore, amyloid fibril formation cannot be discussed separately from protein folding. As the energy landscape describes the relation between various conformations or states of the peptide chains [100], the properties and transitions associated with protein folding are conveniently discussed in terms of this landscape (Chapter 1 and Figure 2.9). In Figure 2.9 a schematic energy landscape is displayed, where the protein folding (light) and the aggregation (dark) energy landscape are combined [89]. As shown in the figure such a landscape portrays typically a multitude of conformational states accessible to the peptides in solution. The stability of each state strongly depends on the primary amino acid sequence and the environmental conditions, which also applies to the barriers separating different states. Hence the dominant structures and pathways can vary strongly between peptide species and environmental conditions, as will be discussed later.
In the energy landscape depicted in Figure 2.9, seven different states can be distinguished [89, 101]. Firstly, there are three states for the monomeric peptide; those which can access a native state (N), intermediate partially folded states (I) and unfolded conformations (U). Secondly, three classes of peptide aggregates can be observed, showing various degrees of order, such as amorphous aggregates (AA), ordered aggregates (OA) and oligomers (O). The final and lowest energy state is the amyloid fibril (F), the structure of which has been discussed in the previous section.

Usually, a structural transformation of the peptide is a prerequisite for a free monomer to transform into a building block of a fibril. While a single molecule is unlikely to acquire the $\beta$-strand conformation characteristic of an amyloid fibril, this structural rearrangement is possible after aggregation, as the peptide can enter a distinct region of the energy landscape upon aggregation (dark grey area in Figure 2.9), displaying other stable secondary structure motifs due to close proximity of the other peptides. This effect shows an analogy with environmental changes, where different solvents can affect the shape of the energy landscape.

**Molecular detail of amyloid formation: The monomer**

Conformational fluctuations of the monomeric peptide, leading to the population of aggregation-prone structures, are a key event in the amyloid fibril formation. For example, the resistance of full-length A$\beta$-peptide (A$\beta$40 and A$\beta$42) to structural refinement with for instance solution state NMR indicates a highly degenerate monomeric...
structural ensemble. A similar characteristic is observed for fibrillogenic fragments of \( \text{A}\beta \)-peptide, which do not fold into a unique native state [102–104]. Consequently there is an enhanced chance for the peptide to populate an amyloid fibril promoting state [105]. On the other hand the \( \text{A}\beta_{21-30} \) peptide does not form fibrils when dissolved, although it can exist in a stable fibril structure [80]. The minimum energy structure is characterized by a large propensity to form a loop that is stabilized by a strong interaction between residues Val-24 and Lys-28 [80, 106–108]. The resulting deep free-energy minima explains the resistance of \( \text{A}\beta_{21-30} \) solutions to form fibrils.

Differences in the time spent in the nucleation stage or the rate of fibril growth between different peptides or under different conditions can often be explained by the population density of the aggregation prone monomer states. [104, 109] For instance mutations can promote monomeric conformations that have a high propensity to form amyloid fibrils [110], thereby decreasing the lag-phase and increasing the fibril growth rate. Similarly, it has been shown that the C-terminus of \( \text{A}\beta 42 \) shows a much higher rigidity and \( \beta \)-strand content than the C-terminus of \( \text{A}\beta 40 \) [111, 112]. The frequent occurrence of this rigid \( \beta \)-strand structure facilitates amyloid fibril formation and growth, which explains why \( \text{A}\beta 42 \) has a much higher aggregation rate and a shorter lag-phase than \( \text{A}\beta 40 \).

**Molecular detail of amyloid formation: Small aggregates**

The different populations of the conformational states of the monomer clearly will have implications for the early folding pathways in the nucleation phase. However, characterization of the peptide aggregates that are formed, prior to amyloid fibril formation, is extremely difficult. First, isolating the distinct aggregates in an experiment is currently impossible, posing a challenging problem to experimentalists. In addition the time scale of the formation and transition of the various types of aggregates makes detailed computer simulations such as molecular dynamics intractable. Only the development of coarse-grained simulation methods in the last five years has provided mechanistic insight into the early stages of aggregation [113], although the simplifications and assumption made in these methods result in a high uncertainty in the findings [114].

Several coarse grained simulations show that nucleation is a two-step process, where the formation of structured oligomers is preceded by random aggregates [115, 116]. The structured oligomers show high content of \( \beta \)-strands, observed in double layered \( \beta \)-sheets and sometimes in \( \beta \)-barrel type structures as an off-pathway product [117–120]. These observations correspond very well with AFM experiments [121,122].

Also for the aggregates in the nucleation phase, the equilibrium between different states depends strongly on the environmental conditions and the specific aminoacid sequence [123, 124]. For instance \( \text{A}\beta 40 \) shows a preference for monomers and dimers, while \( \text{A}\beta 42 \) prefers higher number aggregates, such as pentamers and hexamers [116,124]. In addition the dominant pathways of aggregation for two distinct fragments of the \( \text{A}\beta \)-peptide are different. On the one hand, the weakly hydrophobic fragment \( \text{A}\beta_{25-35} \) forms cross-\( \beta \)-sheet structures in one step without any intermediate
unstructured aggregate formation. On the other hand, fibril formation of the more hydrophobic \( \text{A}\beta_{16-22} \) is preceded by a collapse into an unordered aggregate, supporting a two-step mechanism. The relative strength of hydrophobic interactions thus seems to be one factor affecting the choice for a specific pathway [125].

**Molecular detail of amyloid formation: Fibril growth**

The formation of a stable nucleus from these structured and random aggregates initiates the growth phase (Figure 2.7) [32, 127]. Here, the nucleus is defined as the starting structure from which a fibril can grow, but at present the specific nature of the nucleus
Chapter 2. Amyloidogenesis

is still subject of debate [128–131]. With the definition for the nucleus given here, the minimum nucleus size for \( \text{Ab}_{16-22} \) is 8-16 monomers [132] and \( \text{Ab}_{40} \) requires a minimum assembly of 16 monomer chains for a stable nucleus. The structural stability of the 16-mer nucleus correlates with a plateau in the hydrophobic residue density and a decrease in the likelihood of losing hydrophobic interactions by rotating the fibril subunits or protofilaments. [133]

After the formation of stable nuclei, mature fibrils will emerge rapidly during the growth phase. Fibrils can grow laterally through merging of protofilaments or clusters of protofilaments. Furthermore, fibrils can grow in length by the addition of monomers to the tip of the fibril [94]. The latter has been shown to be a thermodynamic process [134], happening either through a two step dock and lock mechanism when hydrophobic interactions are dominant [126,135,136], or in one step driven by hydrogen bond formation for weakly hydrophobic peptides [125].

In the first step of the dock-lock mechanism, a monomer docks to the amyloid fibril template. The docking is driven by the hydrophobic interactions between side chains, similar to the hydrophobic collapse in protein folding. After docking the monomer slowly rearranges in the locking stage. In this phase the hydrogen bonds between the monomer and the fibril template are formed in the characteristic cross-\( \beta \)-sheet pattern (Figure 2.10) [126, 135, 136].

Goals of this thesis

With the progress of technological advances also our understanding of the amyloid structure and aggregation mechanism has grown over the years. The structure of most amyloid fibrils can be elucidated by a combination of experimental and computational techniques, although it still presents a challenging task. The mechanism of amyloid formation on the other hand is still not completely understood, despite the fast increase in our knowledge on this subject.

There are still many unanswered questions regarding amyloidogenesis. To name a few: What are the structures of the distinct oligomeric species; is there one specific oligomeric structure that can be characterized as the required transition state; are there any off-pathway oligomers (i.e. not leading to fibril formation); if so, why are they off-pathway and what is their structure; what is the binding strength of peptides in a fibril; what are the effects of specific amino acids on the structure of the amyloid fibril; what is the structural basis for amyloid formation; why is the fibril width always 7-12 nm, irrespective of the peptide; how do the environmental conditions affect the final fibril structure?

Many of these questions require a very detailed understanding of the system under consideration. Computer simulations at atomistic scale can provide the required detail (see Chapter 1). Unfortunately the computer power to date is still a limiting factor in simulation studies of biomolecular systems, such as amyloidogenesis. Consequently the search for faster algorithms, methods showing better performance and faster computers is still very important.
Goals of this thesis

The high computational demand to simulate a biomolecular system has two causes, namely the ensemble sampling problem and the search problem. The first problem arises when the ensemble (all relevant states of the system) is degenerate, i.e. there are multiple relevant low-energy structures. To compare the stability of these different structures, both the structures themselves and the transitions between them must be sampled sufficiently. However, generally the barriers separating the low-energy structures are preventing frequent transitions and consequently the ensemble is not sampled correctly.

When for example answering the question what the binding strength of peptides in a fibril is, the complete set of states describing the dissociation of one peptide from the fibril has to be sampled. However in a standard simulation, the ensemble sampling problem prevents correct sampling of the relevant states.

Part II of this thesis is focused on solving this ensemble sampling problem when analyzing the thermodynamics of amyloid fibril growth. Therefore umbrella sampling (Chapter 1) is combined with replica exchange steps (Chapter 1) to calculate association constants and free energies that are important in the self-assembly of peptides. This leads to a more than 10-fold acceleration over conventional umbrella sampling, thereby providing a practical method to calculate these free-energy differences. The method is then used to predict the fibril forming propensity and strength of four related model peptides.

The second issue, the search problem manifests itself when simulations are used to search for the low-free-energy structures of a biomolecular system without any input information on the structure. The number of possible conformations for such a biomolecular system is enormous, making it intractable to sample all of them, even with powerful computers. For example, the number of different backbone conformations of a very small peptide composed of 16 amino acids can be approximated by $3^{16} \cdot 2^{16} \approx 10^{15}$, where 3 is the number of rotation minima of the peptide bonds, 16 the number of residues and 2 the number of backbone peptide bonds per residue that can be rotated.

This search problem poses a serious obstacle when assessing the distinct oligomeric structures accessible to a particular peptide solution with computer simulations. In Part III of this thesis a novel approach to overcome this problem is developed in the context of protein folding, as protein folding and aggregation show great similarity. With this new approach protein folding is accelerated in all-atom molecular dynamics simulations by introducing alternating hydrogen bond biasing potentials as a supplement to the force field, which leads to quick formation of secondary structure elements.

The initial implementation of this method results in non-equilibrium sampling of the conformational space. To measure the thermodynamic properties of the system, such as the relative stability of different states, equilibrium sampling is required. Therefore the hydrogen bond biasing potentials are combined with Hamiltonian replica exchange, which is used to assess the folding landscape of a heptapeptide in methanol and a 16-residue poly-alanine in water. Ultimately, when protein folding can be
assessed efficiently, the method can also be applied to examine the different states in the nucleation phase.
PART II
PREDICTING AMYLOID FIBRIL GROWTH PROPERTIES
Rapid free energy calculation of peptide self-assembly by REMD umbrella sampling


**Abstract** We extend umbrella sampling with replica exchange steps to calculate free energies that are important in the self-assembly of peptides. This leads to a more than 10-fold speed up over conventional umbrella sampling, thereby providing a practical method to calculate these free-energy differences. This approach can also observe first order phase transitions and pinpoint the location of the concomitant boundary. When conformational changes are involved this method can handle peptides up to \(~7\) residues, providing a rapid and accurate assessment of the thermodynamic properties of model systems, and can thus be used to answer fundamental questions about peptide self-assembly. When no major conformational changes are involved we expect the size limit to be equal to that of standard molecular dynamics.

**Introduction**

Many biological entities are composed of peptides that self-assemble into well-defined structures. The self-assembly of proteins into microtubules and the cytoskeleton plays a pivotal role in the structure and functioning of various cell types [137]. In higher organisms, the self-assembly of peptides into amyloidal fibrils is generally associated with malfunction, *i.e.* Alzheimers disease and Huntingtons disease [29]. The specific characteristics of the self-assembly process can also be used in rational design of engineered materials, such as the formation of nanoassemblies [38, 138].

Elucidation of the mechanisms and properties underlying the self-assembly pro-
cess is thus of fundamental importance. For the rest of this paper we will focus on the self-assembly of peptides into amyloidal fibrils. This process has been the subject of both experimental and computational studies [32, 51, 139, 140], however most of these experiments explore the mesoscopic level. A detailed description of self-assembly of polypeptides into fibrils on the atomic level is still not available.

The peptides in these self-assembled structures are held together by non-covalent interactions, e.g. hydrogen bonds and hydrophobic interactions. Since these interactions are strongly dependent on the macroscopic environment, properties like the pI of the peptide and the pH of the medium, as well as the presence of other species in the surrounding medium can modulate the balance in the self-assembly process [141, 142]. It has been argued that (almost) every peptide can be made to self-assemble into fibrils given the right conditions [143].

At present, the mechanism of peptide self-assembly into fibrils is considered to be a nucleation-growth mechanism, commonly indicated by a lag-phase [32, 140, 144]. Formation of a stable nucleus precedes rapid growth of the assembly until equilibrium is reached between the monomers in solution and in the assembly. Although the first step is under kinetic control, the final equilibrium composition is under thermodynamic control [53, 90, 145, 146] and can be described by free energy differences.

Similarly free-energy differences can be used to predict the final equilibrium composition of self-assemblies and it is related to the equilibrium constants (through \( \Delta G = -RT \ln K_e \)) providing a link with experiments. In addition the calculated free-energy differences can be coupled to the free-energy parameters used in predicting equilibrium properties from statistical thermodynamic models [147].

Here we introduce distance replica exchange molecular dynamics simulations (REMD), previously described by Lou and Cukier to study conformational fluctuations [148]. We use this replica exchange molecular dynamics (REMD) as a sophisticated umbrella sampling technique, to extract a potential of mean force (and the corresponding free-energy difference and equilibrium constant) from molecular dynamics trajectories that describe the dissociation of one peptide from the assembly. A more general formulation of the replica exchange umbrella sampling can be found in [149]. The method is applied to the study of a small peptide, KFFE, specifically designed to form amyloid structures in vitro based on the \( \beta \)-propensity of its middle residues and the charge attraction of its outer residues [79]. This system is suitably small, with concomitantly low computational cost, and it serves as a model system for amyloid formation in general.

REMD umbrella sampling as presented here is a considerable improvement over existing methods, showing a more than 10-fold speedup over standard umbrella sampling. It provides a fast and accurate method to calculate free-energy differences important in peptide self-assembly. This allows for systematic studies on the thermodynamic properties of peptide self-assembly in model systems. The peptide size limit in studies where the structure of the monomer is different from that in the assembly is ~7 residues, because the conformational changes of larger peptides are too slow. We are confident that in studies where these structures are similar this limit is set by the
simulation method, in this case standard molecular dynamics.

**Method**

*Replica exchange umbrella sampling*

The easiest way to record the free energy difference between two states, A and B, is to measure the probabilities to find the system in state A and in state B. Direct sampling of both probabilities with the required accuracy in one simulation is only feasible when marginally different free energies and low transition barriers are involved. For that reason more sophisticated techniques are needed to obtain the free energy differences in complex systems where these conditions are not met.

One of these techniques is umbrella sampling. In a general umbrella sampling simulation a path between two states is chosen that can be explored exhaustively. To that end a number of independent simulations (windows) is performed with the addition of different biasing (umbrella) potentials $V_{bias}$ to the standard potential of the system $V_{standard}$:\

$$V_{total}(q) = V_{standard}(q) + V_{bias}(q)$$ (3.1)

Here $V_{total}$ is the total potential energy in one simulation and $q$ denotes all particle positions. The biasing potential $V_{bias}$ in one simulation restrains the system around a point on the chosen reaction path. The simulation samples the phase space around this point on the reaction path sufficiently to calculate the potential of mean force for that part of the reaction path. All simulation results can be effectively combined using the Weighted Histogram Analysis Method (WHAM) to generate a potential of mean force [13–15]. The details of the biasing potential and WHAM can be found further down in this section.

The efficiency of umbrella sampling can be increased by adding replica exchange steps [148–150]. Obtaining the potential of mean force with umbrella sampling already requires $N$ copies of the system with different biasing potentials. As a standard umbrella sampling already requires sufficient overlap in the energy distribution of neighboring windows, using the windows as replicas in a Hamiltonian-REMD [18] is straightforward and can enhance sampling without additional computational cost [148, 149].

An attempt to exchange the Hamiltonian for replica $i$ and $j$ is accepted according to the following Metropolis criteria:

$$p_{acc}(i, j) = \begin{cases} 
1 & \Delta(i, j) \leq 0 \\
\frac{1}{e^{-\Delta(i, j)}} & \Delta(i, j) > 0 
\end{cases}$$ (3.2)

$$\Delta(i, j) = \beta [[V_{total,i}(q_i) - V_{total,j}(q_j)] + [V_{total,j}(q_j) - V_{total,i}(q_i)]]$$ (3.3)

with $p_{acc}$ the acceptance probability and $\beta = 1/k_BT$, where $k_B$ is Boltzmann’s constant and $T$ the temperature. This exchange depends on the potential energy difference when the potential energy functions are swapped between system $i$ and $j$, i.e. when the particle positions currently subjected to the potential energy function $i$ and
are then subjected to the potential energy function of \( j \) and \( i \) respectively. As the potential energy of replica \( i \) and \( j \) only differ in the contribution of the biasing potential equation 3.3 can be reduced to

\[
\Delta(i, j) = \beta \{ [V_{bias,j}(q_i) - V_{bias,j}(q_j)] + [V_{bias,i}(q_j) - V_{bias,i}(q_i)] \}
\] (3.4)

**System**

In most amyloids described in the literature the fundamental structural element is the cross-\( \beta \)-sheet, which is a \( \beta \)-sheet running across different peptide chains \([49, 151]\). Extended cross-\( \beta \)-sheets are called protofilaments, which can combine by lateral interactions to form fibrils. Note that some authors use the term protofilament to indicate two intertwined cross-\( \beta \)-sheets \([64]\).

Unfortunately, assemblies that are large enough to represent experimentally observed protofilaments or fibrils are intractable in the detailed simulations that are required to obtain reliable free-energy differences for the assembly or dissociation process. Therefore we study smaller systems that may reproduce the characteristics of the full-length assembly upon extrapolation. In order to explore the feasibility of this approach, we study the tetrapeptide KFFE that has been specially designed for its amyloidogenic properties. The thermodynamic properties of three systems will be studied containing two, five and ten peptides, which will be referred to as dimer, pentamer and decamer respectively, in the remaining part of this chapter.

The dimer is the smallest system that represents the primary building block of an assembly with a well-defined structure. In this system we calculate the potential of mean force linking two peptides in a cross-\( \beta \)-sheet complex to two peptides in solution (Figure 3.1a). The pentamer is a cross-\( \beta \)-sheet of five peptide units representing a (truncated) protofilament. The decamer is composed of two cross-\( \beta \)-sheets of five peptide units involved in a lateral interaction representing a fibril. In all cases we calculate the potential of mean force describing the removal/association of one peptide from/to the assembly (Figure 3.1b and 3.1c).

**Equilibrium constants**

The three systems discussed in the previous paragraphs are summarized in the following reaction equation

\[
A_n + A \rightleftharpoons A_{n+1}
\] (3.5)

with \( n = 1 \), \( n = 4 \) and \( n = 9 \) for the dimer, pentamer and decamer, respectively.

To obtain the potential of mean force associated with these reactions from the separate windows we need to solve the WHAM equations \([14, 15]\) 3.6 and 3.7 self-consistently.

\[
\rho^u(r) = \sum_{i=1}^{N_{rep}} \frac{\rho^b_i(r)}{\sum_{j=1}^{N_{rep}} e^{-\beta(V_{bias,j}(r) - f_i)}}
\] (3.6)

\[
e^{-\beta f_i} = \int dr e^{-\beta V_{bias,i}(r)} \rho^u(r)
\] (3.7)
Figure 3.1 — Overview of the calculated free-energies differences. The dimer \textbf{a} and the pentamer \textbf{b} are shown in top view to emphasize the cross-β-sheet interaction. The decamer \textbf{c} is shown in side view to emphasize the lateral interactions.

In these equations $r$ is the distance between two particles $i$ and $j$ (the reaction coordinate), $N_{rep}$ the number of replicas and $f_i$ a constant that represents the free energy associated with the introduction of the biasing potential of replica $i$. With equations 3.6 and 3.7 the biased probability density, $\rho_i^b(r)$, from the different replicas are combined into one unbiased probability density, $\rho^u(r)$. Before we calculate the potential of mean force through $W(r) = -k_B T \ln \rho^u(r)$ the unbiased probability density is scaled so that $\rho^u(r) = 1$ for large $r$.

The potential of mean force is connected to the equilibrium constant $K$ through the following relation [152]

$$K = \frac{\sigma_{int} n_{int}}{\sigma_{ij}} \int_{r_{low}}^{r_{up}} e^{-W_{ij}(r)/k_B T} 4\pi r^2 dr$$

(3.8)
Chapter 3. Rapid free energy calculation by REMD umbrella sampling

with $K$ the association constant, $r_{up}$ and $r_{low}$ the upper and lower limit of the potential of mean force, respectively, $\sigma_{ij}$ the symmetry number (1 when $i \neq j$ and 2 when $i = j$), $n_{int}$ the number of interaction spaces and $\sigma_{int}$ the percentage of a full sphere that is used by the interaction space.

Here we introduced the interaction space, which is the volume over which the potential of mean force acts. For a spherically symmetric system, like the dimer reaction, the whole volume is the interaction space and $n_{int} = \sigma_{int} = 1$. When considering fibril growth we can not assume spherical symmetry, but it is reasonable to describe the interaction site as a cone along the fibril axis between the dissociated peptide and the closest cross-$\beta$-sheet bonded peptide. Therefore we estimate $\sigma_{int} = 1/4$ and since every fibril has two tips $n_{int} = 2$.

**Simulation details**

All simulations were performed using the GROMACS [11] software package version 3.3.1 and the GROMOS96 56a6 force field [153] in combination with SPC water molecules [154]. This forcefield has been developed to give accurate free energies of solvation and thus appears very suitable for our purpose. An integration timestep of 2 fs was used, with all bonds constrained using the LINCS algorithm [155]. Van der Waals interactions were ignored outside a cut-off of 1.2 nm. Electrostatic interactions were treated with the PME method applying a real-space cut-off of 0.9 nm [156]. The system temperature was coupled to a Berendsen thermostat, no pressure coupling was used.

The cross-$\beta$-sheet structures in the simulations were generated manually and were dissolved in a box of approximately 3500, 2600 and 3200 SPC water molecules for the dimer, pentamer and decamer, respectively, to obtain a density of 0.99 kg/l. Initial velocities were generated randomly to create a system at 300 K after which the systems were equilibrated for 5 ns.

In the simulations a harmonic biasing potential was chosen that acts on the center of mass (COM) distance between two peptides (equation 3.9).

$$V_{bias} = \frac{1}{2} f_c (d_{i,j} - d_{COM})^2$$  \hspace{1cm} (3.9)

where $d_{i,j}$ is the COM distance between the terminal peptide in the assembly and the closest neighbor cross-$\beta$-sheet bonded peptide (Figure 3.1), $d_{COM}$ is the reference COM distance and $f_c$ the force constant. After testing several values we chose a force constant of 5000 kJ/mol for our simulations. A higher value will decrease the width of the energy distribution and thus decrease the overlap of the potential energy between neighboring replicas thereby lowering the exchange acceptance rate. A smaller value is not able to hold $d_{i,j}$ around $d_{COM}$ on the steep slope of the potential of mean force and thus some parts of the reaction coordinate will not be sampled sufficiently. We used the pull-code provided in GROMACS version 3.3.1 with a modification to allow for a radial COM distance restraint.

For the different windows the $d_{COM}$-values are separated by 0.04 nm starting from 0.42 nm in the smallest replica to a total of 28, 28 and 40 replicas for the dimer, pentamer and decamer respectively in both the standard and REMD umbrella sampling.
Results and discussion

**Figure 3.2 — Visitation of the replicas by the systems.** a The minimum (lower solid) and maximum (upper solid) replica visited by the systems enclosing the visited area (grey). This is averaged over 3 runs of 10 ns of the dimer with system X starting in replica X. Two distinct regions are seen: 1) small entropy and large enthalpy and 2) large entropy and small enthalpy. b Two systems that migrate through the whole range of replicas.

In the REMD umbrella sampling we used an all-pair exchange [157] with exchange attempts every picosecond. To generate the starting structures for all the different copies the biasing potentials were consecutively introduced starting from the equilibrated cross-β-sheet structure and the smallest COM distance. Every window was equilibrated for another 1 ns before production runs started.

In the dimer system we used a radial COM distance restraint while in the pentamer and decamer the COM distance was restrained along the direction of the fibril axis. The protofilament part of the pentamer and the fibril part of the decamer where kept in place by a position restraint to mimic the lack of mobility in a full-grown fibril.

**Results and discussion**

**REMD umbrella sampling of biopolymer assembly**

A good understanding of the difference between a replica and a system is crucial in the use of REMD: the term system here refers to a specific box of atoms that is monitored in time and replica refers to a specific Hamiltonian. When a swap is accepted both systems will be subjected to another Hamiltonian, in other words they will be in different replicas.

In a REMD simulation the different systems should migrate through the replicas, however, it was observed that the majority of the systems do not visit all available replicas. We will discuss this for the dimer, but similar behavior is observed for the pentamer and the decamer. Figure 3.2a shows the range of replicas visited by the different system averaged over 3 runs of 10 ns. This Figure shows two distinct regions: 1) replicas biased to a small COM distance exchange with each other and 2) replicas biased to a large COM distance exchange with each other. Exchanges between these two regions are rare, although not excluded (Figure 3.2b).
Clearly a poor exchange acceptance rate can cause this behavior. However the spacing of 0.04 nm between neighboring replicas is chosen to have an exchange acceptance rate of 30 %, which is close to the optimal 38 % [158]. This rate is reached throughout the whole simulation between all neighboring replicas and the exchange acceptance rate can thus be ruled out as a cause.

Next, we examined both the intermolecular interactions and the conformational freedom of the peptides as a function of the COM distance (set by the Hamiltonian (replica)) to determine the causes of this behavior. As an illustration of the intermolecular interactions, the probability distribution of the number of backbone hydrogen bonds as a function of the COM distance in one REMD simulation is shown in Figure 5.1. The dashed line in this Figure indicates the boundary between the two regions. This Figure shows a high probability to find hydrogen bonds in region 1 (left) and a very low probability to find hydrogen bonds in region 2 (right). The conformational fluctuations are illustrated in Figure 3.4 by the probability distribution of the root mean square deviation (rmsd) as a function of the COM distance. The rmsd calculations are performed per replica and the reference structure per replica is the middle structure after clustering all structures corresponding to that replica. Since a low value of the rmsd indicates a strong structural resemblance, the high probability to find low rmsd values in region 1 (left) indicates a single rather well-defined structure. The high probability to find high rmsd values in region 2 (right) is indicative of the occurrence of many different structures.

Whereas the single well-defined structure in region 1 is characterized by many intermolecular interactions, the structures in region 2 show almost no intermolecular interaction. From a thermodynamic point of view this means that the system tends to segregate into a low-entropy/high-enthalpy and a high-entropy/low-enthalpy region. Transitions between two such regions resemble a first order phase transition which agrees well with the experimentally observed first-order kinetics in seeded fibril growth [92].

Crossing the barrier associated with the first order phase transition does occur, albeit with a low frequency. However, considering the approach adopted here, this does not present a problem, assuming that both the unfolded and the folded basins are sufficiently sampled, with enough overlap between the two.

REMD umbrella sampling thus seems extremely well suited for this type of study since a) the migration of the systems through the replicas ensures that the effect of conformational changes will be spread over all the replicas and b) the REMD acceptance rule ensures that at the boundary associated with the first order phase transition mixing of the two states is under thermodynamic control.

Standard umbrella sampling vs REMD umbrella sampling

To assess the efficiency of the REMD umbrella sampling method in peptide self-assembly processes we compared our results with a standard umbrella sampling simulation. We expected that convergence would occur more rapidly, similar to what has been reported by Lou and Cukier [148].
Results and discussion

Figure 3.3 — Intermolecular interactions to illustrate the enthalpy contribution. Probability distribution describing the chance to find Y backbone hydrogen bonds at COM distance X between the two peptides in a dimer REMD simulation.

Figure 3.4 — Conformational fluctuations to illustrate the entropy contribution. Probability distribution depicting the chance to find a root mean square deviation (rmsd) value Y after rotation and orientation fit of the separated peptide at a COM distance X between the two peptides in a dimer REMD simulation. The rmsd calculations are performed per replica and the reference structure is the middle structure found for that replica after clustering.

Figure 3.5 — Performance of the standard (dashed) and the REMD (solid) umbrella sampling. The free energy is plotted as a function of total simulation time.
Figure 3.6 — Conformational transitions in the standard (black) and REMD (red) umbrella sampling. The C- to N-terminus distance in one peptide (either ~1 nm or ~0.6 nm) is plotted as a function of time for one biasing potential of the dimer. The simulation times were 200 and 10 ns for a standard and a REMD umbrella sampling, respectively.

Figure 3.5 shows the development of the free-energy difference as a function of the total simulation time used to calculate it for both a standard and a REMD umbrella sampling experiment. It is obvious that the standard umbrella sampling simulations show poor convergence and even over 100 ns of simulation time clear convergence of the free-energy difference is not observed. On the other hand the REMD umbrella sampling shows convergence of the free energy within 10 ns.

This large difference in convergence time can be attributed to a small number of windows in the standard umbrella sampling experiment, where conformational changes are very infrequent. For example in Figure 3.6 the termini distance of one peptide in a standard umbrella sampling window shows two transitions in 200 ns, while the same window in the REMD umbrella sampling experiment shows 6 transitions in 10 ns. Consequently long simulations are required in a standard umbrella sampling experiment to accurately sample the part of the potential of mean force simulated in these windows. In the REMD umbrella sampling the conformational changes are spread over the different replicas subject to Boltzmann weighting and the complete reaction path can equilibrate much faster. Remarkably, the windows that suffer most from these infrequent conformational changes are right at the boundary of the first-order phase transition mentioned above.

The reason for the large difference in convergence time and the infrequent transitions can be explained by the free-energy landscape of the distance between the termini that are in contact in the fibril versus the COM distance between the dissociated
Results and discussion

Figure 3.7 — Free energy landscape associated with fibril formation. The free energy is plotted as a function of the distance between the interacting termini in the fibril versus the COM-distance between the dissociated and the closest cross-β-sheet bonded peptide in a pentamer simulation. The black line is the COM-distance to which the peptides in the 7th replica are restrained.

and the closest cross-β-sheet bonded peptide (Figure 3.7). As the COM distance is increased we also expect the termini distance to increase. In Figure 3.7 the black line corresponds to the COM distance the peptides are restrained to in the 7th window. On this line there are clearly two free energy minima (termini distance of 0.45 nm and 1.0 nm) separated by a barrier at termini distance 0.7 nm. This can be explained by the fact that the dissociated peptide will be removed either N-terminus first or C-terminus first. The free-energy barrier of 2 k_B T, associated with swapping the dissociated terminus, is not easily overcome in a standard umbrella sampling. In the REMD umbrella sampling however the system can go around this barrier by migrating through the replicas.

From our results it is clear that a static method like standard umbrella sampling can result in kinetic traps, which slow down the convergence of the system, while in a more dynamic method like REMD umbrella sampling a system migrates through the replicas, allowing it to go around barriers and removing the kinetic trap.
Figure 3.8 — Potential of mean force of fibril formation. Potential of mean force for the dissociation of the dimer (black), of a single peptide from the pentamer (red) and from the decamer (green). The COM distance is between the dissociated peptide and the closest cross-$\beta$-sheet bonded peptide. For every dissociation study two potentials of mean force are plotted to show the reproducibility of the method.

**Free energies of fibril formation**

Figure 3.8 displays the potential of mean force from the REMD umbrella sampling simulations. For each system (dimer, pentamer and decamer) the potential of mean force was calculated twice, with different starting structures and with different starting velocities. From Figure 3.8 it is clear that these potentials of mean force show good convergence.

From these potentials of mean force the associated equilibrium constant can be calculated using eq 4.8. For the association reactions (eq 3.5) this results in association constants of $59 \text{ M}^{-1}$, $2.7 \cdot 10^{2} \text{ M}^{-1}$ and $3.2 \cdot 10^{4} \text{ M}^{-1}$ for the dimer, pentamer and decamer, respectively. Although experimental data is not yet available the equilibrium constants calculated here appear to be within the expected range. For example fibril growth will take place as long as the monomer concentration is larger than $1/K_{\text{decamer}} = 31 \mu\text{M}$, which is well below the initial 200 $\mu\text{M}$ monomer concentration used to generate the KFFE fibril [79]. In addition these results predict that in a 200 $\mu\text{M}$ monomer concentration the lag-phase is characterized by the formation of a first fibril, since both dimer and protofilaments are unstable species with $1/K_{\text{decamer}}$ of 17 mM and 3.7 mM, respectively.

In a recently developed statistical thermodynamic model of fibril formation [147] the interaction free energies are deduced from free-energy differences. In the case of amyloid formation there are two basic interactions: a) between the peptides in the
Concluding remarks

A dimer linked by a cross-β-sheet interaction is the primary building block of a fibril and probably its formation is the first step in fibril formation. This first step involves a double entropic penalty, because both peptides sacrifice a large part of their conformational freedom to engage in a cross-β-sheet interaction. In contrast fibril growth only experiences the entropic penalty of the added peptide and on top of that a fibril will benefit from the lattice energy. These distinctions are reflected in the difference between the free-energy difference observed for the dimer and pentamer simulation, which are -4.1 k_BT and -5.6 k_BT, respectively.

Concluding remarks

In this paper we show that the REMD umbrella sampling is a very efficient strategy to calculate free-energy differences that are important in self-assembly of small amyloidogenic peptides accurately. More general we expect this method to generate reliable free-energy data in any system where association or dissociation interaction, involving non-covalent bonds, occur, like for example docking of a ligand to its receptor or the formation of enzyme-substrate complexes.

The increase in efficiency for REMD umbrella sampling over conventional umbrella sampling can be attributed to barriers that appear in specific windows. In a conventional umbrella sampling thermal fluctuations drive the barrier crossing, while in a REMD umbrella sampling the migration through the replicas allows for alternative routes around this barrier. Consequently there are more transitions between free-energy minima in REMD umbrella sampling and the sampling (and thus the potential of mean force) converges faster.

The association or dissociation of peptides usually involves a first order phase transition, characterized by a compensating jump in entropy and enthalpy. Migration of the replicas in a REMD through the phase boundary that goes with the phase transition is a rare event, while the migration of a replica on one side of the boundary is very efficient. We still expect the potential of mean force to be correct, because for an accurate potential of mean force we only require sufficient sampling along the reaction path and this criterion is still fulfilled.

The maximum size of the peptide that can be studied by REMD umbrella sampling is limited by the efficiency with which the phase space that matters to the potential of mean force can be sampled. If this phase space is characterized by a rugged free-energy landscape perpendicular to the reaction coordinate, exhaustive sampling of the
phase space is very time consuming due to slow barrier crossing. On the other hand if this phase space has a smooth free-energy landscape perpendicular to the reaction coordinate sampling can be short. In practice the size limit in a dissociation or association reaction that involves conformational changes is determined by the efficiency of sampling the complete conformational phase space. At present, the conformational phase space of peptides larger than \(\sim 7\) residues is generally too large, \textit{i.e.} they have too many relevant conformations, to be sampled exhaustively with high resolution computations. This will set the maximum size of the peptide in the current implementation of the method to \(\sim 7\) residues. However if there are no conformational changes involved, the relevant part of the free-energy landscape is smooth and we expect the size limit to be equal to the size limit of standard molecular dynamics.

The REMD umbrella sampling as presented here can be extended to study peptides involved in conformational changes during association or dissociation that are larger than \(\sim 7\) residues. For instance, the simulation time can be increased, thereby giving the system more time to cross the barriers and visit all the minima. Another approach is to introduce a second set of potentials for every umbrella potential, to enhance conformational sampling perpendicular to the umbrella potential, extending the current one-dimensional method to a two-dimensional REMD.
Quantitative prediction of amyloid fibril growth of short peptides: Calculating association constants to dissect side chain importance


<table>
<thead>
<tr>
<th>Peptide</th>
<th>$K_I$ ($M^{-1}$)</th>
<th>$\Delta G_\beta$ (RT)</th>
<th>$\Delta G_{lat}$ (RT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KFFE</td>
<td>$2.9 \cdot 10^4$</td>
<td>-5.4</td>
<td>-4.9</td>
</tr>
<tr>
<td>KLEE</td>
<td>$2.5 \cdot 10^3$</td>
<td>-3.3</td>
<td>-4.5</td>
</tr>
<tr>
<td>KVVE</td>
<td>$1.5 \cdot 10^4$</td>
<td>-4.3</td>
<td>-5.4</td>
</tr>
</tbody>
</table>

$K_I$ = association constant  
$\Delta G_\beta$ = $\Delta G$ of $\beta$-sheet interaction  
$\Delta G_{lat}$ = $\Delta G$ of lateral interaction
Abstract Quantitative prediction of the fibril growth properties of different peptides is conducted with a molecular dynamics approach. Association constants of small peptides used as a model for amyloid formation are calculated and the results show very good agreement with experiments. Also the free-energy differences associated with two important interactions that characterize fibril growth, namely cross-$\beta$-sheet and lateral interactions, are obtained. These two interactions show different dependencies on the physicochemical properties of the side chains, explaining the variation in fibril morphologies between different peptides.

Introduction

Amyloid fibrils are organized peptide aggregates with high $\beta$-sheet character. Their occurrence under pathological conditions, e.g. Alzheimer and Huntington disease, has gained this class of structures widespread attention [29, 32]. Functional amyloid fibrils like silk [29] have been known for a long time. Recent interest has been raised in the field of materials science, i.e. for the fabrication of nanowires on amyloid fibril templates [34, 37].

An amyloid fibril is composed of multiple peptides that are linked through intermolecular $\beta$-sheet interactions, which extend along the fibril axes to form a cross-$\beta$-sheet. Perpendicular to the fibril axes the cross-$\beta$-sheets interact laterally by hydrophobic clustering of the side chains. The structural details of an amyloid fibril are sequence specific and although the fibril structures can be solved experimentally, this still presents a challenging task [56, 67].

It appears that every peptide or protein has the ability to form amyloid fibrils, however, the propensity varies strongly for different peptides [143]. Peptides as small as four residues have been shown to possess the properties necessary to form fibrils at physiological conditions [87]. Analyses of these small peptides by computer simulations [50, 51, 82, 159], together with a broad analysis of natural occurring mutations in disease related amyloid fibrils [160, 161] show that physicochemical interactions such as charge, hydrophobicity and secondary structure preference direct fibril formation. Although these considerations are successfully used to make a qualitative assessment of the propensity of (part of) a peptide to form fibrils [162–164], a quantitative assessment requires more detail.

At present, the mechanism of amyloid fibril formation is considered to be a nucleation-growth mechanism [32, 140, 144]. Formation of a stable nucleus precedes rapid growth of the amyloid fibril until equilibrium is reached. Although the first step is under kinetic control, fibril growth is under thermodynamic control [53, 90, 145, 146] and can be evaluated quantitatively in terms of equilibrium properties such as association constants. Here we examined the propensity of different polypeptides for amyloid fibril growth calculated by the association constant using all-atom MD with explicit solvent.

We examined the fibril growth properties of four tetrapeptides, KFFE, KVVE, KLLE and KAAE. Previously it was shown that KFFE and KVVE form amyloid fibrils, while KLLE and KAAE do not [79]. A dimerization study (nucleation) of these tetrapeptides
showed that the fibril forming propensity of KFFE and KVVE is the result of a hydrophobic collapse. Although KLLE also benefits from this collapse it suffers from a large entropy penalty upon formation of the dimers [159]. While KLLE and KAAE clearly face a higher kinetic barrier to form a nucleus, the question remains whether they can still form full length amyloid fibrils once the nucleus is formed. Therefore we assessed the growth propensity of mature fibrils.
Results and discussion

We started by constructing hypothetical amyloid fibrils from ten peptides for each tetrapeptide and tested the stability by a 10 ns standard MD simulation. The ten peptides have been arranged in two cross-β-sheets of five peptides each that interact through a lateral interaction (Figure 4.1). Within this simulation the KFFE, KLLE and KVVE fibrils appeared to be stable, while the KAAE fibril quickly converted to a random aggregate.

For the three stable fibrils we sampled the distance between the dissociated peptide and the closest cross-β-sheet bonded peptide in an umbrella sampling simulation [13] combined with replica exchange to accelerate convergence [148,149]. From this simulation we calculated the potential of mean force (PMF) that describes the association-dissociation of one peptide from the fibril (Figure 4.2) with the weighted histogram analysis method [14, 15]. From the PMF we calculated the association constants $K_1$, listed in table 4.1. To obtain quantitative results we used the Gromos force field 56a6 specially parameterized to calculate accurate solvation free energies [153].

We repeated this procedure with a hypothetical amyloid protofilament composed of five peptides in a cross-β-sheet conformation (Figure 4.1) to calculate PMF’s (Figure 4.2). The corresponding association constants $K_2$ are listed in table 4.1.

If we assume the number of fibrils in solution to be constant the critical monomer concentration for fibril formation is $1/K_1$. We found critical monomer concentrations of 34, 400 and 67 µM for KFFE, KLLE and KVVE respectively. For a 200 µM monomer concentration, used in the experiments of Tjernberg et al. [79], we expect KFFE and KVVE to form fibrils as opposed to KLLE, in very good agreement with their findings.

The association constant is related to the free-energy difference through
Results and discussion

\[ \Delta G = -RT \ln K. \] The free-energy difference associated with the cross-\(\beta\)-sheet interaction \(\Delta G_\beta\) can thus be estimated from the protofilament association constant. Similarly the fibril association constant is related to \(\Delta G_\beta + \Delta G_{lat}\), with \(\Delta G_{lat}\) the free-energy difference of the lateral interaction between two peptides. \(\Delta G_\beta\) and \(\Delta G_{lat}\) are given in table 4.1.

The origin of the overall negative value of \(\Delta G_\beta\) and \(\Delta G_{lat}\) is the hydrophobic collapse as indicated by the hydrophobic solvent accessible surface (SAS). We observed a reduction in the hydrophobic SAS corresponding to \(\Delta G_\beta\) of 2.4, 2.1 and 2.0 nm\(^2\) per peptide of KFFE, KLLE and KVVE respectively. For \(\Delta G_{lat}\) we found a reduction of 0.6, 0.8 and 0.3 nm\(^2\) per peptide of KFFE, KLLE and KVVE respectively. However, the hydrophobic collapse alone cannot explain the order of \(\Delta G_\beta\) and \(\Delta G_{lat}\).

The propensity to form cross-\(\beta\)-sheets is KFFE > KVVE > KLLE (\(\Delta G_\beta\) in table 4.1). To explain this order the side chain orientation was compared between the monomer and the protofilament in addition to an evaluation of the hydrophobic SAS. KFFE then ranks first, as it has the largest burial of the hydrophobic SAS and the orientation of the F side chains does not require any adjustment (Figure 4.3a,b). Next, the hydrophobic SAS slightly favors KLLE over KVVE while both peptides change side chain orientation (Figure 4.3d,e + 4.3g,h), to obtain an optimal packing. For KVVE this change only results in a small entropy loss as one orientation is now favored over another, whereas for KLLE the favorable interaction resulting in the orientation preference of the monomer has to be disrupted in addition to the entropy loss.

The order for \(\Delta G_{lat}\) is KVVE > KFFE \(\approx\) KLLE. Although both KFFE and KLLE show a larger reduction in hydrophobic SAS than KVVE, they also suffer from excluded volume effects. The excluded volume of the F and L side chains forces them to adopt a less favorable orientation (Figure 4.3a,c and 4.3d,f respectively). This is not observed for KVVE (Figure 4.3g,i).

The quantitative assessments made here are in very good agreement with experimental results. In addition we show that the the physicochemical properties used to make a qualitative assessment are a consequence of the hydrophobic collapse (fibril growth promoting), conformational entropy and excluded volume effects (both decreasing fibril growth propensity).

The hydrophobic collapse, conformational entropy and excluded volume effect all contribute to both \(\Delta G_\beta\) and \(\Delta G_{lat}\). As the relative contributions of these three interactions varies strongly for different side chains, so do the relative strengths of the cross-\(\beta\)-sheet and lateral interaction. This explains why the morphology and details of an

<table>
<thead>
<tr>
<th></th>
<th>(K_1 (\text{M}^{-1}))</th>
<th>(K_2 (\text{M}^{-1}))</th>
<th>(\Delta G_\beta \ (\text{RT}))</th>
<th>(\Delta G_{lat} \ (\text{RT}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>KFFE</td>
<td>(2.9 \cdot 10^4)</td>
<td>(2.5 \cdot 10^2)</td>
<td>(-5.4)</td>
<td>(-4.9)</td>
</tr>
<tr>
<td>KLLE</td>
<td>(2.5 \cdot 10^2)</td>
<td>26</td>
<td>(-3.3)</td>
<td>(-4.5)</td>
</tr>
<tr>
<td>KVVE</td>
<td>(1.5 \cdot 10^4)</td>
<td>73</td>
<td>(-4.3)</td>
<td>(-5.4)</td>
</tr>
</tbody>
</table>

Table 4.1 — Association constants and free energies related to amyloid fibril growth.
**Chapter 4. Predicting fibril growth properties of short peptides**

**Figure 4.3 — Side chain orientation.** Probability distributions of the $\chi_1$ dihedral angle of the KFFE, KLLE and KVVE peptides in a monomer, protofilament and fibril environment.

Amyloid fibril are sequence specific.

**Method**

**Theory**

We use replica exchange molecular dynamics (REMD) umbrella sampling in order to sample the fibril and protofilament growth. To that end a number of independent simulations (windows) are performed with the addition of different biasing (umbrella) potentials $V_{\text{bias}}$ to the standard potential of the system $V_{\text{standard}}$.

\[
V_{\text{total}} = V_{\text{standard}} + V_{\text{bias}}
\]

\[
V_{\text{bias}} = \frac{1}{2} f_c (d_{ij} - d_{\text{COM}})^2
\]

Here $V_{\text{total}}$ is the total potential energy in one simulation, $d_{ij}$ is the COM distance between the terminal peptide in the assembly and the closest neighbor cross-$\beta$-sheet bonded peptide, $d_{\text{COM}}$ is the reference COM distance and $f_c$ the force constant. The biasing potential $V_{\text{bias}}$ in one simulation restrains the COM-distance $d_{ij}$ around the reference COM distance $d_{\text{COM}}$, which is chosen on the reaction path.
The efficiency of umbrella sampling can be increased by adding replica exchange steps [148–150]. Obtaining the potential of mean force with umbrella sampling already requires N copies of the system with different biasing potentials. As a standard umbrella sampling already requires sufficient overlap in the energy distribution of neighboring windows, using the windows as replicas in a Hamiltonian-REMD [18] is straightforward and can enhance sampling without additional computational cost [148, 149].

An attempt to exchange the Hamiltonian for replica $i$ and $j$ is accepted according to the following Metropolis criteria:

$$p_{acc}(i, j) = \begin{cases} 1 & \Delta(i, j) \leq 0 \\ e^{-\Delta(i, j)} & \Delta(i, j) > 0 \end{cases}$$ (4.3)

$$\Delta(i, j) = \beta \left[ V_{total,i}(q_i) - V_{total,j}(q_j) \right] - \left[ V_{total,j}(q_j) - V_{total,i}(q_i) \right]$$ (4.4)

with $p_{acc}$ the acceptance probability and $\beta = 1/k_B T$, where $k_B$ is Boltzmann’s constant and $T$ the temperature. This exchange depends on the potential energy difference when the potential energy functions are swapped between system $i$ and $j$, i.e. when the particle positions currently subjected to the potential energy function $i$ and $j$ are then subjected to the potential energy function of $j$ and $i$ respectively. As the potential energy of replica $i$ and $j$ only differ in the contribution of the biasing potential equation 4.4 can be reduced to

$$\Delta(i, j) = \beta \left[ V_{bias,i}(q_i) - V_{bias,j}(q_j) \right] - \left[ V_{bias,j}(q_j) - V_{bias,i}(q_i) \right]$$ (4.5)

To obtain the potential of mean force associated with these reactions from the separate windows we need to solve the WHAM equations [14, 15] 4.6 and 4.7 self-consistently.

$$\rho^u(r) = \sum_{i=1}^{N_{rep}} \frac{\rho^b_i(r)}{\sum_{j=1}^{N_{rep}} e^{-\beta V_{bias,j}(r) - f_i}}$$ (4.6)

$$e^{-\beta f_i} = \int dr e^{-\beta V_{bias,i}(r)} \rho^u(r)$$ (4.7)

In these equations $r$ is the distance between two particles $i$ and $j$ (the reaction coordinate), $N_{rep}$ the number of replicas and $f_i$ a constant that represents the free energy associated with the introduction of the biasing potential of replica $i$. With equations 4.6 and 4.7 the biased probability density, $\rho^b_i(r)$, from the different replicas are combined into one unbiased probability density, $\rho^u(r)$. Before we calculate the potential of mean force through $W(r) = -k_B T \ln \rho^u(r)$ the unbiased probability density is scaled so that $\rho^u(r) = 1$ for large $r$.

The potential of mean force is connected to the equilibrium constant $K$ through the following relation [152]

$$K = \frac{\sigma_{int} n_{int}}{\sigma_{ij}} \int_{r_{low}}^{r_{up}} e^{-W_{ij}(r)/k_B T} \cdot 4\pi r^2 dr$$ (4.8)

with $K$ the association constant, $r_{up}$ and $r_{low}$ the upper and lower limit of the potential of mean force, respectively, $\sigma_{ij}$ the symmetry number (1 when $i \neq j$ and 2 when $i = j$),
Chapter 4. Predicting fibril growth properties of short peptides

\( n_{\text{int}} \) the number of interaction spaces and \( \sigma_{\text{int}} \) the percentage of a full sphere that is used by the interaction space.

Here we introduced the interaction space, which is the volume over which the potential of mean force acts. For a spherically symmetric system, like the dimer reaction, the whole volume is the interaction space and \( n_{\text{int}} = \sigma_{\text{int}} = 1 \). When considering fibril growth we cannot assume spherical symmetry, but it is reasonable to describe the interaction site as a cone along the fibril axis between the dissociated peptide and the closest cross-\( \beta \)-sheet bonded peptide. Therefore we estimate \( \sigma_{\text{int}} = 1/4 \) and since every fibril has two tips \( n_{\text{int}} = 2 \).

**Model systems**

To quantitatively predict the fibril growth propensity of a mature fibril we use a model system. Therefore we created a decamer, with ten peptides arranged in two cross-\( \beta \)-sheets of five peptides each, which interact through a lateral interaction (Figure 4.1). Every peptide in this decamer has all the characteristic interactions of a peptide in a mature fibril, viz. cross-\( \beta \)-sheet and lateral interactions. As we are only interested in the relative stability of different mature fibrils, less-ordered oligomers and aggregates that are at the basis of fibril formation are not taken into consideration.

To dissect the free-energy differences associated with the cross-\( \beta \)-sheet and lateral interaction we also assessed the growth propensity of a mature protofilament. This protofilament was modeled by five peptides in a cross-\( \beta \)-sheet arrangement (Figure 4.1).

To show that these two model systems are sufficiently large to represent a mature fibril and protofilament, we also calculated the potential of mean force (PMF) of a protofilament modeled by six peptides. As the PMF calculated from a model of five peptides does not show any marked differences from the PMF calculated from six peptides (Figure 4.4), we conclude that our models are sufficiently large to represent mature fibrils and protofilaments.

**Computational details**

All simulations were performed using the GROMACS [11] software package version 3.3.1 and the GROMOS96 56a6 force field [153] in combination with SPC water molecules [154]. An integration timestep of 2 fs was used, with all bonds constrained using the LINCS algorithm [155]. Van der Waals interactions were ignored outside a cut-off of 1.2\( \text{\textnormal{nm}} \). Electrostatic interactions were treated with the PME method applying a real-space cut-off of 0.9\( \text{\textnormal{nm}} \) [156]. The system temperature was coupled to a Berendsen thermostat, no pressure coupling was used.

The protofilament and fibril structures were generated manually and were dissolved in a rectangular box filled with approximately 2600 and 3200 SPC water molecules, respectively, to obtain a density of 0.99 kg/l. For the monomer structure an extended chain was used, which was dissolved in a dodecahedron box filled with approximately 1200 SPC water molecules at a density of 0.99 kg/l. For all systems initial
Method

Figure 4.4 — Dependence of PMF on the size of the model. The PMF calculated from a model of six peptides (red) does not differ significantly from that of five peptides (black). Five peptides are thus sufficient to calculate the association constants and free-energy differences associated to fibril growth of a mature protofilament.

velocities were generated randomly to create a system at 300 K after which the systems were equilibrated for 5 ns. Finally production runs of 10 ns were performed.

For the umbrella sampling the pull-code provided in GROMACS version 3.3.1 was used. After testing several values a force constant of 5000 kJ/mol was chosen for the biasing potential (eq 4.2). A higher value will decrease the width of the energy distribution and thus decrease the overlap of the potential energy between neighboring replicas thereby lowering the exchange acceptance rate. A smaller value is not able to hold $d_{ij}$ around $d_{COM}$ on the steep slope of the potential of mean force and thus some parts of the reaction coordinate will not be sampled sufficiently.

The spacing between the windows was chosen to have an average exchange between neighboring replicas of 30%, which leads to a separation between the $d_{COM}$-values of 0.04 nm. The complete reaction coordinate could be sampled in 28 and 40 replicas for the protofilament and fibril, respectively, starting from 0.42 nm in the smallest replica. An all-pair exchange [157] was used with exchange attempts every picosecond.

To generate the starting structures for all the different replicas the biasing potentials were consecutively introduced starting from the equilibrated cross-$\beta$-sheet structure and the smallest COM distance. Every window was equilibrated for another 1 ns
before production runs started. The protofilament and the fibril were kept in place by a position restraint to mimic the lack of mobility in a full-grown fibril.
PART III

NOVEL COMPUTATIONAL METHODS THAT ACCELERATE CONFORMATIONAL CHANGES IN PEPTIDES TO EXAMINE FOLDING AND AGGREGATION
Fast in silico protein folding by introduction of alternating hydrogen bond potentials

M.G. Wolf and S.W. de Leeuw, Biophysical Journal 94 (2008), 3742-3747

Abstract We accelerate protein folding in all atom molecular dynamics simulations by introducing alternating hydrogen bond potentials as a supplement to the force field. The alternating hydrogen bond potentials result in accelerated hydrogen bond reordering, which lead to quick formation of secondary structure elements. The method does not require knowledge of the native state, but generates the potentials based on the development of the tertiary structure in the simulation. In protein folding the formation of secondary structure elements, especially α-helix and β-sheet, is very important and we show that our method can fold both efficiently and with great speed.

Introduction

Computational studies and in particular molecular dynamics (MD) simulations are widely applied to study protein folding processes, providing detailed insight at atomic resolution. Unfortunately, molecular dynamics can only sample a system for a time period of nanoseconds up to a microsecond, while protein folding generally takes microseconds to milliseconds. Attempts to overcome this sampling problem involve the use of simplified models that require many assumptions [165, 166] and consequently lack microscopic detail required to monitor the various interactions that lead to folding. Alternatively, large scale distributed computing can be used [16, 167–169]. However this approach is not accessible to everyone and with explicit solvent the sampling problem cannot be overcome completely.

Folding of a protein into the native state cannot be described by a random search.
Chapter 5. Alternating hydrogen bond potentials

through all the degrees of freedom, but is believed to be a guided process [170]. Expressed in terms of free-energy landscapes, a protein traversing the free energy landscape is funneled from the high-energy unfolded conformations into the low-energy native state [100, 171, 172].

The free-energy landscape of a protein is determined by two major contributions: hydrophobic interactions and hydrogen bonding. Hydrophobic interactions, joining the hydrophobic side chains (hydrophobic collapse), are generally viewed as one of the driving forces of protein folding [173]. The role of hydrogen bonding is more debated, due to the difficulty to quantify hydrogen bond energies. As the formation of hydrogen bonds between protein atoms results in the loss of hydrogen bonds formed with water it is still unclear whether in an aqueous environment protein intramolecular hydrogen bond formation contributes favorably [174–176] or is negligible to the free energy [177, 178]. In any case during the hydrophobic collapse intramolecular hydrogen bond formation is necessary to compensate for the high free energy cost associated with burying unsatisfied hydrogen bonding groups [179]. Since the number of intramolecular hydrogen bond partners is limited, the necessity of these bonds limits the number of allowable protein conformations and thus primarily provides specificity [173, 180].

The free-energy landscape of a protein is rugged, displaying many valleys and energy barriers [181, 182]. Consequently a protein traversing this landscape is very likely to encounter valleys corresponding to local free-energy minima, before reaching the valley corresponding to the native state. As the valleys are generated by an accumulation of individual interactions (hydrophobic and hydrogen bonds) crossing a barrier requires some of these interactions to be disrupted. Even a very simple peptide, like penta-alanine, possesses a rugged free-energy landscape, with many valleys of comparable free energy [183]. Studies have shown that a barrier height of only 2 $k_B T$ is sufficient to prevent the folding of a protein to occur in a downhill manner [184]. The free energy barrier associated with a single intramolecular hydrogen bond is approximately 3 $k_B T$ in a protein in water [185–187], preventing straightforward folding.

Nature resolves the slow folding problem of proteins by using chaperones to assist in the folding of proteins inside cells. Chaperones prevent association with other proteins and limit the number of accessible conformations [188, 189]. In addition they can actively stimulate proteins to (partially) unfold, allowing the protein to escape from a local free-energy minimum and have another attempt at folding. In this way transitions in the free-energy landscape are facilitated and thus the folding rate is increased [190]. Unfortunately these systems are too large for efficient use in MD simulation and in silico folding experiments.

Here we propose a novel computational method based on the idea that occasional (partial) unfolding of a protein enhances the frequency of barrier crossing and the folding rate of proteins. We perform MD simulations during which we periodically introduce temporary additional forces that alternatingly stimulate unfolding and folding. These forces act on the intramolecular hydrogen bonds. The first reason for this is because distinct hydrogen bonds in a similar context contribute equally to the free
energy, but a free energy barrier separates all the possible hydrogen bonds. In other
words, hydrogen bonds provide kinetic stability both in the global minimum and in lo-
cal minima rather than thermodynamic stability. This has important implications: Un-
folding and folding can be stimulated by reimbursing the activation energy set by the
kinetic barrier of a hydrogen bond. In addition the hydrogen bonds provide specificity
rather than stability with respect to the tertiary structure of a protein, which means that
the interactions that provide thermodynamic stability are unaltered and still guide the
folding process of the protein into its native state, while the time in free-energy min-
ima is decreased. A second more technical reason for influencing the intramolecular
hydrogen bonds is that the number of required additional forces is minimal. This is
because the number of donor-acceptor pair combinations in a protein is limited and
the hydrogen bonds are orientation dependent [191], requiring introduction of only a
few relevant hydrogen bond potentials.

The manipulation of the hydrogen bonds is performed within a single MD sim-
ulation, where alternatingly attractive or repulsive hydrogen bond potentials are in-
troduced in addition to the standard force field potentials. The repulsive potential
destabilizes the hydrogen bonds and lifts the protein to a higher free-energy level.
The attractive potential in turn facilitates hydrogen bond formation to enable a fast
identification of the conformational regions of free-energy minima. Such local un-
folding/folding mechanism would be comparable with the barrier crossing effect of a
chaperone protein. In this method we do not need a priori information on the native
state; rather we use the structure of the protein as it develops during the simulation to
determine which potentials are introduced.

We show that manipulation of hydrogen bonds during an MD simulation can accel-
erate the folding of a protein. The two secondary structure elements appearing most,
α-helix and β-sheet, can be folded efficiently. This is demonstrated by the folding of
a 16 residue polyalanine to the α-helical native state and the 16 residue C-terminal of
the 1GB1 protein to the β-hairpin native state.

**Method**

*Alternating hydrogen bond potentials (AHBP)*

We introduce a hydrogen bond potential $V_{hb}$ as a supplement to the standard force
field, which acts on the atoms involved in hydrogen bonding in order to accelerate pro-
tein folding in MD simulations. This is implemented as a staged molecular dynamics
protocol, where we distinguish three stages, the repulsive stage (R), the attractive stage
(A) and the relaxation stage (E). These three stages each treat hydrogen bonds differ-
ently. In R a potential stimulates hydrogen bond breakage, in A a potential facilitates
hydrogen bond formation and in E the system is allowed to relax. In our simulations
each stage is active for 0.5 ps in the order -(R–E–A–E-)–n.

When a stage is active, every 0.1 ps all intramolecular donor-acceptor pairs of the
protein are evaluated. The relevant pairs are selected and potentials are introduced
that will result in a force acting on the atoms. During selection we avoid targeting
strong hydrogen bonds, because they can be native and try to introduce only one potential per atom to minimize the manipulation of the system. Therefore a pair is excluded from selection if it (a) is a strong hydrogen bond (characterised by a donor-acceptor distance less than 0.35 nm and a donor-hydrogen-acceptor angle larger than $120\degree$) (b) the atoms of the pair are involved in another strong hydrogen bond and (c) the atoms in the pair are already targeted in another hydrogen bond potential (e.g. from a previous evaluation). For the remaining donor-acceptor pairs those with the largest hydrogen bond potential energy (eq 5.1) are selected, with the rule that the atoms in a pair may only be selected once. At the end of each stage all hydrogen bond potentials are removed.

The hydrogen bond potential $V_{hb}(q, t)$ is given in (eq 5.1).

$$V_{hb}(q, t) = f_c(q, t) \cdot E_d(q(t_{ev})) \cdot E_\theta(q(t_{ev}))$$  (5.1)

It is a function of time $t$ and consists of a distance potential $E_d(q(t_{ev}))$, an angle potential $E_\theta(q(t_{ev}))$, a gradually changing force constant $f_c(q, t)$ and the positions of the atoms in the hydrogen bonds $q$.

In the repulsive stage the distance potential $E_d(q(t_{ev}))$ is determined by the distance $d$ (nm) between donor and acceptor (Figure 5.1) at the evaluation time $t_{ev}$. Cutoff distances $d_{min}$ and $d_{max}$ of 0.35 and 0.40 nm are used respectively. For the attractive stage the distance between hydrogen and acceptor (Figure 5.1) is considered and the cutoff distances $d_{min}$ and $d_{max}$ are 0.23 and 0.40 nm respectively. The values of the cutoff distances ensure that only weak to very weak hydrogen bonds are targeted, so that in the repulsive stage the hydrogen bond is pushed up the last part of the free-energy barrier, and in the attractive stage the formation of hydrogen bonds that would possibly be formed in the future is accelerated.

$$E_d(q(t_{ev})) = \begin{cases} 
1 - \frac{d(t_{ev}) - d_{min}}{d_{max} - d_{min}} & d_{min} \leq d(t_{ev}) < d_{max} \\
0 & d_{max} \leq d(t_{ev}) 
\end{cases}$$  (5.2)

The angle potential $E_\theta(q(t_{ev}))$ depends on the angle $\theta$ (degrees) of the donor hydrogen acceptor (Figure 5.1) at activation time $t_{ev}$. The cutoff angle $\theta_{bound}$ in the repulsive stage is set to $120\degree$, which ensures targeting all weak hydrogen bonds, and in the attractive stage to $60\degree$, allowing generation of many hydrogen bonds.

$$E_\theta(q(t_{ev})) = \begin{cases} 
1 & \theta(t_{ev}) \geq \theta_{bound} \\
0 & d_{bound} > d(t_{ev}) 
\end{cases}$$  (5.3)

The potential is slowly introduced and removed from the system to avoid cut-off effects at the boundaries of the distance and angle potential. This is achieved by growing or shrinking the force constant in 50 increments to a maximum force constant or zero respectively. Each time step the counter $n_{inc}$ is incremented by one if the derivative of the distance and angle potential is non-zero until the maximum value of 50 is reached and one is subtracted when this derivative is zero until the counter is zero. The force constant is then given by (eq 5.4), with $0 \leq n_{inc} \leq 50$. The number of 50 increments
Method

Figure 5.1 — A schematic view of a Donor-Acceptor Pair including the corresponding Hydrogen. The relevant quantities in the hydrogen bond potential and the resulting forces are displayed.

to reach the maximum force constant is chosen arbitrarily within the idea of gradually introducing the forces in the system to its maximum.

\[ f_c(q, t) = \frac{1}{50} \cdot n_{inc} \cdot f_{c\text{max}} \] (5.4)

To obtain the maximum force constant several values were tested and the values showing a good response, i.e. many unfolding and folding events, were used. The hydrogen bond potential leads to the introduction of the following force acting on the acceptor atom (Figure 5.1).

\[ F_A = f_c(q, t) \cdot \begin{cases} \frac{1}{d_{max}^2 - d_{min}^2} \cdot d_{X,A(t_{ev})} & d_{min} \leq d(t_{ev}) < d_{max}; \theta(t_{ev}) \geq \theta_{bound} \\ 0 & \text{rest} \end{cases} \] (5.5)

The balancing force is \( F_x = -F_A \). In these equations the X refers to the donor atom in the repulsive stage and to the hydrogen atom in the attractive stage (Figure 5.1). Both introducing the AHBP potentials only at evaluation time as well as the gradual introduction/ removal of the potential means that the energy is not conserved. The possible numerical instability is resolved in a constant temperature ensemble by means of the heat bath.

Simulation protocol

All simulations were performed using the GROMACS [11] software package version 3.3.1 and the GROMOS96 43a1 forcefield [192] in combination with SPC water [154].

73
Chapter 5. Alternating hydrogen bond potentials

Figure 5.2 — Average hydrogen bond lifetime. The autocorrelation function of the intramolecular hydrogen bonds in a 16-residue polyalanine shows that the average hydrogen bond lifetime in an AHBP simulation is shorter than in a standard simulation. The autocorrelation function is averaged over all simulations.

timestep of 2 fs was used, with all bonds constrained using the LINCS algorithm [155]. Van der Waals interactions were ignored outside a cut-off of 1.2 nm. Electrostatic interactions were treated with the PME method applying a real-space cutoff of 0.9 nm [156]. The system temperature was coupled to a Berendsen thermostat, no pressure coupling was used.

The starting structures of the simulations both for the α-helix and for the β-hairpin simulations were extended conformations. The proteins were dissolved in a box of approximately 2000 SPC water to obtain a density of 0.99 kg/l and initial velocities were generated randomly to generate a system at 300 K. In the case of the α-helix this system was equilibrated for 1 ns in order to allow the chain to collapse. For every production run new random velocities corresponding to a system at 300 K where generated.

Results

The method presented above aims to accelerate in silico protein folding. This is achieved by manipulating the intramolecular hydrogen bonds, leading to an increase in the number of barrier transitions. To show that this is indeed the case, the time behavior of a 16-residue polyalanine was examined with standard MD (4 simulations of 30 ns) and with AHBP-MD (5 simulations of 10 ns). The simulations were started from a collapsed coil, which represent a structure in a local minimum possessing many hydrogen bonds. The maximum force constant used in the MD simulation including AHBP were -600 kJ mol\(^{-1}\) nm\(^{-1}\) for the attractive potential and 450 kJ mol\(^{-1}\) nm\(^{-1}\) for the repulsive
To test if the faster and broader sampling of the conformational space of a protein by the AHBP-MD simulations leads to fast formation of secondary structure elements two systems were tested. The polyalanine simulations used to show enhanced barrier crossing in AHBP-MD were also used to test the ability of the AHBP method to form \( \alpha \)-helical secondary structure. To test the \( \beta \)-sheet secondary structure formation we investigated the folding of the 16 residue C-terminus of the protein G (PDB-code 1GB1), which adopts a \( \beta \)-hairpin conformation in an aqueous environment. We performed 10 standard MD simulations of 50 ns and 10 AHBP-MD simulations of 30 ns, which all started from an extended conformation. In these AHBP-MD simulations of the \( \beta \)-hairpin we used a maximum force constant of -300 and 900 kJ mol\(^{-1}\) nm\(^{-1}\) for the attractive and the repulsive potential respectively.

**Hydrogen bond reordering**

The hydrogen bond lifetime and the average number of hydrogen bonds in the polyalanine simulations were evaluated to show that the intramolecular hydrogen bond reordering is accelerated. To compare the hydrogen bond lifetime in standard MD and AHBP-MD the autocorrelation function of the intramolecular hydrogen bonds is displayed in Figure 5.2. Although the autocorrelation function does not converge within the simulated time and thus an accurate estimate of the hydrogen bond lifetime is im-
possible, it is clear that this function decreases significantly faster in the AHBP-MD simulation than it does in the standard MD simulation, showing that the lifetime of hydrogen bonds is much smaller due to AHBP. The average number of intramolecular hydrogen bonds per molecule under AHBP conditions is larger than under standard conditions, having values of 4.7 and 3.9 respectively. So a smaller hydrogen bond lifetime and a larger average number of hydrogen bonds indicate that formation of new hydrogen bonds and opening of old ones is faster in the AHBP-MD simulations than in standard MD simulations.

To show that this fast reordering of hydrogen bonds also leads to a fast and broad sampling of the conformational space, we constructed RMSD matrices for all trajectories. An element in this matrix consists of the RMSD value between the structure of the protein at time x and time y. This value is calculated after a translational and a rotational fit on the backbone of the proteins. In an RMSD matrix the similarity of each conformation in the trajectory with all the other conformations is measured by their RMSD value, with a low value indicating strong structural similarity. In these matrices white (= low RMSD) squares along the diagonal indicate time intervals where the protein structure remains very similar. Comparing the white squares in the RMSD matrix of the standard MD simulation to those of the AHBP-MD simulations shows that a standard MD simulation samples longer in a free-energy minimum, characterized by relatively similar structures, while the AHBP-MD simulation increases the sampling of the conformational space of a protein (Figure 5.3).
Results

Figure 5.5 — Dominating $\alpha$-helix secondary structure in polyalanine. Secondary structure as a function of time for one of the polyalanine in an AHBP-MD simulation. Below the graph some representative structures are depicted.

$\alpha$-helix

For the polyalanine simulations the average number of residues in an $\alpha$-helical conformation is plotted versus time in Figure 5.4. The N- and C-terminus are not taken into account since they are too mobile. From this Figure it is clear that within the very short time of the AHBP simulation fast formation of $\alpha$-helix secondary structure occurs. The fastest formation of a full helix is observed within 6 ns (Figure 5.5) and all simulations show formation of $\alpha$-helical structure elements. In our four standard MD simulations we observe only one short instance of $\alpha$-helix formation (data not shown), confirming that $\alpha$-helix formation is much faster and more abundant when AHBP is turned on.

Although folding of a full $\alpha$-helix is very fast with the AHBP method, the structure is not stable for the rest of the simulation (Figure 5.5). Clearly AHBP provides an efficient way to escape free-energy minima, but this includes the global minimum as well. Especially in the case of a polyalanine, which gains only minor stability in the global minimum [183], the protein is expected to leave this minimum quickly. On average however the $\alpha$-helix is the most visited conformation arising 39 % of the simulation time.
β-sheet

To test for β-sheet formation in the simulation of the folding of 1GB1 β-hairpin, we plotted the average number of residues in a β-sheet conformation versus simulation time (Figure 5.6). In the AHBP-MD simulations a steady rise of the number of residues in a β-sheet conformation is observed, while in the standard MD simulations this number is not as high and not as consistent. So in addition to α-helix formation, AHBP-MD simulations can also lead to fast formation of β-sheet secondary structure.

To establish if the β-sheet structures formed resemble the native state, the root mean square deviation (RMSD) between the folding trajectory and the NMR structure is shown in Figure 5.7. An RMSD value below 0.28 nm indicates a very good structural overlap and simulations reaching this value are folded into the native state. Figure 5.7 shows the best performing, i.e. lowest RMSD value, AHBP and standard MD simulations. It is clear that the AHBP-MD results in fast folding to the native state, which does not occur in standard MD simulations. Four out of ten simulations reach a structure similar to the NMR structure [193] within 30 ns of folding simulation. Some of the remaining folding simulations also yield a β-hairpin conformation, but non-native side-chain interactions lead to an RMSD value of approximately 0.5 nm. However these structures still show a very good structural overlap of the backbones.

Figure 5.6 — Formation of β-sheet structural elements by AHBP and standard MD simulations starting from an extended chain. Average number of amino acid residues in a β-sheet as a function of time for the AHBP-MD (black) and the standard MD (grey) simulations.
Figure 5.7 — Comparison of the β-sheets folded using AHBP-MD with the experimental NMR-structure. Root mean square deviation (RMSD) between the simulation trajectory and the NMR structure as a function of time. The best performing AHBP-MD (black) and standard MD (grey) simulation are plotted.

Discussion

We introduced alternating hydrogen bond potentials in an MD simulation as a supplement to the force field in order to accelerate in silico protein folding. This method proves capable of fast folding of the two most important secondary structure elements, as shown by folding of the α-helix of polyalanine and the β-hairpin of the C-terminus of protein G.

Many folding studies rely on very simple models [165, 166] to sample a longer time-span in a simulation. Especially implicit solvent models [194, 195] are widely applied, however the clear importance of an explicit solvent model in protein folding cannot be ignored [196–198]. With our method microscopic detail is retained, as well as an explicit description of the solvent molecules. These requirements can also be achieved with parallel molecular dynamics schemes [169], however these schemes are limited to small fast folding proteins due to the slow dynamics of each individual simulation. The proposed method increases this speed and thus requires far less computer time to obtain similar folding ensembles. Other methods that increase the dynamics of individual simulations are for instance a self-guided MD approach [199], but this results in an irreversible folding path. AHBP is a reversible method as shown by the unfolding events in the polyalanine simulations.
Reordering hydrogen bonds using Hamiltonian replica exchange enhances sampling of conformational changes in biomolecular systems

J. Vreede, M.G. Wolf, S.W. de Leeuw and P.G. Bolhuis, submitted
Abstract Hydrogen bonds play an important role in stabilizing (meta-)stable states in protein folding. Hence, they can potentially be used as a way to bias these states in molecular simulation methods. Previously, Wolf et al. showed that applying repulsive and attractive hydrogen bond biasing potentials in an alternating way significantly accelerates the folding process (M.G. Wolf and S.W. de Leeuw, Biophys. J. 2008, 94, 3742). As the biasing potentials are only active during a fixed time interval, this alternating scheme does not represent a thermodynamic equilibrium. In this work we present a Hamiltonian REMD scheme that aims to shuffle and reorder hydrogen bonds in the protein backbone. We therefore apply adapted hydrogen bond potentials to apply them in a Hamiltonian REMD scheme, which we call hydrogen bond switching (HS). To compare the performance of the HS to a standard REMD method, we performed HS and temperature REMD simulations of a β-heptapeptide in methanol. Both methods sample the conformational space to a similar extent. As the HS simulation required only five replicas, while the REMD simulation required 20 replicas, the HS method is clearly significantly more efficient. We tested the HS method also on a larger system, 16 residue polyalanine in water. Both the simulations starting from a completely unfolded and a folded conformation resulted in an ensemble with, apart from the starting structure, similar conformational minima. We can conclude that the HS method provides an efficient way to sample the conformational space of a protein, without requiring knowledge of the folded states beforehand. In addition, these simulations revealed that convergence was hampered by replicas having a preference for specific biasing potentials. As this sorting effect is inherent to all Hamiltonian REMD method, finding a solution will result in an additional increase in their efficiency in general.

Introduction

Molecular simulation is a powerful tool to perform detailed investigations of molecular processes in biological systems, such as protein conformational changes. The study of a protein folding reaction at atomic resolution poses a problem as it involves time scales that exceed the range of conventional atomistic methods, such as Molecular Dynamics (MD). The folding of a protein involves traversing a rugged free energy landscape with many high barriers. Consequently, many (local) free energy minima exist including unfolded states, partially folded states and the native state, which are not trivial to escape from. Nevertheless, the folding of a protein is overall a guided process [170], in which several types of interactions play an important role. Hydrophobic interactions are the driving force behind the collapse of a polypeptide chain [173], partially caused by the entropy gain of the solvent molecules. During this collapse, the formation of intramolecular hydrogen bonds can compensate the high free energy cost associated with burying unsatisfied (partially) charged groups and loss of solvent interactions [179]. The free energy barrier associated with breaking a single intramolecular hydrogen bond is approximately $3 k_B T$ in a protein in water, which is not a high barrier in itself [185–187]. The formation of one non-native hydrogen bond, however, may lead to additional non-native contacts, eventually resulting in a misfolded state. Such a misfolded structure may have to unfold again for a new attempt to find the native state. As such, even one incorrectly formed hydrogen bond may lead to kinetic traps.
that slow down the folding process enormously.

Introducing potentials that target the breaking and formation of backbone hydrogen bonds can therefore achieve a significant speed-up when simulating protein folding in explicit solvent, as was recently shown [200]. When applying hydrogen bond forming and hydrogen bond breaking potentials in an alternate way, fast reordering of the backbone hydrogen bonds occurs, leading to accelerated folding of the protein [200]. In this alternating hydrogen bond potential scheme, the time interval during which the hydrogen bond breaking/forming potential acts on the system is tuned to optimize folding times. Also, alternating hydrogen bond potentials do not conserve the Hamiltonian. These two aspects prevent obtaining thermodynamic properties directly from such simulations.

Alternatively, replica exchange MD (REMD), also known as parallel tempering [17], is a method often used to escape local minima in biomolecular systems [16]. In REMD, a number of MD simulations run simultaneously at different temperatures. By letting the replicas exchange temperatures, according to a Metropolis criterion, a system can diffuse through temperature space and overcome barriers at high temperature, while sampling the stable states at the temperature of interest. Afterwards, the free energy can be obtained by constructing a histogram along any order parameter. Although the application of this method to protein folding has led to several new insights [103, 201], the gain in efficiency is not as large as expected [202]. As neighboring replicas must have sufficient overlap in potential energy for an efficient exchange, many replicas (20 to 70, depending on system size) are required between the temperature of interest and high temperatures that facilitate fast conformational changes. A possible solution to this problem is letting the replicas switch (part of) their Hamiltonians instead of their temperatures [18, 149, 203]. A smooth interpolation between the regular Hamiltonian and one that enables swift conformational changes of the protein allows for efficient sampling of the conformational space. Such Hamiltonians target for example hydrophobic interactions [204] or torsional interactions [205]. In this work we present continuous hydrogen bond potentials that we use in a Hamiltonian replica exchange scheme: hydrogen bond switching (HS).

Hydrogen bond switching is in principle applicable to any system containing hydrogen bonds. The potentials are derived in such a way that only the atoms in the weakest hydrogen bond will feel a repulsive interaction, leading to the breaking of that hydrogen bond. Similarly, the atoms closest to forming a hydrogen bond will undergo an attractive interaction, leading to the formation of that hydrogen bond. In this work, we only targeted the hydrogen bonds in the protein backbone. Before setting up an HS simulation, no knowledge of hydrogen bonds in stable states or the order in which hydrogen bonds will break is required.

As a test system, we use a $\beta$-heptapeptide solvated in methanol. This system is small enough to enable a direct comparison between conventional MD and temperature REMD, in favor of REMD, as Periole et al. have shown in great detail [206]. We show that there is a significant gain in efficiency, compared to REMD, when using HS simulations to extensively sample the conformational space of this system. We also
study 16-residue-polyalanine, \(A_{16}\), a system with more hydrogen bonds. As shown by solid state NMR, polyalanine tends to adopt an \(\alpha\)-helical conformation [207]. In solution, polyalanine forms fibrillar aggregates when containing more than 15 alanine residues [208], indicating that the \(\alpha\)-helical conformation is not stable. Using HS simulations, we have investigated the conformational space of the (un)folding of polyalanine, thus elucidating that \(A_{16}\) favors conformations with \(\alpha\)-helical content, but also collapsed coil conformations occur. The HS simulations of polyalanine also uncovered a problem in Hamiltonian REMD methods that has not been explicitly described up to now. As replicas have a preference for a specific biasing potential, diffusion through \(\lambda\)-space is hampered. This problem is inherent to all Hamiltonian REMD methods that involve entropic barriers, presenting a challenge for future research.

Methods

Continuous hydrogen bond biasing potentials

Molecular simulations can provide thermodynamic properties, if the underlying methods, \(i.e.\) molecular dynamics (MD), sample the canonical ensemble correctly. As shown in previous work, applying attractive and repulsive hydrogen bond biasing potentials in an alternating way induces fast conformational changes in protein systems [200]. As these simulations are out of equilibrium, they cannot provide correct thermodynamic properties. In this work, we aim to apply attractive and repulsive hydrogen bond biasing potentials in a replica exchange scheme. To integrate correctly the equations of motion, using the standard Verlet algorithm, all potentials including the hydrogen bond biasing potentials have to be continuous. The biasing potentials as described in Wolf \textit{et al.} are discontinuous [200] and require modification. In addition, to reorder hydrogen bonds successfully, the biasing potentials must fulfill two other criteria. The first criterion is that the energy and the forces of the hydrogen bond biasing potentials must result in efficient hydrogen bond reordering. Reordering requires a large change in the hydrogen bond distance rather than in the hydrogen bond angle. Large changes in the hydrogen bond angle would result in the occurrence of unlikely structures. Secondly, as a hydrogen bond is a very specific interaction between a donor and an acceptor, excluding other interaction partners, the number of additional biasing potentials acting on the donor/acceptor must be restricted to one. For general applicability, the selection of these hydrogen bond interaction pairs must be based on the conformation in the simulation. The continuous hydrogen bond biasing potential \(V_{hb}(\mathbf{r})\) (Equation 6.1) matches these criteria:

\[
V_{hb}(\mathbf{r}) = f_c \cdot \sum_k^{D,A} \text{max}[u_{hb,X}(r_D, r_{HD}, r_A)]
\]

(6.1)

Here \(f_c\) is the force constant, \(\mathbf{r}\) represents the coordinates of all particles in the system, and \(u_{hb,X}(r_D, r_{HD}, r_A)\) is a positive definite hydrogen bond interaction function as a function of the positions of the hydrogen bond donor \((r_D)\), the hydrogen \((r_{HD})\) and the
hydrogen bond acceptor \((r_A)\). X represents either the donor \(D\) or the hydrogen bonded to the donor \(H_D\). This potential can be either attractive, aimed at forming hydrogen bonds, or repulsive, aimed at breaking hydrogen bonds.

The repulsive (rep) and attractive (att) potentials are:

\[
V_{\text{rep}}(r) = f_c \cdot \sum_{k}^{D,A} \max[h_{hb,D}(r_D, r_{HD}, r_A)]
\]
\[
V_{\text{att}}(r) = -f_c \cdot \sum_{k}^{D,A} \max[h_{hb,H_D}(r_D, r_{HD}, r_A)]
\] (6.2)

The hydrogen bond biasing potential \(V_{hb}(r)\) is a summation over all donors \((D)\) and acceptors \((A)\) in the protein backbone. The \(\max\) function indicates that for each donor (acceptor) only the hydrogen bond pair with the largest value for the interaction function is selected to contribute to the potential. Apart from providing specificity, selecting only one interaction per donor (acceptor) minimizes the bias force introduced in the system when compared to including all possible donor-acceptor interactions. As the system evolves these interacting pairs vary, i.e. one particular hydrogen bond interaction is lost while another is activated as the structure of the protein evolves during the simulation. At such a crossing point the energy of both interactions is equal and therefore swapping the interaction potential is continuous. Also, the total momentum is conserved due to the use of pair interactions.

The hydrogen bond interaction function \(h_{hb,X}(r_D, r_{HD}, r_A)\) comprises a distance function \(u_d(d_{XA})\) and an angle function \(u_\theta(\theta_{DHDA})\):

\[
u_{hb,X}(r_D, r_{HD}, r_A) = u_d(d_{XA}) \cdot u_\theta(\theta_{DHDA})\] (6.3)

The distance function \(u_d(d_{XA})\) depends on the distance \(d_{XA} = |r_X - r_A|\). X represents either the donor \(D\) or the hydrogen bonded to the donor \(H_D\) and \(A\) represents the acceptor. For the repulsive interaction \(d_{XA}\) is the distance between \(D\) and \(A\) \((d_{DA})\).

To ensure that hydrogen bonds are formed correctly for the attractive biasing potential, \(d_{XA}\) is the distance between \(H_D\) and \(A\) \((d_{HD})\). This way the hydrogen atom is correctly oriented during formation of a hydrogen bond.

The explicit expression for the distance function \(u_d(d_{XA})\):

\[
u_d(d_{XA}) = \begin{cases} 1 & d_{XA} < d_{min} \\ \frac{2(d_{XA}-d_{min})}{(d_{max}-d_{min})^3} - \frac{3(d_{XA}-d_{min})^2}{(d_{max}-d_{min})^2} + 1 & d_{min} \leq d_{XA} < d_{max} \\ 0 & d_{max} \leq d_{XA} \end{cases}\] (6.4)

This cubic spline function smoothly varies from the maximum value 1 for distance \(d_{XA}\) smaller than a minimum cut-off distance \(d_{min}\) to the minimum value 0 for distance \(d_{XA}\) larger than a maximum cut-off distance \(d_{max}\). Using this particular spline function ensures that the derivative of this function, the force function, is continuous, preventing the occurrence of spurious impulses.
Chapter 6. Hydrogen bond switching

The angle function $u_\theta(\theta_{DHDA})$ in Equation 6.3 depends on the angle $\theta_{DHDA}$ between $D$, $H_D$ and $A$. This is schematically shown in Figure 6.1 a. The angle function $u_\theta(\theta_{DHDA})$ is given by:

$$u_\theta(\theta_{DHDA}) = \begin{cases} 1 & \theta_{DHDA} < \theta_{\text{min}} \\ (1 - 0.6) \cdot \left[ \frac{2(\theta_{DHDA} - \theta_{\text{min}})^3}{(\theta_{\text{max}} - \theta_{\text{min}})^3} - \frac{3(\theta_{DHDA} - \theta_{\text{min}})^2}{(\theta_{\text{max}} - \theta_{\text{min}})^2} \right] + 1 & \theta_{\text{min}} \leq \theta_{DHDA} < \theta_{\text{max}} \\ 0.6 & \theta_{\text{max}} \leq \theta_{DHDA} \end{cases}$$

This spline function affects the $D$-$H_D$-$A$ angle and varies smoothly from its maximum value of 1 for angle $\theta_{DHDA}$ smaller than a minimum cut-off angle $\theta_{\text{min}}$ to its minimum value of 0.6 for angle $\theta_{DHDA}$ larger than a maximum cut-off angle $\theta_{\text{max}}$. When using a minimum value of zero, rather than 0.6, the angle potential would generate very high forces. This could lead to the combination of a very small $d_{DA}$ with a large $\theta_{DA}$, resulting in the occurrence of unlikely or even non-physical conformations. Again, with this function the force function is continuous.

The values for the cut-off distances $d_{\text{min}}$ and $d_{\text{max}}$ should be chosen such that the repulsive biasing potentials will act on weak hydrogen bonds and that the attractive biasing potentials act on donor-acceptor pairs that are close to forming a hydrogen bond. The cut-off angles $\theta_{\text{min}}$ and $\theta_{\text{max}}$ should ensure the direction in which the donor/acceptor pairs are pulled apart or drawn together.

Figure 6.1 b displays a graphical representation of the repulsive hydrogen bond interaction function $u_{hb,D}$ as a function of the position of the acceptor $A$ for fixed donor $D$ and hydrogen $H$ positions. For a proper hydrogen bond the acceptor will be at the red plateau ($u_{hb,D} = 1$) and consequently no force resulting from the hydrogen bond potential acts on $A$ and $D$. When the hydrogen bond is not in an optimal configuration, $u_{hb,D}$ will decrease. Then, the repulsive hydrogen bond potential results in forces exerted on both $A$ and $D$, causing further disruption of the hydrogen bond. As the angle potential has a lower bound of 0.6, there is an intermediate configuration, in which no forces act on the donor and acceptor. This plateau has no physical meaning and the probability of finding a hydrogen bond in this area is very low, due to steric hindrance in the backbone. The purple plateau indicates the conformation with the lowest value for the biasing potential, in which the hydrogen bond is completely broken.

**Hydrogen bond switching**

The continuous hydrogen bond biasing potentials we have developed can be used in an MD simulation of any (bio) molecular system to enhance hydrogen bond reordering. We can also use these potentials in a Hamiltonian REMD scheme in which the replicas exchange their Hamiltonians. As the Hamiltonian of a replica also includes the hydrogen bond biasing potential, exchanging two replicas $i$ and $j$ has the following acceptance probability $p_{\text{acc}}$ [18]:

---
Methods

**Figure 6.1 — Continuous hydrogen bond potential.** a Schematic representation of a donor-acceptor (D-A) pair including the hydrogen (H_D). The labels indicate parameters used in the hydrogen bond potentials. In the repulsive potential the distance between the donor D and the acceptor A (d_{DA}) is used and in the attractive potential the distance between the hydrogen H_D and A (d_{H_D A}) is used. \( \theta \) is the angle between D, H_D and A. b Hydrogen bond interaction function corresponding to the repulsive hydrogen bond interaction (u_{hb,D}). The interaction function u_{hb,D} is a function of the acceptor (A) position in the plane around a fixed donor (D) and hydrogen (H). The color indicates the value of u_{hb} and the contour lines connect similar values of u_{hb,D} at intervals of 0.2.

\[
p_{acc}(i, j) = \begin{cases} 
1 & \Delta(i, j) \leq 0 \\
e^{-\Delta(i, j)} & \Delta(i, j) > 0 
\end{cases} 
\] (6.6)

\[
\Delta(i, j) = \beta [(H_j(r_i) - H_j(r_j)) - (H_i(r_j) - H_i(r_i))] 
\] (6.7)

with \( H \) the Hamiltonian, dependent on the coordinates \( r \), and \( \beta \) equal to \( 1/k_BT \). The Hamiltonians in replicas \( i \) and \( j \) differ only in the contribution from the hydrogen bond biasing potential energy \( E_{hb} \), reducing Equation 6.7 to:

\[
\Delta(i, j) = \beta [(E_{hb,j}(r_i) - E_{hb,j}(r_j)) - (E_{hb,i}(r_j) - E_{hb,i}(r_i))] 
\] (6.8)

The hydrogen bond biasing potential energy \( E_{hb} \) can have two forms: repulsive, aimed at breaking hydrogen bonds (rep), and attractive, aimed at forming hydrogen bonds (att). In our Hamiltonian REMD scheme, a replica can have either a repulsive potential, an attractive potential or no hydrogen bond biasing potential at all. The hydrogen bond biasing potential energy for replica \( i \), \( E_{hb} \) is given by:

\[
E_{hb,i}(r_i) = \lambda_{rep,i} V_{hb,rep}(r_i) + \lambda_{att,i} V_{hb,att}(r_i) 
\] (6.9)
Chapter 6. Hydrogen bond switching

with $V_{hb,rep}$ and $V_{hb,att}$ the repulsive and attractive hydrogen bond biasing potentials respectively. The weighting constants $\lambda_{rep,i}$ and $\lambda_{att,i}$ tune the strength of the repulsive and attractive biasing potentials in the replica. Equation 6.8 then becomes:

$$
\Delta(i, j) = \beta \left[ (\lambda_{rep,i} - \lambda_{rep,j}) (V_{hb,rep}(r_j) - V_{hb,rep}(r_i)) \\
+ (\lambda_{att,i} - \lambda_{att,j}) (V_{hb,att}(r_j) - V_{hb,att}(r_i)) \right]
$$

(6.10)

In our Hamiltonian REMD scheme, a replica can have an additional repulsive hydrogen bond interaction ($0 < \lambda_{rep,i} \leq 1, \lambda_{att,i} = 0$) or an additional attractive hydrogen bond interaction ($\lambda_{rep,i} = 0, 0 < \lambda_{att,i} \leq 1$). Also, there is a neutral replica in which both biasing potentials are switched off ($\lambda_{rep,i} = 0, \lambda_{att,i} = 0$). As no biasing potentials are active, all configurations that visit this neutral replica sample the correct canonical distribution.

Using the configurations from the neutral replica we can calculate order parameters $Q$ for which it is possible to construct a probability histogram $P(Q)$ and compute the corresponding (relative) Landau free energy:

$$
F(Q) = -k_B T \ln P(Q)
$$

(6.11)

Including properties of the rejected Monte Carlo moves in a replica exchange scheme significantly improves the accuracy of the histograms [209, 210]. This means that the contributions of all other replicas to the histogram of order parameters $Q$ in the neutral replica are scaled with a Metropolis factor. The probability distribution $P(Q)$ can then be estimated as:

$$
P(Q) = \sum_{j=1, j \neq i}^{N} [1 - \min(1, \exp(-\Delta))] \delta(Q_i - Q) \\
+ \sum_{j=1, j \neq i}^{N} [\min(1, \exp(-\Delta))] \delta(Q_j - Q)
$$

(6.12)

with $\Delta$ from Equation 6.10, $Q_i$ the value of the order parameter of replica $i$, $\delta(x)$ the Dirac delta-function, and $N$ the total number of replicas. The min function returns the smaller of its arguments. Note that this approach only improves the quality of the histograms by including all replicas, but it does not improve the sampling itself [209, 210].

Results and discussion

Continuous hydrogen bond biasing potentials

In this work we present a Hamiltonian REMD scheme using hydrogen bond biasing potentials. One condition for using biasing potentials in a replica exchange scheme is that the biasing potential must be continuous. The hydrogen bond potentials as used
Results and discussion

in Ref. [200] are discontinuous and must be adapted, as described in the Theory section. Before using these continuous hydrogen bond potentials in a replica exchange scheme, we first need to establish that swapping the hydrogen bond interactions does not affect the continuity of the energy functions and conserves linear momentum. To this end we performed three NVE simulations of polyalanine in a folded conformation, \( A_{16} \), at 300 K. One simulation did not contain any biasing potentials, while in the other two simulations either attractive or repulsive hydrogen bond biasing potentials were applied. For all simulations the average total energy was 101·10\(^3\) kJ/mol. The simulation without any biasing potential had fluctuations of 1.8 kJ/mol in the total energy, representing only 1% of the fluctuations in the potential and the kinetic energy. The simulations in which attractive or repulsive hydrogen bond biasing potentials were active, the fluctuations in total energy were 3.1 and 4.2 kJ/mol respectively. These fluctuations represent still only 2% of the fluctuations in the potential and kinetic energy. This means that the total energy is conserved very well, when applying the hydrogen bond biasing potentials. Consequently, we can conclude that our new hydrogen bond biasing potentials are indeed continuous and conservative.

When applying hydrogen bond biasing potentials in an alternating way, *i.e.* switching the potentials on and off at fixed time intervals, fast reordering of hydrogen bonds occurs, thus significantly accelerating the folding process [200]. To test whether this alternating scheme using the continuous potentials would also result in fast folding, we performed a simulation of unfolded polyalanine, the \( A_{16}^u \) system. The force constant of both the attractive and the repulsive potentials is 50 kJ/mol. Figure 6.2 shows a running average of 100 ps of the number of helical hydrogen bonds, between residues \( n \) and \( n+4 \), as a function of the simulation time. Within 2.5 ns, four \( \alpha \)-helical hydrogen bonds are formed, indicating that part of the backbone has assumed an \( \alpha \)-helical conformation. For lower values of the force constant we also observed formation of \( \alpha \)-helical structure, although on a longer time scale. Without biasing potentials, the probability of finding such a conformation after 10 ns is very low, when starting from an unfolded conformation [200]. Therefore, we can conclude that our continuous hydrogen bond biasing potentials speed up folding when applied in an alternating way.

**Efficiency of hydrogen bond switching**

As the alternating method is inherently a non-equilibrium method and cannot reproduce true thermodynamic equilibrium, we try to emulate the alternating scheme by using the hydrogen bond biasing potentials in a Hamiltonian REMD scheme. This requires the REMD scheme to facilitate fast switching of the replicas between the attractive and repulsive biasing potentials. Consequently, the probability of exchange between neighboring replicas must be relatively high. When using a force constant of 50 kJ/mol, as we did in the alternating scheme simulation of \( A_{16}^u \), at least 16 replicas are required to achieve an acceptance probability of about 0.3, provided that the energy difference between neighboring replicas equals one hydrogen bond or less. For almost any peptide and/or protein system, the difference in the number of hydrogen bonds of folded and unfolded conformations is more than one. As a consequence, fast
Chapter 6. Hydrogen bond switching

Figure 6.2 — Helix formation when continuous hydrogen bond biasing potentials are alternated. Helix formation of polyalanine $\text{A}_{16}$ in a simulation with alternatingly continuous hydrogen bond biasing potentials. The number of $\alpha$-helical hydrogen bonds, between residues $n$ and $n + 4$, is plotted as a function of the simulation time, as a running average of 100 ps.

switching between attractive and repulsive biasing potentials is impossible, when using such a high value for the force constant. Therefore, in the hydrogen bond switching simulations, we used a much lower value for the maximum force constant, 1 kJ/mol, for both the attractive and repulsive hydrogen bond biasing potentials in the hydrogen bond switching simulations. At room temperature, the probability of exchanging two neighboring replicas at a $\lambda$-difference of 0.5 is around 0.6, if the difference in number of hydrogen bonds is 4. This way, exchange between not-neighboring replicas can also occur occasionally. Even when using such weak biasing potentials, the sampling of conformational space will be accelerated, compared to conventional MD.

We performed a HS simulation of a $\beta$-heptapeptide in methanol, using five replicas with a $\lambda$-spacing of 0.5. Figure 6.3 a shows the value of $\lambda$ as a function of the simulation time for each replica. The $\lambda$-value is shown as a running average of 100 ps for clarity. Note that the $\lambda$ used here is a condensed version of $\lambda_{\text{at}}$ and $\lambda_{\text{rep}}$, with $-1 \leq \lambda < 0$ representing $\lambda_{\text{at}}$ and $0 < \lambda \leq 1$ representing $\lambda_{\text{rep}}$. After 1 ns, most replicas have switched at least once between attractive and repulsive hydrogen bond biasing potentials, and also visited the neutral replica. Figure 6.3 b shows every exchange step for 200 ps, indicating that the exchange is very fast. Because all replicas traverse the whole range of biasing potentials, they also visit the replica without any biasing potential. As the contributions of different systems to the unbiased replica are Boltzmann weighted, the probabilities of finding specific conformations still reflect equilibrium. We can therefore use the HS scheme to efficiently explore the free energy landscape of polypeptides. To this end, we calculated free energy profiles, using Equation 6.12, as a function of the number of backbone hydrogen bonds ($N_{HB}$) and the positional RMSD of the back-
Results and discussion

Figure 6.3 — Exchange in the HS simulation of the β-heptapeptide. a For each replica a running average of 100 ps of $\lambda$ is plotted as a function of the simulation time. The $\lambda$ used here is a condensed version of $\lambda_{att}$ and $\lambda_{rep}$, with $-1 \leq \lambda < 0$ representing $\lambda_{att}$ and $0 < \lambda \leq 1$ representing $\lambda_{rep}$. b For each replica every exchange step is plotted for the time slice indicated in (a).

Bone atoms of residues 2 to 6 from the NMR structure [211], denoted as RMSD. The first nanosecond of the simulation was excluded from the free energy profile calculation to allow for equilibration. In Figure 6.4 a we can identify three free energy wells: $F$, $I$, and $U$. $F$ is the deepest minimum at ($N_{HB} = 3$, RMSD = 0.02 nm) and represents the folded state. A barrier of $3 k_B T$ separates $F$ from the shallow minimum $I$ at ($N_{HB} = 2$, RMSD = 0.08 nm). At higher values for the RMSD, between 0.17 and 0.28 nm, $U$ is a broad well, ranging from 0 to 2 backbone hydrogen bonds. The free energy profile shows a clear correlation of decreasing number of hydrogen bonds with increasing deviation from the folded structure. Our results agree well with the simulations described in Ref. [206] that also indicated the low-RMSD configuration $F$ as the most stable state at a temperature of 300 K.

To address the performance of our hydrogen bond switching scheme, we performed 20 ns of temperature REMD simulation of this system, using the settings described in Ref. [206]. Excluding the first ns, conformations at $T = 298$ K were combined into trajectories, of which we calculated two order parameters, $N_{HB}$ and RMSD. Figure 6.4 b shows the resulting free energy profile. Indeed, the two free energy profiles are similar, both containing three free energy minima, $F$, $I$ and $U$, at approximately identical locations. The resemblance extends to the height of the barrier separating free energy wells $F$ and $I$, which is $3 k_B T$ in both simulations. Although the profiles are very similar, they show small differences in the unfolded region of the conformational space. Minimum $U$ is slightly shallower in the temperature REMD simulation and contains fewer conformations without any hydrogen bonds. These differences are probably caused by convergence issues. Despite this difference, the free energy profile from the HS simulation, requiring only five replicas, remarkably resembles that of the temperature REMD simulation, which required twenty replicas. As such, hydrogen bond switching represents a significant improvement in efficiency over temperature REMD.
Chapter 6. Hydrogen bond switching

Figure 6.4 — Free energy profiles of the $\beta$-heptapeptide. These profiles were obtained by a simulation with a HS and b temperature REMD. The profiles are calculated as a function of the number of backbone hydrogen bonds $N_{HB}$ and the positional RMSD of the backbone atoms of residues 2 to 6 from the NMR structure [211]. The contour lines indicate levels of $k_B T$, decreasing with 1 going from yellow to red. The red star in a indicates the starting conformation. The labels F, I and U indicate free energy wells. In c typical conformations representing the free energy wells are displayed as stick models. Backbone carbon atoms are dark grey and carbon atoms in side chains are colored light grey. Hydrogen bonds are shown as yellow dotted lines.
The efficiency of the hydrogen bond switching scheme originates from the fast rate of conformational change, induced by the repulsive and attractive biasing potentials. The HS simulation started from a misfolded conformation (red star in Figure 6.4 a) containing three hydrogen bonds, but deviating significantly from the NMR structure. The repulsive hydrogen bond biasing potentials flattened the free energy surface, thus facilitating many barrier crossings. In addition, the attractive hydrogen bond biasing potentials enabled the sampling of free energy minimum $F$.

Recently, several other Hamiltonian replica exchange schemes have become available, that also require significantly fewer replicas [204, 205, 212–214]. Most of these schemes have reduced non-bonded interactions [204,212,214] and one approach lowered rotational barriers by affecting torsional potentials [205]. Similar in all these approaches is that the barriers separating different protein structures are decreased to facilitate faster conformational sampling, by introducing additional biasing potentials or lowering the strength of (specific) interactions. Lowering the free energy barriers effectively flattens the free energy surface, resulting in faster sampling of the conformational space. In such a flattened free energy landscape, entropic barriers have increased, as more conformations can be sampled. Visiting free energy minima is then less probable and therefore, sampling folded conformations is hampered [215]. Introducing biasing potentials that favor folded conformations will decrease these entropic barriers again. The HS scheme combines both the advantage of flattening the free energy surface with making the low energy states more attractive. Moreover, the HS scheme as presented here is not fully optimized, as the cut-off distances and angles in the biasing potentials can be tuned further. Also, as the exchange is very fast, the value for the repulsive and attractive force constants can be higher, enabling faster sampling of the conformational space.

**Increasing system size**

The β-heptapeptide contains at most five hydrogen bonds and as such represents one of the simplest systems to sample with hydrogen bond switching simulations. To investigate the performance of our Hamiltonian REMD scheme when applied to a protein system containing more hydrogen bonds, we performed hydrogen bond switching simulations of sixteen residue polyalanine, $A_{16}$, starting from two initial conformations. The first conformation is a folded, α-helical structure, further referred to as $A_{16}^f$, and the second structure is an unfolded, extended structure, further denoted as $A_{16}^u$. We used 11 replicas with a spacing of 0.2 in $\lambda$, using a maximum force constant of 1 kJ/mol. With these conditions, the probability of exchanging two neighboring replicas at room temperature is around 0.6, if the difference in number of hydrogen bonds is 12.

Figure 6.5 a displays the exchange of the replicas for the $A_{16}^u$ simulation. The value for $\lambda$ for each replica is plotted as a function of the simulation time, as a running average of 100 ps to enhance visibility. Each line represents one replica, and most replicas visit each value of $\lambda$ often. There are two replicas that stay at extreme values of $\lambda$ often. Figure 6.5 b shows the exchange in the $A_{16}^f$ system. In this simulation the replicas spend
Figure 6.5 — Exchange in the HS simulations of $A_{16}$. For each replica, $\lambda$ is plotted as a function of the simulation time as a running average of 100 ps for a $A_{16}^{u}$ and b $A_{16}^{f}$. The $\lambda$ used here is a condensed version of $\lambda_{att}$ and $\lambda_{rep}$, with $\lambda < 0$ representing $\lambda_{att}$ and $\lambda > 0$ representing $\lambda_{rep}$.

most of the time in either the repulsive or attractive $\lambda$-region, as indicated by the white gaps between $\lambda = -0.2$ and $\lambda = 0.6$. There are a few replicas that cross from repulsive to attractive and vice versa, but most of these replicas quickly return. There is one replica that remains in the repulsive region after crossing for longer than 100 ps, this replica crosses from attractive to repulsive at around 7.5 ns.

In developing hydrogen bond switching, we aimed for fast diffusion through the entire range of biasing potentials. In the HS simulations we observe that every replica visits each value of the biasing potential, although on average it prefers a specific bias. Because all replicas traverse the whole range of biasing potentials, they also visit the neutral replica without any biasing potential. Using the conformations from the neutral replica we calculated free energy profiles, using Equation 6.12, as a function of the number of helical hydrogen bonds and the contact order (See Computational details, Equation 6.13). Figure 6.6 a and b show the free energy profiles obtained from respectively the $A_{16}^{u}$ and the $A_{16}^{f}$ simulations. In the $A_{16}^{f}$ free energy profile we identified three regions: $F$ ($HB_{n-n+4} \approx 11, CO \approx 4.4$), $P$ ($HB_{n-n+4} \approx 7, CO \approx 6.6$) and $C$ ($HB_{n-n+4} \approx 0, CO \approx 6.4$). We found the latter two also in the $A_{16}^{u}$ free energy profile, as well as free energy minimum $U$ ($HB_{n-n+4} \approx 0, CO \approx 2.4$). $F$ represents the fully folded, $\alpha$-helical conformation, whereas $U$ represents the fully unfolded, extended chain. $C$ and $P$ contain intermediate conformations and respectively represent a collapsed coil and a partially formed $\alpha$-helix. Figure 6.6 c shows representative conformations of these four regions. It is clear that after 20 ns the two simulations have not converged to identical free energy profiles. Nevertheless, the profiles do overlap substantially. In both profiles, free energy minimum $C$ occurs, although for the $A_{16}^{u}$ this region covers a larger contact order interval. Partially folded intermediate $P$ is not a minimum in the $A_{16}^{f}$ free energy profile, but in the free energy profile of $A_{16}^{u}$ it is.

Both the partial helical conformations of state $P$ and the collapsed coils in state $C$ may be part of the structural ensemble of polyalanine $A_{16}$. Experimentally, the struc-
Results and discussion

Figure 6.6 — Free energy profiles obtained by HS simulations. a the $A_{16}^U$ and b the $A_{16}^f$ systems. The profiles are calculated as a function of the number of helical hydrogen bonds and the contact order (Equation 6.13). The contour lines indicate levels of $k_B T$, decreasing with 1 with darker shading. The labels indicate free energy minima. c Snapshots of conformations of $A_{16}$ representing the various free energy minima, shown as ribbon representation.

ture of polyalanine in solution, and $A_{16}$ in particular, could not be determined as it readily forms aggregates, prohibiting structural determination at high resolution [208]. Solid state NMR experiments on a polyalanine stretch, capped by glycine residues, shows that a polyalanine stretch can adopt a helical conformation [207]. From a physicochemical point of view, the $C$ and the $P$ state are favorable compared to the extended conformations in state $U$ and the fully formed $\alpha$-helical conformations in state $F$. In both $C$ and $P$, the solvent accessibility of the hydrophobic regions, comprising $\alpha$ and $\beta$ carbon atoms, are minimized, and as such, these states maximize the solvent entropy.

Several research groups have investigated polyalanine in varying compositions using temperature REMD. Polyalanine $A_{21}$ required at least 32 replicas ranging from 275 to 475 K [216]. Using the temperature distribution webtool [217] polyalanine $A_{16}$ as
used in this work would require 60 replicas ranging from 275 to 475 K for an acceptance probability of 0.25. In the HS scheme, only eleven replicas are required for $A_{16}$. This is a significant improvement.

**Sorting of replicas**

Although the hydrogen bond switching scheme requires fewer replicas than temperature REMD, we have encountered a problem in the exchange between replicas that slows down convergence. Once a replica has adjusted to a certain biasing potential, it becomes significantly more difficult to switch and adjust to another biasing potential. This is evident from the fraction of simulation time, the residence, each replica spends at each value of $\lambda$, as shown in Figure 6.7 a and b. Ideally, each replica would spend $1/11 (\approx 0.09)$ of the simulation time at each value of $\lambda$. For the $A_{16}^{u}$ simulation (Figure 6.7 a) nine of the eleven replicas show this ideal behavior. Two replicas spend significantly more time at extreme values of $\lambda$. The inset in Figure 6.7 a shows the exchange of these two replicas for a short time interval. Surprisingly, the two replicas that remain at high $\lambda$-values do exchange with other replicas, even with replicas at opposite $\lambda$. Nevertheless, in the next exchange step they return back to a high value of $\lambda$, but not necessarily the same value. This effect is even stronger for the $A_{16}^{f}$ simulation (Figure 6.7 b), where none of the replicas spends equal amounts of time at all $\lambda$-values. Instead, each replica favors specific values of $\lambda$. For extreme $\lambda$-values, the replicas are more trapped than replicas at $\lambda$-values close to zero. The inset in Figure 6.7 b shows the exchange steps of the replicas that spend most of their time at extreme $\lambda$-values. Most exchanges occur between replicas that are close in $\lambda$, but there are exchanges with replicas at $\lambda$-values close to zero and even opposite sign. Similar to the $A_{16}^{u}$ system, the replica returns to an extreme $\lambda$-value after such an exchange.

Examination of the conformations in the replicas that are stuck at extreme $\lambda$-values shows that at high $\lambda$ (strong repulsive hydrogen bond biasing potential) most conformations are unfolded. At low $\lambda$ (strong attractive hydrogen bond biasing potential) the conformations are folded. A folded conformation with many hydrogen bonds favors $\lambda$-values corresponding to attractive biasing potentials, as its energy is lowest in this region. Similarly, unfolded conformations with little to no hydrogen bonds favor $\lambda$-values corresponding to repulsive biasing potentials. There are exchanges that result in unfolded structures at attractive hydrogen bond potentials and vice versa, but in the next exchange attempt, the unfolded replica goes back to a repulsive hydrogen bond potential (and vice versa). The time replicas spend at opposite potentials is not long enough to induce conformational changes, *i.e.* adapt to the new potential and thus, the replicas are sorted. Also in other Hamiltonian REMD schemes sorting of replicas occurs. The distribution of replicas visiting the different force fields in Ref. [204] is skewed, and as such, shows similar behavior as the HS scheme.

Convergence is slowed down, as the diffusion of the replicas through the $\lambda$ values is hampered by sorting, which is a common trait in Hamiltonian REMD schemes that target protein conformational changes. We expect a much faster convergence if this sorting can be prevented, which requires the replicas to adapt to the new potentials
Results and discussion

Figure 6.7 — Sorting of the replicas. Fraction of simulation time, the residence, of each replica at each value of $\lambda$, for a $A_{16}^u$ and b $A_{16}^f$. The insets show each exchange step for replicas that are indicated by the lines. The $\lambda$ used here is a condensed version of $\lambda_{att}$ and $\lambda_{rep}$, with $-1 \leq \lambda < 0$ representing $\lambda_{att}$ and $0 < \lambda \leq 1$ representing $\lambda_{rep}$.

within the time between exchange attempts. However, adjusting to the new potential might take a very long time, depending on the system and the force constant. There are two straightforward ways to deal with this problem. The first solution is increasing the force constant. This would require many more replicas to maintain reasonable acceptance probability, thus losing the gain in efficiency. In addition, using very strong potentials will lead to very high forces. Correctly addressing those forces requires a shorter time step, leading to a higher computational cost for the simulations. Moreover, sorting would occur with many more replicas as well. The second solution to facilitate faster adaptation to the new potential is increasing the time between exchange attempts. Diffusion through the replicas is then severely slowed by the lower frequency of exchange attempts, and as a consequence, the increase in efficiency is less. As gaining efficiency is the motivation behind developing these Hamiltonian REMD methods, clearly these solutions are not very satisfactory. Preventing this sorting effect is thus a challenging problem, that, when solved, can lead to an additional increase in the efficiency of Hamiltonian REMD methods.

Conclusions

In this work we present a Hamiltonian REMD scheme, using biasing potentials that either break or form hydrogen bonds in the protein backbone, called hydrogen bond switching (HS). To this end, we developed continuous hydrogen bond biasing potentials based on potentials used in a non-equilibrium alternating scheme aimed at fast
Chapter 6. Hydrogen bond switching

folding [200]. These potentials still result in conservation of the total energy in the NVE ensemble, and when applied in the alternating scheme, rapid sampling of the conformational space of polyalanine occurs.

To address the efficiency of the HS method, we performed an HS and a temperature REMD simulation of a β-heptapeptide in methanol. Both simulation methods resulted in very similar free energy profiles and required 5 ns or less to sample most of the conformational space of the β-heptapeptide. As the HS simulation required only five replicas, while the temperature REMD required twenty replicas, the HS method is significantly more efficient.

In addition, we performed HS simulations of polyalanine $A_{16}$ in explicit water, starting from both a fully folded α-helical structure and an extended chain. Both simulations sampled similar intermediate states, representing collapsed coil-like structures and partially folded α-helices. As the HS simulations require only eleven replicas to sample the conformational space of $A_{16}$, our method presents a significant improvement in efficiency over temperature REMD. When setting up an HS simulation, no knowledge of folded states is required, similar to temperature REMD. To conclude, HS presents an efficient way of sampling protein conformational space by reordering hydrogen bonds in the protein backbone.

Our simulations also revealed that replicas have a preference for specific biasing potentials, hampering the diffusion of replicas through $\lambda$-space thereby slowing down convergence. This sorting of replicas also occurs in other Hamiltonian REMD schemes that target protein conformational changes. As straightforward solutions require a transition to much less efficient REMD protocols, future research should focus on a more sophisticated solution to the sorting effect. Preventing this sorting will presumably further increase the efficiency and utilization of Hamiltonian REMD methods in general.

Computational details

Systems

In all simulations we used the GROMACS software package [11], in combination with the GROMOS96 43a1 force field [218, 219]. We investigated two systems: 1. a β-heptapeptide ($\beta$VAL-$\beta$ALA-$\beta$LEU-$\beta$ALA($\alpha$-methyl)-$\beta$VAL-$\beta$ALA-$\beta$LEU) in methanol and 2. a sixteen residue polyalanine, $A_{16}$ in water. The beta-heptapeptide system was provided by Periole and required no further preparation [206]. The system was 3028 atoms in size in a cubic periodic box with an edge of 3.96 nm. Non-bonded interactions were treated with a twin-range cut-off (1.0 - 1.4 nm). Interactions within the short-range cut-off were considered every time step, whereas interactions within the longer-range cut-off were evaluated every ten steps, together with the updating of the pair list. Long-range electrostatic interactions beyond the cut-off were treated with a reaction-field correction [206].

The polyalanine in SPC water was initialized in an α-helical conformation, $A_{16}^{f}$, and a fully extended conformation, $A_{16}^{u}$. The α-helical conformation was generated man-
ually by taking residues 129-145 from PDB entry 101M and removing all side chains up to $C\beta$. This conformation was solvated in a periodic cubic box of 2950 SPC water molecules [154] resulting in a system of 8952 atoms. This system was equilibrated in the $NpT$ ensemble for 100 ps at 300 K and 1 bar, resulting in a box with an average edge length of 4.53 nm. To obtain an unfolded conformation for $A_{16}^f$, we performed an MD simulation of the $A_{16}^f$ system at a temperature of 600 K and constant volume. Polyalanine unfolded into a fully extended chain after 100 ps, after which we equilibrated the system again at 300 K and 1 bar for 100 ps. We used a time step of 2 fs throughout the preparation procedure. The van der Waals interaction cut-off was 1.2 nm. The electrostatic interactions were treated with the PME method with a cut-off of 0.9 nm [156].

For both the $\beta$-heptapeptide and polyalanine, all bonds were constrained using LINCS [155]. Temperature was kept constant at 300 K using the Nose-Hoover thermostat [220, 221]. Pressure was kept constant at 1 bar using the Parrinello-Rahman barostat [222, 223].

**NVE simulations**

To test if swapping the hydrogen bond interactions does not affect the continuity of the energy functions we performed three simulations in the NVE ensemble, of the $A_{16}^f$ system, of 50 ps, with a time step of 0.1 fs. The three simulations had respectively no biasing potentials, repulsive and attractive hydrogen bond biasing potentials. Both the repulsive and attractive potentials had a force constant of 10 kJ/mol. All parameters used for the hydrogen bond potentials are listed in Table 6.1.

**Alternating hydrogen bonds simulations**

We performed a simulation of $A_{16}^u$ using the alternating hydrogen bond scheme [200]. For 2.5 ns, the system underwent alternating cycles of 2 ps. During such a cycle repulsive hydrogen bond biasing potentials were active for 0.5 ps and were switched off for the next 0.5 ps. Subsequently, attractive hydrogen bond biasing potentials were switched on for 0.5 ps, that were switched off for the last 0.5 ps of the alternating cycle. Both repulsive and attractive potentials had a force constant of 50 kJ/mol. All parameters used for the hydrogen bond potentials are listed in Table 6.1.

**Hydrogen bond switching simulations**

We performed hydrogen bond switching (HS) simulations for the $\beta$-heptapeptide, the $A_{16}^f$ and the $A_{16}^u$ systems. The $\beta$-heptapeptide simulation consisted of 5 replicas, while the polyalanine simulations contained 11 replicas. For all three, we used maximum repulsive and attractive force constants of 1 kJ/mol. The scaling factor $\lambda$ jump between neighboring replicas was 0.5 for the $\beta$-heptapeptide and 0.2 for both $A_{16}^f$ and $A_{16}^u$. During each HS simulation all-pair exchange attempts [157] were performed every 1 ps. These 1 ps MD runs were performed using GROMACS [11] extended with the continuous hydrogen bond biasing potentials, using a time step of 1 fs. The constant volume
Chapter 6. Hydrogen bond switching

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Attractive</th>
<th>Repulsive</th>
<th>Simulation</th>
<th>$f_{c_{\text{rep}}}$ (kJ/mol)</th>
<th>$f_{c_{\text{att}}}$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{min}}$ (nm)</td>
<td>0.32</td>
<td>0.23</td>
<td>NVE</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$r_{\text{max}}$ (nm)</td>
<td>0.40</td>
<td>0.40</td>
<td>alternating</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>$\theta_{\text{min}}$ (rad)</td>
<td>0.85 (145°)</td>
<td>0.85 (145°)</td>
<td>HS</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\theta_{\text{max}}$ (rad)</td>
<td>0.50 (60°)</td>
<td>0.50 (60°)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1 — Parameters for the hydrogen bond biasing potentials. The variables $r_{\text{min}}$, $r_{\text{max}}$, $\theta_{\text{min}}$ and $\theta_{\text{max}}$ indicate the range in which the potential introduces a force to the system. For the different types of simulations the (maximum) force constants are listed.

MD simulations were kept at a temperature of 298 K with the Nose-Hoover thermostat [220, 221]. Table 6.1 lists the parameters used for the repulsive and attractive biasing potentials at their maximum values. All HS simulations also contained a neutral replica, in which no biasing potentials were active.

**Temperature REMD simulations**

To compare the HS method to temperature REMD, we performed a temperature REMD simulation of the $\beta$-heptapeptide using 20 replicas. Each replica started from a different conformation, ranging in temperature from 275.34 K to 399.58 K. These settings were obtained from Ref. [206]. Exchanges between all replicas were attempted every 1 ps using a home-made Perl script [201]. The 1 ps MD runs were performed using GROMACS [11], with a time step of 2 fs. The Nose-Hoover thermostat imposed a temperature on the replicas [220, 221].

**Analysis**

Conformations at $\lambda = 0$ for the HS simulations, or at $T = 298$ K for the temperature REMD simulation, were combined into trajectory files. Analysis of these files consisted of tracking several order parameters using the tools included in the GROMACS software package, in combination with Perl scripts. For the $\beta$-heptapeptide we used the number of backbone hydrogen bonds and the positional RMSD of the backbone atoms of residues 2 to 6 from the NMR structure [211]. The order parameters used to investigate the conformational free energy barriers of the $A_{16}$ system were the backbone $\alpha$-helical hydrogen bonds between residue $n$ and $n + 4$, and the contact order:

$$CO = \frac{1}{LN} \sum_{i<j}^{N} \Delta S_{ij} \quad (6.13)$$

with $L$ the number of residues in the polypeptide, $N$ the total number of contacts and $\Delta S_{ij}$ the sequence separation, in residues, between contacting residues $i$ and $j$ [224]. Residues were considered to be in contact if their $C_\alpha$ atoms were within a distance of 0.6 nm.
PART IV
GENERAL DISCUSSION
Perspective and future directions

Abstract The relation of amyloid fibrils to pathological conditions as well as the fibrils useful properties to nano-technology and material science, requires a detailed understanding of the structure and formation of these fibrils. Unfortunately this understanding is currently not complete. In this chapter we will review the additions to this understanding made by the new methods that were developed and the examinations that were conducted as described in this thesis. In addition we will give an outlook for the future, including some readily feasible projects.

Amyloidogenesis: Current status and longterm goals

The relevance of amyloidogenesis has increased over the last years and will continue to do so in the coming years. On the one hand, because age related pathological conditions that are connected to amyloidogenesis, such as Alzheimer disease, will occur more frequently, as the average life-expectancy of people increases. On the other hand materials scientists show a growing interest in amyloid fibrils in the search for ever smaller materials with a variety of applications, as well as the need for novel materials manufactured from renewable materials under mild conditions. Both these fields require a clear understanding of the structure and formation of amyloid fibrils. The first for the development of therapeutic strategies and the second in rational design of fibrils with specific features or morphology.

Structure and stability of fibrils

Structurally, amyloid fibrils consist of multiple peptide chains that form cross-β-sheets with the hydrogen bonds directed parallel to the fibril axes. These cross-β-sheets interact through lateral interactions (usually by a hydrophobic collapse), resulting in rope-
Chapter 7. Perspective and future directions

like fibrils. The exact structural details of the amyloid fibril are different for each individual peptide and depend strongly on the nucleation conditions. Peptide chains that form cross-β-sheets with the hydrogen bonds directed parallel to the fibril axes. These cross-β-sheets interact through lateral interactions (usually by a hydrophobic collapse), resulting in rope-like fibrils. The exact structural details of the amyloid fibril are different for each individual peptide and depend strongly on the nucleation conditions.

The experimental assessment of the amyloid fibril structure formed by a random peptide is far from trivial [67], although the structural details of some fibrils have largely been resolved [52, 81]. In addition, computational methods have not advanced sufficiently to predict this structure from the primary amino acid sequence alone. Rather they are used to assess the stability of proposed fibril structures [50] or predict the fibrillogenic nature of a specific amino acid sequence [160]. As a consequence, the underlying cause for the variety in the molecular details of amyloid fibrils, composed of different peptides or formed under different conditions, is poorly understood.

Unfortunately, efficient rational design of functional fibrils requires that the interactions responsible for the variety in the fibril details are comprehended. For example, if the fibrillogenic nature of every amino acid is characterized, the strength of a fibril can be modulated by mutating the right amino acid. Furthermore, as the morphology of a fibril (e.g. branched or unbranched) is an important asset when fibrils are used as a template in fabrication of nano-structures, insight in the structural changes upon variations in sequence or conditions is required to create templates with a desired morphology.

Thus, although rational design of functional materials is the ultimate goal in nanotechnology and material science, the application of functional amyloid fibrils is minimal. Designing a peptide that will assemble into a fibril is within our abilities [79], however designing one with specific properties or morphology is currently hampered by the limited knowledge on the basis of the variety in the molecular details and morphology of fibrils. Therefore, the next milestone is to quantify the important physicochemical interactions in fibril formation and analyze their effect on the details and morphology of amyloid fibrils.

**Formation of amyloid fibrils**

Amyloid fibrils are formed through a nucleation-growth mechanism. During nucleation the peptides in the solution form a minimal-sized structure with amyloid fibril characteristics, the nucleus. This nucleus can then grow into a mature fibril, by assembling with other nuclei or protofilaments through the lateral interaction or by addition of monomers to the fibril tips.

The interest in the nucleation mechanism and the intermediate states preceding fibril growth has increased recently. An important reason is the accumulating evidence that these early aggregates, rather than mature fibrils, are the pathological agents in amyloid related diseases [28–31]. Currently, the development of therapeutic strategies for these amyloid related diseases is one of the major goals in medicine.
Understanding the physico-chemical properties crucial to the stability of fibrils

In both the nucleation and the growth phase the peptides undergo major conformational changes to adopt the configuration as found in mature fibrils. To examine these conformational changes is very difficult. Experimentally, the resolution of many techniques is too low, such as cryo electron microscopy and Förster (or fluorescence) resonance energy transfer. Another problem is that in many experimental techniques an ensemble average is measured. In other words, although all proteins in the solution produce a different signal corresponding to their individual situation (including configuration and possible neighbors), the instrument can only measure the average of all these signals. As the peptides have many different conformations and neighbors during the nucleation phase this poses a serious problem. Computational methods on the other hand can capture the individual peptide configurations and transitions, but cannot sample all relevant structures within a reasonable amount of time. If not for this sampling problems, computational methods are obviously the preferred approach to obtain high resolution information on the nucleation mechanism.

As the view on the pathological agent in amyloid diseases is shifting from the fibril to the oligomeric species, the starting point in the development of therapeutic strategies has also shifted to the nucleation phase. Unfortunately, the key events and molecular details of this mechanism are largely unresolved. Although there are some low-resolution methods available, currently there are no experimental or computational techniques that can provide the molecular details of the nucleation mechanism. Developing these high-resolution methods is the first step toward a better understanding of the nucleation mechanism.

Understanding the physico-chemical properties crucial to the stability of fibrils

The characteristic interactions in amyloid fibrils are the cross-β-sheet and lateral interactions. Although the peptide backbone is vital to form cross-β-sheets, the structural details of a fibril are accounted for by the side chain interactions (and to a lesser extent the peptide length) \[60\], as the backbone interactions (i.e. cross-β-sheet hydrogen bonds) are similar to all peptides. In a first step to quantify the important physico-chemical interactions in fibrils we focused on the thermodynamic stability of amyloid fibrils composed of different peptides to answer the question how the side chains influence the fibril structure.

Although the nature of amyloid fibril formation is still controversial, i.e. is fibril formation a kinetic or a thermodynamic process, there are two reasons that make a thermodynamic assessment valid. Firstly, why should one process apply to the formation of all amyloid fibrils. It is much more reasonable to expect the process of fibril formation to depend on the specific peptide. For example, fibrils formed by peptides with a well-defined monomer native state are likely to be a kinetic product, as only products requiring slight unfolding are readily accessible. On the other hand monomeric unfolded peptides have access to a wide variety of structures, making it very likely that the fibril is a thermodynamic product. Secondly, a kinetic product is still the thermody-
Chapter 7. Perspective and future directions

naturally most stable product within the kinetic trap. So the thermodynamic stability of amyloid fibrils, either overall or within a kinetic state, is important.

As fibril growth is a thermodynamic process [134], irrespective of the formation process, the thermodynamic properties of amyloid fibrils, i.e. the free energy of the cross-β-sheet and lateral interaction, can be derived from the association constant of a monomer to a fibril. In chapter 3 a combination of computational methods, viz. REMD and umbrella sampling, is used to calculate these thermodynamic properties for model peptides. The application of this replica exchange umbrella sampling allows for the comparison of association constants and free-energy differences between different peptide sequences. By comparing carefully chosen sequences the important physico-chemical properties per amino acid can be deduced, as for example described in chapter 4. This approach has provided two major results.

First, a comparison between the fibril growth propensities of four closely related tetrapeptides, KFFE, KLLE, KVVE and KAAE, provided new insights into the important factors affecting the stability of amyloid fibrils. We concluded that, when considering hydrophobic side chains, the fibril structures are stabilized by the hydrophobic collapse, which is opposed by entropy contributions and excluded volume effects. The contribution of these opposing forces varies strongly per amino acid and per interaction (cross-β-sheet or lateral). For example phenylalanine has a stronger cross-β-sheet interaction, but a weaker lateral interaction than valine. Consequently, we expect that there is less force required to break a fibril composed of valine-rich peptides than one composed of phenylalanine-rich peptides, although the stability of both fibrils is very similar. These properties are very important to material scientists in the search for novel materials.

Second, the association constants that we calculated are quantities that are accessible to experimental techniques. As generally the properties measured in experiments are very complicated or even impossible to obtain through computer simulations, currently the dialogue between these two fields is minimal. By providing a computational method that can produce experimentally measurable quantities for a wide variety of systems, it is now up to the experimental community to assess systems that are within reach of this method. This way the cooperation and discussion between the communities of experimentalists and computationalists can be improved, offering a broader basis for the conclusions drawn from both experiments and simulations.

A limitation of this REMD umbrella sampling is that the simulations are limited to ~7 amino acids. This limitation is caused by the search problem also encountered when assessing protein folding with simulations. In general, a peptide larger than ~7 amino acids can not sample every accessible conformation in the course of a simulation. Hence important conformations may be missed, thereby creating an incomplete free-energy landscape. Nevertheless, REMD umbrella sampling as described in chapter 3 can provide important new insight as shown in chapter 4. Clearly combining REMD umbrella sampling with methods that overcome the search problem will increase the size limit. An example of such a method is described in chapter 5 and 6 and will be discussed in the next section.
Assuming that, with careful design, the contribution of individual amino acids to the lateral and cross-β-sheet interaction is additive, i.e., the lateral and cross-β-sheet interactions of a peptide is the sum of the individual amino acid contributions, the interactions and properties of mature fibrils can be controlled by the amino acid sequence. The rational design of amyloid fibrils with specific properties thus requires knowledge of the interaction strength for pairs of amino acids and the effect of peptide sequence order. The work described in this thesis covers only a very small portion of the amino acids encountered frequently in the core of amyloid fibrils. Clearly, for efficient design knowledge on the fibril interactions of all amino acids is required as well as the effect of various environmental conditions.

The method described in chapter 3 provides a technique to obtain information on the amino acid or environment effects on fibril stability. It is straightforward to perform similar experiments as described in chapter 4 on small fibrillogenic peptides containing other amino acids. A survey of the lateral and cross-β-sheet interaction free energies for all amino acids thus only requires a suitable small fibrillogenic peptide containing this amino acid. The many small fibrillogenic peptide fragments, isolated from larger fibrillogenic peptides, including possible known mutants [87], seem a suitable start for this survey.

In addition, the effects of environmental conditions on the thermodynamic stability of fibrils can be assessed with the aforementioned method. For example, elevated temperatures or elevated pressures are known to influence the final fibril structure. Simulating a system under different conditions is very simple. Hence, a set of simulations can be performed to reveal the effect of different conditions on the thermodynamic stability of a specific fibril structure.

Also, the relative stabilities of different fibril morphologies of the same peptide, e.g., parallel or antiparallel, can be examined. Since the free-energy is a state function, the free-energy difference between two different morphologies can be obtained by calculating the free-energy difference of both fibrils to a common state. Fortunately, the dissociated state used in the calculation of the association constant is a common state to both morphologies, and the relative stability is thus the difference between the two association free energies.

All these proposed computational experiments can already be conducted with the current state of computer technology and the available computational techniques, and presents a suitable project for the next 3-5 years. We expect that these and other investigations on the structural properties of amyloid fibrils will give rise to a level of understanding that will allow for predictions on the structural change in the fibril upon simple changes to the system and allow for design of simple amyloid fibrils with a specific desired property in the next 5-10 years.
Assessment of the structures and transitions in the nucleation phase requires new methods

The recent evidence supporting oligomeric species as the pathogenic agent in amyloid diseases demands a clear understanding of the nucleation mechanism. During the nucleation phase the fibrillogenic peptides adopt many different conformations and form various aggregates and oligomers. A detailed characterization of these species and evaluation of all the transitions involved is a challenging task, impractical with the currently available methodologies. Although computer simulations are capable of providing the desired level of detail, they cannot sample all the important aggregates and transitions within a reasonable amount of time. In order to assess the conformational changes in the nucleation phase or during protein folding with computer simulations this search sampling problem has to be tackled.

As the sampling problem is a result of the ruggedness of the energy landscape, either this landscape has to be flattened or the rate of barrier crossing has to be increased in order to overcome the search problem. In protein systems the minima in the energy landscape are mainly due to hydrogen bonds and hydrophobic interactions. By targeting one of these interactions the transitions between valleys in the energy landscape can be affected and faster conformational sampling can be induced.

In chapter 5 a new method is described, which influences the hydrogen bond interaction by adding alternating hydrogen bond biasing potentials to the standard Hamiltonian of the system. This new method (AHBP) results in a fast reordering of the backbone hydrogen bonds and accelerates in silico protein folding. Although hydrogen bonds are a key element of the structure of a protein, this is the first account of targeting hydrogen bonds to increase the folding rate in computer simulations.

The addition of alternating hydrogen bond biasing potentials has been applied with great success in the study of two peptides of 16 amino acids, one with an α-helix native state and the other with a β-hairpin native state. The dominant conformations in the simulations where similar to the native state obtained through experiments. In addition the use of alternating hydrogen bond biasing potentials provided these conformations 10-100 times faster than other computational methods.

Nevertheless, problems are expected when increasing the peptide size, since the alternating hydrogen bond biasing potentials cannot accelerate the conformational transition from one hydrophobic collapse to another. In an AHBP simulation the transition from a non-native hydrophobic collapse to the native hydrophobic collapse therefore depends on thermal fluctuation, just as in standard simulations. This clearly increases the in silico folding time and will reduce the efficiency of the AHBP method.

In order to maintain the efficiency of AHBP for larger protein systems the hydrophobic interactions should be targeted as well. For example, similar potentials as acting on the hydrogen bonds can be devised to act on the hydrophobic atoms or the hydrophobic side chain. This will lead to fast formation of hydrophobic clusters, although the internal structure of the hydrophobic cluster then strongly depends on the initial conditions. Consecutively, this problem can be solved by using soft-core poten-
Assessment of the nucleation phase requires new methods for the interactions between the hydrophobic atoms in the side chains, which will allow for an internal relaxation of the hydrophobic cluster.

Before taking this route, careful consideration of the desirability of this approach is in order. As the hydrophobic interactions are generally considered to be the driving force of protein folding, it is very important to make sure that the potentials do not alter the relative depths of the different minima in the free-energy landscape, but only affect the barriers separating them. This will probably require extensive tuning of the potentials, which is a time-consuming task with no guarantee of success.

Another problem is that AHBP is a non-equilibrium method, hence thermodynamic data is not available, i.e. the relative stabilities of different states, such as folding intermediates or small aggregates during the nucleation phase, cannot be calculated from AHBP simulations. To solve this we combined the hydrogen bond biasing potentials with REMD, named hydrogen bond switching, as described in chapter 6. This allows for efficient conformational sampling in thermodynamic equilibrium. Hydrogen bond switching has been applied successfully to examine the folding landscape of a β-heptapeptide. A 4-fold reduction in computational effort was achieved with respect to temperature REMD, while the resulting free-energy landscapes are very similar and fulfilling the experimental constraints.

Nevertheless, a problem arises when applying hydrogen bond switching to explore the folding landscape of larger peptides, such as the polyalanine α-helix also used to test the AHBP method. A separation in the replicas is observed, as a result of the large difference in the enthalpy and entropy of folded and unfolded states. A folded state, having high enthalpy and low entropy, favors folding promoting replicas as its energy is lowest in this region. Similarly unfolded states, with low enthalpy and high entropy, favor unfolding promoting replicas. If the energy gaps between the replicas are close enough, exchanges occur that result in folded states visiting unfolding promoting replicas and vice versa. At the next exchange step however, the replicas have not yet adapted to the new situation because of slow conformational changes and have a high probability to return to their original state. Consequently these transitions are only successful when a folded state spontaneously adopts an unfolded state or vice versa, which is a rare event. This entropy-enthalpy problem is a general concern in Hamiltonian replica exchange as it is observed in both the REMD application in chapter 3 and 6 and in other studies [204]. Finding a solution is thus important for Hamiltonian REMD methods in general.

To force the conformations to adapt to the new replica before the next exchange attempt in Hamiltonian REMD schemes, it would take either a very large force or a very long time. Both solutions indicate a considerable increase of simulation time. The first requires very long time intervals, whereas the second requires the addition of many replicas to maintain reasonable acceptance probability. Another possibility might be to devise elaborate REMD schemes in which sampling of the characteristic large entropy of the unfolded state is avoided. For example a high solute temperature that results in unfolding, can be combined with an attractive hydrogen bond potential to force folding again. These schemes require careful tuning, as a favorable replica for...
a specific state will prohibit efficient migration through all the replicas, similar to the effect of the entropy-enthalpy problem.

A more promising approach to the entropy-enthalpy problem is to force transitions to different replica conditions. Ideally, a system will then exchange quickly back and forth between unfolding and folding conditions, similar to the scheme of the AHBP method. For example, in a simulation one system and a set of Hamiltonians, of which only one is active, are selected. At set time intervals an attempt is made to change the Hamiltonian, which is accepted or rejected according to a Metropolis criteria. Unfortunately, in a standard Monte Carlo sampling only the lowest energy states will be sampled and consequently the unfolding Hamiltonians will be avoided. By assigning an artificial free-energy term to every Hamiltonian, which is incremented for every step spent in that Hamiltonian, similar to metadynamics, the Monte Carlo sampling can be biased toward the other Hamiltonians. Consequently, the system will be forced to another Hamiltonian when it spends too much time in one Hamiltonian by the slowly increasing free-energy term. Clearly, at the end of the simulation the data has to be unbiased in a method similar to the WHAM method described in chapter 3.

Whether one of these approaches really can solve the entropy-enthalpy problem encountered in REMD requires testing. Performing these investigations will take another 2-4 years, presenting for example a challenging PhD-project. As a consequence of these and other approaches aimed at a fast assessment of folding landscapes, in addition to the ever growing computer power, we expect that within 5-10 years it is possible to examine complex folding landscapes of peptides up to approximately 50 residues in atomic detail (including explicit solvent). This will also allow assessment of the small aggregates and oligomers present during fibril nucleation of model peptides such as KFFE. However, as the peptides associated to pathological conditions are much larger (more than 10 fold), a comprehensive picture in atomic detail of the nucleation and growth of the fibrils related to amyloid diseases lies much further in the future.
Bibliography


[33] Sigma-Aldrich. *Neurodegenerative disease: Alzheimer’s, Parkinson’s, Huntington’s, Prion disease and more*. Sigma-Aldrich Co., St. Louis, Missouri, USA, 2005.


[73] L. Miravalle, T. Tokuda, R. Chiarle, G. Giaccone, O. Bugiani, F. Tagliavini, B. Frangione,


[88] D. Zanuy, B. Ma, and R. Nussinov. Short peptide amyloid organization: Stabilities and


[105] W. Han and Y.D. Wu. A strand-loop-strand structure is a possible intermediate in fibril


[140] D. Thirumalai, D.K. Klimov, and R.I. Dima. Emerging ideas on the molecular basis of pro-


Bibliography


Summary

Amyloidogenesis is the process of amyloid formation. The basic structure of an amyloid is a rope-like fibril with a diameter of $\sim 10$ nm and indefinite length. Amyloid formation is an important hallmark in many pathological conditions, including Alzheimer disease. In addition, the small size and highly ordered nature of amyloid fibrils are favorable attributes for applications in nano-technology. Furthermore, the high tensile strength amyloid fibrils can withstand before breaking as well as the fact that the formation and growth of amyloid fibrils is a self-assembly process, are highly favored properties by material scientists.

An amyloid fibril is composed of many units, usually of the same peptide. Every peptide interacts with its neighboring peptides through intermolecular $\beta$-sheet interactions extending parallel to the fibril axes, forming cross-$\beta$-sheets. These cross-$\beta$-sheets can engage in a lateral interaction perpendicular to the fibril axes to form rope-like fibrils. Amyloid fibrils are formed through a nucleation-growth mechanism. In the initial stage, nucleation, the peptides in the solution form a minimal-sized structure with amyloid fibril characteristics, the nucleus. This nucleus can then grow into a mature fibril by assembling with other nuclei or with monomers either by lateral addition or by addition to the fibril tips.

Despite the considerable interest in amyloidogenesis raised by the aforementioned fields and the high-resolution structural data that have become available recently, the underlying molecular basis for amyloid fibril formation is poorly understood. As a result of the small length scales of molecular interactions, experiments addressing molecular details are extremely difficult. Therefore, the approach taken in this thesis is to apply computer simulations, which are well suited to address these small length scales. Nevertheless, as time scale problems limit the application of computer simulations to assess biomolecular systems, advanced simulation methods have to be applied and/or developed in order to analyse the molecular interactions important in amyloid fibril formation.

When examining the molecular basis of amyloid fibril formation it is important to distinguish two different processes. On the one hand, there is the growth of nuclei, protofilaments and small fibrils into mature fibrils. This process is mainly driven by thermodynamics and hence can be discussed in terms of fibril stability. On the other hand, there is the formation of the initial nuclei. The lag phase associated with this process indicates a kinetic process, making dynamical studies more appropriate.
The molecular interactions that contribute to the thermodynamic stability of fibrils are conveniently discussed by examining the free-energy difference associated with the dissociation/association of one peptide unit from/to the fibril. This free-energy difference is related through the association constant to the potential of mean force, which is commonly obtained by a basic umbrella sampling calculation. In such a calculation the umbrella window will ensure sampling around a specific distance between the dissociated peptide and the last peptide of the fibril by a harmonic potential.

Unfortunately, standard umbrella sampling requires long sampling times to obtain convergence, making these calculations intractable. The convergence of these calculations is hampered by the ensemble sampling problem, i.e. different minima in the energy landscape are only sampled according to their Boltzmann weight after a long time, as crossing the barriers that separate the minima is a rare event. By combining umbrella sampling with replica exchange steps convergence of the potential of mean force is accelerated considerably, because barriers present in one replica can be circumvented by moving to another replica with a smaller or no barrier. As the convergence in replica exchange umbrella sampling is at least one order of magnitude faster than in standard umbrella sampling, this approach now allows to calculate the free-energy difference associated with fibril growth quantitatively for small peptides (∼7 amino acids).

This approach has been applied to examine the effect of different side chains on the fibril stability. Therefore, we have created hypothetical fibrils and protofilaments consisting of 10 and 5 peptides, respectively, for the following peptides: KFFE, KLLE, KVVE and KAAE. In a 5 ns standard molecular dynamics simulation KAAE was unstable. For the remaining peptides, KFFE, KLLE and KVVE, we have calculated the association constant for fibril and protofilament growth. It was shown that these calculations could predict experimental behavior in that KFFE and KVVE form fibrils at a 200 µM concentration as opposed to KLLE. From the calculated association constants the free-energy difference associated with the cross-β-sheet and the lateral interaction could be deduced. The free-energy differences for all three peptides allow for fibril growth, i.e. the free energy differences are negative, but for KFFE the cross-β-sheet interaction is stronger than the lateral interaction while for KLLE and KVVE the opposite relation is found. Furthermore, it is shown that the hydrophobic effect drives fibril growth for these peptides, while loss of side chain conformational entropy and excluded volume effects decrease fibril growth propensity.

The molecular interactions that drive the initial nucleus formation are ideally addressed by evaluating the natural time evolution of nucleus formation from a solution of monomeric peptides. Generally, time evolution of biomolecular systems is examined with molecular dynamics simulations, where the question at hand determines the level of detail applied. Since water molecules can be structurally involved in amyloid fibrils, we wanted to address the nucleation events by all-atom molecular dynamics simulations including explicit water molecules.

Straightforward molecular dynamics, taking all atoms and the solvent molecules explicitly into account, to assess events in biomolecular systems at the microsecond
timescale, such as nucleus formation or protein folding, is currently without reach of the available computational resources. The microsecond timescale in biomolecular systems is related to the search sampling problem. Before reaching the lowest free-energy state the system experiences many folding events, characterized by the concomitant rare barrier crossing from one minimum to another. Therefore we developed a new method that aims to accelerate such folding events in computer simulations, ultimately leading to one of the lowest free-energy structure.

To accelerate protein folding in computer simulations reordering of hydrogen bonds is stimulated by the introduction of alternating hydrogen bond potentials. The rationale behind targeting hydrogen bonds is as follows. There are two major contributions to the structure and stability of a protein structure, namely hydrophobic collapse and hydrogen bonds. The hydrophobic collapse is commonly regarded as the driving force behind protein folding, whereas the hydrogen bonds are necessary to prevent uncompensated (partial) charges in the protein interior. Forming of peptide-peptide hydrogen bonds is not accompanied by a free-energy contribution. Nevertheless, breaking one hydrogen bond and forming another is associated with barriers in the free-energy landscape, thereby decreasing the protein folding rate. Thus, if we can increase the rate of hydrogen bond reordering, we can accelerate protein folding.

To achieve accelerated hydrogen bond reordering, hydrogen bonds need to be broken and subsequently new hydrogen bonds have to be formed. Therefore we created repulsive and attractive hydrogen bond potentials, respectively, that are alternatingly active in the simulation. These alternating hydrogen bond potentials (AHBP) have been applied in computer simulations of two 16-residue peptides, one a native α-helix and the other a native β-sheet. Our new method has been shown to result in a fast and efficient sampling of the conformational space, ultimately leading to the native state.

Although the alternating hydrogen bond potentials induce rapid formation of folded conformations, the nonequilibrium implementation prohibits the quantitative evaluation of the various minima observed. Therefore we reformulated the alternating hydrogen bond potentials into an equilibrium method; hydrogen bond switching.

Hydrogen bond switching (HS) is a hamiltonian replica exchange method. The variation of the hamiltonian over different replicas is accounted for by the addition of repulsive, attractive or no hydrogen bond potential. Similar to the alternating hydrogen bond potentials, in hydrogen bond switching repulsive potentials force breaking of hydrogen bonds, while attractive hydrogen bond potentials facilitate the formation of new hydrogen bonds. Migration through the range of replicas will then induce reordering of the hydrogen bonds.

To test the performance of hydrogen bond switching with respect to temperature replica exchange a β-heptapeptide in methanol was simulated. Both methods sampled the phase space of the β-heptapeptide to a similar extent, although hydrogen bond switching required significantly fewer replicas. Subsequently we simulated a 16-residue polyalanine in water starting from a α-helix and a collapsed coil conformation using hydrogen bond switching. Both simulations display similar conformational minima, but the free-energy landscape did not converge. The replicas appear to be sorted,
Summary

preventing efficient migration through the complete range of replicas and reducing the convergence rate. This effect has also been observed in other hamiltonian replica exchange protocols and hence might be a general problem.

The described methods to increase the rate of hydrogen bond reordering, have been shown to accelerate protein folding events in computer simulations considerably. Nonetheless, application of these methods to assess the initial nucleus formation in amyloidogenesis in atomic detail with an explicit solvent representation requires additional improvements and thus further research.
Samenvatting

Begrijpen van amyloidogenese door computer simulaties. Het aanpakken van bemonsteringsproblemen door gebruik te maken van replica exchange umbrella sampling en afwisselende waterstof-brug potentialen.

Amyloidogenese is het proces van amyloid formatie. De algemene structuur van een amyloidfibril is een touwachtig vezeltje met een diameter van ~10 nm en onbepaalde lengte. Amyloid formatie is een belangrijk kenmerk in veel pathologische aandoeningen, waaronder de ziekte van Alzheimer. Tevens zijn de kleine grootte en de geordende structuur van amyloidvezeltjes gunstige attributen voor toepassingen in nanotechnologie. Verder zijn de hoge trekkraft die amyloidvezeltjes kunnen weerstaan voordat ze breken, evenals het feit dat de groei en de formatie van amyloidvezeltjes een zelfassemblage proces is, erg gewilde eigenschappen voor materiaalwetenschappers.

Een amyloidvezeltje bestaat uit een groot aantal eenheden meestal van hetzelfde peptide. Elk peptide heeft een interactie met de naburige peptiden door intermoleculaire β-sheet interacties, die parallel aan de as van het vezeltje lopen, een cross-β-sheet vormend. Deze cross-β-sheets kunnen laterale interacties loodrecht op de as van het vezeltje aangaan om touwachtige vezeltjes te vormen. Het formeren van amyloidvezeltjes gebeurt in een nucleatie-groei mechanisme. In de initiële fase, nucleatie, vormen de peptiden in de oplossing een structuur van minimale grootte met de kenmerken van het amyloidvezeltje, de nucleus. Deze nucleus kan vervolgens uitgroeien tot een volwassen vezeltje door zich samen te voegen met andere nuclei of met monomeren, hetzij door laterale additie, hetzij door additie aan de uiteinde van het vezeltje.

Ondanks de aanzienlijke interesse in amyloidogenese vanuit de eerdergenoemde kennisgebieden en de hoge-resolutie structuurgegevens die recent beschikbaar zijn gekomen, is de onderliggende moleculaire basis voor de formatie van amyloidvezeltjes nog niet helemaal begrepen. Door de kleine lengteschaal van moleculaire interacties zijn experimenten die de moleculaire details onderzoeken erg moeilijk. Daarom is er in dit proefschrift voor gekozen om computer simulaties te gebruiken, welke zeer
Samenvatting

geschikt zijn om kleine lengteschalen te onderzoeken. Desalniettemin, omdat tijdschalaalproblemen het gebruik van computersimulaties voor de beoordeling van biomoleculaire systemen limiteren moeten geavanceerde simulatie methode gebruikt en/ of ontwikkeld worden om de moleculaire interacties die belangrijk zijn voor de vorming van amyloïdvezeltjes te analyseren.

Bij het onderzoek naar de moleculaire basis voor vorming van amyloïdvezeltjes is het belangrijk twee processen te onderscheiden. Aan de ene kant is er groei van de nuclei, protovezeltjes en kleine vezeltjes tot volwassen vezeltjes. Dit proces is hoofdzakelijk gedreven door thermodynamica en kan daarom besproken worden in termen van de stabiliteit van het vezeltje. Aan de andere kant is er de formatie van de initiële nucleus. De vertragingphase geassocieerd met dit proces wijst op een kinetisch proces, waarvoor dynamische studies meer toepasselijk zijn.

De moleculaire interacties die bijdragen aan de thermodynamische stabiliteit van de vezeltjes zijn eenvoudig te bespreken door de vrije-energieverschillen geassocieerd met de dissociatie/associatie van een peptide eenheid van/ aan het vezeltje te onderzoeken. Dit vrije-energieverschil is via de associatieconstante gerelateerd aan de potentiële van de gemiddelde kracht, welke gewoonlijk verkregen wordt door een standaard umbrella sampling berekening. In een dergelijke berekening zorgt de harmonische potentiële van het umbrella venster ervoor, dat er rond een bepaalde afstand tussen het gedissocieerde peptide en het laatste peptide in het vezeltje bemonsterd wordt.

Helaas vereist standaard umbrella sampling lange bemonsteringstijd om te convergeren, waardoor deze berekeningen onuitvoerbaar zijn. De convergentie van deze berekeningen wordt belemmerd door het ensemble bemonsteringsprobleem, i.e. verschillende minima in het energielandschap worden alleen na een lange tijd volgens hun Boltzmann gewicht bemonsterd, daar het oversteek van de barrières die de minima scheiden zelden voorkomt. Door umbrella sampling met replica exchange stappen te combineren wordt de convergentie van de potentiële van de gemiddelde kracht aanzienlijk versneld, omdat de barrières die in de ene replica aanwezig zijn, omzeild kunnen worden door naar een andere replica te gaan, waar de barrières kleiner zijn of zelfs helemaal niet bestaan. Daar de convergentie in replica exchange umbrella sampling zeker een ordgrootte sneller is dan in standaard umbrella sampling, kan met deze benadering het vrije-energieverschil geassocieerd met de groei van een vezeltje kwantitatief berekend worden voor kleine peptiden (∼7 aminozuren).

Deze benadering is toegepast om het effect van verschillende zijketens op de stabiliteit van het vezeltje te onderzoeken. Daarom hebben we hypothetische vezeltjes en protovezeltjes gecreëerd bestaand uit respectievelijk 10 en 5 peptiden voor de volgende peptiden: KFFE, KLLE, KVVE en KAAE. In een 5 ns standaard moleculaire dynamica simulatie was KAAE onstabiel. Voor de overige peptiden, KFFE, KLLE en KVVE, hebben we de associatieconstante voor de groei van het vezeltje en het protovezeltje uitgerekend. Dit liet zien, dat deze berekeningen het experimentele gedrag, dat KFFE en KVVE, in tegenstelling tot KLLE, vezeltjes vormen bij een 200 µM monomeer concentratie, kan voorspellen. Van de uitgerekende associatieconstanten konden de vrije-
energieverschillen geassocieerd met de cross-β-sheet en de laterale interactie afgeleid worden. De vrije-energieverschillen voor alle drie de peptiden kan tot de vorming van vezeltjes leiden, *i.e.* de vrije-energieverschillen zijn negatief, maar voor KFFE is de cross-β-sheet interactie sterker dan de laterale interactie, terwijl voor KLEE en KVVE de tegenovergestelde relatie gevonden werd. Verder werd aangetoond dat het hydrofobe effect de drijvende kracht is voor vezeltjes groei van deze peptiden, terwijl het verlies van de conformationele entropie van de zijketens en uitgesloten volume effecten de neiging om vezeltjes te vormen verminderd.

De moleculaire interacties die de initiële nucleus formatie drijven, worden in het ideale geval bestudeerd door de natuurlijke tijdsevolutie van nucleus formatie in een oplossingen van monomeer peptiden te evalueren. Normaal gesproken wordt de tijdsevolutie van een biomoleculair systeem bestudeerd met moleculaire dynamica simulaties, waar de vraag in kwestie de mate van detail bepaald. Omdat water moleculen betrokken kunnen zijn bij de structuur van amyloidvezeltjes, wilden wij de nucleaties gebeurtenissen bestuderen met moleculair dynamica op het atomaire niveau inclusief expliciete water moleculen.

Recht door zee moleculaire dynamica, waarbij alle atomen en de moleculen van het oplosmiddel expliciet meegenomen worden, om gebeurtenissen in biomoleculaire systemen zoals nucleusformatie en eiwitvouwing op een tijdschaal van te bestuderen, is momenteel buiten het bereik van de beschikbare computerkracht. De microsecond stijggewijs stimulatie van waterstofbruggen door afwisselende waterstofbrug potentialen te introduceren. De gedachtegang achter het uitlichten van waterstofbruggen is als volgt. Er zijn twee belangrijke contributies aan de structuur en stabiliteit van eiwit structuren, namelijk de hydrofobe instorting en waterstofbruggen. Algemeen wordt aanvaard dat de hydrofobe instorting de drijvende kracht van eiwitvouwing is, terwijl de waterstofbruggen noodzakelijk zijn om ongecompenseerde (partiële) ladingen binnenin het eiwit te voorkomen. Het vormen van waterstofbruggen in het eiwit gaat niet gepaard met een vrije-energiecontribution. Desalniettemin wordt het breken van een waterstofbrug en het vormen van een andere geassocieerd met barrières in het energielandschap, waardoor de snelheid van eiwitvouwing wordt verlaagd. Dus, als we de snelheid van het reorganiseren van waterstofbruggen kunnen verhogen, kunnen we eiwitvouwing versnellen.

Om eiwitvouwing in computersimulaties te versnellen, wordt de reorganisatie van waterstofbruggen gestimuleerd door afwisselende waterstofbrug potentialen te introduceren. De gedachtegang achter het uitlichten van waterstofbruggen is als volgt. Er zijn twee belangrijke contributies aan de structuur en stabiliteit van eiwit structuren, namelijk de hydrofobe instorting en waterstofbruggen. Algemeen wordt aanvaard dat de hydrofobe instorting de drijvende kracht van eiwitvouwing is, terwijl de waterstofbruggen noodzakelijk zijn om ongecompenseerde (partiële) ladingen binnenin het eiwit te voorkomen. Het vormen van waterstofbruggen in het eiwit gaat niet gepaard met een vrije-energiecontribution. Desalniettemin wordt het breken van een waterstofbrug en het vormen van een andere geassocieerd met barrières in het energielandschap, waardoor de snelheid van eiwitvouwing wordt verlaagd. Dus, als we de snelheid van het reorganiseren van waterstofbruggen kunnen verhogen, kunnen we eiwitvouwing versnellen.

Om een versnelde reorganisatie van waterstofbruggen te realiseren moeten bestaande bruggen verbroken worden en nieuwe gevormd. Daarom hebben wij respectievelijk repulsieve en attractieve waterstofbrugpotentialen gemaakt, die afwisselend
Samenvatting

actief zijn in de simulatie. Deze afwisselende waterstofbrugpotentialen zijn toegepast in computer simulaties van twee eiwitten bestaande uit 16 aminozuren, waarvan de één een natieve $\alpha$-helix is en de ander een natieve $\beta$-sheet. We hebben aangetoond dat onze nieuwe methode in een snelle en efficiënte vouwing van het eiwit resulteert, wat uiteindelijk tot de natieve toestand leidt.

Hoewel de afwisselende waterstofbruggen snelle vorming van gevouwen conformaties induceert, verhindert de niet-evenwichtsimplementatie de kwantitatieve evaluatie van de verschillende waargenomen minima. Daarom hebben we de afwisselende waterstofbrugpotentialen geherformuleerd tot een evenwichtsmethode; waterstofbrugwisseling.

Waterstofbrugwisseling is een hamiltonian replica exchange methode. De afwijking in de hamiltoniaan voor de verschillende replica’s wordt verzorgd door de toevoeging van een repulsieve, een attractieve of geen waterstofbrugpotentiaal. Vergelijkbaar met de afwisselende waterstofbrugpotentialen zorgen in waterstofbrugwisseling de repulsieve potentialen voor het breken van waterstofbruggen, terwijl de attractieve potentialen de vorming van nieuwe faciliteert. Migratie door het bereik van replica’s zal dan de reorganisatie van de waterstofbruggen induceren.

Om de prestatie van waterstofbrugwisseling met respect tot temperatuur replica exchange te testen, hebben we een $\beta$-heptapeptide in methanol gesimuleerd. Beide methoden bemonsterden the faseruimte van de $\beta$-heptapeptide in vergelijkbare mate, terwijl waterstofbrugwisseling significant minder replica’s nodig had. Vervolgens hebben we een polyalanine bestaande uit 16 residuen in water gesimuleerd met waterstofbrugwisseling, met respectievelijk een $\alpha$-helix en een opgevouwen keten als beginconformatie. Beide simulaties toonden vergelijkbare conformationele minima, maar het vrije-energiediagram was niet geconvergeerd. De replica’s lijken gesorteerd te zijn, wat een efficiënte migratie door het gehele bereik van de replica’s verhindert en de convergentiesnelheid reduceert. Dit effect is ook in andere hamiltonian replica exchange methoden waargenomen en kan daarom een algemeen probleem van replica exchange technieken zijn.

De beschreven methoden om de reorganisatie van waterstofbruggen te versnellen, zorgen voor een aanzienlijke versnelling van de eiwitvouwingsgebeurtenissen in computersimulaties. Desalniettemin benodigd het gebruik van deze methoden, om de initiële nucleus formatie in amyloidogenese te bestuderen op atomair niveau met een expliciete oplosmiddel, additionele verbeteringen en dus verder onderzoek.
Dankwoord


Ook ben ik Jocelyne erg dankbaar. Toen de vakgroep in Delft opgeheven werd, heeft zij ervoor gezorgd dat ik mij deels kon aansluiten bij de vakgroep in Amsterdam. Naast dat zij ervoor heeft gezorgd dat ik in een wetenschappelijk stimulerende omgeving bleef, heeft zij mij ertoe aangezet om mijn eerste artikel te schrijven.

Mijn ouders, Bep en Ben, en mijn zus, Evelien, wil ik bij deze heel erg bedanken voor hun motiverende gesprekken en de gezellige etentjes op dinsdag.

Als laatst wil ik natuurlijk ook mijn vrouw bedanken. Zij was een luisterend oor tijdens menig wandeling, waarin ik verhandelde over kralenkettingen, berglandschappen en hoe je van de ene vallei in de andere moest komen. Eline, bedankt voor je onvoorwaardelijke steun.
CV van Maarten Gijsbrecht Wolf

geboren te Heemskerk, op 6 mei 1980


1998-2003 Studie scheikunde aan de Universiteit van Amsterdam met als afstudeerrichting Fysische Chemie. Het onderwerp van het afstudeerproject was computersimulaties aan colloidale deeltjes in polymeeroplossingen, begeleid door Dr. Evert-Jan Meijer en Prof. Dr. Peter Bolhuis. Het onderwerp van de eindscriptie was theoretische beschrijvingen van polymeren, beoordeeld door Prof. Dr. Peter Bolhuis.

2004-2008 Promotieonderzoek aan de Technische universiteit Delft. Dit onderzoek was gericht op computersimulaties om de moleculaire basis van amyloid formatie te begrijpen, wat in dit proefschrift beschreven is. Tijdens deze periode zijn de volgende vakken gedoceerd: Inleiding in de scheikunde en Een introductie in chemische thermodynamica. Tevens zijn Bachelor en Master studenten begeleid.
Publications

M.G. Wolf and S.W. de Leeuw
Fast in silico protein folding by introduction of alternating hydrogen bond potentials.
Biophysical Journal, 2008, 94, 3742-3747

M.G. Wolf, J. Vreede, P.G. Bolhuis and S.W. de Leeuw
Improved hydrogen bond potentials to assess in silico protein folding.
CBSB08 Conference proceedings, 2008, NIC-Directors, Jülich Germany, 413-416

M.G. Wolf, J. van Gestel and S.W. de Leeuw
In Nanostructure Design: Methods in Molecular Biology, (E. Gazit and R. Nussinov, Eds.)
Humana Press, 2008, Totowa, New Jersey, USA, 153-179

M.G. Wolf, J.A. Jongejan, J.D. Laman and S.W. de Leeuw.
Rapid free energy calculation of peptide self-assembly by REMD umbrella sampling.

M.G. Wolf, J.A. Jongejan, J.D. Laman and S.W. de Leeuw
Quantitative Assessment of Amyloid Fibril Growth of Short Peptides from Simulations:
Calculating Association Constants to Dissect Side Chain Importance.
Journal of the American Chemical Society, 2008, 130, 15772-15773

J. Vreede, M.G. Wolf, S.W. de Leeuw and P.G. Bolhuis
Hydrogen bond switching: An efficiënt and general approach to investigate protein folding.
submitted

Patent

M.G. Wolf, J. Flohil and S.W. de Leeuw
A method for generating information of a 3-dimensional molecular structure of a molecule.
NL 47482 - MP (2007)