2010

A study of the elongation of fibrils of the Alzheimer's amyloid beta protein with a coarse-grained force field

Ana Rojas
Louisiana State University and Agricultural and Mechanical College, arojas1@lsu.edu

A STUDY OF THE ELONGATION OF FIBRILS OF THE ALZHEIMER'S AMYLOID BETA PROTEIN WITH A COARSE-GRAINED FORCE FIELD

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Physics And Astronomy

by
Ana Rojas
Licenciada en Física, Universidad Nacional de Córdoba, 2001
December 2010

Acknowledgments

It is a pleasure to thank my advisor Professor Dana Browne for his support in my journey into the field of computational biology. For his willingness to learn new topics, for the time he dedicated, for his advice, and his questions to always make sure that I considered the physics of my research.

I would like to thank Professor Harold Scheraga from Cornell University for receiving me in his group, for his advice and patience, and for the great care he has put in carefully revising each of my manuscripts.

I am very grateful to Professor Adam Liwo from University of Gdansk in Poland, who patiently taught me the details of the UNRES force field, and helped me to develop the software for my simulations. From him I learned how to sort the day to day problems that I encounter during my research.

This work was supported by the Economic Development Assistantship Program and by the CCT Graduate Assistantship Program from Louisiana State University.

This research was conducted by using the resources of (a) the Beowulf cluster at Baker Laboratory of Chemistry, Cornell University, (b)the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdansk, and (d) the resources of the Center for Computation and Technology at Louisiana State University.

Table of Contents

Abstract

Alzheimer's disease (AD) is a brain disorder named after the German psychiatrist Alois Alzheimer, who first described the disease in 1906[1, 2]. AD is a form of dementia, characterized by impairement of memory and other intellectual abilities, caused by the loss of neurons and synapses (the structures that permit comunication between neurons) in certain regions of the brain[3]. Although the mechanism by which neurons are affected is not yet well understood, research points to a small protein, the β-amyloid peptide (Aβ), as the first suspect[4, 5, 6, 3, 7]. Aβ chains are present in the brains of healthy individuals, but in AD patients, they associate and form clumps that deposit outside neurons, and are believed to trigger the disease[6].

In this dissertation, I use computational methods to study the behavior of the structures formed by Aβ chains, when they associates with each other, and how these structures grow. The goal of this work is to use computational methods to complement experimental results by filling the gaps about structural and association mechanism that cannot be accessed by experiments.

In chapter 1, I will guide the reader through several concepts needed to understand this work. In section 1.1, I will describe what proteins are. In section 1.2, I will outline the protein folding problem, a broad problem that encompasses the one that I treat in this dissertation. Section 1.3, includes a summary of the computational and experimental methods applied today to study the folding of proteins, and the motivation for choosing the method used in this work.

Chapter 2 describes the equations and the software that I developed to study protein association in general, and Aβ in particular and the tests that I carried out to assure that the method was capable of treating systems formed by several protein chains.

In chapter 3, I will describe the state of the research on AD, and the role and mechanism of Aβ in the disease. The Aβ-AD connection is a complex problem, and I have chosen to study one aspect of this problem, namely the stability and growth of Aβ fibrils (structures formed by Aβ, which are described in section 3.4). Also in chapter 3, I will explain the motivation for choosing this particular aspect of the problem. I will also describe the methods used in my study and the results obtained. In section 3.7, I will elaborate on the implications of the results and on how they complement experiments.

Chapter 1
Introduction

1.1 Proteins

1.1.1 Amino Acids

Proteins are crucial to virtually all biological processes; they transport molecules, catalyze chemical reactions, form structures for cellular organization, transmit information between cells and compartments, control gene expression, and provide immune functions in complex organisms. Proteins are formed by amino acids arranged in long chains, often referred to as polypeptide chains. There are 20 different amino acids that can be combined to form proteins. Each protein in a living organism has its unique sequence of amino acids, with the template for the the sequence encoded in the organism's DNA.

All amino acids have the general structure shown in Figure 1.1. They differ only in the composition of the side chain (denoted by the letter R in Figure 1.1). The side chain determines the amino acid's specific physical properties, such as size, flexibility, charge, polarity, etc.

The amino acids in a protein chain are linked by peptide bonds (a strong chemical bond) formed between the carboxyl carbon of one amino acid and the amino nitrogen of the second amino acid (see Figure 1.2). Because these peptide bonds have a partial double bond character, rotation around this bond is restricted. Thus, the peptide unit is planar, and rotation is restricted to the bonds involving the α carbons. The end of the chain with the free amino nitrogen is called the N-terminus, while the end with the free carboxyl carbon is called the C-terminus, and the amino acid sequence is numbered starting from the N-terminus.

Depending on the polarity of the side chain, amino acids are classified into polar or non polar. While polar amino acids will tend to interact with a polar environment (such as water), nonpolar amino acids will rather interact with each other or with a nonpolar environment (such as the lipid bilayer of a cell). This property plays a major role in determining the arrangement that a protein chain adopts.

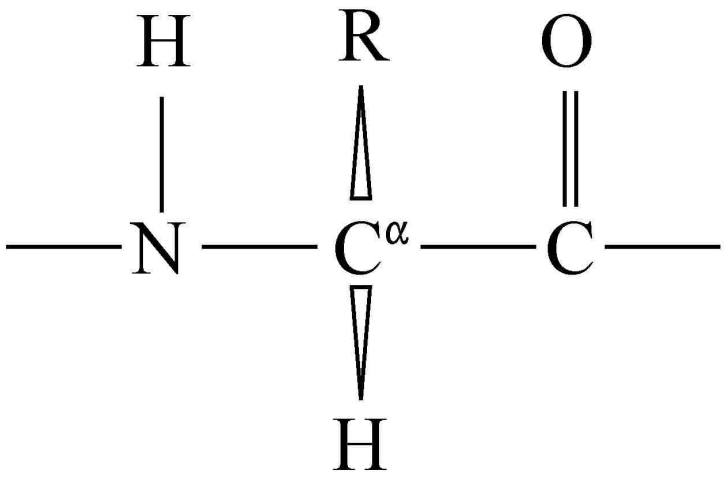

FIGURE 1.1. The general structure of an amino acid.

1.1.2 Protein Structure: The α-Helix and the β-Sheet

In order to perform its biological function, a protein must adopt a specific shape. Different proteins (*i.e.*, different sequences of amino acids) adopt different structures, giving rise to a great diversity of tools, each with a specific task. The shape into which a protein must fold is known as its native conformation. The native conformation of a protein is that with the lowest free energy. *i.e.*, that conformation which minimizes the free energy of the protein chain and its environment.

Although combinations of the 20 amino acids allow for a great diversity of conformations, these structures are mainly composed of two motifs: the α-helix and the β-sheet. In an α-helix [see Figure 1.3 (A)], as the name suggests, the amino acids are arranged forming a right-handed helix with 3.6 residues (amino acids in a peptide chain) per turn, giving a pitch (distance between consecutive turns along the axis of the helix) of 5.4 Å. In a β-sheet [see Figure 1.3 (B)], regions of the protein (strands) align adjacent to each other in parallel or anti-parallel orientation. Both these structures are mainly stabilized by hydrogen bonds. A hydrogen bond is formed when a pair of electronegative atoms, such as oxygen and nitrogen, share a hydrogen between them. In the α-helix, the hydrogen bonds are formed between the N-H group of one turn and the C=O group of the neighboring turn [see Figure 1.3 (A)], and in the β-sheet, they are formed between these same atoms but from consecutive strands, as shown in Figure 1.3 (B).

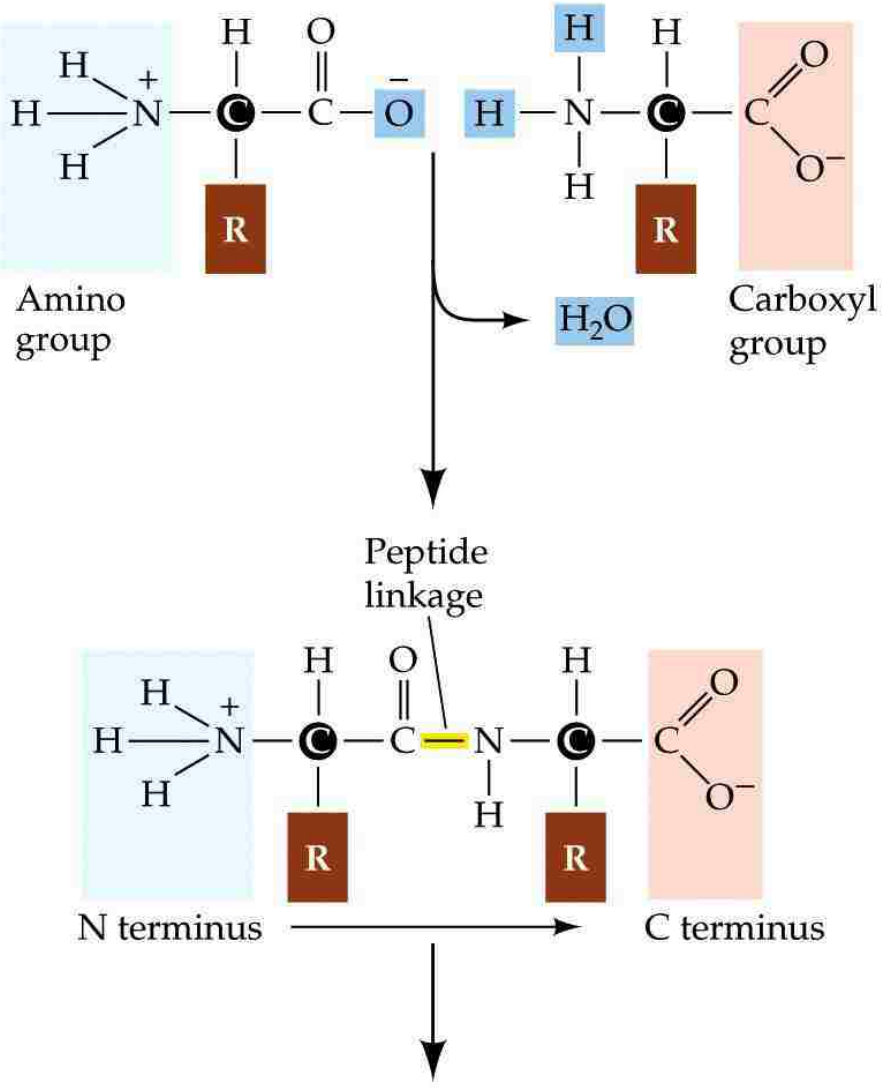

FIGURE 1.2. Proteins are chains of amino acids joined together by peptide bonds. When a peptide bond is formed a molecule of water is released.

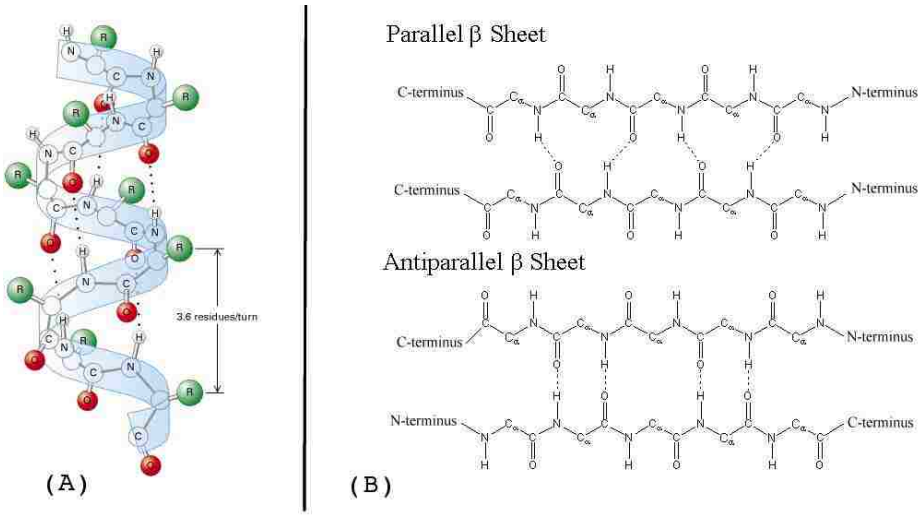

FIGURE 1.3. Secondary structure of proteins. (A) Representation of an α-helix. The side chains are represented by green spheres and denoted by the letter R. The hydrogen bonds between backbone atoms are represented by dotted lines. (B) Representation of parallel and antiparallel β-sheets. The hydrogen bonds between the backbone atoms in the different strands are represented by dotted lines. The side chains have been omitted.

Four structural levels are used to characterize the structure of proteins:

- **Primary structure** refers to the amino acid sequence of the chain.

- **Secondary structure** refers to the sub-structures, α-helix and β-sheets, in the peptide chain. To describe the secondary structure of a protein, one should know whether the chain adopts the α-helix or β-sheet conformation, or a combination of both.

- **Tertiary structure** is the three-dimensional structure of a single protein chain; the spatial arrangement of the secondary structure elements. To describe the tertiary structure of a protein, one should know the relative orientation and positions of the α-helices or β-sheets in the protein.

- **Quaternary structure** is the highest level used to describe protein structures, and it is used to describe molecules composed of several protein chains. Many proteins exist in multimeric form, meaning that two or more chains bind to each other to form a molecule complex. Each chain in the complex is referred to as a monomer, and the complex as a multimer or an oligomer. To describe the quaternary structure of a protein, one should know the relative arrangement of the monomers conforming the complex.

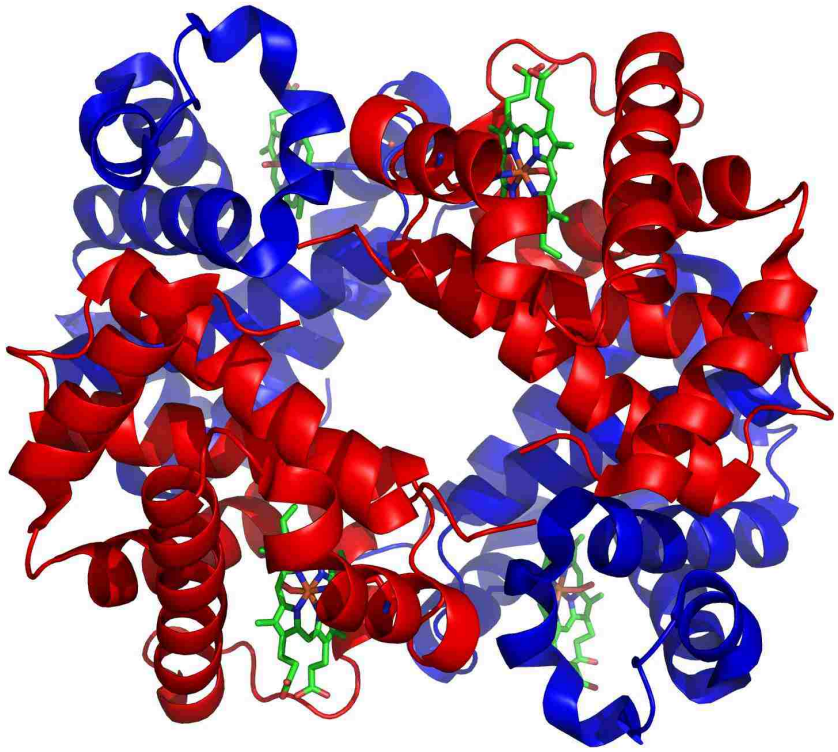

FIGURE 1.4. The three-dimensional structure of human hemoglobin (PDB ID: 1GZX). The α and β subunits are colored in red and blue, respectively. The small, green colored, molecules are the iron-containing heme groups.

Many multimeric proteins are formed by identical subunits, these proteins are called homodimers (two subunits), homotrimers (three subunits), and so forth, depending on the number of subunits. The subunits can also be different, in which case the multimers are referred to as hetero-dimers, hetero-trimers, etc. The most typical example of a an oligomer is hemoglobin, the oxygen carrying protein of the blood (see Figure 1.4). Hemoglobin is hetero-tetramer formed by two α and two β subunits (both with α-helical structures) arranged with the structure shown in Figure 1.4.

1.1.3 Misfolding and Disease

Although the amino acid sequence and the protein environment dictate the biologically active conformation of a protein, it is possible that a protein would adopt the wrong conformation. The cell has a quality control mechanisms that ensure that misfolded proteins are eliminated. However, this mechanism is not always effective enough, and in recent years, it has been discovered that many diseases are

caused by the aggregation of misfolded proteins. Examples of such diseased are Alzheimer's disease (AD), Parkinson's disease (PD), type II diabetes, and spongiform encephalopathies. Understanding the mechanism that leads a protein towards its misfolded conformation is crucial to the design of therapeutic agents for these diseases. Protein misfolding is therefore an active area of research in the protein field.

1.2 The Protein Folding Problem

More than 40 years ago, Cyrus Levinthal[8] posed a paradox that challenges us to understand more deeply the physical processes involved in protein folding. Levinthal suggested that a protein cannot fold by exhaustively searching through all possible conformations, but instead it must have a more efficient way to find its native conformation. To illustrate the paradox, let us assume that a single amino acid, through varied positioning of its backbone atoms and side-chain atoms, can adopt five distinct but stable conformations. If we consider a moderately sized protein of 100 amino acid residues, this yields $5^{100} \simeq 10^{70}$ conformations of the chain. If a newly synthesized protein were to spend 10^{-15} seconds in each conformation, fully searching all the possible conformations for the correct, native one would take 10^{55} seconds, many orders of magnitude longer than the age of the universe. Yet, actual proteins fold in the millisecond to second range.

Levinthal's paradox lead to the view that each protein fold following a specific pathway. Although the modern view is that proteins choose among many competing pathways, it is clear that protein folding is guided by the physical forces between the different atoms in the protein and between these atoms and those in the environment. Understanding the physical principles dictating protein folding and predicting the folding pathway of proteins is an intensive area of research that has been approached by computational and experimental methods.

1.3 Methods Used to Study Native Structures and Folding of Proteins

1.3.1 Experimental Methods

The two main methods utilized to determine the structure of proteins with atomic resolution are X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. In X-ray crystallography, a protein crystal is grown, and the interference pattern produced as X-rays pass through the crystal lattice is used to determine the arrangement of atoms within the lattice. In NMR spectroscopy, a

magnetic field is applied to the sample. The magnetic field causes the degenerate state of proton's spins to split. The degree of this splitting depends on the electronic environment of the proton-containing atoms, and can be used to identify atoms in the protein. Further, two-dimensional NMR reveals coupling between protons that are linked by a small number of covalent bonds or are separated by small spatial distances. The coupling patterns thus supply a set of distance constraints on the positions of atoms in the protein. Using models for the protein, structures can be determined to satisfy the constraints.

The two most important techniques used to study the protein folding process have been fluorescence spectroscopy and circular dichroism (CD). Fluorescence spectroscopy is used when any of the fluorescent amino acids tryptophan or tyrosine are present in the protein. The technique consist in exciting the electrons in the sample by a beam of light. The emission spectra of the fluorophore (in this case the fluorescent amino acid) is very sensitive to the local environment, hence, the spectra can be used to determine the conformational state of the protein. CD spectroscopy makes use of a property of chiral molecules known as optical activity. When left and right circularly polarized light travels through a chiral molecule, each one experiences a different refractive index. As a consequence, the direction of the polarization plane is changed. The two main motifs in the secondary structure of proteins, the α-helix and the β-sheet, are chiral, and each structure produces a different CD spectrum. CD spectroscopy is often used to monitor the formation of secondary structure as a function of time.

1.3.2 Computational Methods

Predicting the native structure of a protein remains one the most difficult problems in contemporary computational biology. The two main approaches to the problem have been *comparative* and *ab initio* modeling. Comparative modeling refers to techniques such as homology modeling[9] or threading[10], which use information from databases, of previously solved structures, to predict the native structure of another protein. In the Ab initio approach, the three-dimensional structure of the protein is determined based on physical principles. A Hamiltonian for the protein system is proposed, and different computational techniques are used to find the global free energy minimum of the system.

An even more challenging problem is the prediction of the folding pathway of a protein. The main technique used to study the folding process of a protein is molecular dynamics (MD). MD

algorithms calculate the time evolution of the system by numerical integration of the equations of motion. When MD is applied, the ab initio approach has the advantage that it provides a physical Hamiltonian from which the equations of motion can be derived. By carrying out MD simulations, thermodynamic and kinetic information about the different stages of the folding process as well as the final structure are obtained. To accomplish such predictions, it is necessary to simulate the folding process in real time, starting from a statistical coil (unfolded) conformation, until the native structure is reached. For such a simulation to be accurate, the energy function of the system should ideally include the interactions between all the atoms in the protein and the solvent[11]. Models with such resolution are called all-atom models, and when they include the atoms in the solvent, they are referred to as explicit-solvent all-atom models. Although these models can give very detailed information of the system, with today's computational power, explicit-solvent all-atom molecular dynamics (MD) algorithms can simulate only events that range up to nanoseconds for typical proteins or microseconds for very small ones[11, 12, 13]. These time scales are at least 1 order of magnitude smaller than the folding times of proteins. To overcome this problem, all-atom simulations either implement alternative sampling methods, such as umbrella sampling[14], or simulate the unfolding process and some aspects of its refolding[11, 12].

Initially, simulations primarily treated single-chain proteins, but with the increase in computational power, studies of oligomeric proteins have also been possible [15, 16, 17, 18, 19, 20]. In general, simulations of oligomers either study the stability of a specific structure[16, 20] or the kinetics of folding and/or assembly[15, 17, 18, 19] of the subunits. Stability studies are usually carried out by all-atom MD[16, 17, 20], but this technique is computationally too expensive to study the kinetics of the folding process. To reduce the computational cost, the main approach has made use of minimal models; a minimal model is one for which each amino acid is represented by a few interaction sites, reducing the dimensionality of the problem. Although the information that they can provide is not as detailed as that obtained by all-atom models, they can achieve longer simulation times. Some minimal models have further reduced the computational cost by using a Gō-type potential[17, 18, 19], which creates a funnel-like landscape biased towards the native structure, thereby speeding up the folding process. Although Gō-type potentials can fold proteins in very short time, the energy landscape has been oversimplified by the biased introduce, and therefore, the information that they provide is not

physical. It has also been possible to study the kinetics of oligomeric proteins without including any structural knowledge of the particular protein of interest. For example, Vieth *et al.*[15] used a lattice model with a statistical potential (*i.e.*, biased towards structures in a library, but not towards the particular structure being studied) and Monte Carlo (MC) dynamics to study the folding pathway of the GCN4 leucine zipper from randomly generated initial structures.

In recent years, a molecular dynamics algorithm has been developed[21, 22, 23, 24] for the physics-based united-residue (UNRES) force field[25, 26, 27, 28, 29, 30]. I will refer to this implementation of UNRES as UNRES/MD. UNRES was originally designed and parameterized to locate native-like structures of proteins, but with the implementation of UNRES/MD it can also be used to predict the folding pathway of proteins. The latest version of UNRES force field, referred to as 4P[30], was optimized on four training proteins: 1GAB (all-α), 1E0L (all-β), 1E0G ($\alpha+\beta$) and 1IGD ($\alpha+\beta$). It performed well in the CASP6 exercise[31]; the largest molecule that was folded with this force field contained 225 amino acid residues.

Since the degrees of freedom corresponding to the fastest motions are averaged out in UNRES[26], UNRES/MD was able to simulate events that fall into the microsecond time scale[23]. After the success of UNRES/MD with single-chain proteins, it seemed an excellent choice to study the association of multiple-chain proteins.

Chapter 2

Implementing MD in UNRES to Treat Multichain Proteins*

2.1 Summary

This chapter describes the implementation of MD with the united-residue (UNRES) force field for multichain proteins. In section 2.2, I describe the UNRES force field, its extension to multichain proteins, and the implementation of MD. I describe with great detail the equations of motion and the algorithms used for the MD simulations. Therefore, section 2.2 might be useful to anyone trying to reproduce the work. Once MD was implemented, the software was tested on multimeric proteins with relatively simple architectures. Section 2.3 elaborates on the selection of the appropriate system for testing the implementation. The test simulations were carried out at constant temperature by means of Berendsen or Langevin thermostats. With both thermostats the method was capable of finding the conformations corresponding to UNRES global minimum of energy. Section 2.4, describes the simulations carried out to test the implementation and the results obtained. This chapter ends with section 2.5, which is a short summary of the results.

2.2 The United Residues Force Field (UNRES)

UNRES (see Fig 2.1) is a coarse-grained model[25, 26, 27, 28, 29, 30, 31, 32] in which the backbone is represented as a sequence of α-carbon (C^{α}) atoms linked by virtual bonds designated as **dC**, with united peptide groups (p's) in their centers. United side chains (SC) are connected by virtual bonds designated as **dX** to the backbone at the C^{α} positions with the center of mass of the side chain at the end of **dX**. The geometry of the protein is then fully described by the virtual bond vectors **dC**'s and **dX**'s. Since the forces in UNRES are exerted on the peptide groups and side chains, hereafter I will use the term interacting sites to refer to both united peptide groups and side chains. The complete UNRES potential energy function for a single chain is given by equation 2.1

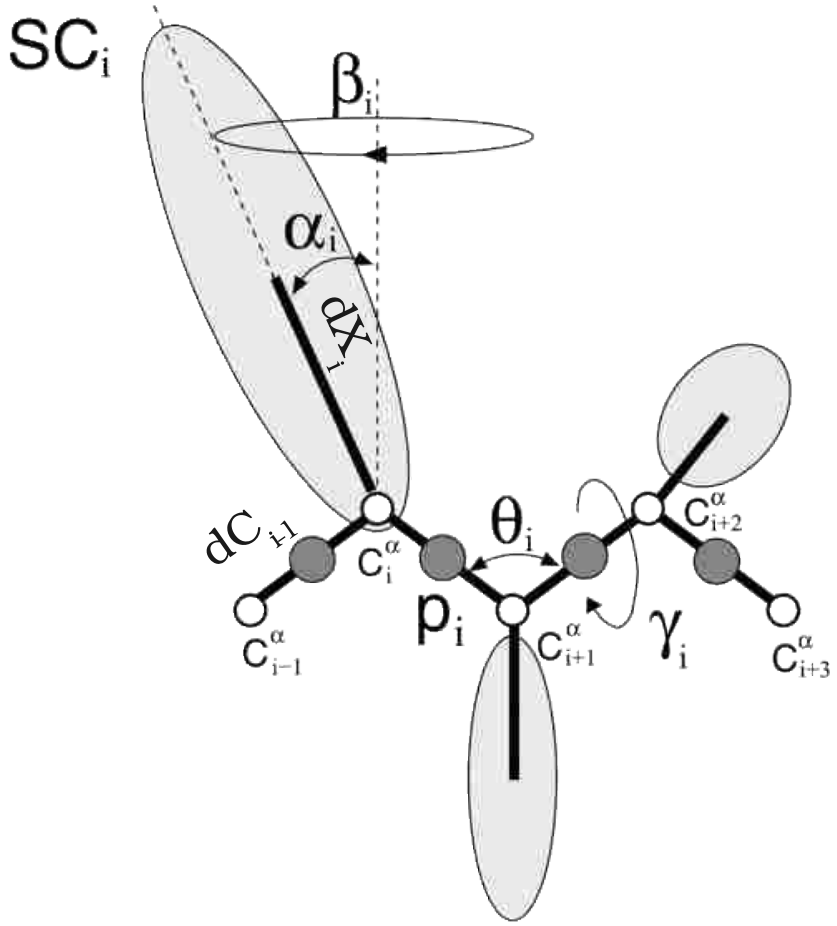

FIGURE 2.1. UNRES representation of a polypeptide chain. Filled circles represent the united peptide groups (p's), and open circles represent the C^{α} atoms, which serve as geometric points. Ellipsoids with their centers of mass at **dX** positions represent UNRES side chains (SC's). The p's are located halfway between two consecutive C^{α} atoms, at positions (1/2)**dC**. The conformation of the polypeptide chain can be described fully by either the coordinates of all the **dC** and **dX** vectors or by the virtual bond angles θ, the virtual bond dihedral angles γ, and the angles α and β defining the orientation of the side chain with respect to the backbone.

$$
\begin{aligned}
U_{\text{single chain}} &= \sum_j \sum_{i<j} U_{\mathrm{SC}_i\mathrm{SC}_j} + \sum_{ss} U_{\mathrm{Cys}_{iss}\mathrm{Cys}_{jss}} + w_{\mathrm{SCp}} \sum_j \sum_{i\neq j} U_{\mathrm{SC}_i\mathrm{p}_j} \\
&+ w_{\mathrm{pp}}^{\mathrm{VDW}} \sum_j \sum_{i<j-1} U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{VDW}} + w_{\mathrm{pp}}^{\mathrm{el}} \sum_j \sum_{i<j-1} U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{el}} + w_{\mathrm{tor}} \sum_i U_{\mathrm{tor}}(\gamma_i) \\
&+ w_{\mathrm{tord}} \sum_i U_{\mathrm{tor}}(\gamma_i, \gamma_{i+1}) + w_{\mathrm{b}} \sum_i U_{\mathrm{b}}(\theta_i) + w_{\mathrm{rot}} \sum_i U_{\mathrm{rot}}(\alpha_{\mathrm{SC}_i}, \beta_{\mathrm{SC}_i}) \\
&+ \sum_{m=2}^{6} w_{\mathrm{corr}}^{(m)} U_{\mathrm{corr}}^{(m)} + w_{\mathrm{vib}} \sum_i U_{\mathrm{vib}}(\mathrm{d}_i)
\end{aligned} \tag{2.1}
$$

where the indices i and j run over the residues. The terms $U_{SC_iSC_j}$ (derived and parametrized in ref [25]) correspond to the mean free energy of hydrophobic (hydrophilic) interactions between the side chains. These terms implicitly contain the contributions from the interactions of the side chain with the solvent. The terms $U_{Cys_{iss}Cys_{jss}}$ (derived and parametrized in ref [33]) account for the energy of disulfide bonds, with ss running through all those pairs of half-cysteines that are known a priori to form disulfide bonds[33]. $U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{VDW}}$ and $U_{\mathrm{p}_i\mathrm{p}_j}^{\mathrm{el}}$ correspond to Lennard-Jones and electrostatic interaction energies between peptide groups, respectively[27]; The terms $U_{\mathrm{SC}_i\mathrm{p}_j}$ correspond to the excluded volume potential of the side chain-peptide group interactions. The terms U_{tor} and U_{tord} (derived and parametrized in ref [29]) are the torsional and the double-torsional potentials, respectively, for the rotation about a given virtual bond or two consecutive virtual bonds. The terms U_{b} and U_{rot} (derived and parametrized in ref [26]) are the virtual-angle bending and side-chain-rotamer potentials, respectively. The terms $U_{\mathrm{corr}}^{(m)}$ (m) (derived in ref [27] and parametrized in ref [30]) correspond to the correlations (of order m) between peptide group electrostatic and backbone-local interactions. The terms $U_{\mathrm{vib}}(\mathrm{d}_i)$ (derived and parametrized in ref [21]), d_i being the length of the ith virtual bond, are simple harmonic potentials defined by eq 2.2

$$
U_{\mathrm{vib}}(\mathrm{d}_i) = \frac{1}{2} k (\mathrm{d}_i - \mathrm{d}_i^{\circ})^2 \tag{2.2}
$$

where k is a force constant, currently set at 500 kcal/(mol Å^2) and d_i° is the average length (corresponding to that used in the fixed-bond UNRES) of the ith virtual bond. The w's in eq 2.1 are the weights of the respective terms.

The UNRES force field has also been extended to multiplechain proteins[34]. In the present work, the interchain interaction energies (and their form, parameters, and weights) were taken to be the same as those of the intrachain terms in the treatment of single chains. However, since the interacting sites between chains are not backbone-connected, not all the terms present in eq 2.1 contribute to the interchain energy. The interaction energy between two different chains (identified by superscripts k and l, respectively) can be expressed by

$$\begin{aligned} U^{k,l}_{\text{inerchain}} &= \sum_i \sum_j U_{\text{SC}^k_i \text{SC}^l_j} + \sum_{ss^{k,l}} U_{\text{Cys}^k_{iss} \text{Cys}^l_{jss}} + w_{\text{SCp}} \sum_i \sum_j U_{\text{SC}^k_i \text{p}^l_j} \\ &+ w_{\text{SCp}} \sum_i \sum_j U_{\text{p}^k_i \text{SC}^l_j} + w^{\text{VDW}}_{\text{pp}} \sum_i \sum_j U^{\text{VDW}}_{\text{p}^k_i \text{p}^l_j} + w^{\text{el}}_{\text{pp}} \sum_i \sum_j U^{\text{el}}_{\text{p}^k_i \text{p}^l_j} \\ &+ \sum_{m=2}^{6} w^{(m)}_{\text{corr,nonadj}} U^{(m)}_{\text{corr,nonadj}} \end{aligned} \tag{2.3}$$

where $U^{(m)}_{\text{corr,nonadj}}$ represents the correlation terms, or order m, corresponding to interactions between nonadjacent residues. The different terms in eq 2.3 have the same form, and the weights have the same values, as those in eq 2.1. Detailed descriptions of each of the terms in eqs 2.1 and 2.3 can be found in refs [25, 26, 27, 29, 30], and [33].

It should be noted that eq 2.3 accounts only for the interaction between two chains in the system. Hence, eq 2.3 is only part of the contribution to the complete multiple-chain potential energy. It should also be mentioned here that, for the force field used in this work (4P force field[32]), the weights of the fifth- and sixth-order correlation terms, $w^{(5)}_{\text{corr}}$ and $w^{(6)}_{\text{corr}}$ in eq 2.1 and $w^{(5)}_{\text{corr,nonadj}}$ and $w^{(6)}_{\text{corr,nonadj}}$ in eq 2.3, are zero[32], but these terms have been included in the equations for completeness.

To mimic peptide concentrations, the system was confined within a soft sphere. This was done by adding another term, U_{conf}, to the potential energy, causing each interacting site (either a peptide group or a side chain) to feel an attractive force toward the center of the sphere whenever it is outside the boundary of the sphere. This potential, which is added to eqs 2.1 and 2.3, is defined by eq 2.4

$$U_{conf} = \sum_k \sum_i u_{\mathrm{conf}_i^k} \tag{2.4}$$

where $u_{\mathrm{conf}_i^k}$, the confining potential acting on interacting site i in chain k, is given by

$$u_{\mathrm{conf}_i^k} = \begin{cases} 0 & \text{if} \quad r_i^k \leq R_0 \\ \mathrm{k}_c (r_i^k - R_0)^4 & \text{if} \quad r_i^k > R_0 \end{cases} \tag{2.5}$$

where k_c is a force constant with unit value ($\mathrm{k}_c = 1$ kcal/(mol Å^4)), r_i^k is the distance from interacting site i to the center of the sphere (placed at the center of mass of the initial conformation), and R_0 is the radius of the sphere. The radius of the sphere determines the volume of the system (volume $= 4\pi(R_0)^3/3$). Therefore, the value of R_0 and the number of peptide chains in the solution determine the peptide concentration of the simulated solution (see section 2.4 for details of the concentrations used in the simulations); in all simulations, the number of chains was taken as the number of chains in the multichain complex.

Combining eqs 2.1, 2.3, and 2.4, we obtain the multiple-chain UNRES potential energy (eq 2.6)

$$U = \sum_k U^k_{\text{single chain}} + \sum_k \sum_{l>k} U^{k,l}_{\text{inerchain}} + U_{\text{conf}} \tag{2.6}$$

where the indices k and l run through the different chains.

2.2.1 Equations of Motion

To find the time evolution of a system, it is necessary to solve the equations of motion of the system. In general, for a system with generalized coordinates q_1, q_2, ..., q_n and generalized momenta $\dot{q}_1$, ..., $\dot{q}_n$, this is equivalent to solving the set of Lagrange's equations

$$\frac{d}{dt}\Big[\nabla_{\dot{q}_i} L(q_1, q_2, \ldots, \dot{q}_1, \dot{q}_2, \ldots)\Big] - \nabla_{q_i} L(q_1, q_2, \ldots, \dot{q}_1, \dot{q}_2, \ldots) = Q_i \tag{2.7}$$

where $i = 1, ..., n$, L is the Lagrangian of the system, and the Q_i's are the generalized dissipative (Rayleigh) forces acting on the system.

The Q_i's are nonconservative forces and, therefore, cannot be derived from the potential energy of the system. For our system, these nonconservative forces are the friction and stochastic forces; they represent collisions with the solvent molecules due to a net motion of the system and random impact of the fluctuating solvent molecules on the solute molecules, respectively, as well as the net effect of averaging out the internal secondary degrees of freedom of the protein molecule. Each Cartesian component of each generalized force will have the form

$$Q_i = -\gamma_i v_i(t) + f_i^{rand} \tag{2.8}$$

with γ_i and $v_i(t)$ being the friction coefficient and velocity related to the ith coordinate and f_i^{rand} being a stochastic force with zero mean and intensity given by[35] eq 2.9

$$< f_i^{rand}(t) f_j^{rand}(t+\tau) >= 2\gamma_i R T_0 \delta(\tau) \delta_{ij} \tag{2.9}$$

where R is the universal gas constant, T_0 is the temperature of the bath, $\delta(\tau)$ is the Dirac delta function (evaluated at an arbitrary time interval τ), and δ_{ij} is the Kronecker delta function. When the Q_i's are identified with the sum of the stochastic and friction forces, they account for the coupling of the protein chain(s) under study to the solvent, which in turn acts as a thermostat, thereby maintaining an average constant temperature of the system.

Following previous work[21], I chose to describe each chain by a set of virtual bond vectors $\mathbf{dC}_i^k$ and $\mathbf{dX}_i^k$, with $\mathbf{dC}_i^k$ being the vector pointing from $C_{i^k}^\alpha$ to $C_{i+1^k}^\alpha$ (Figure 2.1), except for $\mathbf{dC_0^k}$ which points from the origin to the first C^α in the chain, and $\mathbf{dX}_i^k$ being the vector pointing from $C_{i^k}^\alpha$ to SC_i (Figure 2.1). The superscript k indicates the chain to which reference is being made. The entries corresponding to glycine residues are omitted from the list of $\mathbf{dX}$'s since they have zero length. A dummy C^α atom is introduced at the beginning (end) of the chain if the first (last) residue is not glycine and if the chain is unblocked[36].

To simplify the notation, the $\mathbf{dC}_i^k$ and $\mathbf{dX}_i^k$ vectors will be grouped in a single vector $\mathbf{q}^k = (\mathbf{dC}_0^k, \mathbf{dC}_s^k, \mathbf{dC}_e^k, \ldots, \mathbf{dX}_1^k, \mathbf{dX}_2^k, \ldots, \mathbf{dX}_m^k)^{\mathrm{T}}$. The indices s and e correspond to the first and last real residue, *i.e.*, $s = 1$ if the first residue is Gly and $s = 2$ otherwise. Likewise, if the last residue is a dummy one, then the index $e = n - 1$, with n being the number of residues in the chain, and $e = n$ otherwise. The index m is the number of non-glycine residues in the chain. It should be noted that, although I have omitted the superscripts, the values of s, e, n, and m might in principle be different for different chains within the complex.

The coordinates $\mathbf{x}_{\mathrm{p}i^k}$ and $\mathbf{x}_{\mathrm{SC}_{i^k}}$ of the united peptide groups and side chains can be reconstructed from the $\mathbf{dC}_i^k$ and $\mathbf{dX}_i^k$ vectors through eqs 2.10 and 2.11

$$\mathbf{x}_{p_i^k} = \mathbf{dC}_0^k + \sum_{j=s}^{j=i-1} \mathbf{dC}_j^k + \frac{1}{2}\mathbf{dC}_i^k \tag{2.10}$$

$$\mathbf{x}_{SC_i}^k = \mathbf{dC}_0^k + \sum_{j=s}^{j=i-1} \mathbf{dC}_j^k + \mathbf{dX}_i^k \tag{2.11}$$

Defining vectors $\mathbf{x}^k = (\mathbf{x}_{ps}^k, \ldots, \mathbf{x}_{pe}^k, \mathbf{x}_{SC1}^k, \ldots, \mathbf{x}_{SCm}^k)^{\mathrm{T}}$, eqs 2.10 and 2.11 can be expressed in matrix form, obtaining a single equation for each chain

$$\mathbf{x}^k = \mathbf{A}^k \mathbf{q}^k \tag{2.12}$$

where $\mathbf{A}^k$ is the matrix that transforms from the generalized coordinates q^k of the kth chain to the Cartesian coordinates of the interacting sites, $\mathbf{x}^k$, of the same chain. The same relation holds for the velocities $\mathbf{v}^k = (\mathbf{v}_{ps}^k, ..., \mathbf{v}_{pe}^k, \mathbf{v}_{SC1}^k, ..., \mathbf{v}_{SCm}^k)^{\mathrm{T}}$

$$\mathbf{v}^k = \mathbf{A}^k \dot{\mathbf{q}}^k \tag{2.13}$$

Then, when writing Lagrange's equations, we obtain a relation, for each chain, of the form

$$\frac{d}{dt}\left[\nabla_{\dot{\mathbf{q}}^k} K^k(\mathbf{q}^k, \dot{\mathbf{q}}^k)\right] + \nabla_{\mathbf{q}^k} U(\mathbf{q}^1, \mathbf{q}^2, \ldots, \mathbf{q}^N) = \mathbf{f}^{fric_k} + \mathbf{f}^{rand_k} \tag{2.14}$$

where k indicates the chain in question, K^k is its kinetic energy, $\mathbf{f}^{fric_k}$ and $\mathbf{f}^{rand_k}$ are the friction and random forces acting on that chain, and N is the total number of chains in the protein.

The different chains are coupled only through the UNRES potential energy U, which also includes the free energy of the solvent implicitly in the $U_{\mathrm{SC}_i\mathrm{SC}_j}$ terms. The kinetic energy of a specific chain does not contain any dependence on the coordinates from a different chain. This enabled us to easily generalize the single-chain equations derived in refs [21] and [22] to the multichain problem.

Since both peptide groups and side chains are rigid bodies, their kinetic energies have translational and rotational contributions. The kinetic energies of a peptide group and a side chain, K_{pi^k} and K_{SCi^k}, are given by equations 2.15 and 2.16, respectively.

$$K_{pi^k} = \frac{1}{2} m_p \|\mathbf{v}_{pi^k}\|^2 + \frac{1}{2} \frac{I_{pi^k}}{\|\mathbf{dc}_i^k\|^4} (\mathbf{dc}_i^k \times \mathbf{v}_{pi^k})^2 \tag{2.15}$$

$$K_{SCi^k} = \frac{1}{2} m_{SCi} \|\mathbf{v}_{SCi^k}\|^2 + \frac{1}{2} \frac{I_{SCi^k}}{\|\mathbf{dX}_i^k\|^4} (\mathbf{dX}_i^k \times \mathbf{v}_{SCi^k})^2 \tag{2.16}$$

where I_{pi^k} is the moment of inertia of the ith peptide group and I_{SCi^k} is the moment of inertia of the ith side chain, in chain k. In eqs 2.15 and 2.16, the first term corresponds to the translational kinetic energy, while the second term corresponds to the internal, rotational, kinetic energy.

The moment of inertia of a peptide group is $I_{pi^k} = (1/12) m_p (\|\mathbf{dc}_i^k\|)^2$, and that of a side chain is $I_{SCi^k} = (1/12) m_{SCi^k} (2\|\mathbf{dX}_{i^k}\|)^2 = (1/3) m_{SCi^k} (|\mathbf{dX}_i^k\|)^2$ (since the length of a side chain is twice its virtual-bond length).

We now define a diagonal matrix $\mathbf{H}^k$, which contains the moments of inertia of petide groups and side chains divided by their bond lenght

$$\mathbf{H}^k = \begin{bmatrix} \frac{1}{12}m_p & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{12}m_p & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{3}m^k_{SC1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{3}m^k_{SC2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & m & \cdots & \frac{1}{3}m^k_{SCm} \end{bmatrix} \tag{2.17}$$

and the diagonal matrix $\mathbf{M}^k$, containing the masses of the interacting sites

$$\mathbf{M}^k = \begin{bmatrix} m_p & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & m_p & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m^k_{SC1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & m^k_{SC2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & m & \cdots & m^k_{SCm} \end{bmatrix} \tag{2.18}$$

Then, using the definitions from equations 2.12, 2.13, 2.17 and 2.18, the kinetic energy of chain k can be expressed as follows

$$K^k = \frac{1}{2}(\mathbf{A}\dot{\mathbf{q}}^k)^T\mathbf{M}^k(\mathbf{A}\dot{\mathbf{q}}^k) + \frac{1}{2}(\dot{\mathbf{q}}^k)^T\mathbf{H}(\dot{\mathbf{q}}^k) \tag{2.19}$$

From equation 2.19, it follows that Lagrange's equations (equation 2.14) can be expressed by equation 2.20

$$\mathbf{G}^k\ddot{\mathbf{q}}^k = -\nabla_{\mathbf{q}^k}U(\mathbf{q}^1, \ldots, \mathbf{q}^N) - (\mathbf{A}^k)^T\mathbf{\Gamma}^k\mathbf{A}^k\dot{\mathbf{q}}^k + (\mathbf{A}^k)^T\mathbf{f}^{rand_k} \tag{2.20}$$

where $\mathbf{G}^k$ is the inertia matrix, defined by eq 2.21

$$\mathbf{G}^k = (\mathbf{A}^k)^T(\mathbf{M}^k)(\mathbf{A}^k) + \mathbf{H}^k; \tag{2.21}$$

$\mathbf{\Gamma}^k$ is a diagonal matrix containing the friction coefficients of the interacting sites (peptide groups and side chains), and the components of the vector $\mathbf{f}^{rand_k}$ of random forces are calculated from a normal distribution according to [37, 38, 39]

$$(\mathbf{f}^{rand_k})_i = \sqrt{\frac{2\gamma_i RT}{\delta t}}\mathbf{N}(0,1) \tag{2.22}$$

where $(\mathbf{f}^{rand_k})_i$ is the random force acting on the ith interacting site from chain k, γ_i is the friction coefficient associated with that site, R is the universal gas constant, T is the temperature of the bath, δt is the integration time step, and $N(0,1)$ is a tridimensional normal distribution with zero mean and unit variance.

2.2.2 Integrating the Equations of Motion

The time evolution of the system is obtained by numerical integration of the equations of motion (equation 2.14). This is done using the velocity Verlet algorithm[40]. The algorithm[40] consist of two steps:

- step 1. Compute the coordinates at time $t + \delta t$ (where δt is the time step)

$$\mathbf{q}(t + \delta t) = \mathbf{q}(t) + \dot{\mathbf{q}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{q}}(t)(\delta t)^2 \tag{2.23}$$

- step 2. Compute the momenta at time $t + \delta t$

$$\dot{\mathbf{q}}(t + \delta t) = \dot{\mathbf{q}}(t) + \frac{1}{2}[\ddot{\mathbf{q}}(t) + \ddot{\mathbf{q}}(t + \delta t)](\delta\ t)^2 \tag{2.24}$$

- go back to step 1.

2.2.3 Simulations in the Microcanonical Ensemble

Following the procedure for single chain proteins[21], I first carried out MD calculations in the microcanonical ensemble. In this case, the stochastic and the friction forces are set to zero; therefore, the total energy of the system should be conserved. To check that the total energy condition was satisfied, I carried out simulations on two chains of an unblocked Ala10 polypeptide with the variable time step as described in section 3 of ref [21]. The simulations showed that the fluctuations in the total energy are negligible when compared with those in the kinetic and potential energies. The total energy is conserved, although only to the extent that it is conserved in ref [21]. The results of the microcanonical simulations are not shown here since that is an issue that has already been addressed in ref [21].

2.2.4 Berendsen Dynamics (BD)

There are several methods that can be used to carry out constant temperature simulations. These methods can be classified in two large groups: extended Lagrangian methods[41, 42] and rescaling of velocities[35, 43]. The method that I chose for the MD simulations belongs to the second category and is known as the Berendsen thermostat[35]. The idea behind this method is that the system is forced to have the same kinetic energy as if it were subject to the forces in eq 2.8. To accomplish this, the velocities are rescaled by a factor

$$\lambda = \left[1 + \frac{\delta t}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)\right] \tag{2.25}$$

at every simulation step, where δt is the time step, T_0 is the reference temperature, τ_T is an adjustable parameter (known as the time constant of the thermostat), and $T(t)$, the instantaneous temperature of the system at time t, is given by equation 2.26

$$T(t) = \frac{2K(t)}{RD} \tag{2.26}$$

where $K(t)$ is the kinetic energy of the system, R is the universal gas constant, and D is the number of degrees of freedom of the system. As a result, the system is globally coupled to a heat bath at temperature T_0. The integration algorithm can be summarized as follows:

- step 1. Compute the accelerations at time t according to equation 2.27

$$\ddot{\mathbf{q}}^k(t) = -[\mathbf{G}^k]^{-1}\nabla_{\mathbf{q}^k(t)}U(\mathbf{q}^1(t),\ldots,\mathbf{q}^N(t)) \tag{2.27}$$

were the supraindices k indicate the chain number.

- step 2. Compute the coordinates at time $t+\delta t$

$$\mathbf{q}^k(t+\delta t) = \mathbf{q}^k(t) + \dot{\mathbf{q}}^k(t)\delta t + \frac{1}{2}\ddot{\mathbf{q}}^k(t)[\delta t]^2 \tag{2.28}$$

- step 3. Compute the accelerations at time $t+\delta t$

$$\ddot{\mathbf{q}}^k(t+\delta t) = -[\mathbf{G}^k]^{-1}\nabla_{\mathbf{q}^k(t+\delta t)}U(\mathbf{q}^1(t+\delta t),\ldots,\mathbf{q}^N(t+\delta t)) \tag{2.29}$$

- step 4. Compute the momenta at time $t+\delta t$

$$\dot{\mathbf{q}}^k(t+\delta t) = \dot{\mathbf{q}}^k(t) + \frac{1}{2}[\ddot{\mathbf{q}}^k(t) + \ddot{\mathbf{q}}^k(t+\delta t)(\delta t)^2 + \tag{2.30}$$

- step 5. Rescale the momenta

$$\dot{\mathbf{q}}^k(t+\delta t) \rightarrow \lambda\dot{\mathbf{q}}^k(t+\delta t) \tag{2.31}$$

- go back to step 1.

Although this method has not been proven to generate a true canonical ensemble, it has the advantage that the coupling can be made as weak as desired by manipulating the constant τ_T. It has been shown[35] that small values of τ_T (strong coupling) reduce the fluctuations in the kinetic energy K at the expense of increasing fluctuations in the total energy E. On the basis of earlier work[21], I set $\tau_T = 48.9fs = 1mtu$ (mtu = molecular time unit) and $\delta t = 0.05mtu$. These values were tested by carrying out MD simulations with Berendsen dynamics on a system composed of two chains of an unblocked Ala10 polypeptide at a concentration of 1 mM. During the simulations, the fluctuations in the total (E), kinetic (K), and potential (U) energy were monitored. The simulations showed that the parameters used for the single chain were appropriate for the multichain complex as well.

2.2.5 Langevin Dynamics (LD)

As pointed out in section 2.2.1, the system can be kept at a constant temperature by inserting stochastic and friction terms in the equations of motion, yielding a Langevin equation, namely, eqs 2.8 and 2.9. The trajectory of the system is obtained by numerical integration of eq 2.20. The integration algorithm can be summarized as follows[22]:

- step 1. Compute the accelerations at time t according to equation 2.32

$$\begin{aligned} \ddot{\mathbf{q}}^k(t) &= -[\mathbf{G}^k]^{-1}\nabla_{\mathbf{q}^k(t)}U(\mathbf{q}^1(t),\ldots,\mathbf{q}^N(t)) \\ &- [\mathbf{G}^k]^{-1}(\mathbf{A}^k)^T\mathbf{\Gamma}^k\mathbf{A}^k\dot{\mathbf{q}}^k(t) \\ &+ [\mathbf{G}^k]^{-1}(\mathbf{A}^k)^T\mathbf{f}^{rand_k}(t) \end{aligned} \tag{2.32}$$

  were the supraindices k indicate the chain number.

- step 2. Compute the coordinates at time $t+\delta t$

$$\mathbf{q}^k(t+\delta t) = \mathbf{q}^k(t) + \dot{\mathbf{q}}^k(t)\delta t + \frac{1}{2}\ddot{\mathbf{q}}^k(t)[\delta t]^2 \tag{2.33}$$

- step 3. Compute the accelerations at time $t+\delta t$ according to equation 2.34

$$\begin{aligned} \ddot{\mathbf{q}}^k(t+\delta t) &= -[\mathbf{G}^k]^{-1}\nabla_{\mathbf{q}^k(t+\delta t)}U(\mathbf{q}^1(t+\delta t),\ldots,\mathbf{q}^N(t+\delta t)) \\ &- [\mathbf{G}^k]^{-1}(\mathbf{A}^k)^T\mathbf{\Gamma}^k\mathbf{A}^k\dot{\mathbf{q}}^k(t) \\ &+ [\mathbf{G}^k]^{-1}(\mathbf{A}^k)^T\mathbf{f}^{rand_k}(t) \end{aligned} \tag{2.34}$$

- step 4. Compute the momenta at time $t+\delta t$

$$\dot{\mathbf{q}}^k(t+\delta t) = \dot{\mathbf{q}}^k(t) + \frac{1}{2}[\ddot{\mathbf{q}}^k(t) + \ddot{\mathbf{q}}^k(t+\delta t)(\delta t)^2 + \tag{2.35}$$

- go back to step 1.

2.2.6 Optimal Temperature for MD Simulations

The parameters and weights in UNRES have been determined by a hierarchical optimization method[28, 30, 31, 44]. The idea behind this method is to reproduce a funnel-like energy landscape with energy decreasing as the number of nativelike elements in a structure increases[28, 44]. Because the 4P force field was designed to find nativelike structures as global minima in the potential energy surface, the free-energy gaps between the nativelike structures and the lowest-energy non-native structure of the training protein were overemphasized in the optimization process[32]. Consequently, the optimal folding temperature for the MD simulations with the UNRES 4P force field turned out to be 800 K[23]. This value gave the best compromise between folding time and stability of the nativelike structures for several benchmark proteins[23]. This high temperature was not a problem while carrying out single-chain simulations because the internal forces acting on a polypeptide chain were tuned to this high temperature. However, in multichain simulations, the chains move with respect to each other, and the external motions are too strong to allow association. Therefore, I rescaled all energy term weights by a factor of 3/8 to reduce the folding temperature to 300 K. This operation changes only the energy scale but not the structure of the energy landscape.

2.3 Choosing the Appropriate System to Test UNRES/MD Multichain

Because the goal was to test the multichain MD implementation, and not the force field, I chose systems that the 4P force field could treat. For this porpuse, I carried out simulations on the following three α-helical proteins of known native structure: 1G6U (two chains, 48 residues each), 2ZTA (two chains, 33 residues each), and 1C94 (four chains, 38 residues each). The complexity and size of these proteins is similar to that of the α-helical proteins tested with the single-chain UNRES/MD[23], and the size of the smallest of them (2ZTA) is within the average size of structural segments of α-helical proteins that can be predicted successfully with the 4P force field[32]. These systems are, therefore, appropriate to test the UNRES/MD approach for multichain proteins, given the limitations of the present force field. I did not use β or $\alpha + \beta$ proteins because earlier work[23] showed that UNRES/MD generally produces non-native α-helical structures for such proteins, even though the native structures are global energy minima in the UNRES energy surface; this happens

because the conformational entropy was neglected in the parametrization of the 4P force-field. This issue has been addressed in newer versions of the UNRES force filed[45].

2.4 Testing UNRES/MD Multichain

To study different aspects of the UNRES/MD multiple-chain implementation, I carried out a number of tests. I compared Langevin (LD) and Berendsen dynamics (BD) by carrying out multiple-chain simulations with the same initial conditions with each method. To test whether the presence of other chains was a necessary condition to fold the monomers, I also carried out single-chain simulations with BD and LD and compared the structures obtained with those of the monomers in the crystal structures of the oligomers. Finally, since the method failed to predict the native structure of 1C94, additional simulations starting from the PDB structure were carried out for this protein. This was done to check whether the native structure was not found because of insufficient simulation time or because the force field was not good enough to properly represent the energy landscape of this protein.

All the runs (both single-chain and multichain), except those starting from the PDB structure, were started with the chains in an extended conformation. In all simulations, the initial velocities of the peptide groups and side chains were randomly generated. In the multichain runs, the chains were placed parallel to each other, separated by a distance large enough (20 Å for GCN4- p1 and 1C94 and 40 Å for 1G6U) to allow them to rearrange independently. Since the chains rapidly adjust to an equilibrium ensemble, after starting from extended conformations, the simulations are practically independent of the starting condition. The initial velocities were selected from a Gaussian distribution corresponding to the average kinetic energy at the simulation temperature, and the temperature was held constant at 300 K during all of the simulations.

In the multichain runs, for those starting from the extended conformation, the radius of the confining sphere was initially set large enough to fit the extended chains. After the first 24 ps of simulation, the radius of the sphere was decreased slowly until the desired concentration (1 mM for the dimers and 10 mM for the tetramer) was reached. This concentration, although higher than those concentrations used in the experiments[46, 47, 48] was chosen because it resulted in a volume large enough to fit

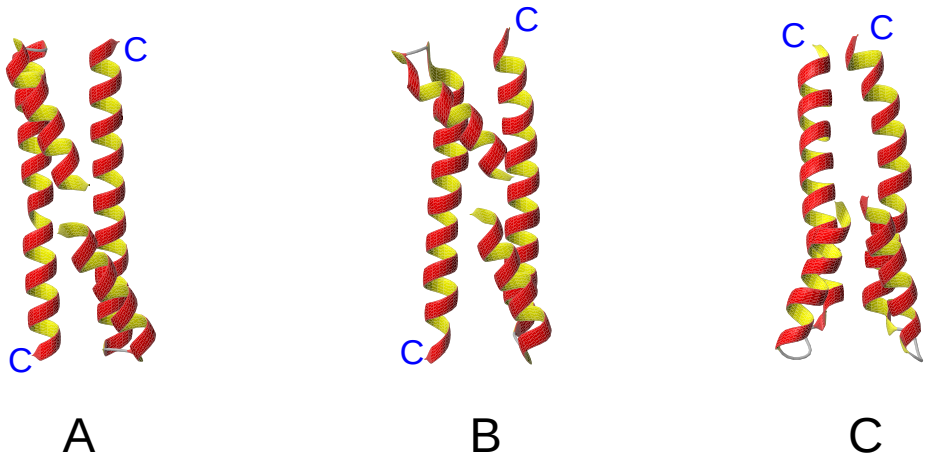

FIGURE 2.2. (A)Experimental structure of 1G6U. (B) The most nativelike structure (C^{α} RMSD = 1.79 Å) obtained with BD UNRES/MD . (C) An example of a misfolded structure. The C-terminus of each chain is marked.

the chains without altering their structures and small enough for the monomers to find each other and interact in a short period of time.

To classify the runs into success and failure, we monitored the C^{α} root-mean square devation (RMSD) between the computed structures and the crystal structure. If this value, hereafter referred to as ρ, fell below a cutoff value, ρ_{cut}, then the protein was considered to have folded. The folding time τ_f, defined as the time at which ρ fell below the cutoff ρ_{cut} for the first time, and the residence time τ_{res}, defined as the fraction of the total time that ρ was below ρ_{cut}, were also computed. For 1G6U, ρ_{cut} was 5 Å for the monomers and 7 Å for the dimers, for GCN4, ρ_{cut} was 3.4 Å for the monomers and 4.8 Å for the dimers, and for 1C94, ρ_{cut} was 4 Å for the monomers, 5.6 Å for the dimers, and 8 Å for the tetramers. If the monomers were folded by this criterion and were stable, and the arrangement of the chains was stable but not native, then the overall structure was classified as misfolded. If this criterion was not met, then the structure was classified as nonfolding.

2.4.1 Domain Swapped Dimer (PDB Code 1G6U)

1G6U is a synthetic α-helical homodimer with 48 residues per chain[48]. Each monomer consists of two α-helix segments, with the shortest (14 residues) helix packed against the longest (28 residues) helix. The monomers assemble forming a three-α-helix bundle with the long helices in the antiparallel position (Figure 2.2A). I will refer to the shortest helix as H1 and the longest helix as H2 (Figure 2.3). To provide a better description of the folding trajectories, I monitored the RMSD (Table 2.1) with respect to the native structure for the entire protein, for each of the monomers, and for each

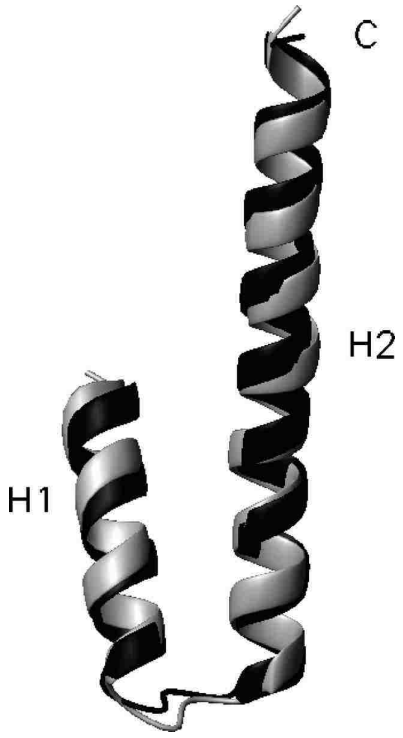

FIGURE 2.3. Superposition of one of the monomers in the 1G6U experimental dimer structure (black) on the most nativelike structure (gray) (C^{α} RMSD = 1.22 Å) obtained with the UNRES/MD simulations of the monomer using BD. The N-terminal helix H1 and the C-terminal helix H2 are indicated as well as the C-terminus.

of the helices (H1 and H2). To determine the folding times of H1 and H2, I set their cutoff RMSD's at 1.5 and 4 Å respectively.

2.4.1.1 Monomers

As can be seen in Table 2.1, all of the simulations of the monomers converged to nativelike structures, showing that dimerization is not necessary for the folding and stabilization of the individual chains. The most nativelike structure, 1.22 Å from native, was produced by BD. A superposition of this structure and the native structure is shown in Figure 2.3.

Figure 2.4 shows potential energy and ρ values for an LD trajectory (panels A and B, respectively) and a BD trajectory (panels C and D, respectively) for an isolated monomer of 1G6U. As can be seen from Figure 2.4, the native basin was very stable, and with both methods, once the peptide adopted nativelike structures, the fluctuations in the potential energy and ρ became smaller, and the peptide remained in the native basin.

TABLE 2.1. Summary of Trajectories for **1G6U**

	N_f[a]	$\langle\tau_f\rangle$[b] (ns)	$\langle\tau_f(H1)\rangle$[c] (ns)	$\langle\tau_f(H2)\rangle$[d] (ns)	ρ_{min}[e] (Å)	$\langle\tau_{res}\rangle$[f]	$\langle E\rangle_f$[g] (kcal/mol)	N_{mf}[h]	$\langle E\rangle_{mf}$[i] (kcal/mol)	CPU[j] (h)
					Dimer					
Berendsen	9(20)	4.8(0.30)	0.14	0.16	1.79	49%	-402	1	-401	2.9
Langevin	2(19)	14.9(4.0)	0.35	0.20	2.38	36%	-403	2	-398	3.9
					Monomer					
Berendsen	10	0.92	0.18	0.21	1.22	86%	-186			
Langevin	10	2.6	0.25	0.24	1.28	69%	-188			

[a]Number of trajectories (out of 10) that folded to nativelike structures. In the dimer simulations, the number of monomers (out of 20, since there were 2 monomers on each of the 10 dimer simulations) that folded to a nativelike structure is indicated between parentheses;
[b]Average folding time. The folding time was defined as the time at which the RMSD with respect to the crystal structure fell below the cutoff value (7 Å for the dimers and 5 Å for the monomers). In those runs for which the RMSD never went below the cutoff, the folding time was considered to be the simulation time (12 ns for Berendsen and 16 ns for Langevin). In the dimer simulations, the average folding time of the monomers is indicated between parentheses;
[c]Average folding time for the N-terminal helix, H1. The folding time was defined as the time at which the RMSD with respect to the crystal structure fell below 1.5 Å;
[d]Average folding time for the C-terminal helix, H2. The folding time was defined as the time at which the RMSD with respect to the crystal structure fell below 4 Å;
[e]The lowest RMSD in all of the fluctuating trajectories;
[f]Fraction of the time that the peptide spent in the native basin averaged over all of the folding trajectories;
[g]Average potential energy over all structures in the native f basin;
[h] Number of trajectories (out of 10) that yielded misfolded structures;
[i]Average potential energy over all structures in the misfolded mf basin;
[j] Average CPU time (in hours) per 1 ns of simulation on a single 3.06 GHz Intel Pentium IV Xeon processor.

2.4.1.2 Dimers

In the simulation of dimers, the initial separation distance between chains was 40 Å, the initial arrangement was parallel, and the simulation time was approximately 12 ns for BD and 16 ns for LD. The final concentration of 1 mM was achieved within the first nanosecond. The results are summarized in Table 2.1. Both algorithms, BD and LD, folded the protein. In general the folding times with BD were shorter than those with LD, as observed in earlier work on single-chain proteins[22]. BD also produced the most nativelike structure, which is shown in Figure 2.2B. From the simulations, it became evident that the energy landscape generated by the 4P UNRES force field has two basins with low free energy. One of these basins corresponds to the native structure, and the other one to a structure that differs from the native in that the long helices are parallel to each other instead of antiparallel (Figure 2.2C). I will refer to the latter structure as a misfolded one. Both structures were very stable, and once the protein fell into one of these basins, it would not escape within the simulation time (12 ns for BD and 16 ns for LD). The difference in average potential energy between the native and the misfolded basin is very small (Table 2.1). Thus, it is natural to expect that, for some trajectories, the forces will drive the system to the native basin and, for some others, to the misfolded basin. Indeed, this is what was observed in these simulations. Presumably, improvement of the 4P UNRES force field will stabilize the native basin to a greater extent compared to the non-native basin.

Snapshots of a successful trajectory obtained with LD are shown in Figure 2.5. For the same trajectory, the values of ρ and the potential energy as a function of time are shown in Figure 2.6. The snapshots show that helix formation takes less than 1 ns, and for this particular example, the packing of the helices on both monomers takes about 3 ns. Also for this example, the monomers fold independently, but they are close enough so that, after the subunits have folded, they can overcome the friction forces to turn around (since the initial orientation of the helices is parallel, but in the native structure the orientation is antiparallel) and assemble in less than 2 ns. The folding of the dimer is completed in a total time of 5 ns. The two LD trajectories that converged to the native basin (Table 2.1) showed the folding mechanism illustrated in Figure 2.5.

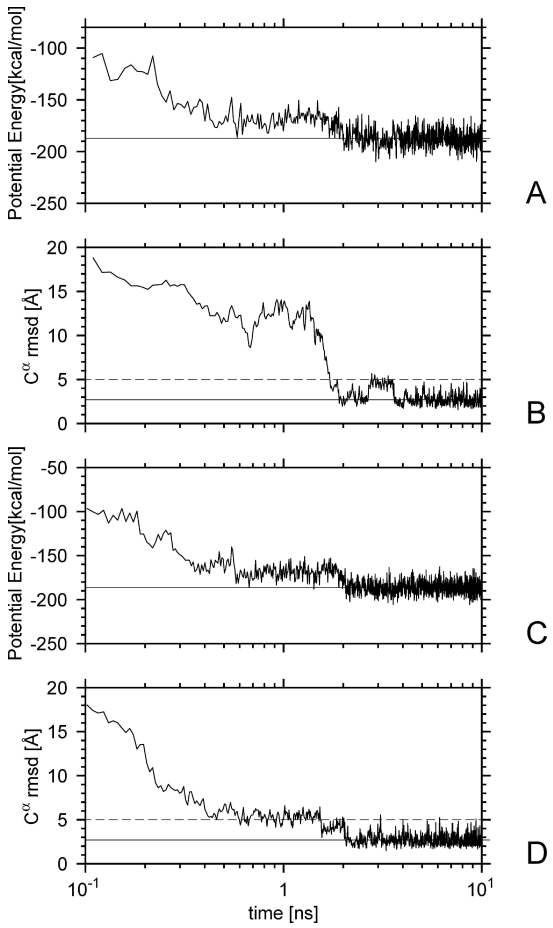

FIGURE 2.4. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure of the monomer in the dimer during the folding of an isolated monomer of 1G6U obtained with Langevin dynamics. The solid horizontal line at -187.1 kcal/mol in panel A is the mean value of the energy after the monomer has reached the native basin. In panel B, the dashed horizontal line at 5 Å corresponds to the cutoff RMSD above which the monomer structure is considered to have left the native basin, and the solid horizontal line at 2.7 Å indicates the mean C^{α} RMSD of the monomer inside the native basin. Panels C and D contain the same information as panels A and B, respectively, for a trajectory obtained with Berendsen dynamics. The solid horizontal line at -187.1 kcal/mol in panel C is the mean value of the energy after the monomer has reached the native basin, and the solid horizontal line at 2.7 Å in panel D is the mean C^{α} RMSD inside the native basin of the monomer from the monomer in the native structure of the dimer.

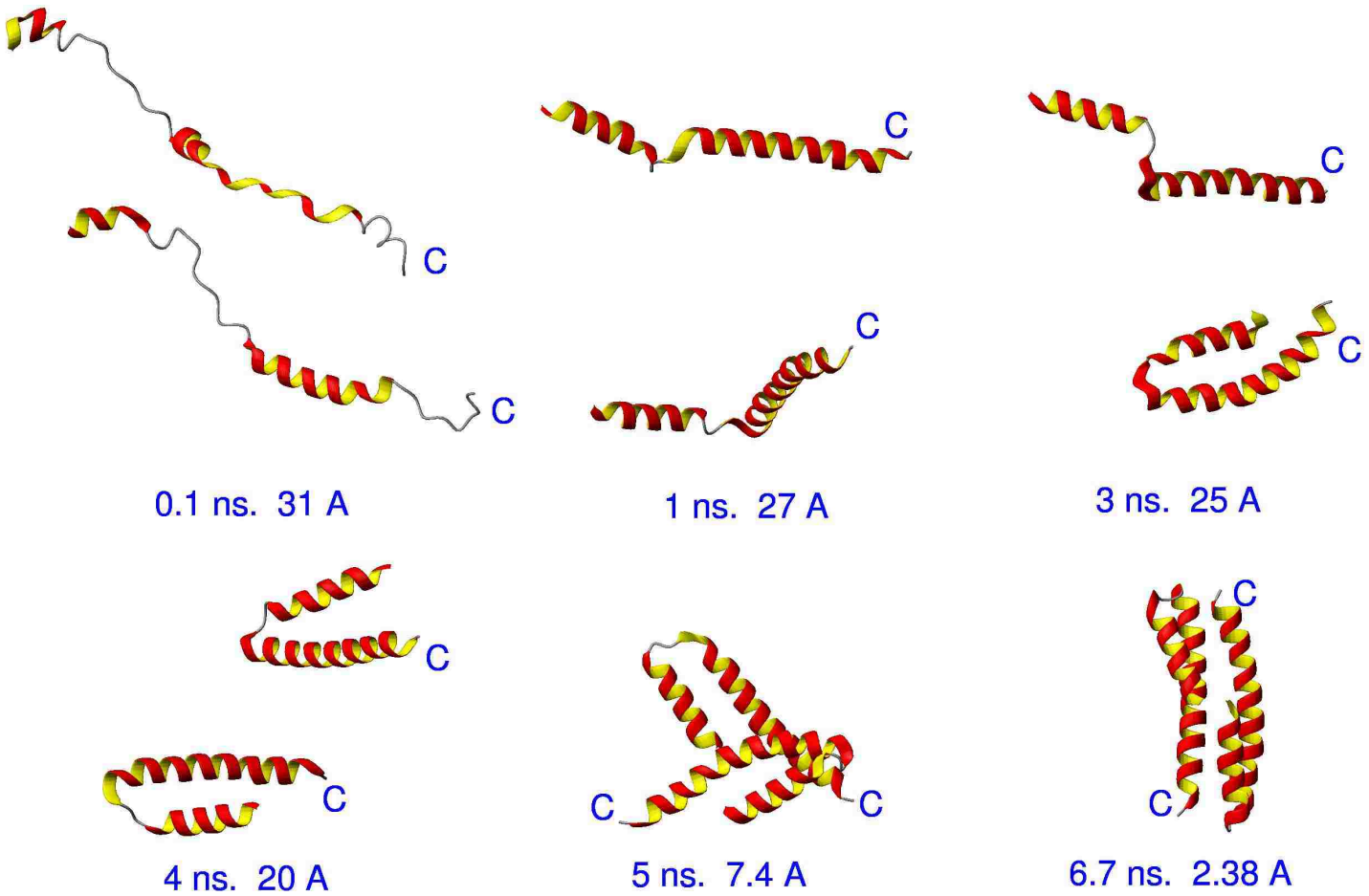

FIGURE 2.5. Example of a successful trajectory of 1G6U obtained with Langevin dynamics. The C-terminus of each chain is marked.

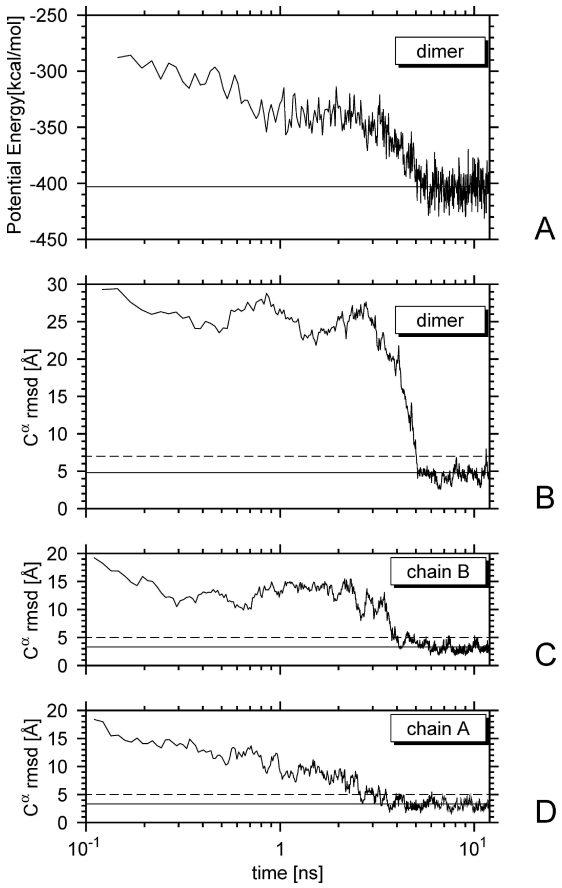

FIGURE 2.6. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for the dimer in a successful trajectory of 1G6U obtained with Langevin dynamics. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native for each of the monomers. The solid horizontal line at -403 kcal/mol in panel A is the mean value of the energy after the dimer has reached the native basin, and the solid line at 4.8 Å in panel B is the mean C^{α} RMSD inside the native basin of the dimer. The dashed horizontal line in panels B, C, and D corresponds to the cutoff RMSD (7 Å for the dimer and 5 Å for the monomers) above which a structure is considered to have left the native basin. The solid horizontal line at 3.3 Å in panels C and D is the mean C^{α} RMSD inside the native basin of the monomer.

Figure 2.7 shows snapshots of an LD trajectory leading to a misfolded structure. The values of ρ and the potential energy for this trajectory are shown in Figure 2.8. For this particular trajectory, chain A folds first (cf. panels C and D), and chain B folds while it binds to form the dimer (cf. panels B and C). The formation of the dimer in Figure 2.7 corresponds to the stabilization of ρ around 15.6 Å in panel B of Figure 2.8. For the other LD trajectory that converged to the misfolded basin, the assembly mechanism was similar to that described in Figure 2.5, in the sense that the monomers folded completely before they assembled. Thus, folding of the monomers followed by their assembly does not always lead to the native basin.

Those LD trajectories that did not converge to the native or misfolded basin reached a state (called nonfolded) in which either one or both monomers were folded, but they had not yet assembled within the 16 ns simulation time. Their structures were similar to either the 3- or the 4-ns snapshot in Figure 2.5. With BD, all of the simulations converged to either the native or the misfolded basin (Table 2.1).

Among those runs that converged to the native basin, two different pathways were observed, one on which the subunits fold before their assembly (lock-and-key mechanism) and another one on which the subunits fold simultaneously with their assembly (induced-fit mechanism). Although only a few runs followed the latter assembly mechanism (3 out of 9 folding trajectories), this pathway seems to be 2.5 times faster on average than the assembly of already folded subunits, which is not surprising since, after the monomers are folded, they might collide several times until they find the right orientation, which will in general slow down the process. Figures 2.9 and 2.10 illustrate these two folding pathways. For the trajectory shown in Figure 2.9 (fast folding pathway), folding and association of the chains occurs simultaneously, with the dimer folding in less than 0.4 ns, while for the trajectory shown in Figure 2.10 (slow folding pathway), although the chains collide several times (snapshots at 0.23, 0.33, 5.83, and 6.44 ns), only the last collision results in the formation of the dimer. There is a long period between the snapshots at 0.56 and 5.83 ns (this period is not shown in the snapshots) during which the chains remain folded but they do not collide at all. Figure 2.11 contains the values of ρ and the potential energy as a function of time, corresponding to the trajectory shown in the snapshots in Figure 2.11. In Figure 2.11, two pronounced drops in energy can be seen (panel A). The first one corresponds to the folding of the monomers (ρ below 5 Å in

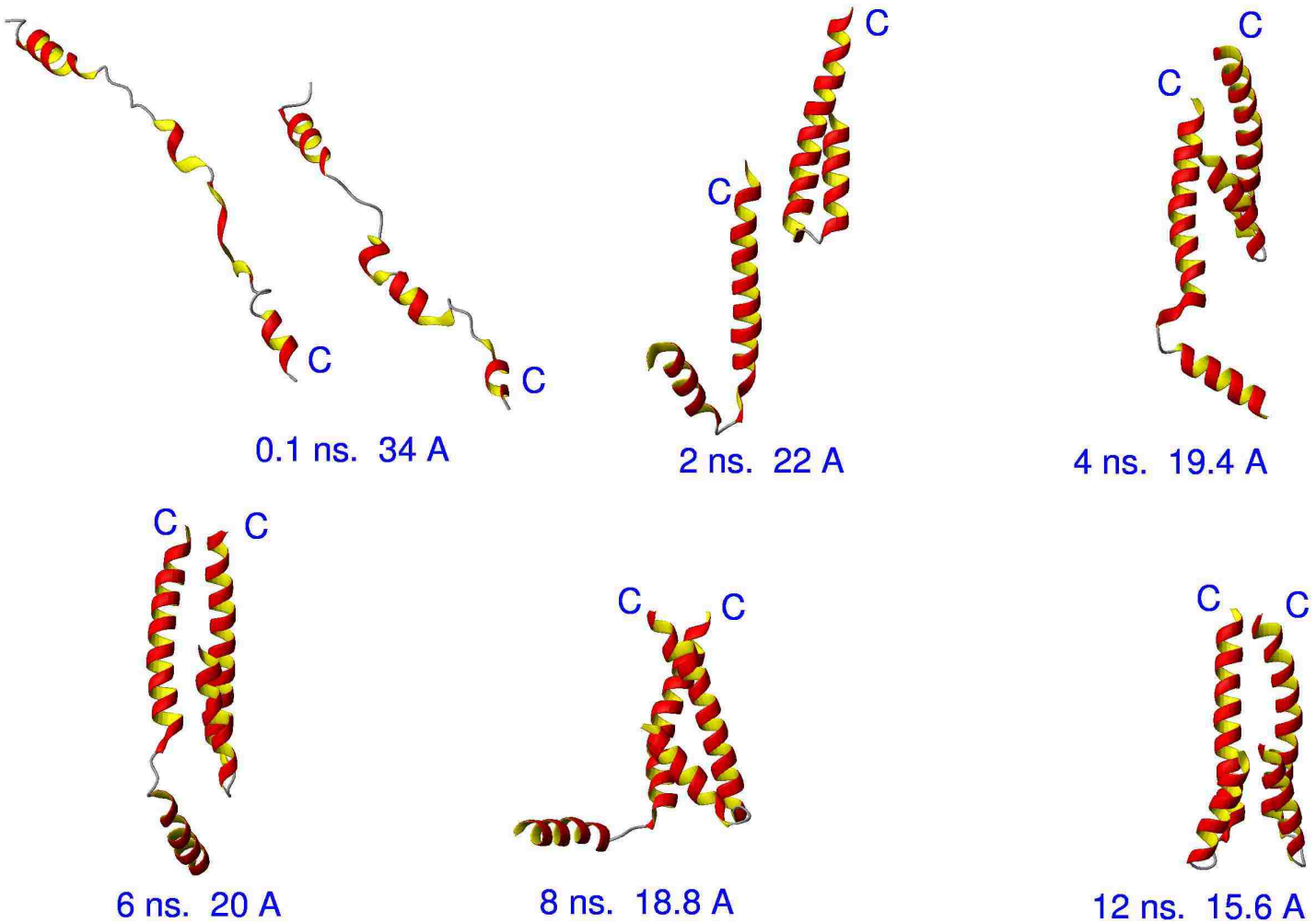

FIGURE 2.7. Example of a trajectory of 1G6U, obtained with Langevin dynamics, leading to the misfolded structure. The C-terminus of each chain is marked.

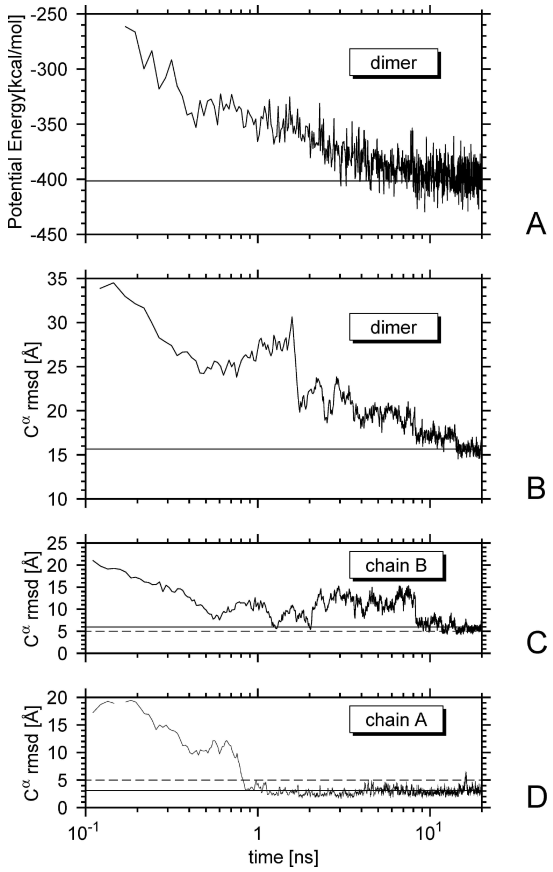

FIGURE 2.8. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for the dimer for a misfolding trajectory of 1G6U obtained with Langevin dynamics. The misfolded structure differs from the native in that the long helices are parallel to each other instead of antiparallel. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native for each of the monomers. In panels A and B, the solid horizontal line (at -401 kcal/mol in panel A and 15.7 Å in panel B) is the mean value of the energy and the C^{α} RMSD from the native, respectively, after the protein has fallen into the misfolded basin. The dashed horizontal line in panels C and D corresponds to the 5 Å cutoff RMSD, above which the monomers are considered to have left the native basin; i.e., the monomers folded but the overall structure was misfolded. The solid horizontal line in panels C and D (at 5.9 Å in panel C and 3.1 Å in panel D) is the mean C^{α} RMSD inside the native basin of the monomer.

panels C and D), and the second one corresponds to the assembly of the dimer (ρ below 7 Å in panel B).

The folding mechanism of the only BD trajectory that converged to the misfolded basin was similar to the one described in Figure 2.9 (fast folding pathway), except that the orientation of the chains was parallel instead of antiparallel as in the native structure.

When comparing the folding of the isolated monomers of 1G6U in the single- and multichain simulations, I found that, with LD, the average folding time of the monomers in the single-chain simulations was shorter than that in the multichain simulations (Table 2.1), which suggests that the interactions between chains might slow down the folding of the individual chains. To further elucidate whether this delay occurs in the formation of helices H1 and H2 or in their packing, I compare the folding times of H1 and H2 in the single-chain simulations with their folding times in the multichain simulations. I found almost no difference in the average folding time of H2, and in the case of H1, the formation of the helix seems to be slightly faster for the single-chain simulations (Table 2.1). This suggests that, for 1G6U with LD, the interactions between the chains can hinder the packing of helices H1 and H2 and can also slow down the formation of the shortest helix (H1).

With BD, on average, the monomers folded 3 times faster in the multichain simulations than in the single-chain simulations (0.30 ns compared to 0.92 ns) (Table 2.1). Further analysis of the folding times of helices H1 and H2 showed that H1 and H2 fold at approximately the same rate for single-chain and multichain simulations (Table 2.1). This indicates that interactions between chains enhance the packing of H1 and H2 but have no substantial effect on the formation of the helical structures.

The fact that the packing of H1 and H2 is favored by multichain interactions with BD and hindered with LD might be explained as follows: With BD, in which the friction forces are absent, the chains can move very fast, and if a collision that does not favor the packing of H1 and H2 has taken place, then the chains can quickly rearrange to find a better orientation while, with LD, the reorientation of the chains is much slower due to the friction forces from the solvent. With both methods, collisions will sometimes favor the packing of H1 and H2 and other times hamper it, the only difference is that, with BD, the chains can collide more frequently, and overall (when averaged over several trajectories) the presence of another chain will favor single-chain folding.

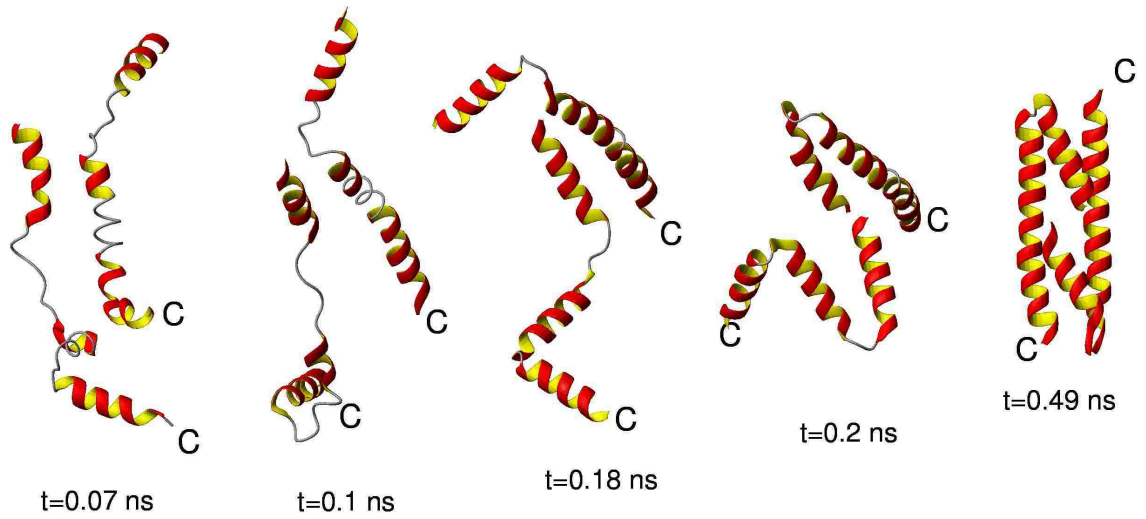

FIGURE 2.9. Example of a fast folding trajectory of 1G6U obtained with Berendsen dynamics. The C-terminus of each chain is marked.

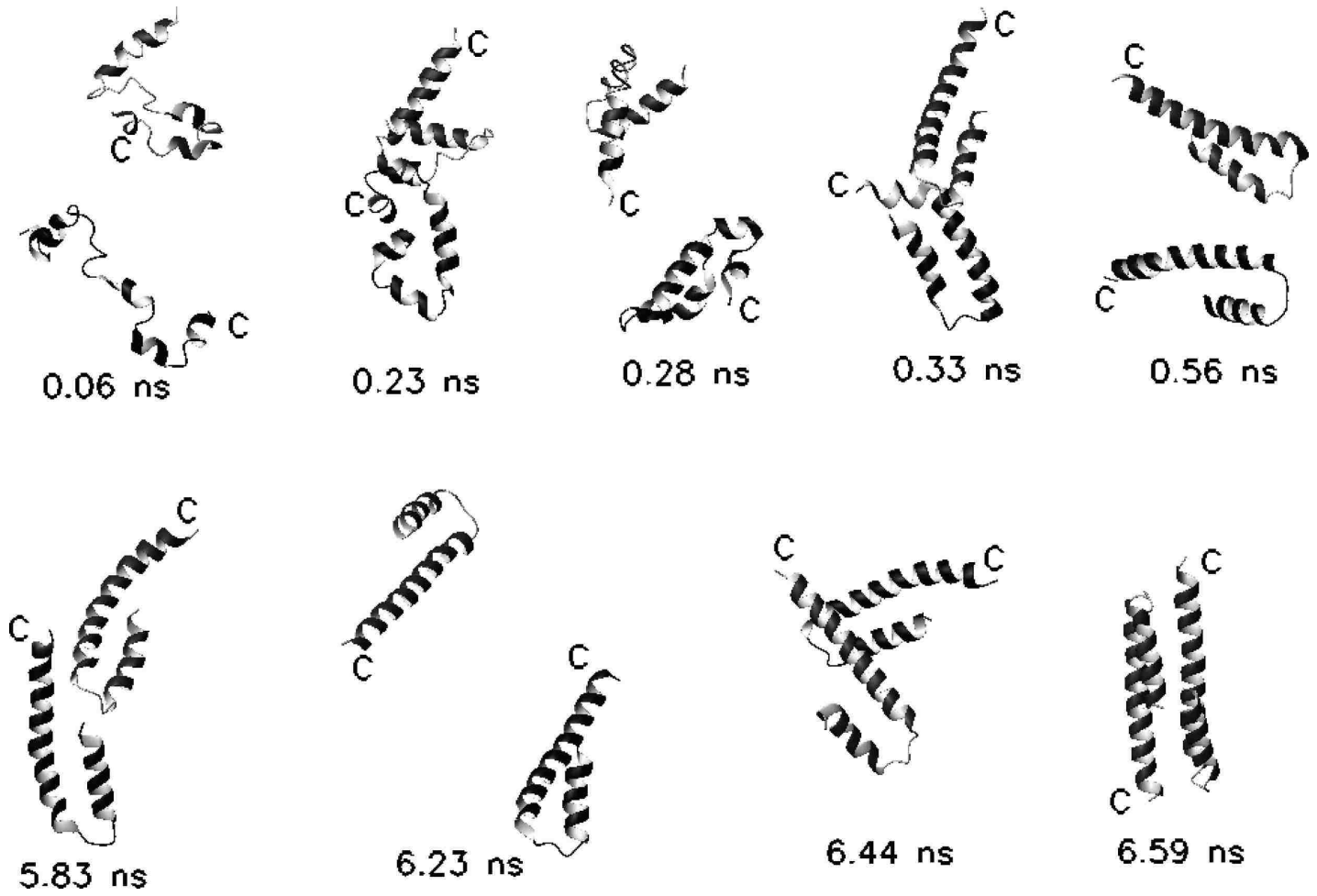

FIGURE 2.10. Example of a slow folding trajectory of 1G6U obtained with Berendsen dynamics. T he C-terminus of each chain is marked.

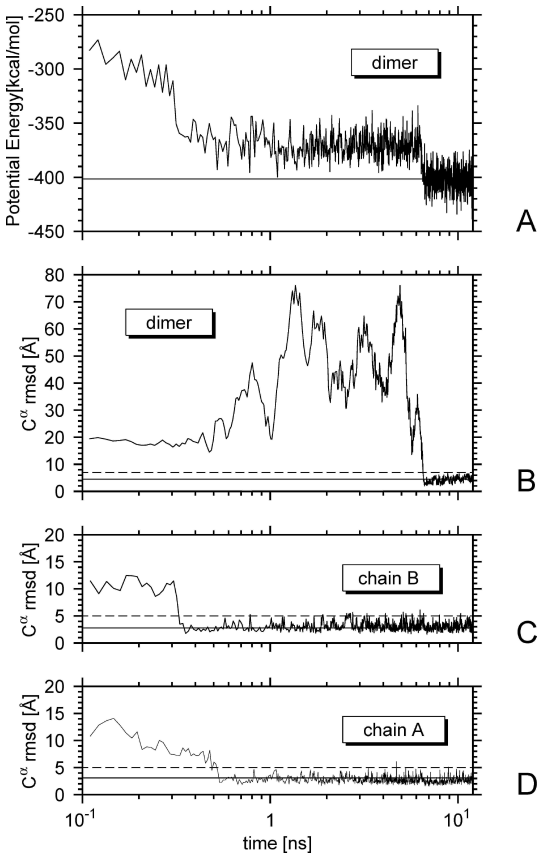

FIGURE 2.11. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for the dimer in a successful trajectory of 1G6U obtained with Berendsen dynamics. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native structure for each of the monomers. The solid horizontal line at -401 kcal/mol in panel A is the mean value of the energy after the protein has reached the native basin. The dashed horizontal line in panels B, C, and D corresponds to the cutoff RMSD (7 Å for the dimer and 5 Å for the monomers) above which a structure is considered to have left the native basin. The solid horizontal line at 4.5 Å in panel B is the mean C^{α} RMSD inside the native basin of the dimer. The solid horizontal line in panels C (2.8 Å) and D (3.1 Å) is the mean C^{α} RMSD inside the native basin of the monomer.

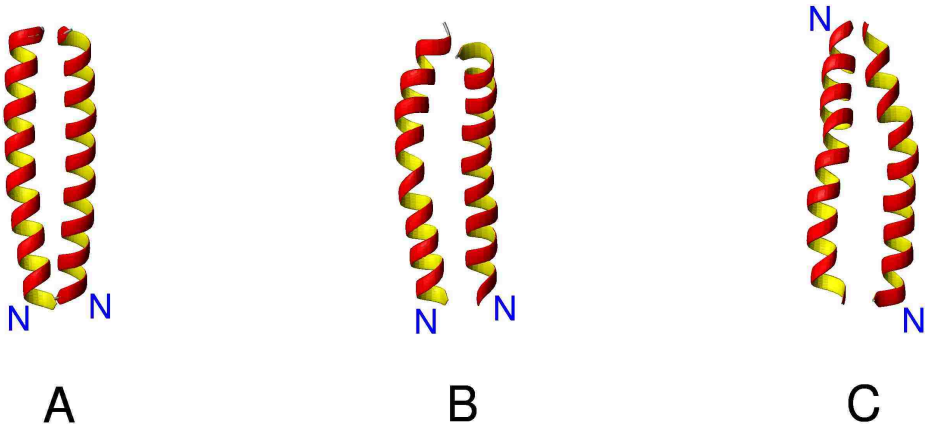

FIGURE 2.12. (A) Experimental structure of GCN4-p1. (B) The most nativelike structure (C^{α} RMSD = 1.19 Å) obtained with LD UNRES/MD. (C) An example of a misfolded structure. The N-terminus of each chain is indicated.

2.4.2 GNC4 Leucine Zipper (PDB Code 2ZTA)

The GCN4 leucine zipper (GCN4-p1), derived from the yeast transcriptional activator GCN4, is an α-helical homodimer consisting of two parallel chains with 33 residues per chain[46] (Figure 2.12A). Since the helices in GCN4-p1 wrap around each other, its motif is known as a coiled coil. The coiled coil motif is found in many proteins, and for this reason, GCN4-p1 and its mutants have been the subject of numerous studies[46, 49, 15]. In particular, simulations of the folding pathway of GCN4-p1 have been carried out by Vieth et al.[15], as mentioned in section 1.3.2.

2.4.2.1 Monomers

With both BD and LD methods, 9 out of 10 monomer trajectories converged to nativelike structures, as can be seen from Table 2.2. Moreover, these nativelike structures were quite stable, indicating that dimerization is not necessary for the folding and stabilization of the individual chains. A superposition of the most nativelike structure, obtained with BD, and the experimental structure is shown in Figure 2.13A. Those BD and LD trajectories that did not find the native basin by the end of the simulation showed structures with ρ values around 11 Å in which the helix was bent, packing against itself, as shown in Figure 2.13B.

The structure shown in Figure 2.13B was also found along the pathway of some of the trajectories that converged to nativelike structures. Potential energy and ρ values as a function of time, for an LD trajectory showing such a behavior, are shown in Figure 2.14 (potential energy in panel A and ρ values in panel B). During the first 3 ns of simulation of this trajectory, the peptide adopts structures similar to those shown in Figure 2.13B, which corresponds to the plateau in ρ values

TABLE 2.2. Summary of Trajectories for **GCN4-p1**

	dimer								monomer				
	$N_f^{\mathbf{a}}$	$\langle\tau_f\rangle^{\mathbf{b}}$ (ns)	$\rho_{min}^{\mathbf{c}}$ (Å)	$\langle\tau_{res}\rangle^{\mathbf{d}}$	$\langle E\rangle_f^{\mathbf{e}}$ (kcal/mol)	$N_{mf}^{\mathbf{f}}$	$\langle E\rangle_{mf}^{\mathbf{g}}$ (kcal/mol)	$CPU^{\mathbf{h}}$ (h)	$N_f^{\mathbf{a}}$	$\langle\tau_f\rangle^{\mathbf{b}}$ (ns)	$\rho_{min}^{\mathbf{c}}$ (Å)	$\langle\tau_{res}\rangle^{\mathbf{d}}$	$\langle E\rangle_f^{\mathbf{e}}$ (kcal/mol)
Berendsen	4(17)	6.6(3.2)	1.22	29%	-214	6	-217	1.5	9	1.5	0.59	69%	-104
Langevin	3(16)	9.1(3.4)	1.19	81%	-218	1	-225	1.9	9	2.7	0.70	74%	-97

[a]Number of trajectories (out of 10) that folded to nativelike structures. In the dimer simulations, the number of monomers (out of 20, since there were 2 monomers on each of the 10 dimer simulations) that folded to a nativelike structure is indicated between parentheses;
[b]Average folding time. The folding time was defined as the time at which the RMSD with respect to the crystal structure fell below the cutoff value (4.8 Å for the dimers and 3.4 Å for the monomers). In those runs for which the RMSD never went below the cutoff, the folding time was considered to be the simulation time (12 ns). For both BD and LD, in the dimer simulations, the average folding time of the monomers is indicated between parentheses;
[c]The lowest RMSD in all of the fluctuating trajectories;
[d]Fraction of the time that the peptide spent in the native basin averaged over all of the folding trajectories;
[e]Average potential energy over all structures in the native f basin;
[f] Number of trajectories (out of 10) that yielded misfolded structures;
[g]Average potential energy over all structures in the misfolded mf basin;
[h] Average CPU time (in hours) per 1 ns of simulation on a single 3.06 GHz Intel Pentium IV Xeon processor.

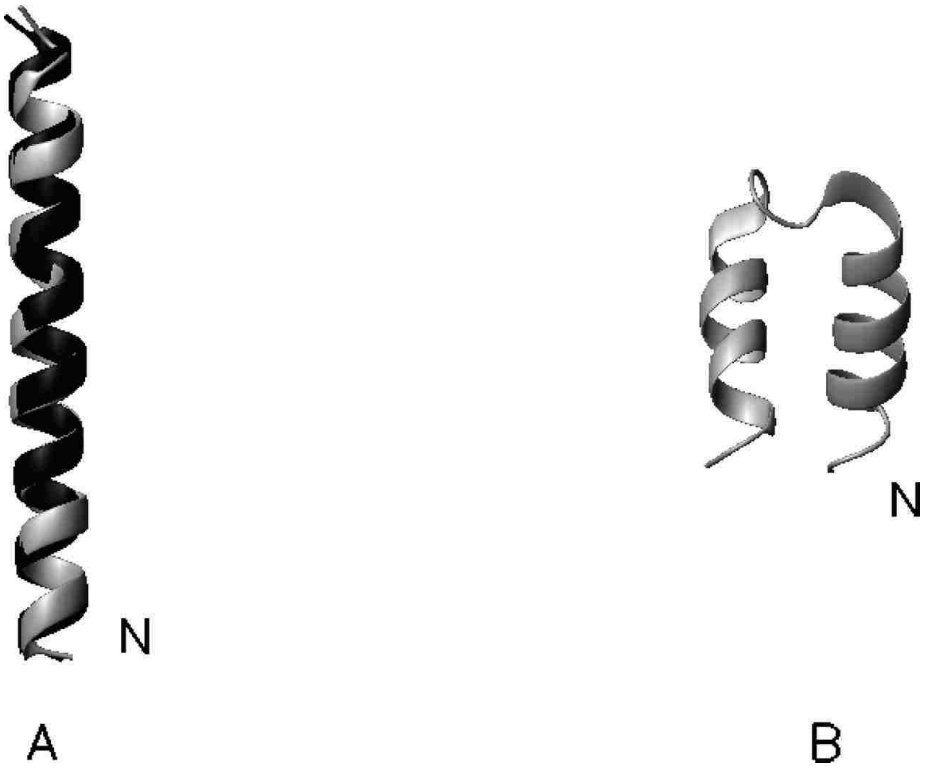

FIGURE 2.13. (A) Superposition of one of the monomers from the experimental structure of GCN4-p1 (black) on the most nativelike structure (gray) (C^{α} RMSD = 0.59 Å) obtained with BD UNRES/MD. (B) A structure that was often found during the folding pathway of GCN4-p1 (with both BD and LD) and was the final structure of those trajectories that did not f ind the native basin. The N-terminus is indicated.

around 11 Å in panel B. At the third nanosecond of simulation, the monomer finds the native basin (ρ falls below the 3.4 Å cutoff in panel B), and the energy drops considerably (panel A), showing that the structure in Figure 2.13B is only a local minimum and does not compete with the native structure.

Not all the trajectories that converged to the native basin exhibited the folding pathway described in the previous paragraph. In other simulations, a fast folding pathway was observed, with the monomer rapidly finding the native basin without spending time in any intermediate structure. An example of such behavior can be seen in the BD trajectory shown in panels C and D of Figure 2.14 (potential energy in panel C and ρ values in panel D). This behavior was the most commonly observed among all the runs (both BD and LD).

In general, with either BD or LD, the native basin was very stable, which can be inferred from the behavior of ρ in panels B and D of Figure 2.14; once ρ crossed the 3.4 Å RMSD cutoff (equivalent to finding the native basin), it remained within this cutoff most of the time.

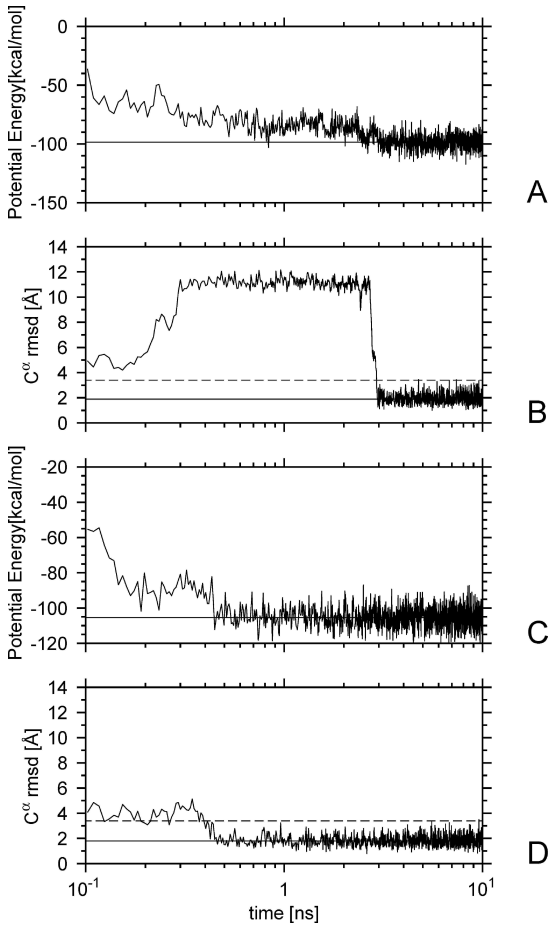

FIGURE 2.14. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure of the monomer in the dimer during the folding of an isolated monomer of GCN4-p1 obtained with Langevin dynamics. The solid horizontal line at -98 kcal/mol in panel A is the mean value of the energy after the monomer has reached the native basin. In panel B, the dashed horizontal line at 3.4 Å corresponds to the cutoff RMSD above which the monomer structure is considered to have left the native basin, and the solid horizontal line at 1.9 Å is the mean C^{α} RMSD of the monomer inside the native basin. Panels C and D contain the same information as panels A and B, respectively, for a trajectory obtained with Berendsen dynamics. The solid horizontal line at -105 kcal/mol in panel C is the mean value of the energy after the monomer has reached the native basin, and the solid horizontal line at 1.8 Å in panel D is the mean C^{α} RMSD inside the native basin of the monomer from the monomer in the native structure of the dimer.

2.4.2.2 Dimers

The initial separation distance between chains was 26 Å, and the initial arrangement was parallel. Both methods, BD and LD, generated trajectories leading to nativelike structures within 12 ns of simulation. The results are summarized in Table 2.2. The equilibrium concentration of 1 mM was reached during the first 24 ps of simulation. Again, as for 1G6U, two families of stable structures (corresponding to basins with low free energy) were found; one of them was nativelike, and the other one differed from the native structure in that the orientation of the helices was antiparallel instead of parallel. The most nativelike structure generated by UNRES/MD as well as an example of a misfolded structure are shown in Figures 2.12B and 2.12C, respectively.

When running in the LD mode, two different pathways were observed, one on which folding and assembly of subunits were coupled, induced-fit mechanism, and another one in which the subunits folded before they assemble, lock-and-key mechanism. Of the three LD trajectories that converged to the native basin, two of them folded by the induced fit mechanism and the remaining one by the lock-and-key mechanism. Snapshots from one of the runs that folded by the induced fit mechanism are shown in Figure 2.15, and the potential energy and ρ values for the same trajectory are shown in Figure 2.16. In Figure 2.15, dimerization starts with the association of the small helical segments at the N-termini and propagates toward the C-termini simultaneously with formation of the helices. The two trajectories folding by this mechanism folded in less than 0.3 ns, which was 10 times faster than the trajectory folding by the lock-and-key mechanism.

Snapshots from the trajectory folding by the lock-and-key mechanism are shown in Figure 2.17, and the corresponding potential energy and ρ values as a function of time are shown in Figure 2.18. In Figure 2.17, the folding of the helices is almost completed at the 0.20 ns snapshot, but the chains fail to bind and move apart. It takes almost 5 ns more for the chains to find the right orientation and form the dimer. This folding mechanism will in general lead to a larger folding time since, once the individual chains adopt their native structure, moving through the solvent to find the proper packing is difficult, while if the subunits are already attached (in the right place) the rate of folding is limited only by the folding of the individual chains.

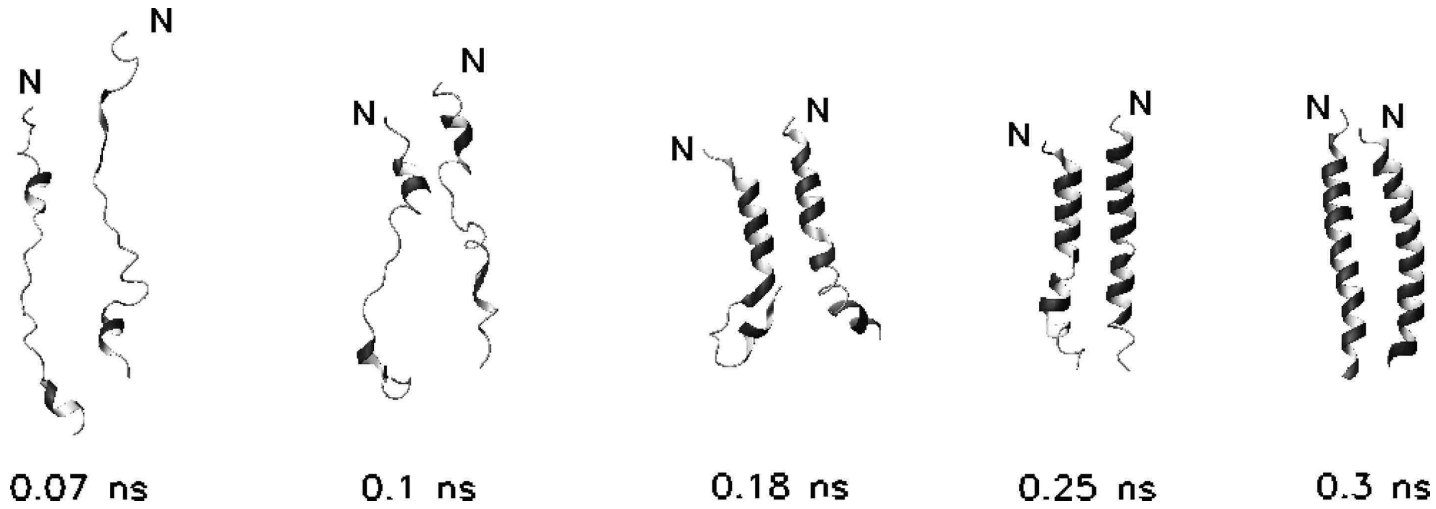

FIGURE 2.15. Example of a fast folding trajectory of GCN4-p1 obtained with Langevin dynamics. The N-terminus of each chain is marked.

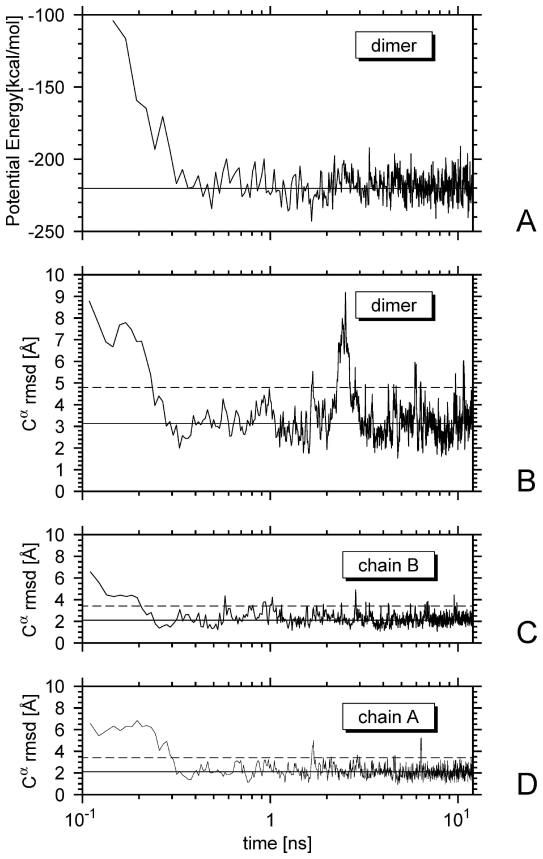

FIGURE 2.16. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for the dimer in a fast folding trajectory of GCN4-p1 obtained with Langevin dynamics. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native structure for each of the monomers. In panel A, the solid horizontal line at -220 kcal/mol is the mean value of the energy after the dimer has reached the native basin. The dashed horizontal line in panels B, C, and D corresponds to the cutoff RMSD (4.8 Å for the dimer and 3.4 Å for the monomers) above which a structure is considered to have left the native basin. The solid horizontal line at 3.1 Å in panel B is the mean C^{α} RMSD inside the native basin of the dimer. The solid horizontal line in panels C and D (at 2.1 Å in panel C and 2.2 Å in panel D) is the mean C^{α} RMSD inside the native basin of the monomer.

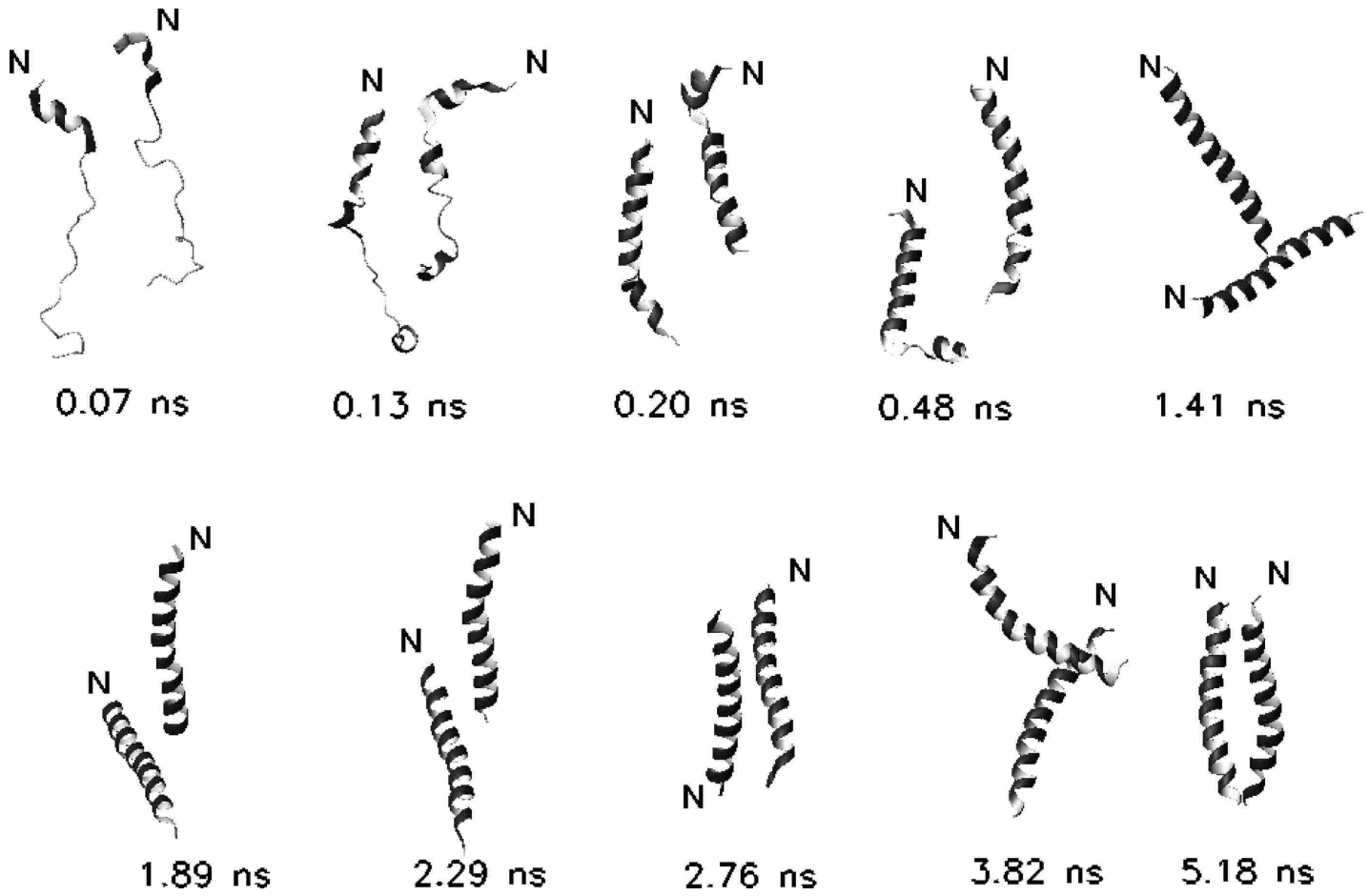

FIGURE 2.17. Example of a slow folding trajectory of GCN4-p1 obtained with Langevin dynamics. The N-terminus of each chain is marked.

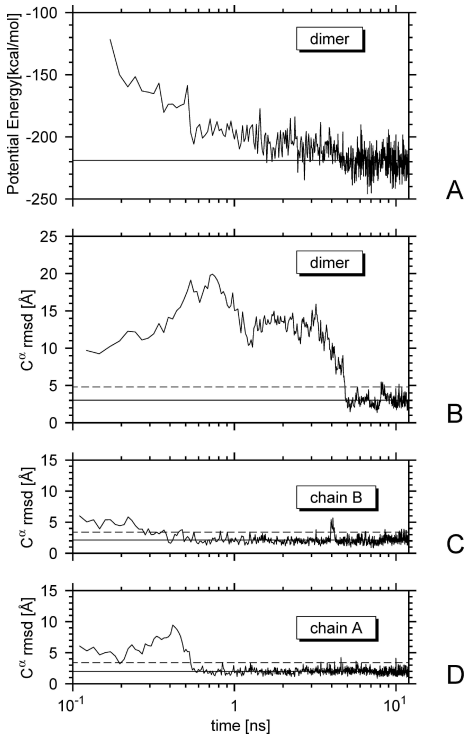

FIGURE 2.18. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for the dimer in the slow folding trajectory of GCN4-p1 obtained with Langevin dynamics. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native structure for each of the monomers. In panel A, the solid horizontal line at -219 kcal/mol is the mean value of the energy after the dimer has reached the native basin. The dashed horizontal line in panels B, C, and D corresponds to the cutoff RMSD (4.8 Å for the dimer and 3.4 Å for the monomers) above which a structure is considered to have left the native basin. The solid horizontal line at 3.0 Å in panel B is the mean C^{α} RMSD inside the native basin of the dimer. The solid horizontal line in panels C and D (at 2.0 Å in panel C and 2.1 Å in panel D) is the mean C^{α} RMSD inside the native basin of the monomer.

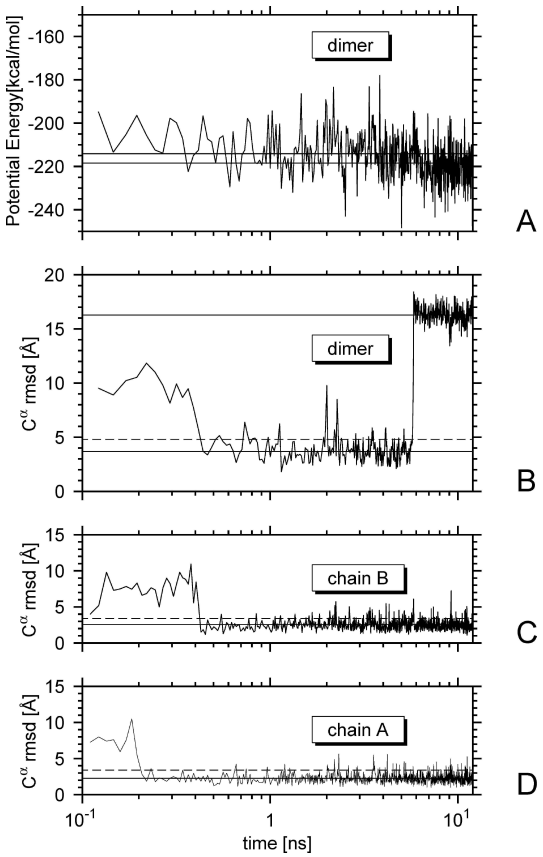

FIGURE 2.19. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for the dimer in a folding trajectory of GCN4-p1 obtained with Berendsen dynamics. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native structure for each of the monomers. The dimer remains in the native basin for almost 5 ns after which it jumps to the misfolded basin. The solid horizontal lines at -214 and -218 kcal/mol in panel A correspond to the mean values of the potential energy inside the native basin and the misfolded basin, respectively. The solid horizontal lines at 3.7 and 16.3 Å in panel B correspond to the mean C^{α} RMSD inside the native basin and the misfolded basin, respectively. The dashed horizontal line in panels B, C, and D corresponds to the cutoff rmsd (4.8 Å for the dimer and 3.4 Å for the monomers) above which a structure is considered to have left the native basin. The solid horizontal line in panels C (at 2.3 Å) and D (at 2.6 Å) corresponds to the mean C^{α} RMSD inside the native basin of the monomer.

When running in the BD mode, for some of the trajectories, the protein jumped from one basin to the other one. The potential energy and ρ values for a representative trajectory presenting this behavior are shown in Figure 2.19. It can be seen that the dimer (panel B) folds and misfolds without affecting the structure of the monomers (panels C and D), which is consistent with the results from single-chain simulations indicating that the monomers are stable by themselves.

As observed for 1G6U, the average potential energies of the native and misfolded basins were very similar (Table 2.2), the slightly lower values for the misfolded structures being within the expected error in the potential function.

When comparing the folding times for the monomers in the multichain simulations with those in the single-chain simulations, I notice that, with both BD and LD, the isolated monomers fold, on average, slightly faster. A closer look at those monomers that, in multichain simulations, have the largest folding times, or did not fold at all, shows that the folding was delayed because the monomers are trapped in structures similar to that shown in Figure 2.13B. In all simulations, the dimers were formed, but one or both chains have this bent structure. As already mentioned, this structure was also found along the pathway of some of the trajectories in the simulations of isolated monomers, but the fact that the isolated monomers were able to find the native structure faster indicates that multichain interactions might stabilize the structure shown in Figure 2.13B.

Those trajectories that did not converge to the native or misfolded basin reached a state (called nonfolded) in which a dimer was formed, but one or both chains had the non-nativelike structure shown in Figure 2.13B.

It should be emphasized that UNRES/MD reflects the energy landscape produced by the UNRES 4P force field. The presence of non-native stable structures is a feature of the force field, not the method. Improvement of the 4P UNRES force field is expected to stabilize the native over the non-native basin to a greater extent.

2.4.3 Retro-GNC4 Leucine Zipper (PDB Code 1C94)

1C94 is a synthetic α-helical homotetramer of 38 residues per chain. The sequence of 1C94 corresponds to the reversed sequence of the leucine zipper portion of GCN4, viz., GCN4-p1 (section 2.4.2). Thus, 1C94 is referred to as the retro-GNC4 leucine zipper. GCN4-p1 consists of 33 residues, and 1C94 consists of the same 33 residues but in reversed order from N- to C-terminus; in addition 1C94 is extended at the N-terminus with the tripeptide sequence Cys-Gly-Gly and at the C-terminus with Gln-Leu[47]. The crystal structure, consisting of four α-helices oriented parallel to each other (Figure 2.20A), was modeled[47] as a dimer of dimers since mass spectroscopic analysis indicated that the chains were covalently linked in pairs by disulfide bonds[47].

2.4.3.1 Monomers

As can be seen from Table 2.3, 9 out of 10 monomer Langevin trajectories and all 10 Berendsen trajectories converged to nativelike structures. The remaining trajectory that did not find the native

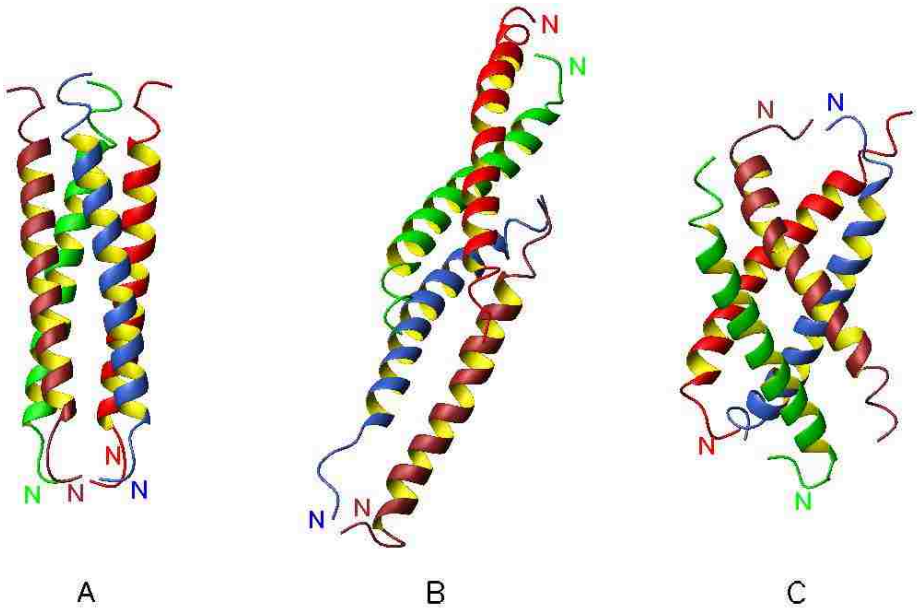

FIGURE 2.20. (A) Experimental structure of 1C94 and (B and C) examples of misfolded structures obtained with LD and BD UNRES/MD. The N-terminus of each chain is indicated.

basin by the end of the simulation showed structures with ρ values around 13 Å where the helix is broken, packing against itself. An example of such a structure is shown in Figure 2.21B. With an older version of the UNRES force field (α_0 force field[50]), Saunders and Scheraga[34] identified a structure of the type shown in Figure 2.21B as the lowest UNRES energy structure. With the force field used in this work (4P force field),23 however, these types of structures have a higher energy than the nativelike structures, as can be seen by comparing the two Langevin trajectories shown in Figure 2.22. Panels A and B show the energy and ρ values, respectively, for the LD trajectory with final structures similar to that shown in Figure 2.21B, and panels C and D show the same information for the LD trajectory converging to the native basin. The mean value of the potential energy in the native basin is indicated with the solid line at -152 kcal/mol in panel C, which is 12 kcal/mol lower than the same quantity in panel A, showing that the UNRES 4P potential energy is lower in the native basin.

Figure 2.23 shows potential energy (panel A) and ρ values (panel B) for a sample trajectory obtained with BD. As can be seen in this example, all Berendsen trajectories showed higher energy values (panel A) and higher fluctuations in the ρ values (panel B) compared to LD runs (panels C and D in Figure 2.22). This could be explained by the fact that, for BD, the absence of friction forces allows for larger conformational changes. No simulations were carried out for 1C94 dimer

TABLE 2.3. Summary of Trajectories for **1C94**

	tetramer							monomer				
	from extended conformation				from crystal structure							
	$N_f^{\mathbf{a}}$	$\langle\tau_f\rangle^{\mathbf{b}}$ (ns)	$N_{mf}^{\mathbf{c}}$	$\langle E\rangle_{mf}^{\mathbf{d}}$ (kcal/mol)	$N_n^{\mathbf{e}}$	$\langle E\rangle_n^{\mathbf{f}}$ (kcal/mol)	CPU time$^{\mathbf{g}}$ (h)	$N_f^{\mathbf{a}}$	$\langle\tau_f\rangle^{\mathbf{b}}$ (ns)	$\rho_{min}^{\mathbf{h}}$ (Å)	$\langle\tau_{res}\rangle^{\mathbf{i}}$	$\langle E\rangle_f^{\mathbf{j}}$ (kcal/mol)
Berendsen	0(18)	2.2	4	-504			6.9	10	1.4	1.36	81%	-107
Langevin	0(18)	2.6	3	-508	3	-510	8.1	9	2.0	1.28	83%	-152

[a]Number of trajectories (out of 10) that folded to nativelike structures, starting from the extended conformation. In the multichain simulations, the number of monomers (out of 40, since there were 4 monomers on each of the 10 simulations of tetramers) that folded to a nativelike structure is indicated between parentheses;

[b]Average folding time of the monomers. The folding time was defined as the time at which the RMSD with respect to the crystal structure fell below 4 Å. In those runs for which the rmsd never went below the cutoff, the folding time was considered to be the simulation time (12 ns for the isolated monomer simulations, 35 ns for the tetramers simulations with LD, and 26 ns for the tetramer simulations with BD). The average folding times for the multichain complex are not calculated since none of the simulations led to nativelike tetramers;

[c]Number of trajectories (out of 10) that yielded misfolded structures;

[d]Average potential energy over all the structures in the misfolded basin;

[e]Number of trajectories, out of 10 simulations started with the crystal structure as the initial conformation that, after 8 ns of simulation, still had nativelike structures (RMSD with respect to crystal structure below 8 Å);

[f]Average potential energy over all those trajectories that, starting with the crystal structure, remained in the native basin after 8 ns of simulation;

[g]Average CPU time (in hours) per 1 ns of simulation on a single 3.06 GHz Intel Pentium IV Xeon processor;

[h]The lowest rmsd in all of the fluctuating trajectories;

[i]Fraction of the time that the peptide spent in the native basin averaged over all of the folding trajectories;

[j]Average potential energy over all structures in the native f basin.

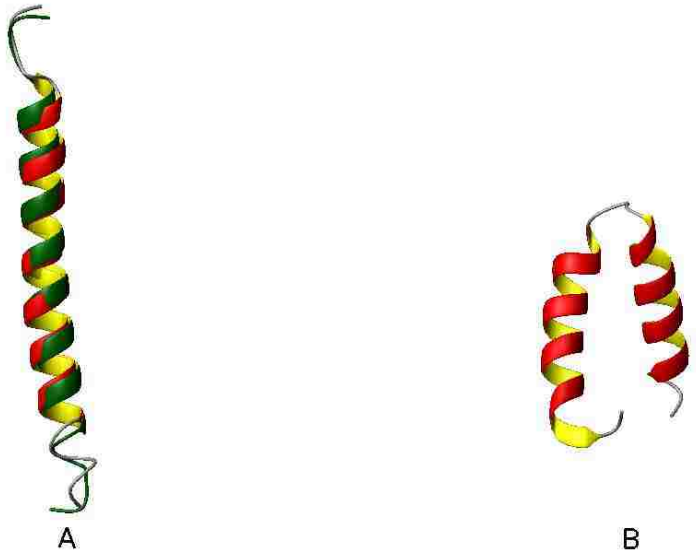

FIGURE 2.21. (A) Superposition of one of the monomers in the experimental structure of 1C94 (green) on the most nativelike structure (red) (C^{α} RMSD = 1.28 Å) obtained with the BD UNRES/MD. (B) A structure that was often found during the folding pathway of 1C94 (either with BD or LD) and was the final structure of the monomer LD trajectory that did not find the native basin. The N-terminus is indicated.

2.4.3.2 Tetramers

Berendsen and Langevin simulations were carried out starting with the four chains in the extended conformation, with each pair of chains cross-linked by disulfide bonds. The chains were in the same plane, parallel to each other and with a 20 Å distance between consecutive chains. On the basis of the experimental data[47], the Cys residue at the first N-terminal position was assumed to form a disulfide bond with the corresponding Cys residue in another chain; however, this residue was never included in the RMSD calculations since it is not resolved in the experimental structure. The simulation time was 35 ns for LD runs and 28 ns for BD runs. The equilibrium concentration of 10 mM was reached during the first 50 ps of simulation.

None of the trajectories obtained with UNRES/MD yielded nativelike structures. On the other hand, both methods found stable structures consisting of two parallel dimers bound together in an antiparallel orientation (instead of parallel as in the native structure), examples of which are shown in Figures 2.20B and 2.20C. In the structure shown in Figure 2.20B, the dimers have native-like structures, but the area of contact between the dimers is very small. However, the structure shown in Figure 2.20C has better packing, but the dimers have non-native-like structures, and the disulfide-linked monomers are not parallel to each other but slightly twisted to align in an antipar-

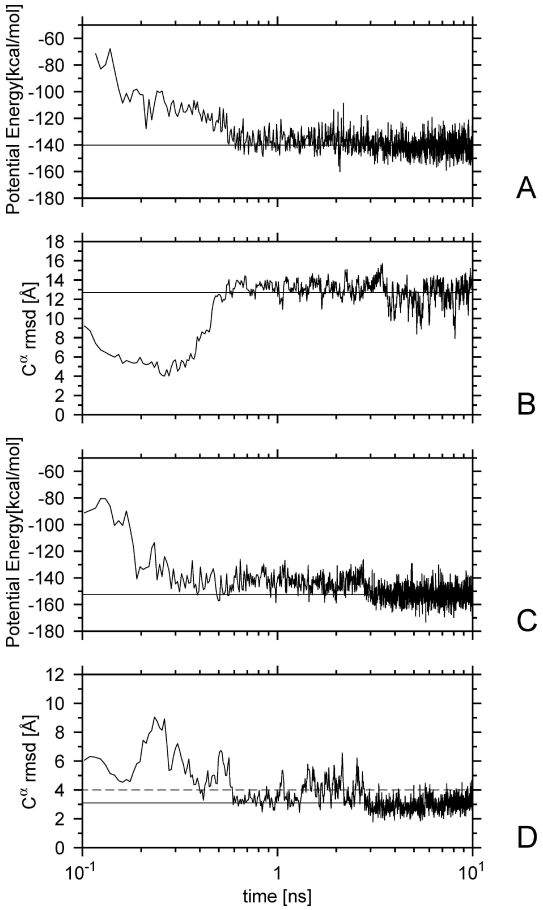

FIGURE 2.22. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure of the monomer in the tetramer as a function of time for an LD trajectory of an isolated monomer of 1C94 converging to a non-native-like structure (which is shown in Figure 2.21B). In panel A, the solid horizontal line at -140 kcal/mol is the mean value of the energy after the monomer has adopted the non-native stable structure. In panel B, the solid horizontal line at 12.8 Å is the mean C^{α} RMSD after the peptide has adopted the non-native structure. Panels C and D contain the same information as panels A and B, respectively, for an LD trajectory converging to the native basin. The solid horizontal line at -152 kcal/mol in panel C is the mean value of the energy after the peptide has reached the native basin. In panel D, the dashed horizontal line at 4 Å corresponds to the cutoff RMSD above which the structure is considered to have left the native basin, and the solid horizontal line at 3.1 Å is the mean C^{α} RMSD inside the native basin. The solid horizontal line at -152 kcal/mol in panel C is the mean value of the energy after the peptide has reached the native basin.

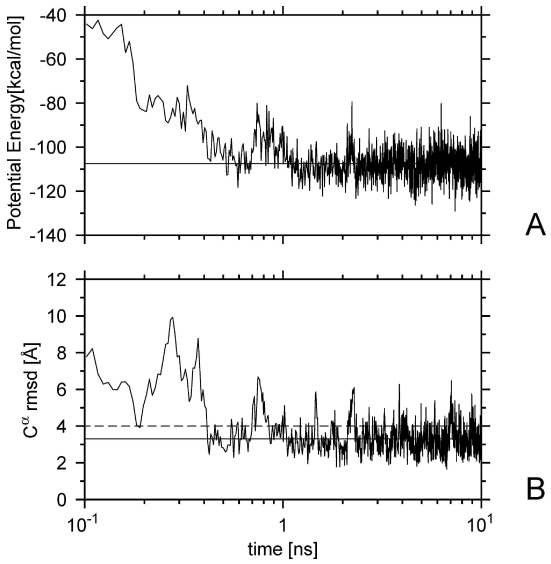

FIGURE 2.23. (A) Variation of the potential energy and (B) the C^{α} rmsd from the native structure during the folding of an isolated monomer of 1C94 obtained with Berendsen dynamics. In panel A, the solid horizontal line at -105 kcal/mol is the mean value of the energy after the protein has reached the native basin. The dashed horizontal line, at 4 Å, in panel B corresponds to the cutoff rmsd above which the structure is considered to have left the native basin, and the solid horizontal line at 3.3 Å in the same panel is the mean C^{α} rmsd inside the native basin.

allel orientation with the monomers from the other dimer. These two structures have approximately the same potential energy (approximately -507 kcal/mol); I will refer to either of them as misfolded structures. Figure 2.24 shows the potential energy (panel A) and ρ values for the tetramer (panel B) and for the dimers (panels C and D) as a function of time for the trajectory leading to the structure in Figure 2.20B. It can be seen that, by the end of the simulation, the ρ values for the tetramer stabilize around 22 Å (indicated by a solid line in panel B) while, for the dimers, it remains below or close to the 5.6 Å cutoff (indicated by the dashed lines in panels C and D). The potential energy also stabilizes by the end of the simulation, with values around -510 kcal/ mol (indicated by a solid line in panel A).

To determine whether the native structure of the tetramer could not be found because of imperfections in the UNRES 4P force field or simply because the simulation times were too short, I carried out a set of 8 ns simulations with the crystal structure as the initial conformation using Langevin dynamics. As can be seen in Table 2.3, 3 out of 10 simulations remained in the native basin. Potential energy and ρ values corresponding to one of the trajectories that did not remain in the native basin are shown in Figure 2.25. It is important to notice that although the tetramer leaves the native basin (ρ values crossing the dashed line at the 8 Å cutoff in panel B) there is no substantial change

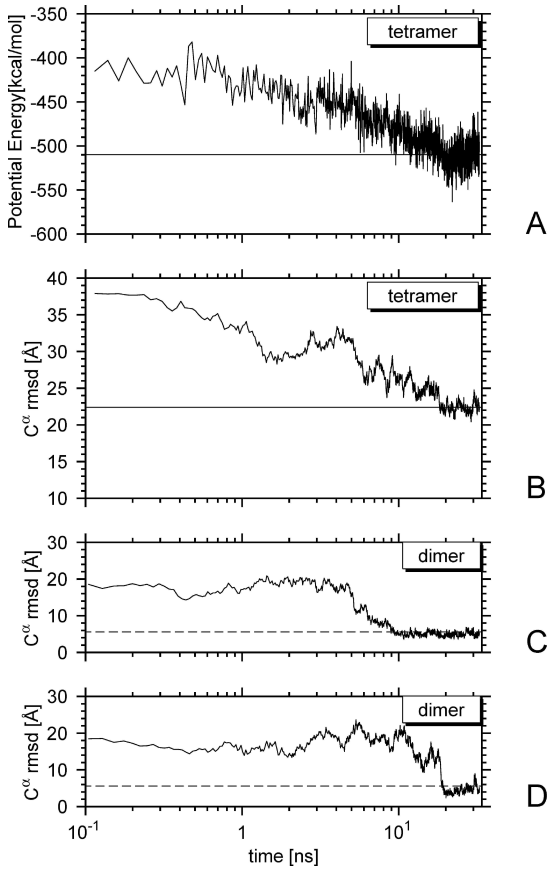

FIGURE 2.24. (A) Variation of the potential energy and (B) the C^{α} RMSD from the native structure for a misfolding trajectory of 1C94, starting with extended chains, obtained with Langevin dynamics. For the same trajectory, panels C and D show the variation of the C^{α} RMSD from the native for each of the dimers. In panels A and B, the solid horizontal line is the mean value of the energy (at -510 kcal/mol) and C^{α} RMSD from the native structure (at 22.4 Å), respectively, after the tetramer has found the misfolded basin. The dashed horizontal line in panels C and D corresponds to the 5.6 Å cutoff RMSD, above which the dimers are considered to have left the native basin; i.e., the dimers folded but the overall structure was misfolded.

in the potential energy (panel A). I calculated the average potential energy among those structures that remained in the native basin and compared it with the average energy among the misfolded structures. The values obtained were almost equal (Table 2.3), indicating that the protein might choose either conformation with the same probability. However, when starting from the extended conformation, none of the simulations led to nativelike structures. Therefore, the energy landscape generated by the UNRES 4P potential makes the antiparallel conformation more easily accessible than the parallel (native) conformation; i.e., the free energy of the misfolded basin has a lower value compared to that of the native basin.

When comparing the folding times of the monomers in the single- and multichain simulations (Table 2.3), I did not find any appreciable difference, indicating that, for this protein, multichain interactions do not play an important role in the folding of the monomers.

It can be concluded that the failure to fold the protein to the native tetramer with the UNRES 4P force field should be attributed to the imperfections in the potential rather than to insufficient simulation time because, first, for the two preceding proteins (1G6U and GCN4-p1), I observed the formation of both the native and the non-native dimers and, second, in a previous implementation of UNRES to search for the native structures of multichain proteins with CSA[34, 51] the native structure of retro-GNC4 could be predicted by global optimization only when native symmetry constraints were imposed. Improvement of the 4P UNRES force field is expected to stabilize the native basin to a greater extent compared to the non-native basin.

2.5 Conclusions

The UNRES/MD implementation described in ref [23] was extended to treat multichain proteins. The method was tested on three α-helical proteins, two dimers and one tetramer.

To simulate a constant temperature bath, two alternative methods were implemented, the Berendsen thermostat (BD) and a method based on the Langevin equation (LD). The latter method includes friction and stochastic forces explicitly as opposed to the former for which these forces are included implicitly. When comparing the time required for each method to find the global minimum of the energy, BD proved to be much faster than LD, as observed in earlier studies on singlechain proteins[22]. However, it should be noted that, despite its predicting efficiency, BD might not reproduce the true

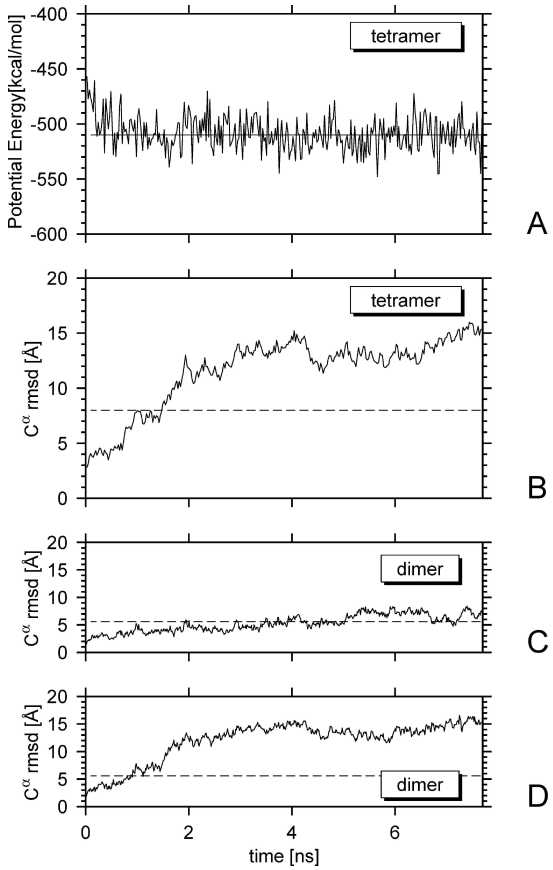

FIGURE 2.25. (A) Variation of the potential energy and (B) the C^α RMSD from the native structure for a trajectory of 1C94 that did not remain in the native basin, obtained with Langevin dynamics, with the crystal structure as the initial conformation. The solid horizontal line at -510 kcal/mol in panel A is the mean value of the energy during the simulation. The dashed horizontal line in panel B corresponds to the 8 Å cutoff RMSD, above which the tetramer is considered to have left the native basin. For the same trajectory, panels C and D show the variation of the C^α RMSD from the native for each of the dimers. The dashed horizontal line in panels C and D corresponds to the 5.6 Å cutoff RMSD, above which the dimers are considered to have left the native basin.

folding pathway. LD, which reproduces a true canonical ensemble, should be used instead when studying the kinetics of the folding process, as in ref [24].

Simulations of single chains and multichain complexes were carried out with BD and LD. Single-chain simulations indicate that, for each of the three α-helical proteins tested in this work, the structure adopted by the monomer in the multichain complex is also the lowest UNRES 4P energy structure of the isolated monomer. In general, the folding times of the monomers in the single-chain simulations were shorter than those in the multichain simulations, which indicates that, with the UNRES 4P force field, the short-range interactions, responsible for the folding of the single-chain α-helices, are impaired by the interactions between different chains. However, the folding of 1G6U with BD (section 2.4.1) was the exception. In these simulations, the monomers folded faster when they were allowed to interact with another monomer; i.e., the correct packing of the two helices on each monomer is favored by the interactions with another monomer. Although the wrong orientation of the monomers with respect to each other can sometimes hinder the packing of the helices, with BD, in which the friction forces are absent, the chains can rearrange quickly to find a more favorable orientation that will aid the packing of each monomer. This behavior is probably an artifact of BD and might not represent the folding mechanism of 1G6U.

It is important to note that, although some of the trajectories led to non-native-like structures, these structures were indeed free-energy minima within the context of UNRES 4P. In the case of the two dimers, the non-native structure was competing with the native one. This competition was reflected in the simulations, especially in the case of GCN4-p1 for which the dimer switched from one structure to the other. In the case of 1C94, the results were poor since none of the trajectories yielded the native structure. The reason for this failure might be found in the defects of the UNRES parameters. Improvement of these parameters is ongoing research in the UNRES developing team.

It must be emphasized that the goal of this work was to test the implementation of UNRES/MD on multichain proteins and not to improve the 4P force field, and therefore, I chose relatively simple systems which the force field could treat to test the approach, as pointed out in the Introduction. The UNRES 4P force field was trained using four proteins with different topologies and tested on 66 proteins with chain lengths from 28 to 144 amino acid residues. The average size of correctly predicted segments of α-helical proteins was approximately 67 residues[32]. The parametrization

procedure and the limitations of the UNRES 4P force field are described extensively in ref [32]. The reason for such limitations must be found in the old parametrization procedure[28, 30, 31, 32], which neglected conformational entropy, an issue that has been addressed, and preliminary results are reported in [45].

Finally, in contrast to earlier calculations of multichain complexes[34, 51] with CSA[52, 53] as a global optimization algorithm, in which symmetry constraints had to be imposed to simulate the experimental structure, no such constraints were imposed here. Apparently, in the time scale achieved in MD with UNRES, the search of the conformational space of a dimer is more efficient than that with CSA.

Chapter 3
Mechanism of Fiber Assembly of the Aβ-Peptide

3.1 Summary

In this chapter, I describe my studies of the mechanism of growth of fibrils of the β-amyloid peptide (Aβ). In section 3.2, I describe the role of Aβ in Alzheimer's disease (AD), summarize some of the most recent discoveries in the field, and introduce the motivation for studying Aβ association. To study different aspect of Aβ fibrils, I carried out different types of simulations. In section 3.3, I study the ensemble of conformations explored by the isolated monomer of Aβ_{1-40}. In section 3.4, I analyze the stability of small oligomers of Aβ_{1-40} with the structure that is characteristic of Aβ_{1-40} fibrils, and determine how their stability is related to the size of the oligomers. In section 3.5, I study the presence of cooperativity in the hydrogen-bond interactions in Aβ_{1-40} fibril templates. In section 3.6, I describe the simulations of the elongation process of Aβ_{1-40} fibrils. Finally, in section 3.7, I summarize all the simulations that I carried out on Aβ_{1-40}, the results I obtained, and elaborate on the implication of these results.

3.2 Amyloids, Aβ and Alzheimer's Disease

Many diseases have been associated with deposits of amyloid plaques, including Alzheimer's disease (AD), Parkinson's disease (PD), type II diabetes, and spongiform encephalopathies. In the particular case of (AD), these plaques contain filamentous forms of a protein known as the β-amyloid peptide (Aβ)[4, 5]. Oligomeric forms of this protein, both fibrilar[6] as well as soluble nonfibrilar Aβ aggregates[7], have been identified as the cause of AD. However, the mechanism(s) by which they may initiate the disease is still unclear[54].

Great progress has been achieved in elucidating the 3D structure of amyloid fibrils[55, 56, 57, 58, 59, 60, 61], and we now know that amyloid fibrils from different species share a characteristic motif, the cross-β structure, in which the polypeptide chains form extended β strands that align perpendicular to the axis of the fibril. Fibrils formed by the Alzheimer's Aβ_{1-40} peptide have been studied extensively by Deco and co-workers[58, 59, 61]. Based on constraints from solid state NMR, structural models of Aβ_{1-40} fibrils have been proposed[59, 61].

Despite progress in understanding the fibrilar state of Aβ, the mechanism by which small oligomers evolve into their fibrilar form is not yet well understood. In recent years, the role of the fibrils in the disease has been questioned. Instead, it has been proposed that smaller soluble oligomers are sufficient to trigger the disease[7], making it urgent to understand these smaller structures and the assembly mechanism of amyloid fibrils. However, the short life-time and noncrystallinity of these intermediate structures have hindered a description of their molecular structure as well as the process by which they might evolve into fibrils or how these fibrils grow[62, 63]. Computer simulations consistent with the available experimental data could provide some insight into the understanding of amyloid formation and growth.

In the laboratory, Aβ_{1-40} fibril formation takes as long as days[64, 65]. Once the fibrils are formed, the growth proceeds by incorporating new monomers at a constant rate of approximately 0.3μm/minute (with a few milli-seconds per monomer incorporated)[64]. These time scales make simulations of fibril formation, or elongation, extremely challenging. To overcome the time limitation, most all-atom studies have focused on small fragments of Aβ[63, 66]. Although these studies[63, 66] have contributed greatly to our understanding of the transition that an unstructured monomer undergoes upon binding to a fibril, they might not reflect the full complexity of the complete Aβ_{1-40} system. Implicit-solvent all-atom simulations of elongation of Aβ_{1-40} have been carried out[67] but, due to their high computational cost, these simulations could not describe the assembly of a completely unstructured and unbound monomer into a fibril template. Another approach has been the use of coarse-grained models, biased towards the desired conformation[68, 69] or simplified models in which the polypeptide chain is represented by a tube, and the interactions between amino acids are derived from geometry and symmetry considerations[70]. These models have the disadvantage that they might not reproduce the complexity of the true energy landscape.

In this work, we have adopted a coarse-grained united-residues (UNRES) model[27, 71] to partially surmount the time-scale problem. The advantage of UNRES over other coarse-grained force fields is that UNRES has been derived on the basis of physical principles. The energy terms are the result of averaging the less important degrees of freedom of the all-atom free energy of a protein and the solvent[27]. The force field ultimately has been parametrized to reproduce the free energy landscape of a small training protein, completely different from Aβ[45]. Therefore, the force field is not biased

towards the Aβ fibril conformation. Moreover, UNRES has been shown to be able to carry out MD simulations of the folding of multichain systems within reasonable time, starting from completely unstructured conformations, and without using any information from the native structure of these systems[71]. Therefore, UNRES has been adopted to simulate the assembly of a free monomer onto a fibril template without imposing any type of restraint on the monomer.

With the UNRES model, we carried out canonical molecular dynamics (MD) and replica exchange MD (REMD) simulations to: a) describe the ensemble of conformations explored by the isolated monomer of Aβ_{1-40}; b) analyze the stability of small oligomers of Aβ_{1-40} with the structure that is characteristic of Aβ_{1-40} fibrils, and determine how their stability is related to the size of the oligomers; and c) study the elongation process of Aβ_{1-40} fibrils.

3.3 Studying the Conformations Adopted by Isolated Monomers of Aβ_{40}

It is extremely difficult to carry out experimental studies of monomeric Aβ because the peptide has a high tendency to aggregate and eventually precipitate. For this reason, it has not yet been possible to study the full-length peptide in water solution. Experiments on fragments of Aβ in water have been possible, and they show that the fragments have little regular structure[72, 73]. To prevent aggregation, many experiments are carried out in a mixture of water and organic solvents, such as trifluoroethanol (TFE) [74, 75, 76, 77] or micellar solutions[78, 79]. Under these conditions, the monomeric Aβ peptide shows substantial helical structure.

All-atom computer simulations of the full-length Aβ_{40} and Aβ_{42} in implicit water indicate that both peptides exist predominantly in two types of conformations, each one possessing significant amounts of either α or β-structure[80]. All-atom simulations of Aβ_{39} showed that the peptide has limited helicity and no β structure as a monomer[81], implying that the β rich structures characteristic of the fibrilar and intermediate conformations might be stabilized upon oligomerization[81].

Although the 3D structure of the full length Aβ in the absence of organic solvent is still not known, the conformational changes accompanying fibril formation have been studied by circular dichroism (CD) spectroscopy[82, 83]. These experiments show that a helical intermediate precedes fibril assembly. Furthermore, Fezoui and Teplow[83] studied Aβ fibril assembly in the presence and absence of TFE, a solvent known to stabilize α helical conformations. Since the peptide adopts a

β structure in the fibril, one might expect that, by stabilizing helical structures, fibril formation would be prevented. However, the study[83] showed that TFE, at low concentrations, while still promoting α-helical conformations, accelerates rather than inhibits amyloid fibril formation. These experiments[82, 83] support the hypothesis that a partially folded intermediate with a certain α-helical content is present during Aβ fibril formation. Moreover, the presence of this intermediate might facilitate the process[82, 83].

The foregoing results indicate that a model suitable for the study of Aβ amyloids should be able to capture an α-helical propensity at the monomer level as well as to form oligomeric structures with high β content. To test whether UNRES could capture the ability of monomers to adopt α-helical and β-sheet conformations, we carried out a set of 40 ns independent canonical MD simulations of an isolated monomer of Aβ_{1-40}, with the temperature of the system held constant at 300 K by means of the Berendsen thermostat[35] (section 2.2.4). The simulations were started with the monomer in the extended conformation. The system was allowed to equilibrate for 20 ns, and the conformations visited during the remaining 20 ns were clustered based on their structures.

Conformations were stored every 150,000 steps and clustered into families by means of the minimal-tree algorithm[84, 85] based on the C^{α} root-mean-square deviation (RMSD) distances between conformations. Three large clusters were identified, accounting for 69% of the conformations. These clusters also corresponded to the lowest energy values calculated with the UNRES force field. The largest cluster, containing 56.5% of the conformations, corresponds to structures with high α-helical content (see Figure 3.1). The second and third largest clusters, accounting for 7.5% and 4.7% of the conformations, have β structures. Figure 3.1 shows the probability of occurrence of conformations populating the three largest clusters as a function of the UNRES potential energy, as well as the representative conformation of each cluster (defined as the one with the lowest average RMSD from all other members of the cluster). The UNRES energy of each cluster is computed as the energy of the representative trajectory of the cluster.

These results indicate that, at the monomer level, UNRES can reproduce the ability of Aβ_{1-40} to adopt helical and β-strand conformations. Furthermore, the UNRES force field, being a coarse-grained one, can facilitate a study of the behavior of large oligomers, a task that is still challenging with an all-atom force field, making UNRES a very good choice to study Aβ amyloids.

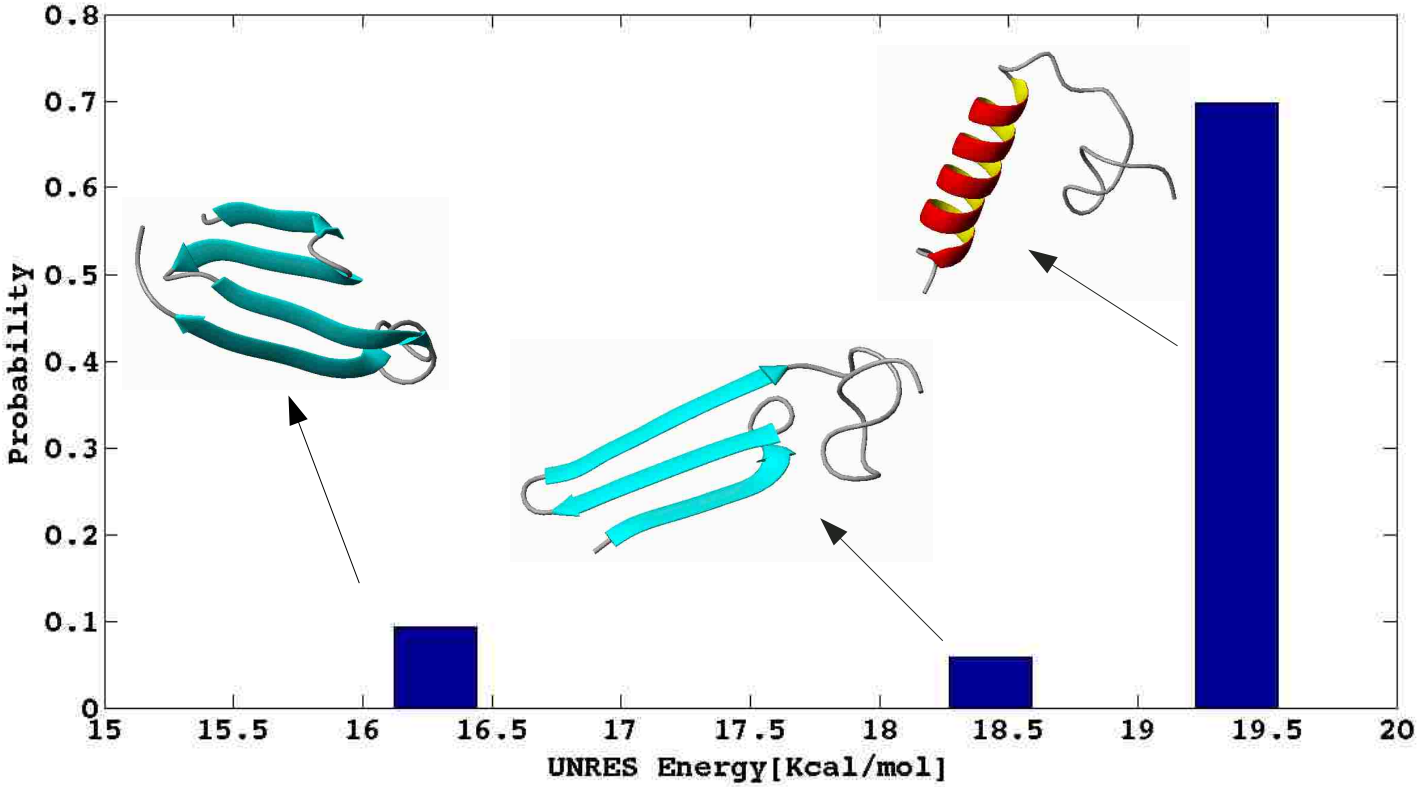

FIGURE 3.1. Probability of occurrence of conformations populating the three largest clusters as a function of the UNRES potential energy of the representative conformation. The representative conformation of a cluster is defined as that with the lowest RMSD from all other members of the cluster. The representative conformation for each cluster is shown, and the correspondence is indicated by arrows.

3.4 Studies of the Stability of Aβ_{1-40} Fibrillar Conformation

3.4.1 Motivation for Studying the Stability of Aβ_{1-40} Fibrillar Conformation

Fibrils formed by Aβ_{1-40} have been studied extensively by solid state NMR[58, 59, 61]. These studies have shown that, in the fibrilar conformation, the peptide adopts the cross-β structure. *i.e.*, it forms β strands that lie in a plane perpendicular to the axis of the fibril (see Figure 3.2), with the β-strands stabilized by hydrophobic interactions in each plane (within hairpins and between hairpins, Figure 3.2 c), but lacking the hydrogen bonds of conventional anti-parallel β-sheets. These double-hairpin structures of Figure 3.2 c form interplane parallel β-sheet-like hydrogen bonds with a similar pair of hairpins in a consecutive layer; the specific type of interactions are described bellow.

Two types of Aβ_{1-40} fibrils have been identified, depending on the growing conditions[58]. Fibrils grown in a quiescent solution exhibit a periodic twist, while fibrils formed in a solution that is gently agitated do not present any resolvable twist, have a smaller diameter, and tend to associate laterally with other fibrils forming flat bundles[58]. These two fibril types have been named *twisted pair* and *striated ribbon*[61], respectively. Although both fibril types present the cross-β structure at the molecular level, twisted pair fibrils consist of stacks of trimers while striated ribbon fibrils are stacks of dimers, as shown in Figure 3.2. Given the simpler architecture of the layers in the striated ribbon fibrils, we selected these as a model for computer simulations. Hereafter, when we refer to Aβ_{1-40} fibrils, we will be referring to the striated ribbon morphology.

NMR data indicated that residues 1-8 were conformationally disordered, and were omitted in the structural model[59]. Therefore, in our simulations we used the Aβ_{9-40} segment, for which the coordinates are available. Regarding the terminology to describe the fibrilar structures, we use the term *layer* to refer to the unit containing the dimer (Figure 3.2 c), perpendicular to the fibril axis. The term *semi-filament* is used to refer to a stack of hydrogen-bonded monomers, parallel to the fibril axis. According to this terminology, a fibril can be seen as formed by two parallel semi-filaments, or by a stack of parallel layers.

The structures of smaller soluble oligomers, intermediate to fibrils, have been more elusive. Incidentally, these intermediates have been identified as the most toxic species in AD[7]. In recent years,

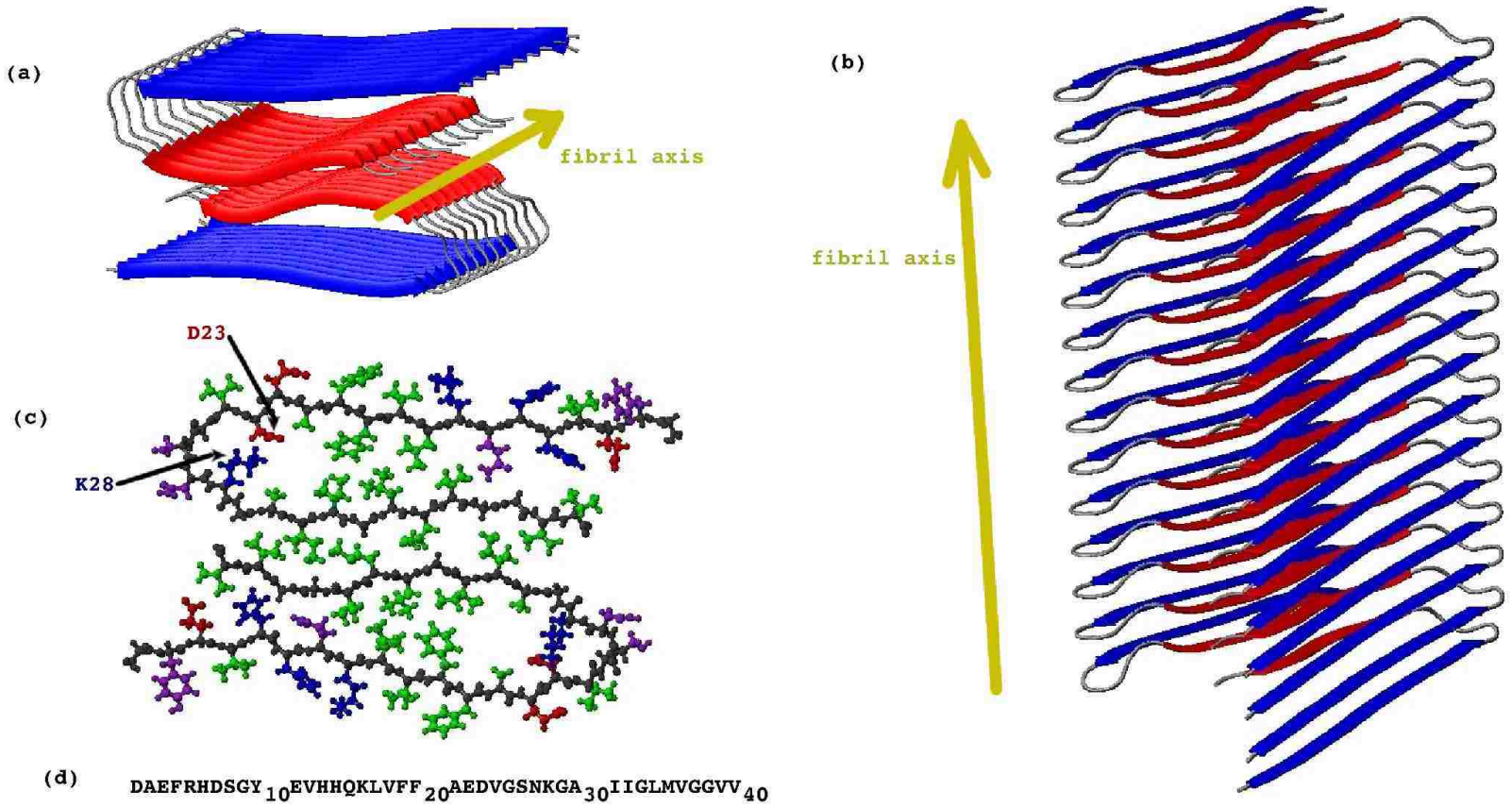

FIGURE 3.2. Structural model for an Aβ_{1-40} fibril with the striated ribbon morphology. The Figure was produced with MolMol [86], based on the coordinates provided by Robert Tycko for the structural model of Petkova *et al.*[59]. Residues 1-8 are omitted from the diagram because they were conformationally disordered in the NMR model[59]. (a) Axial view and (b) side view of the fibril. The fibril axis is indicated by a dark yellow arrow. N-terminal β strands are colored in blue, while C-terminal β strands are colored in red. The fibril is formed by layers of dimers, lying perpendicular to the fibril axis. (c) An all-atom representation of a dimer from a fibril layer. Hydrophobic, polar, negatively charged and positively charged side chains are colored in green, purple, red, and blue, respectively. (d) The sequence of Aβ_{1-40}. Only residues 9-40 were used in the simulations of oligomers.

great effort has been directed towards the characterization of these intermediates[62, 87] and their mechanism for toxicity. However, a detail characterization of their 3D structure has not yet been possible.

Regarding the role of small oligomers in AD, in recent work Ono *et al.*[87] showed that neurotoxicity increases considerably with the structural order of the oligomeric intermediates. A question that arises is whether small oligomers could adopt the conformation characteristic of large fibrils, and what would be the size of the smaller oligomer capable of retaining the fibrilar structure.

If small oligomers with the molecular structure of fibrils do exist, the internal forces holding the structure together might be the same as in the fibrils. NMR data have provided valuable insight into the interactions stabilizing Aβ_{1-40} fibrils [58, 59]. It is now known that residues L17, F19, A21, A30, I32, L34, and V36 create a hydrophobic cluster between the β-strands in each monomer (Figure 3.2 c) and between the β-strands of one monomer and those of a monomer in a consecutive layer within each semi-filament. The structure is further stabilized by salt bridges between oppositely charged residues D23 and K28, within the same or consecutive layers. At the interface between the two monomer in a given plane (Figure 3.2 c), the structure is stabilized by hydrophobic interactions involving residues I31, M35, and V39. In-registry intermolecular hydrogen bonds comprising residues 10-22 and 30-40 are formed between consecutive layers[58, 59].

3.4.2 Studying the Structural Stability of Aβ_{9-40} Oligomers with Different Numbers of Chains

Computer simulations have been very valuable to help gain an understanding of the structural stability of Aβ fibrils. Buchete *et al.*[88] used molecular dynamics (MD) and all-atom force fields to study the behavior of a four-layer Aβ_{9-40} oligomer (*i. e.*, an eight-chain oligomer) with the striated ribbon morphology. This study showed that a system of that size was stable during a 10 ns simulation. On the other hand, with a coarse-grained model, Fawzi *et al.*[69] found that Aβ_{1-40} oligomers were stable only for systems with 8 layers (16 chains) or more.

Regarding the stability of Aβ_{9-40} oligomers, we wanted to answer two questions. First, will the native structure of the Aβ_{9-40} oligomers be stable with the UNRES force field? And second, how will the stability of the oligomers change when their size is changed? To answer these two questions, we carried out canonical MD simulations with the UNRES force field, starting with native

conformation[59], and allowing it to fluctuate freely We studied systems with different numbers of layers, ranging from 2 to 8 (*i. e.*, 4 to 16 chains). For each system, 8 independent canonical MD trajectories (with the Berendsen thermostat), at 300 K, were simulated, and each simulation was 5 ns long.

To assess the extent of the structural changes during the simulations, we measured the C^{α} root-mean-square deviation (RMSD) with respect to the initial conformation. The average RMSD (taken over all the trajectories with the same size) as a function of time for different sizes is shown in Figure 3.3. The 4-chain systems are the least stable; the RMSD grows very quickly and, within the first nanosecond of simulation, these systems find a more stable non-native conformation, where the RMSD stabilizes at about 17 Å. For systems between 6 and 12 chains, although the RMSD grows at a slower rate, the initial conformation is not stable and, by the end of the simulation, the RMSD reaches values around 25 Å for systems between 6 and 10 chains and 15 Å for the 12-chain systems. Only the largest systems, with 14 and 16 chains, retain most of the fibrilar structure during the length of the simulation, with the 16-chain systems being the most stable.

To illustrate the behavior of the different oligomers that do not retain their fibrilar structure, snapshots along the pathway of representative trajectories are shown in Figure 3.4 for a 2-layer (4-chains), a 3-layer (6-chains) and a 6-layer (12-chains) system. All the systems shown lose their structural stability during the simulation. The 2-layer system [Figure 3.4, panel (a)] is the most unstable, and at 0.4 ns it has already lost its structure. The semi-filaments quickly rearrange and assemble one on top of the other in opposite orientation, forming a tube shaped structure with most of the hydrophobic residues buried inside the core. In the 3-layer system [Figure 3.4, panel (b)], the structure of each semi-filament is already lost at 2.5 ns. The β strands on each semi-filament (colored in blue and red) have opened up, and the two semi-filaments have started to separate. At 5 ns the semi-filaments have completely changed their orientation. The 6-layer system [Figure 3.4, panel (c)] retains the structure of the contact area between the semi-filaments during the length of the simulation, but the β strands on each semi-filament have started to go apart at 2.5 ns, and they have completely opened up at 5 ns.

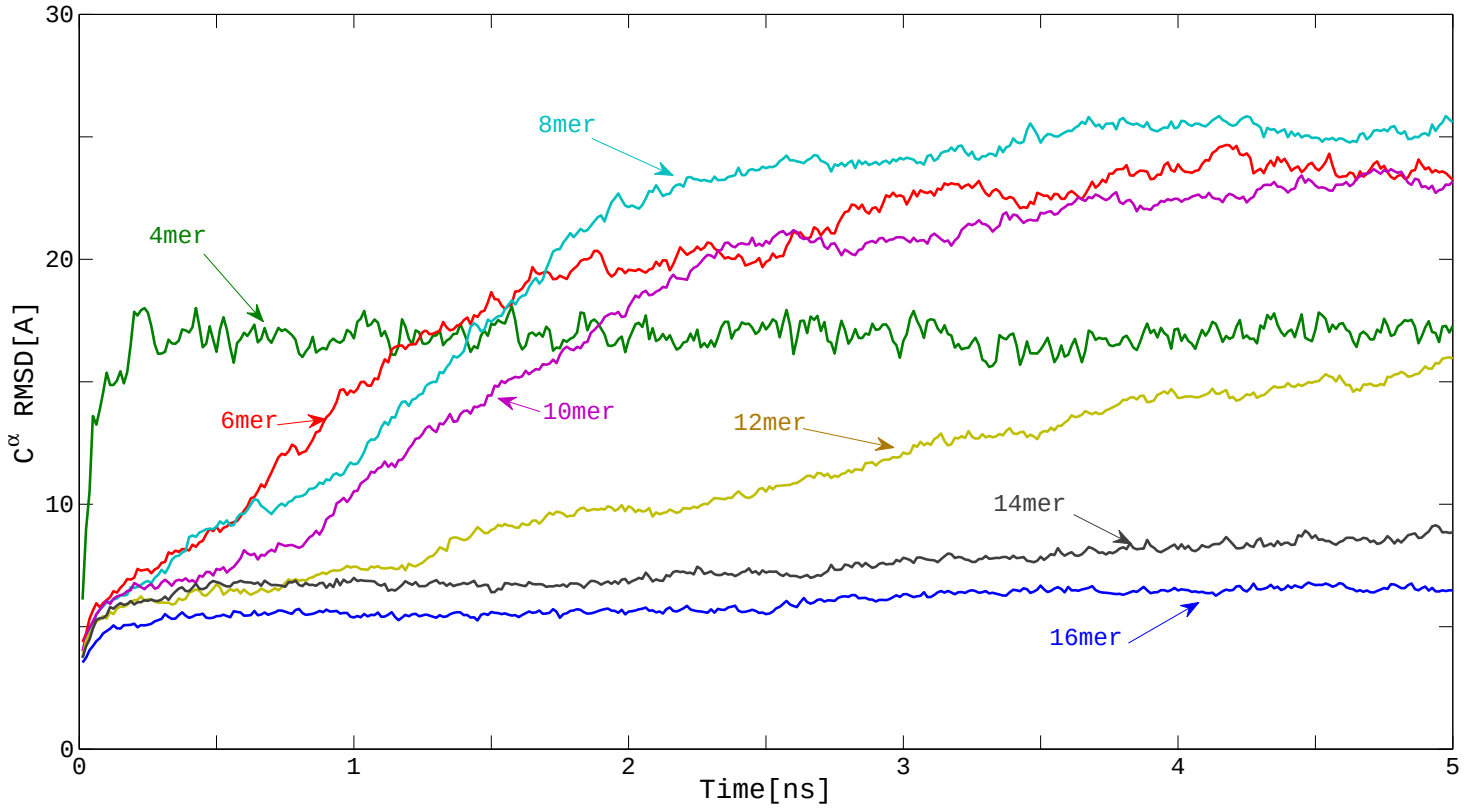

FIGURE 3.3. Average variation of the C^{α} RMSD with respect to the initial structure during constant temperature canonical MD simulations of $A\beta_{9-40}$ oligomers with different numbers of chains per oligomer.

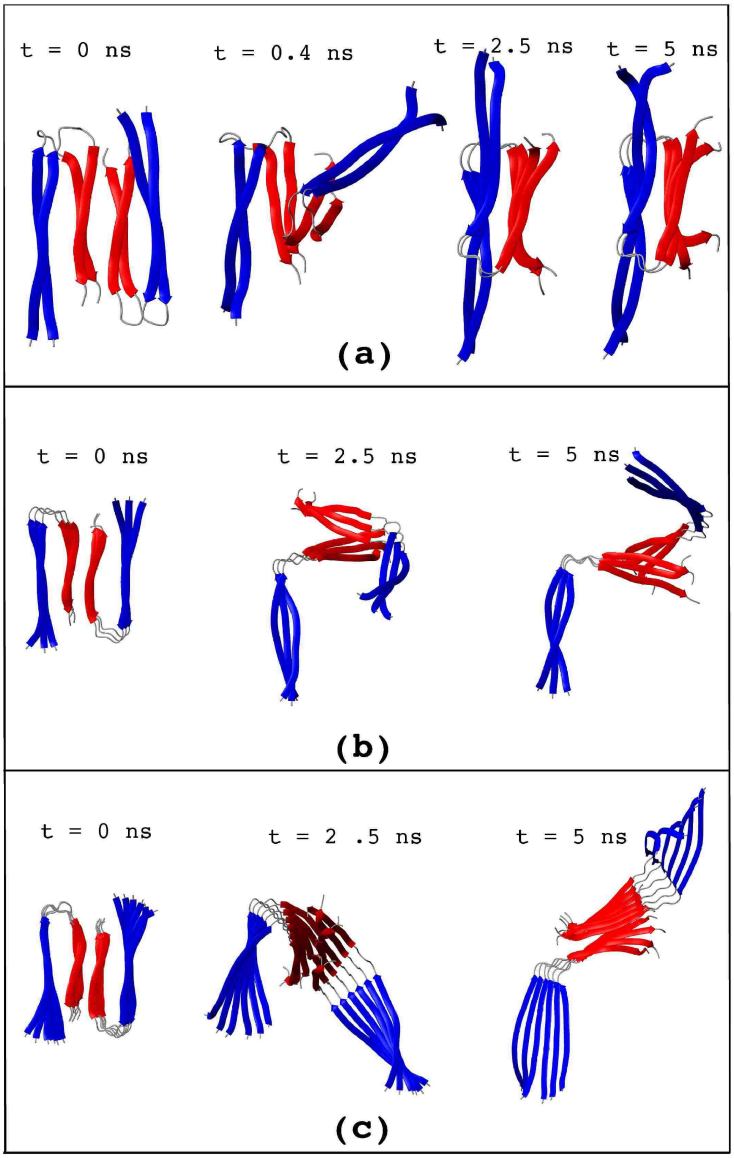

FIGURE 3.4. Representative trajectories from the 2-layer (a), 3-layer (b), and 6-layer (c), systems to show the conformational changes of the different systems with time. The conformations at 0, 2.5 and 5 ns for the selected trajectories are shown. An additional snapshot at t=0.4 ns is shown for the 2-layer system because the conformational changes in this system occur faster. The N- and C-terminals are colored in blue and red, respectively.

To find the reasons for the instability of the different oligomers we analyzed the energetics of the system. In sections 3.4.2.1 to 3.4.2.3 we examine the three main interactions stabilizing Aβ fibrils, hydrophobic interactions, hydrogen bonds and salt bridges.

3.4.2.1 Hydrophobic Interactions

The explanation for the instability of the smaller oligomers of Aβ_{9-40} with the UNRES force field can be found in the competition between hydrophobic interactions and local electrostatic interactions [$U_{corr}^{(3)}$ (see section 2.2)]. In UNRES, the $U_{corr}^{(3)}$ energy term corresponds to the coupling between the dipole moments of two interacting peptide groups and the geometry of the backbone around them[27]. While hydrophobic interactions will help to stabilize this structure, $U_{corr}^{(3)}$ interactions might not.

To investigate this, we examined the influence of both $U_{corr}^{(3)}$ and the term that represents the hydrophobic/hydrophilic interactions in UNRES, the side chain-side chain energies U_{SCiSCj}[27].

Figure 3.5 shows the behavior of U_{SCiSCj} and $U_{corr}^{(3)}$, and the RMSD with respect to the initial structure for the trajectories shown in Figure 3.4. The 2-layer system [Figure 3.5, panels (a) to (c)] has a different dynamics than the larger systems. For this system, there is almost no cost, in terms of U_{SCiSCj}, when the semi-filaments separate at 0.4 ns [coincident with the sharp peak in the RMSD at the beginning of the simulation in Figure 3.5 (c)], which indicates that the nonpolar side chains are poorly buried in the initial conformations of the 2-layer system. After the semi-filaments rearrange, the system manages to find a more favorable conformation in which the nonpolar side chains are better buried, and, thus, both U_{SCiSCj} and $U_{corr}^{(3)}$ decrease.

For the largest systems, 3- and 6-layers [Figure 3.5, panels (d) to (i)], the values of U_{SCiSCj} rise together with the RMSD. This happens because the changes in RMSD are accompanied by the separation of the β strands of each single hairpin, due to weakening hydrophobic interactions in the monomers, exposing the nonpolar residues. On the other hand, $U_{corr}^{(3)}$ is lowered by this conformational change because the $U_{corr}^{(3)}$ term corresponds to the coupling between the dipole moments of two interacting backbone peptide groups and the geometry of the backbone around them[27]. This term stabilizes long extended chains and, since there are no hydrogen bonds between chains of a monomeric hairpin, it tends to favor the opening of the hairpin to try to favor an extended

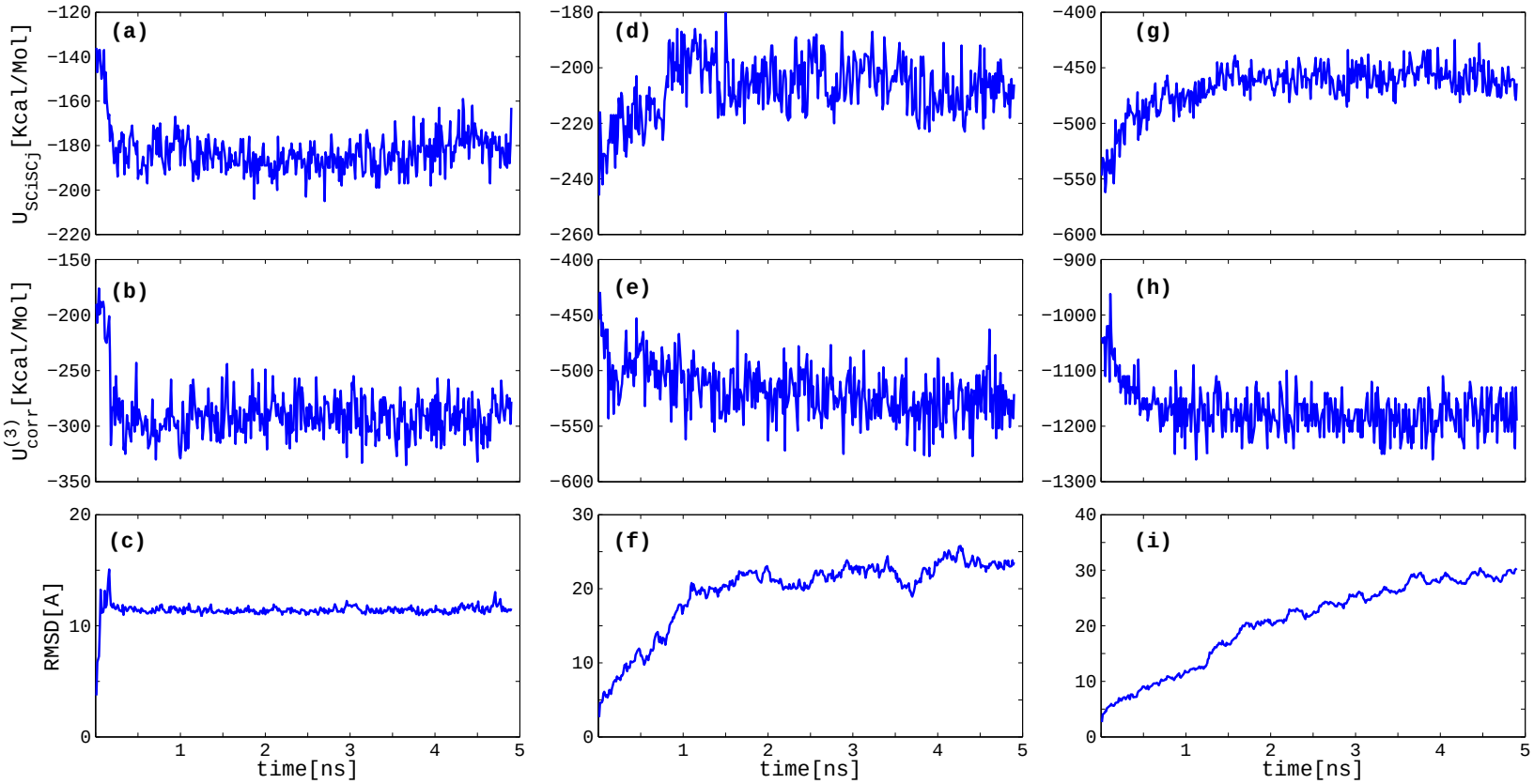

FIGURE 3.5. Behavior of the U_{SCiSCj} interaction energy, the $U^{(3)}_{corr}$ energy, and the RMSD with respect to the initial conformation, for representative trajectories chosen from simulations of 2-layer (a to c), 3-layer (d to f) and 6-layer (g to i) systems.

structure. Therefore, $U_{corr}^{(3)}$ will favor the separation of the β strands and overcome the effect of the hydrophobic interactions.

It is possible that the weight ($w_{corr}^{(3)}$) of $U_{corr}^{(3)}$ has been overestimated during the optimization of the force field[44, 31, 32, 45]. The optimization method[45] was designed to reproduce the sequence of folding events of a training protein. The training protein for this version of the force field is a 28-residue fragment from 1E0L whose native structure is a three-stranded β-sheet. After optimization with this fragment, $w_{corr}^{(3)}$ increased with respect to the weight (w_{SCSC}) of U_{SCiSCj}, which indicates that the force field might favor β-sheets by overemphasizing the role of $U_{corr}^{(3)}$[45].

The explanation for the increase in stability for the larger oligomers can be found in the hydrophobic interactions. As the number of layers in the oligomer increases, the size of the hydrophobic core increases as well, and the nonpolar residues, especially in the center of the structure, are better buried. This becomes evident in Figure 3.6 (a), which shows the average side chain-side chain energy per chain ($\langle U_{SCiSCj}\rangle$), for the different oligomers. As the size of the oligomer increases, the average contribution per chain to U_{SCiSCj} becomes larger. This result is in agreement with that obtained by Fawzi *et al.*[69]. In that work, the authors analyzed the average hydrophobic residue density ($\langle$HpRD$\rangle$), defined as the number of nonpolar residues per unit volume, averaged over all chains (see section 3.8.3). They observed that $\langle$HpRD$\rangle$ increased with the oligomer size, until it reached a plateau around 16 chains[69]. Based on these results, they concluded that this was the minimum size for an oligomer to be stable. After equilibration, the values of $\langle$HpRD$\rangle$ for our initial structures, are shown in Figure 3.6 (b). We obtained the same behavior because our starting conformations were the same as those used by Fawzi*et al.*[69]. $\langle$HpRD$\rangle$ grows with the number of chains, and it levels off at $\approx$16 chains, consistent with the behavior of $\langle U_{SCiSCj}\rangle$ [see Figure 3.6 (a) and (b)].

It should be noted that the behaviors of $\langle U_{SCiSCj}\rangle$ and $\langle$HpRD$\rangle$ do not reflect a cooperative effect. As can be seen in Figure 3.6 (c) and (d), both U_{SCiSCj} energy and the total HpRD change linearly with the number of layers, which shows that adding a layer to a template always contributes with approximately the same interactions, independent of the size of the systems. For example, when calculating the total HpRD of an n-layer system, all the chains contribute with approximately the same value ΔHpRD, except for those in the first layer that contributes with a smaller value, HpRD_1. Then the total HpRD in the n-layer (or $2n$-chain) system will be given by the expression

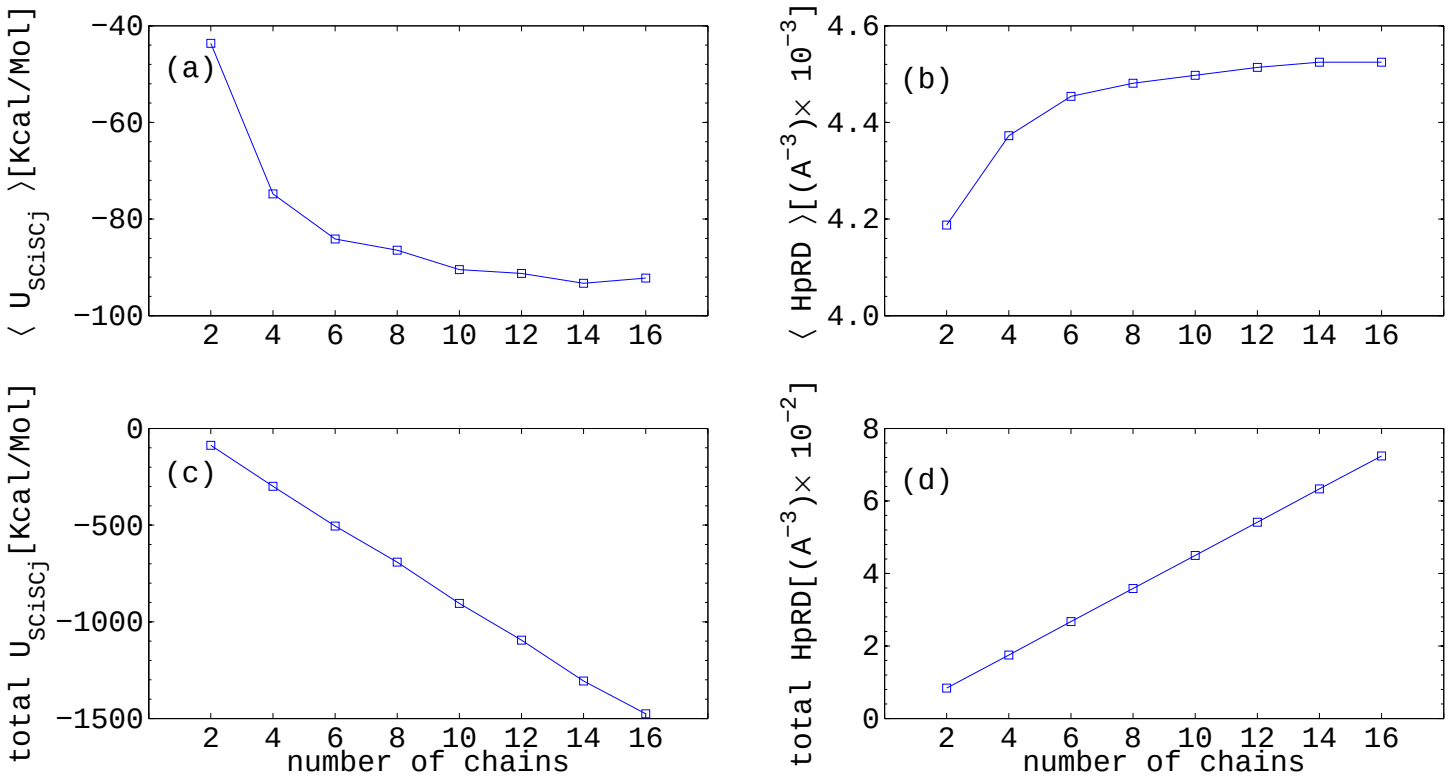

FIGURE 3.6. (a) Average side chain-side chain energy per chain ($\langle U_{SCiSCj} \rangle$) and (b) average hydrophobic residue density per chain ($\langle$HpRD$\rangle$) as a function of the number of chains. (c) Total U_{SCiSCj} energy and (d) total HpRD as a function of the number of chains.

$\mathrm{HpRD}_n = 2\mathrm{HpRD}_1 + 2(n-1)\Delta\mathrm{HpRD}$. It follows that the average HpRD (total HpRD divided by the number of chains) is given by $\langle\mathrm{HpRD}\rangle = \mathrm{HpRD}_n/(2n) = (\mathrm{HpRD}_1 - \Delta\mathrm{HpRD})/n + \Delta\mathrm{HpRD}$. As the size n increases, this number approaches ΔHpRD asymptotically. The first term in this equation represents an edge effect, caused by the first layer not being able to hide the nonpolar residues from the solvent. As the size n increases, the relative size of the edge effect term approaches zero, and the second term, the hydrophobic core, dominates.

3.4.2.2 Hydrogen Bonds

Even when the secondary structure of the monomers is lost, the hydrogen bonds between consecutive layers remain intact. This is expected since the stability of these hydrogen bonds is enhanced by the cooperative nature of the hydrogen bonds along each β-sheet[89], and UNRES is capable of capturing this effect[27]. We conclude that the hydrogen bonds play an important role in stabilizing the structure of the larger Aβ oligomers, although not in the same way as the hydrophobic interactions. The fact that they make the stacking highly stable limits the conformational space available to the peptides in the stack. Being so stable, the hydrogen bonds act as restraints that, by restricting the conformational space of the hydrogen-bonded chains, reduce the conformational entropy of the unfolded state with respect to the folded state. The larger the system, the more limited the conformational space of its unfolded state and therefore, the more stable the system will be.

3.4.2.3 D23-K28 Salt Bridge

Finally, we examined the interactions between the oppositely charged residues D23 and K28, which are buried in the interior of the hydrophobic core in the NMR model, forming a salt bridge that contributes to stabilize the structure. However, the version of UNRES implemented in this study does not favor conformations with residues D and K in close interaction. The interactions between D23 and K28 are slightly repulsive in UNRES, helping to separate the N- and C-terminal strands of the monomers. Although D23-K28 repulsive interactions are not strong enough to destabilize the structure, the absence of an attractive force between them, an interaction that is important in the formation and stability of real $\mathrm{A}\beta_{1-40}$ fibrils[65, 88, 59], hampers the stability of the oligomers.

3.4.2.4 Summary of the Analysis of the Stability of Aβ Fibrillar Conformation

To summarize sections 3.4.2.1 to 3.4.2.3, our data confirm that hydrophobic interactions are very important in stabilizing the structure of the Aβ_{9-40} oligomers. Furthermore, as the size of the oligomer increases, the average contribution per chain of the hydrophobic interactions also increases until it reaches a plateau around 16 chains. Therefore we conclude that the stability of the fibrils is maximized for oligomers of that size. However, that does not mean that smaller oligomers could not be stable as well. Although oligomers with only 16 chains retain their structure with the UNRES force field, we know that the force field was not able to reproduce the salt bridges between residues D23 and K28 that are believed to contribute to the stability of Aβ_{1-40} oligomers[88, 59]. Moreover, as mentioned in section 3.4.2.1, it is possible that the method utilized to obtain the relative weight of the different energy terms in the UNRES force field[44, 32, 31], could have slightly overestimated the weight of the destabilizing $U_{corr}^{(3)}$ term with respect to the stabilizing U_{SCiSCj} term, which contains the hydrophobic interactions. As a consequence of this imbalance, a larger oligomer with a stronger hydrophobic core, will be required to compensate for the the $U_{corr}^{(3)}$ interactions. Finally, the stability of larger oligomers is also enhanced by a network of interlayer hydrogen bonds that, by conserving the stacking of the chains on each semi-filament, restrain the conformational space of the unfolded state, thereby stabilizing the folded state.

3.5 Cooperativity of Hydrogen Bonding in Aβ_{1-40} Fibrillar Conformation

Amyloid fibril formation is known to follow a nucleation dependent mechanism[90, 91, 92]. It has been suggested that cooperativity in the hydrogen bond interactions, along the direction of the fibrils, might contribute to this mechanism[93]. Recently, the crystal structures of several fragments from fibril-forming proteins, also known to form amyloid fibrils, have been resolved[55, 60]. Two independent studies have used quantum mechanical calculations to explore the presence of cooperativity in the hydrogen-bonding interactions for two of these fragments. Tsemekhman *et al.*[93] studied a seven-residue fragment (sequence GNNQQNY) from the yeast prion, Sup35. Their results indicated that hydrogen-bonding interactions are cooperative for the addition of one to three layers, becoming constant for later additions. Plumley and Dannenberg[94] studied the hydrogen-bonding

interactions in the crystals formed by a six-residue fragment (sequence VQIVYK) from the tau protein. That study[94] also showed cooperativity in the hydrogen-bonding interactions. Furthermore, by comparing with hydrogen bonds in β-sheets formed by glutamine (Q) residues, the authors concluded that the formation of hydrogen bonds between the amides in the side chains of the Q residues adds extra strength to the stacking and might explain the observed cooperativity.

A quantum mechanical study of the hydrogen-bonding network of oligomers of the full length Aβ_{1-40} is computationally too expensive to attempt at present. Although the results mentioned above for smaller (six to seven residues) peptides suggest that hydrogen bond cooperativity might also be present in Aβ_{1-40} oligomers, it is not clear whether a much larger peptide, with a more complicated structure and only one glutamine residue involved in the hydrogen-bonding network, will show the same behavior, and if a coarse-grained model will able to reflect it.

We approached the problem with the UNRES force field. To test whether the force field would reflect cooperativity in Aβ_{9-40} oligomers, we calculated the changes in UNRES hydrogen-bonding energy when adding a layer to a preexisting oligomer of n layers. This energy is obtained by computing the difference, $\Delta E_{Hb}(n)$, between the hydrogen-bonding energy of an oligomer with n and $n+1$ layers, $\Delta E_{Hb}(n) = E_{Hb}(n+1) - E_{Hb}(n)$. The values of $\Delta E_{Hb}(n)$ (Figure 3.7) become increasingly negative with the addition of the first three layers, and remain almost constant for subsequent additions, indicating that there is cooperativity between the interlayer hydrogen bonds up to about three layers. Our result shows good agreement with the quantum mechanical calculations by Tsemekhman *et al.*[93] for the seven-residue peptide.

3.6 Studying the Mechanism of Monomer Addition in Aβ_{1-40} Fibrils

The polymerization process of Aβ fibril formation[90, 91, 92] can be described as a nucleation-dependent polymerization process[90, 91, 92]. The process can be divided into three stages, an initial slow lag phase, during which a critical nucleus (seed) is formed, followed by a faster growth phase, during which free monomers are incorporated to the seed, and a last phase in which the fibrils reach an equilibrium size, and no additional fibrilar growth occurs[91]. In vitro experiments have estimated that amyloid fibril formation takes days[65], making computer simulations of the assembly of monomers into fibrils prohibited, even with a coarse-grained approach. However, the lag phase

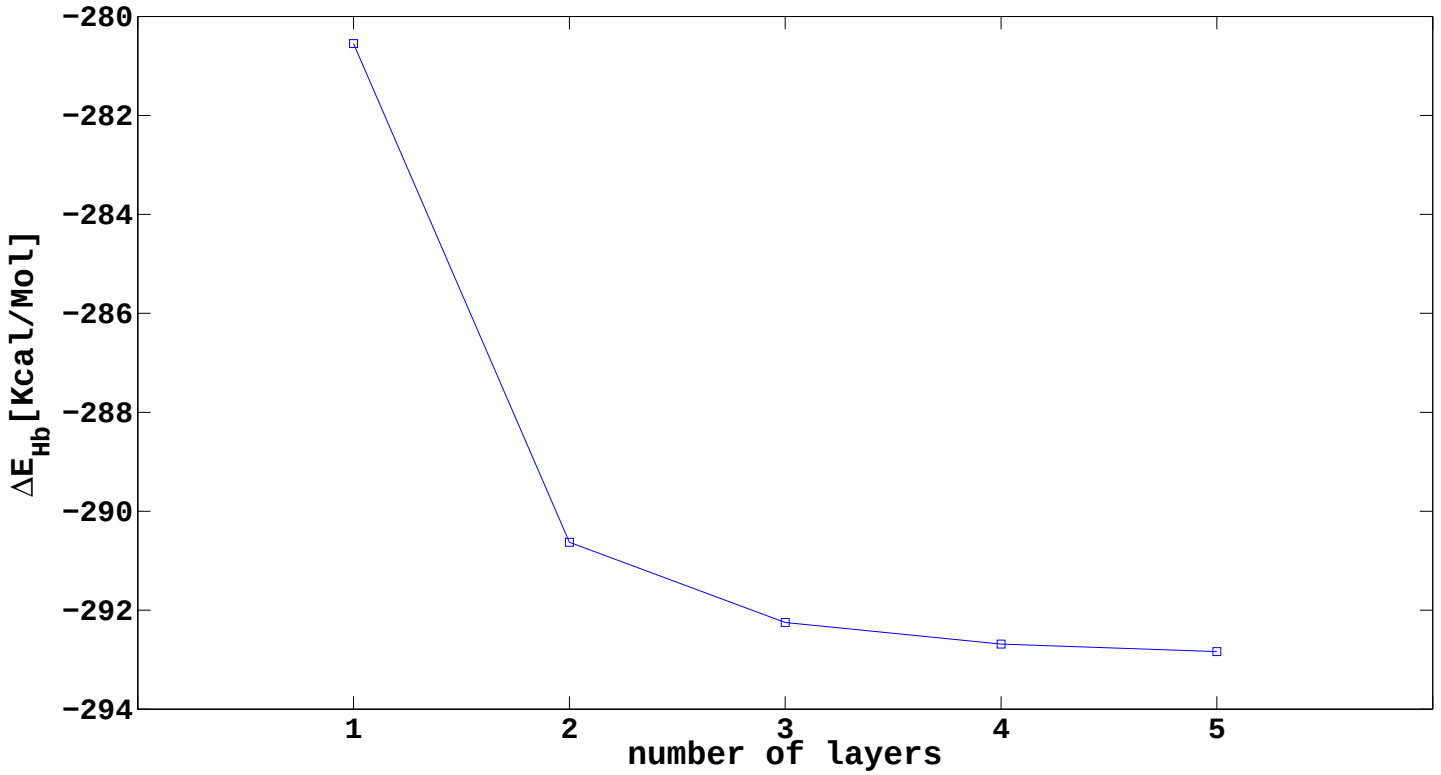

FIGURE 3.7. Difference in UNRES hydrogen-bonding energy, ΔE_{Hb}, when adding a layer to a preexisting oligomer of n layers. The value of $\Delta E_{Hb}(n)$ is obtained by computing the difference between the hydrogen-bonding energy of an oligomer with n and $n+1$ layers. $\Delta E_{Hb}(n) = E_{Hb}(n+1) - E_{Hb}(n)$

can be bypassed if a preformed seed is introduced[91, 65]. There is evidence suggesting that fibrils grow by the addition of one monomer at a time[95], and that the monomers adopt the conformation of the seed, propagating its structure[96]. Based on this information, we focused our studies on the process of addition of one monomer at a time onto a fibril.

It has been proposed that the addition of monomers into Aβ_{1-40} fibrils follows a two-state, "dock-lock" mechanism[97, 98]. In the initial stage, the monomer is docked onto the fibrils, but it can easily dissociate; in the second stage, the monomer is locked into the fibril, *i.e.,* it will rarely dissociate. Studies of deposition of Aβ_{1-40} monomers onto AD brain tissue and synthetic amyloid fibrils[97] identified the transition between docked and locked states as the rate-limiting step. Results from a more recent experiment[98] further revealed a more complex mechanism with two different locked states, the latest having an even slower dissociation rate. *i.e.*, both locked states are very stable, but the final state has the highest stability. Although a mechanism has been proposed, it has not yet been possible to obtain a detailed description of the conformations populating the assembly states. Computer simulations could aid to fill the gaps and provide a more detailed understanding of the process of Aβ_{1-40} elongation.

We studied fibril elongation with the UNRES force field, using the structural model of Petkova *et al.*[59] as a fibril template. Since simulating fibrils of real size would be extremely costly, even with a coarse-grained model, we used templates of 4, 6 and 7 chains (*i.e.* 2, 3 and $3+\frac{1}{2}$ layers). From our studies of the stability of oligomers (section 3.4.2), we knew that template structures of these sizes were not stable with UNRES. Larger templates (16 chains or more) were stable, but it would have been computationally too expensive to use such systems for the simulation of monomer addition. This problem was surmounted by adding a term to the potential energy that stabilized the fibrilar conformation (see section 3.8.1 for details about this energy term), making the smaller templates stable as well. This energy term was applied to the chains of the fibril template, but not to the free monomer. Figure 3.8 shows the initial conformation used in the simulations. The monomer is positioned in the extended conformation at a 20 Å distance from the surface of the fibril template. Preliminary simulations (data not shown) had shown that the monomer can easily become trapped in conformations with a number of energetically favorable contacts, that, although not as stable as the fibrilar conformation (referred to as native here), take a long time to dissociate. To help overcome

these situations, with minimum intrusion, we used replica exchange molecular dynamics (REMD) with a short range of temperatures, between 280 and 320 K (see section 3.8.2 for details about the implementation).

3.6.1 Four-Chain Fibril: Potential of Mean Force (PMF) of Fibril Elongation

The data obtained with the smallest system (4-chain template) were used to calculate the PMF corresponding to monomer addition. For this system, we also carried out additional simulations with the monomer placed in two different initial conformations (one on each end of the template). In this section, we describe the PMF obtained, and in section 3.6.2 we compare these results with the conformations from representative trajectories.

Since each β strand does not lie exactly in a plane in the fibril conformation (see Figure 3.8), the N-terminal strands are more exposed at one of the ends of the fibril (the bottom end in Figure 3.8) relative to the other. Because of this asymmetry, it has been suggested that Aβ fibrils might grow in a unidirectional fashion[69, 67]. We follow the terminology adopted by Takeda and Klimov[67] and refer to the exposed N-terminus as the concave (CV) end, and to the exposed C-terminus as the convex (CX) end. To test whether UNRES would reflect a preferred direction of growth, we carried out two sets of REMD simulations differing only in the initial position of the monomer, *i.e.*, facing either the CV or CX end of the fibril. For each set, we simulated 120, 20 ns long, REMD trajectories. Since we expect that, in solution, the monomer will be free to interact with both fibril ends, we combined the snapshots from the two types of simulations to calculate the PMF at 300 K.

Although the free monomer is more likely to bind to the initially closest fibril end, it is actually free to interact with any of the chains in the template. In fact, for some of the trajectories, the monomer circumvents the template and finally binds on the opposite end. Therefore, there were four possible conformations that the monomer could adopt to propagate the fibril successfully (since there are two exposed chains on each end of the fibril). We needed a reaction coordinate that will not discriminate between these four conformations. Since we observed from the simulations that addition of the monomers is accompanied by formation of hydrogen bonds between the monomer and the fibril, we chose to analyze the data in terms of the number of hydrogen bonds. For each conformation, we computed the number of native and nonnative hydrogen bonds between the monomer and any of the

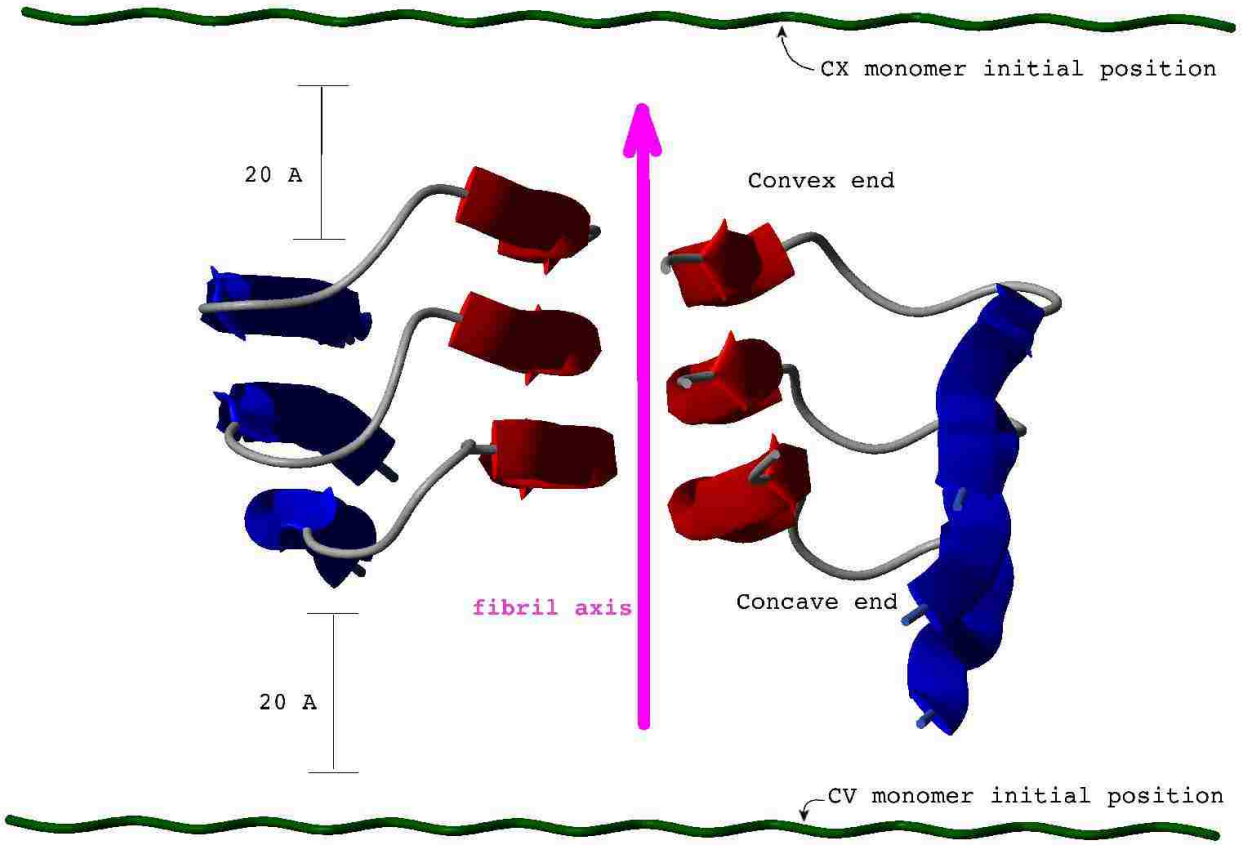

FIGURE 3.8. Representation of the structure of an Aβ_{1-40} fibril. A magenta arrow indicates the direction of the fibril axis. Only three planes along the axis are shown. Due to the asymmetry of the structure on the convex (CX) end, the C-terminal strands (red) are more exposed than the N-terminal strands (blue). The two different initial positions (at the CV and CX ends) of the free monomer (dark green) are shown. In both initial conformations, the monomer is extended, and it is positioned at 20 Å from the closest fibril end.

.

chains on the fibril. A hydrogen bond between peptide groups with indices i and j was considered native if $|i - j| \leq 3$.

A two-dimensional PMF was constructed as a function of the number of native (NHB) and nonnative (nNHB) hydrogen bonds (Figure 3.9). Several basins can be identified. A small basin at NHB = nNHB = 0, contains those conformations in which the monomer has not docked onto the fibril (snapshot A). The basin at NHB $\leq$ 2, and 5 $<$ nNHB $\leq$ 16, corresponds to conformations for which the monomer has bound as an antiparallel β-sheet (snapshots B, C and D). Two more basins can be seen at NHB = 11, and nNHB = 2 (snapshot F) and nNHB = 7 (snapshots G and H), populated by conformations in which the monomer has formed native hydrogen bonds along either the N- or C- terminal strand, but still the other end of the peptide has made few nonnative hydrogen bonds. Finally, a smaller basin can be seen in Figure 3.9 at 25 $\leq$ NHB $\leq$ 27, and nNHB = 0. This latest basin corresponds to native like conformations (snapshot E). Another important remark is the presence of a scarcely populated region at 20 $\leq$ NHB $\leq$ 25 separating the native basin from the rest of the regions, indicating a free-energy barrier.

3.6.2 Four-Chain Fibril: Description of Pathways

To understand the mechanism of monomer addition better, we visually examined the folding pathways of the simulated trajectories. We notice the same pattern in the binding mechanism at both the CV and CX ends of the template. Snapshots from two trajectories leading to successful monomer addition, are shown in Figures 3.10 (starting at the CV end) and 3.11 (starting at the CX end). In Figure 3.10, the first snapshot ($t = 0.76$ ns) shows the monomer before docking onto the template. As expected from our simulations of Aβ monomers, at this point the monomer adopts conformations with significant α-helical content. At $t = 2.62$ ns, the monomer has bound to the template with the wrong (antiparallel) orientation. At $t = 4.77$ ns, the monomer is free again. At $t = 6.89$ ns, it attempts to bind again in a nonnative conformation. Further reorientation leads to the conformation shown at $t = 13.01$ ns, with several native hydrogen bonds along the C-terminal strand. Finally, the N-terminal strand follows and also makes native hydrogen bonds, locking the monomer in the fibrilar conformation ($t = 20$ ns snapshot). The trajectory in Figure 3.11 shows a similar mechanism. Initially the monomer attempts to form nonnative conformations ($t = 0.27$ ns and $t = 1.45$

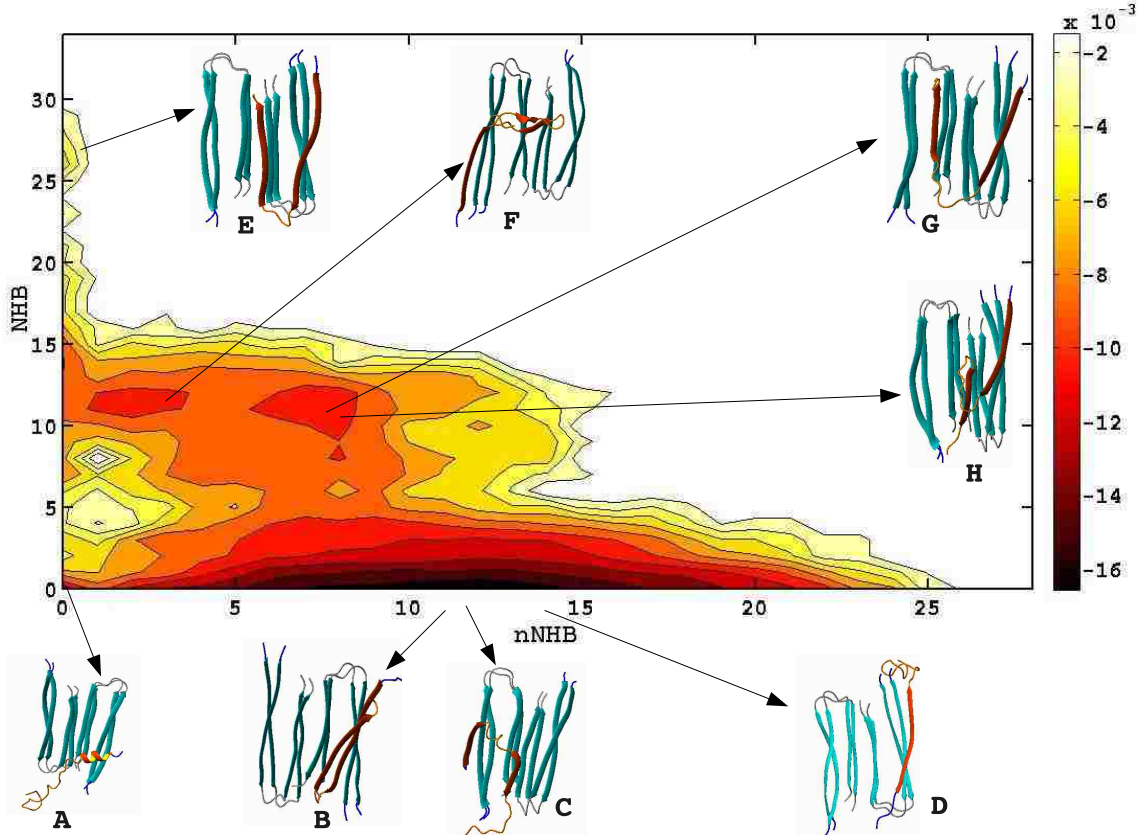

FIGURE 3.9. A two-dimensional PMF as a function of the number of native and nonnative hydrogen bonds (NHB and nNHB), obtained from REMD simulations of a free monomer interacting with a 4-chain fibril template. Snapshots of representative conformations in each basin are shown. The chains in the template are colored in light blue, while the free monomer is colored in orange to red. The small basin at NHB = nNHB = 0 corresponds to unbound conformations (snapshot A). A long basin at NHB $\leq$ 2, and 5 $<$ nNHB $\leq$ 16, corresponds to conformations with nonnative (antiparallel or off-registry for more than 3 amino acids) binding (snapshots B, C and D). Two basins at NHB = 11, correspond to conformations with one strand locked in the fibrilar conformation (snapshots F, G and H). The small basin at 25 $\leq$ NHB $\leq$ 27, and nNHB = 0 is the native basin (snapshot E).

.

ns) that are later disrupted ($t = 3.76$ ns). Native binding starts with the assembly of its N-terminus ($t = 16.75$ ns) and later propagates towards its C-terminus ($t = 20$ ns).

We now look closely at the hydrogen bonds formed between the monomer and the template, along the folding trajectories shown in Figures 3.10 and 3.11. Figures 3.12 (a) and 3.12 (b) show the number of native (NHB) and nonnative (nNHB) hydrogen bonds as a function of time for the trajectories shown in Figures 3.10 and 3.11, respectively. For both trajectories, we can distinguish the three stages in the dock-lock mechanism. During the first (docking) stage, very few native hydrogen bonds are formed. The conformations adopted during this first stage are not very stable and, the monomer binds and unbinds several times (reflected in NHB and nNHB rising and going back to zero several times). In the second stage [starting at $\approx$ 10 ns in Figure 3.12 (a) and $\approx$ 6.5 ns in Figure 3.12 (b)], which corresponds to the first locking state, the monomer makes several native hydrogen bonds (NHB $\approx$ 10), locking only one of the strands, while the other strand is still free to move. The last stage corresponds to the second locking state [starting at $\approx$ 18 ns in Figure 3.12 (a) and $\approx$ 19 ns in Figure 3.12 (b)]. During this stage, the free strand makes the remaining native hydrogen bonds, and the monomer is fully locked in the fibrilar conformation. Once the monomer is locked into this conformation, it can itself serve as a template for subsequent monomer additions.

This assembly mechanism is consistent with the results obtained from experiments of Aβ monomer deposition[97, 98]. We have identified a docking stage, and more remarkably, the two different locking stages. From our simulations, it becomes evident that the first locking stage is a necessary step that, by locking one of the strands, limits the conformational space available to the free strand and facilitates the assembly of the rest of the peptide. We also noticed that, once the still-free strand makes one or two native hydrogen bonds, these bonds quickly propagate along the rest of the strand. This is shown in Figures 3.12 (a) and 3.12 (b) as the abrupt rise in NHB by the end of the simulation. It is also seen in Figure 3.9 as a scarcely-populated region between the native basin (at $\approx$ 26 NHB) and at the region below 20 NHB. This indicates a cooperative binding, in contrast to the non-cooperative effect of hydrophobic interactions in stacking layers on each other, discussed in section 3.4.2.1. This cooperative binding that we observe between the first and second stages of locking (Figures 3.12, 3.10 and 3.11) has also been observed in simulations of the assembly of Aβ fragments[63].

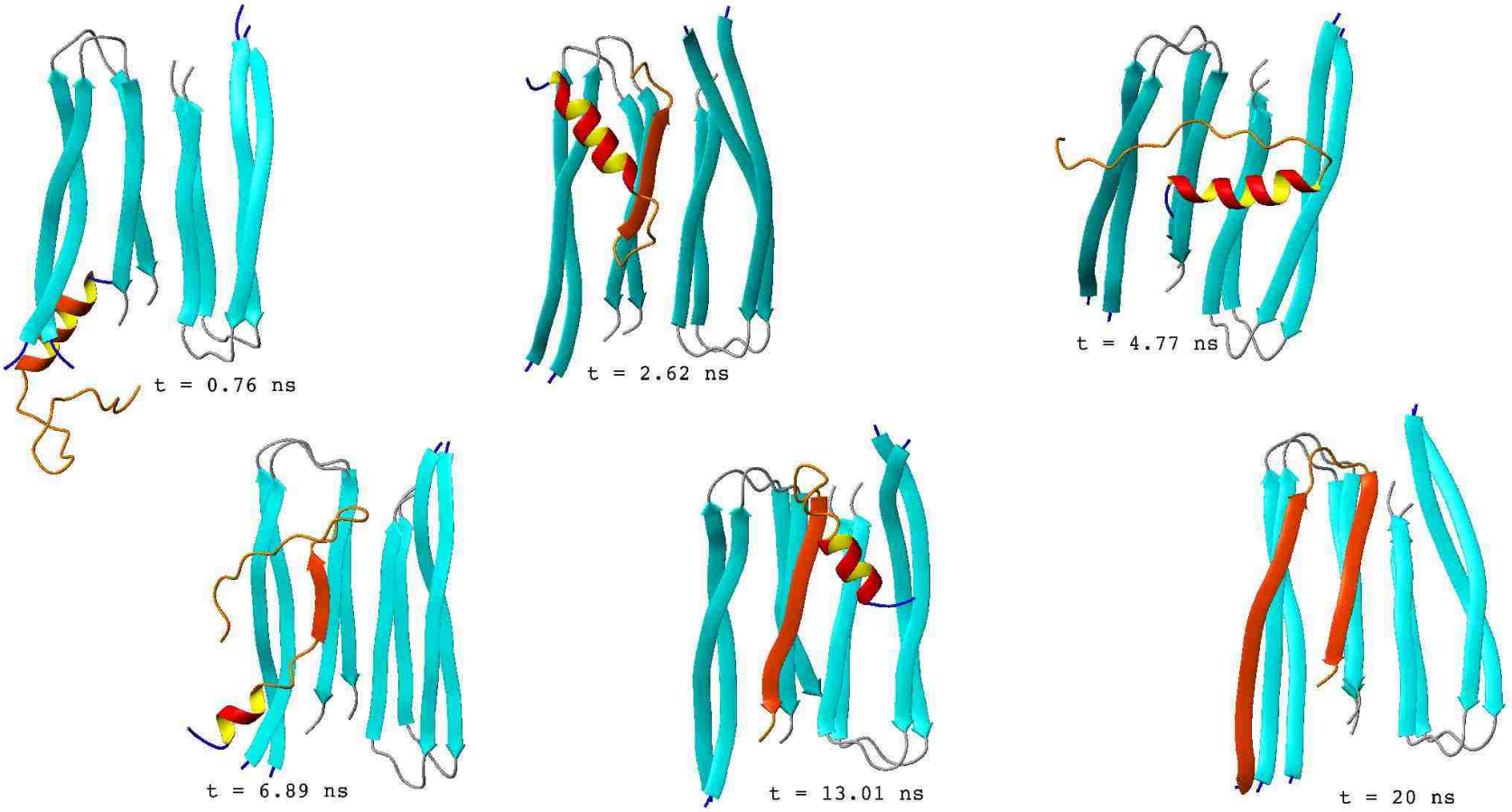

FIGURE 3.10. Selected snapshots along a representative trajectory of a monomer binding to a 4-chain fibril are shown. The monomer is initially placed in the extended conformation, at 20 Å from the CV end of the template. The snapshot at $t = 0.76$ ns shows the monomer before docking onto the fibril in a conformation with significant α-helical content. At $t = 2.62$, the monomer binds forming an antiparallel β-strand along the C-terminus, while the N-teminus forms an α-helix. At $t = 4.77$ ns, the monomer is free from the template again. At $t = 6.89$ ns, the monomers attempts to bind again, but the conformation is still nonnative. The monomer rearranges its position, and at $t = 13.01$ ns, its C-terminus has bound with the native conformation, with the α-helix along the N-terminus still being present. The α-helix unfolds and the N-terminus also binds with the native conformation, locking the monomer into the fibrilar conformation, as shown in the $t = 20$ ns snapshot.

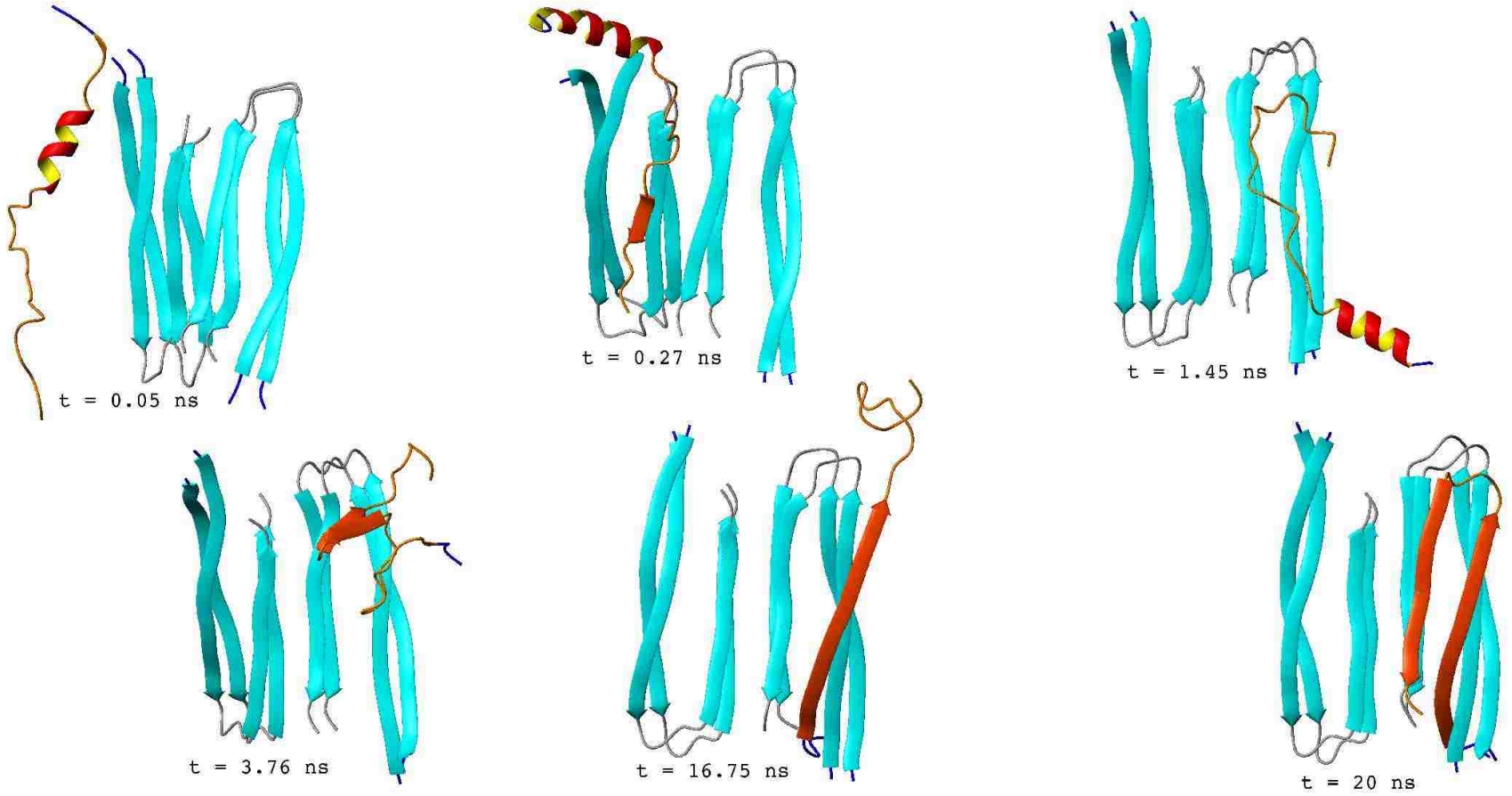

FIGURE 3.11. Same as Figure 3.10, except that the monomer is initially placed in the extended conformation, at a 20 Å from the CX end of the fibril. The snapshot at $t = 0.05$ ns shows the monomer before docking onto the fibril in a conformation with a certain α-helical content. The monomer makes several attempts to bind ($t = 0.27$ ns, $t = 1.45$ ns, and $t = 3.76$ ns), but none of these conformations are native, and the binding is disrupted. Native binding starts with the assembly of the N-terminal strand ($t = 16.75$ ns). The C-terminal strand follows, locking the monomer into the fibrilar conformation as shown in the $t = 20$ ns snapshot.

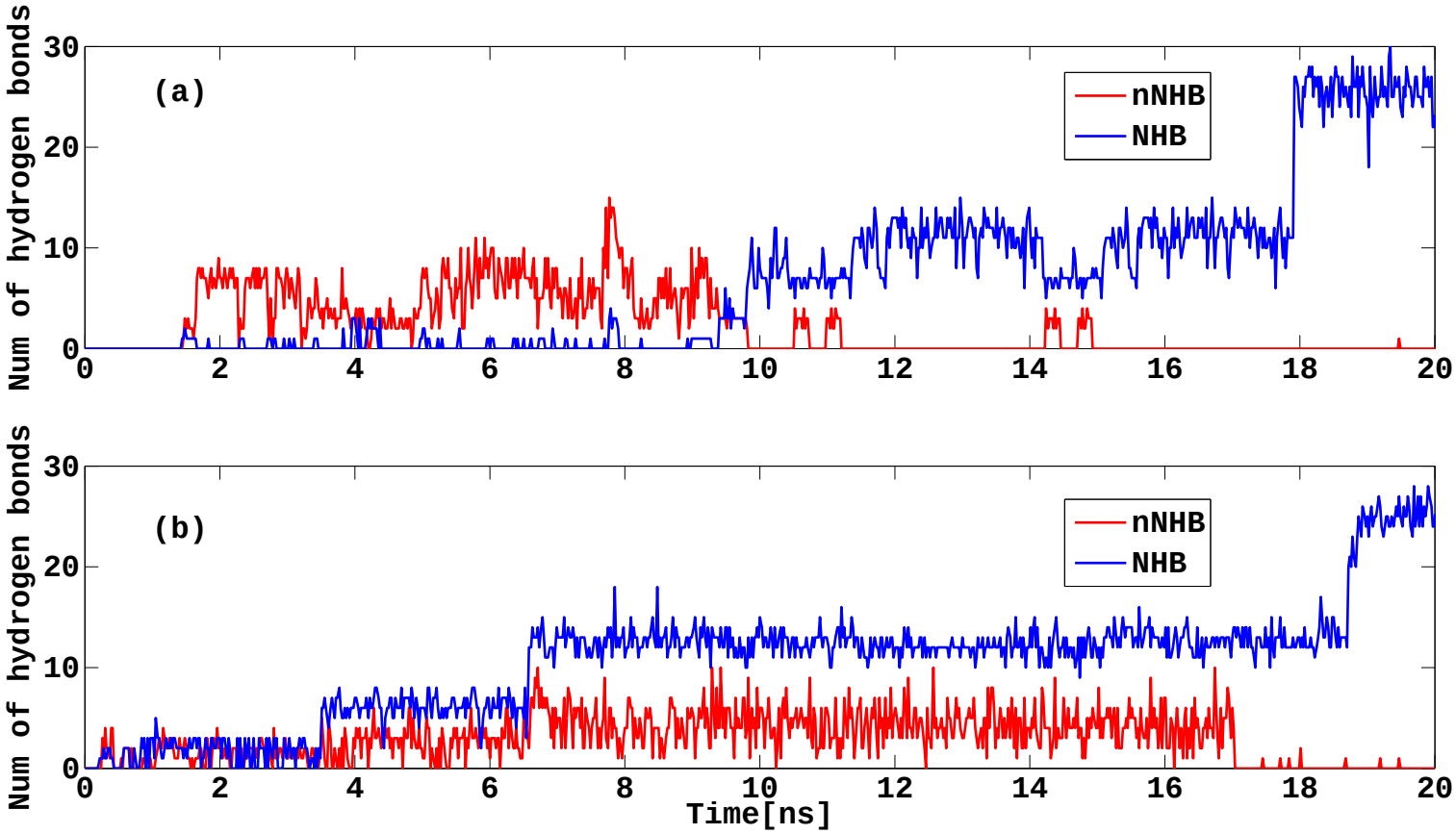

FIGURE 3.12. The number of native and nonnative hydrogen bonds (NHB and nNHB) between monomer and template during a trajectory leading to a full addition starting from the CV end (a) and CX end (b).

In order to describe the ensemble of conformations at the end of our REMD simulations, we adopted the following criteria. If, at the end of the simulation, the monomer has no hydrogen bonds with any of the chains in the template, it is considered undocked. If it has formed less than 10 native hydrogen bonds, it is considered a nonnative addition. If it has formed more than 10, but less than 20 native hydrogen bonds, we consider it a half addition. Finally, if it has formed at least 20 native hydrogen bonds with any of the chains on the fibril, we consider it a full addition. It should be noted that a half addition corresponds to a monomer in the first locking stage, and a full addition to a monomer in the second locking stage. The number of undocked, nonnative, half and full additions are listed in Table 3.1. The data show that the binding can occur at both ends of the fibril (CV or CX). For this system, all the monomers were hydrogen bonded by the end of the simulation, *i.e.*, none was undocked. Of 120 trajectories, we obtained 2 full additions and 14 half additions from the CV end, and 1 full addition and 12 half additions from the CX end. We also noticed that, on several occasions, binding occurred at the opposite end of the fibril, *i.e.*, a monomer, initially facing the CV end, could bind to the CX end, and vice versa. The number of full and half additions and nonnative binding on the opposite end are indicated between parentheses. Although our data show a slightly larger number of half and full additions to the CV end than to the CX end, the numbers are too small to arrive at any conclusions about preferences at the ends. It is, however, important to note that monomers can bind to both ends of the fibril.

3.6.3 Six-Chain and Seven-Chain Fibrils: Description of Pathways

For templates consisting of 6 and 7 chains, 120 REMD simulations, 20 ns long, were carried out. For both systems, the monomer was initially placed at the CX end of the fibril in the extended conformation and 20 Å apart from the end of the fibril. The number of undocked, nonnative, half and full addition are listed in Table 3.1, with the number of full and half additions and nonnative binding on the opposite (CV) end indicated between parentheses.

The mechanism of assembly resembles that observed for the 4-chain templates. For both the 6- and 7-chain templates, only one trajectory led to a full addition. Snapshots from these trajectories are shown in Figures 3.13 (6-chain template) and 3.14 (7-chain template).

In the 6-chain template (Figure 3.13), binding occurs on the opposite end, the CV end. The mechanism of binding is the same as other systems that we have described. Figure 3.13 shows snapshots

TABLE 3.1. Summary of final conformations obtained from 120 REMD simulations

	4-mer+1		6-mer+1	7-mer+1
	From CV end[a]	From CX end[b]	From CX end[c]	From CX end[d]
full additions[e]	2 (0)	1 (0)	1 (1)	1 (0)
half additions[f]	14 (1)	12 (4)	6 (0)	2 (1)
nonnative[g]	104(13)	107(29)	106(24)	91(11)
undocked[h]	0	0	7	26

Number of trajectories that resulted in full additions, half additions, nonnative binding or undocked monomer for the following systems, ^{a} a 4-chain template with the monomer initially positioned facing the CV end, ^{b} a 4-chain template with the monomer initially positioned facing the CX end, ^{c} as in ^{b}, but for a 6-chain template, and ^{d} as in ^{b}, but for a 7-chain template. Trajectories were classified as ^{e} full additions if, by the end of the simulation, the monomer has formed at least 20 native hydrogen bonds with any of the chains on the template, ^{f} half additions if monomer has formed more than 10 but less than 20 native hydrogen bonds, ^{g} nonnative if monomer has formed less that 10 native hydrogen bonds, and ^{h} undocked if monomer has no hydrogen bonds with any of the chains in the template. The number of full and half additions and nonnative binding on the opposite end are indicated between parentheses

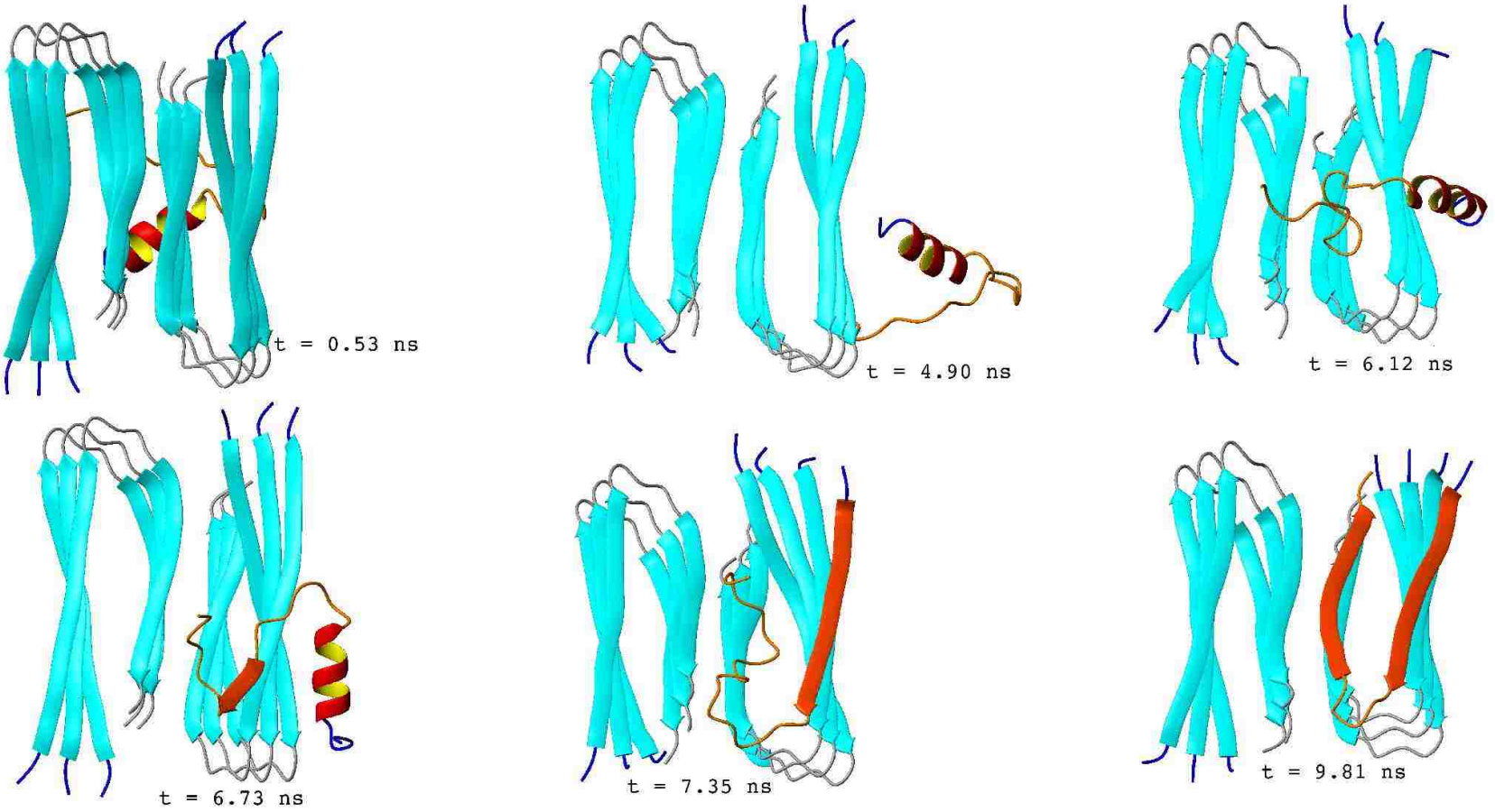

FIGURE 3.13. Selected snapshots along a representative trajectory of a monomer binding to a 6-chain fibril. The template is shown from its CX end. The monomer is initially positioned on the CV end of the fibril. At $t = 0.53$ ns, the monomer still moves free from the template, and it can be seen behind the template, still facing the the CV end. Snapshots at $t = 4.90$ ns and $t = 6.12$ ns show the monomer traveling around the template towards the CX end. At $t = 6.73$ ns, the monomer has bound to the CX end, but in a nonnative conformation. At $t = 7.35$ ns the monomer has locked its N-terminal strand into the fibrilar conformation. At $t = 9.81$ ns, both N- and C-terminal strands are locked into the fibrilar conformation.

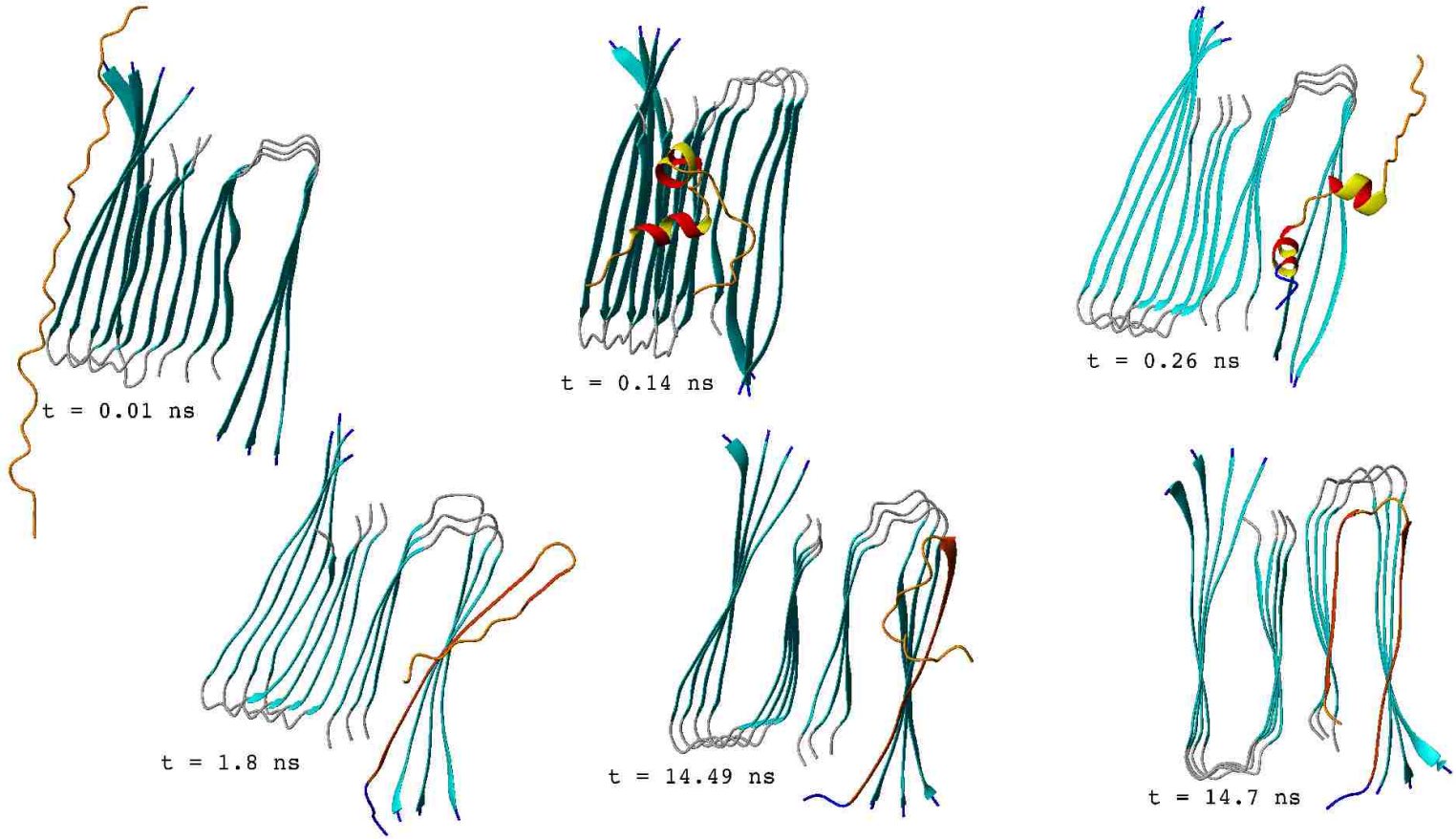

FIGURE 3.14. Selected snapshots along a representative trajectory of a monomer binding to a 7-chain fibril. At $t = 0.26$ ns, the monomer docks with native orientation. At $t = 1.8$ ns, the N-terminal strand is locked into the template. Meanwhile, the C-terminus, which is still free to move, bends and makes a β strand with itself. This conformation is very stable but, at $t = 14.5$ ns, the β strand is finally disrupted. Shortly after that, at $t = 14.7$ ns, the monomer binds with the native conformation.

facing the CV end of the template. The free monomer initially faces the CX end, therefore, in the first snapshot ($t = 0.53$ ns), it is positioned behind the template. The monomer finds its way around the template, and at $t = 6.73$ ns, it binds at the CV end of the template. At $t = 7.35$ ns, the N-terminal strand is locked into the fibrilar conformation and, at $t = 9.81$ ns, the C-terminus follows, and the peptide is fully locked.

For the 7-chain template (Figure 3.14), it can be seen that fibril elongation starts from the native binding of the N-terminal strand, which happens early during the simulation (at $t = 1.8$ ns). In this particular case, native binding of the free strand takes a longer time, and the peptide does not find the native conformation until the 14.7 ns of simulation.

To summarize, the larger systems with 6- and 7-chain templates showed the same dock-lock mechanism as the 4-chain templates. Here too, the two locking states can be distinguished, the first one corresponding to the native binding of one of the strands, and the final locking state corresponding to the native binding of the second strand.

3.7 Conclusions

A coarse-grained model UNRES has been used to study the stability of Aβ_{9-40} oligomers and the process of fibril growth. Using this approach, we succesfully simulated the assembly of free monomers into fibril templates, providing insight into the conformational changes leading to Aβ fibril propagation.

Regarding the stability of oligomers, we found that hydrophobic interactions play an important role in stabilizing their structures, and that these interactions become more important as the size of the oligomer increases, approaching their maximum values at around 16 chains. However, taking into account certain limitations of the force field, we conclude that oligomers smaller than 16 chains might also be stable in the fibrilar conformation.

Our results also showed that the hydrogen bonds, formed between chains in consecutive layers, are extremely stable. These hydrogen bonds act as restraints that, by limiting the conformations that the hydrogen bonded chains can adopt, reduce the conformational entropy of the unfolded state, thereby increasing the stability of the folded state. For larger systems, this effect also becomes

more important because more hydrogen bonded layers will have less energetically favorable states available.

Regarding the hydrogen bonds between consecutive layers, we also studied the increase in their stability when adding a new layer to a preformed oligomer. This was done by computing the differences in the hydrogen-bonding energy between oligomers of different sizes. The results indicate the presence of cooperativity in the interlayer hydrogen bonds when adding one to three layers. For further additions, the energy change becomes constant. The result is in agreement with classical and quantum mechanical calculations with a 7-amino-acid fragment of a fibril-forming peptide from the yeast prion, Sup35[93].

Fibril elongation was studied by allowing a free monomer to interact with a fibril template. The simulations produced trajectories leading to nonnative and native binding (native meaning that the monomer binds, adopting the same conformation as the other chains in the template). By studying those conformations that led to native binding, we observed that they followed a common dock-lock mechanism, and that this mechanism was compatible with that inferred from experiments[97, 98]. During the docking stage, the monomer interacts with the template, often making nonnative hydrogen bonds that later break. The second stage, locking, can be further divided into two consecutive steps. First, the monomer makes native hydrogen bonds along one of the β-strands in the template, and at this point half of the peptide is bound to the template, while the other end can move freely. The final locking step is the native binding of the free end. This final step was highly cooperative, as indicated by the fact that, once one or two native hydrogen bonds are formed between the free end and the template, these hydrogen bonds quickly propagate along the rest of the peptide. This final step locks the monomer into the fibril template. Experiments on monomer deposition[98] have shown the presence of two locking states; however these experiments could not describe the conformations populating these two states. Based on our simulations, we have proposed a description of this mechanism at the molecular level.

3.8 Supplementary Material

3.8.1 Restraining Potential

Fibril elongation was examined by simulating the interaction between a monomer and a fibril template. The fibril template was composed of 4, 6 or 7 chains, and was restrained to the structural

model of Petkova *et al.*[59]. Since systems of such sizes (4 to 7 chains) are not stable with the version of the UNRES force field used here, an additional term, $U_{Restr.}$, was added to the UNRES energy function to restrain the template to the fibrilar conformation. The energy is given by equation 3.1

$$U_{Restr} = w_{Restr} \sum_{l} [Q(l) - 1]^2 \tag{3.1}$$

where the index l runs over all the segments being restrained, w_{Restr} is the weight of the term, set at $5 \times 10^4 Kcal/mol$, and $Q(l)$ is given by equation 3.2

$$Q(l) = \frac{1}{N_{dist_l}} \left[\sum_{i,j} exp \left[-\frac{1}{2} \left(d_{i,j} - d_{i,j}^{nat} \right)^2 \right] \right] \tag{3.2}$$

where $d_{i,j}$ and $d_{i,j}^{nat}$ are the current and native distances between the C^{α} atoms from amino acid i and j, and N_{dist_l} is the total number of distances in segment l. Two types of segments were considered, intrachain and interchain segments. For intrachain segments, the indices i and j run over all the amino acids in the chain, with $i < j$. Interchain segments were considered between adjacent chains (*i.e.*, between chain n and, chain $n+1$ or chain $n+2$). For interchains segments, the indices i and j run over all the amino acids in the corresponding chains.

3.8.2 REMD Simulations

For the simulations of fibril elongation, we used replica exchange molecular dynamics (REMD)[99, 100]. The implementation is described as follows:

- 120 independent canonical MD simulations were started from the same initial conformation but at different temperatures ranging between 280 and 320 K, with intervals of 10 K. Each simulation was run for 20 ns.

- For each trajectory, an exchange of conformations between groups with consecutive temperatures was attempted every 20,000 steps.

- An exchange between conformations $\mathbf{X}_i$ and $\mathbf{X}_j$ is accepted or rejected with a probability Δ given by equation 3.3

$$\Delta = (\beta_j - \beta_i)[U(\mathbf{X}_j) - U(\mathbf{X}_i)] \tag{3.3}$$

where $\beta_i = 1/RT_i$, with R the gas constant and T_i the absolute temperature of the i-th trajectory, and $U(\mathbf{X}_i)$ its UNRES energy. The constant temperature baths were maintained by using the Berendsen thermostat (section 2.2.4).

3.8.3 Hydrophobic Residue Density

The average hydrophobic residue density ($\langle HpRD \rangle$), defined as the number of hydrophobic residues per unit volume, averaged over all chains, was calculated according to equation 3.4

$$\langle HpRD \rangle = \frac{1}{N_{chains}} \sum_{chains} \frac{1}{N_{HpSC}} \sum_{i} N_{HpRes_i} \tag{3.4}$$

where N_{chains} is the number of chains in the system, N_{HpSC} is the number of hydrophobic side chains on each peptide chain, N_{HpRes_i} is the number of hydrophobic residues within a sphere of radius 7.6 Å around side chain i, and the sum is evaluated over all hydrophobic side chains in all chains.

3.8.4 Potential of Mean Forces (PMF)

A two-dimensional PMF was calculated from the 300 K snapshots from the simulations of fibril elongation with four-chain templates. The coordinates were the number of native and nonnative hydrogen bonds (NHB and nNHB). The PMF, $w(NHB, nNHB)$, was computed from equation 3.5

$$w(NHB, nNHB) = -k_B T \, ln(\langle \rho(NHB, nNHB) \rangle) \tag{3.5}$$

where k_B is the Boltzmann constant, T is the temperature (300 K), and $\langle \rho(NHB, nNHB) \rangle$ is the average distribution function calculated as the number of snapshots containing NHB and nNHB divided by the total number of snapshots.

Chapter 4
Conclusions

The main goal of this thesis was to study the folding process of protein systems composed of multiple chains. The first part of the thesis (Chapter 2) focused on building and testing the tools needed to study the folding of multichain proteins from a computational approach. The second part of the thesis (Chapter 3 is the application of these tools to study the Alzheimers Amyloid β (Aβ) fibrils.

Aβ fibrils have been identified as the cause of Alzheimers disease. Understanding the mechanism through which the free harmless monomers assemble to grow the toxic Aβ fibrils can greatly contribute to design drugs that can prevent this process. Experiments have been able to describe some aspect of this transformation[97, 98], but none of these experiments could provide a detailed description of the conformational changes that the monomer undergoes during the process. The simulations presented in chapter 3 describe at the molecular level the process by which free Aβ monomers are incorporated into Aβ fibrils. The simulations show that monomer addition is a two step process. In the first step, half of the monomer binds mimicking the fibrilar structure, while the other half remains free to adopt any conformation. Only after the second half of the fibril also binds with the fibrilar conformation, can fibril propagation continue. This suggests that a peptide that has enough similarity with Aβ to bind to a fibril, but that will not undergo the second conformational change, will be able to stop the elongation process. This result can potentially help to target a peptide that might prevent fibril elongation.

4.1 Further Research

As in any scientific work, the information obtained in this work brings new questions and suggests alternatives to answer those that could not be addressed before. For example, when I first started to study Aβ fibrils, I tried to describe the free energy profile of the monomer while binding onto a fibril, but finding the appropriate parameter to describe the process and a reasonable pathway was not trivial. The results from this thesis not only describe the pathway that the monomer naturally chooses, but also suggest that the hydrogen bonds between the monomer and the fibril will be the most appropriate coordinate to describe the binding. I could now use this information to obtain the

free energy profile of the binding process, and by doing so I can also obtain the free energy barrier between the first and second locking state, which has important implications because the transition between these two states is necessary for fibril propagation to continue.

Another interesting question is the role of the salt bridge between the charged residues on each side of the hairpin structure in the fibrilar conformation (see Figure 3.2 d). It has been shown that linking these two residues speeds up the formation of Aβ fibrils[88, 59], which suggests that this interaction might help the monomer to adopt the fibrilar conformation, but to what extend is a still unanswered question. By carrying out simulations changing the strength of the attraction force between these two residues, one could assess the importance of this salt bridge.

Finally, another question that could be addressed is whether binding is more likely to occur on one of the ends of the fibril (see Figure 3.8). This question could be answered by carrying out restrained simulations that would force the monomer to bind on the CV or the CX end and compute the free energy barrier of binding. The free monomer simulations (Section 3.6) did not produce enough trajectories for which the monomers propagated the fibril, and there were not enough data to address this question. By forcing the monomer to adopt the conformations along the natural pathway (the pathway obtained from the free monomer simulations), I can make sure that enough conformations are sampled and a smooth free energy profile can be obtained for binding on either end.

Bibliography

[1] A. Alzheimer, RA Stelzmann, HN Schnitzlein, and FR Murtagh. An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde". *Clinical anatomy (New York, NY)*, 8:429–431, 1995.

[2] NC Berchtold and CW Cotman. Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. *Neurobiol. Aging*, 19:173–189, 1998.

[3] D.J. Selkoe. Alzheimer's disease: Genes, proteins, and therapy. *Physiol. Rev.*, 81:741–766, 2001.

[4] Colin L. Masters, Gail Simms, Nicola A. Weinman, Gerd Multhaup, Brian L. McDonald, and Konrad Beyreuther. Amyloid plaque core protein in alzheimer disease and down syndrome. *Proc. Natl. Acad. Sci. USA*, 82:4245–4249, 1985.

[5] D.J. Selkoe. The molecular pathology of alzheimers-disease. *Neuron*, 6:487–498, 1991.

[6] A. Lorenzo, M. Yuan, Z. Zhang, P.A. Paganetti, C. Sturchler-Pierrat, M. Staufenbiel, J. Mautino, F.S. Vigo, B. Sommer, and B.A. Yankner. Amyloid β interacts with the amyloid precursor protein: a potential toxic mechanism in Alzheimer's disease. *Nat. Neurosci.*, 3:460–464, 2000.

[7] D.M. Walsh, I. Klyubin, J.V. Fadeeva, W.K Cullen, R. Anwyl, M.S. Wolfe, M.J. Rowan, and D.J. Selkoe DJ. Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. *Nature*, 416:535–539, 2002.

[8] C. Levinthal. ARE THERE PATHWAYS FOR PROTEIN FOLDING ? *J. Chim. Phys.*, 65:44, 1968.

[9] Y. Zhang and J. Skolnick. The protein structure prediction problem could be solved using the current PDB library. *Proc. Natl. Acad. Sci. USA*, 102:1029–1034, 2005.

[10] J.U. Bowie, R. Luethy, and D. Eisenberg. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253:164–170, 1991.

[11] R. Day and V. Daggett. All-atom simulations of protein folding and unfolding. *Adv. Protein Chem.*, 66:373–403, 2003.

[12] A. R. Fersht and V. Daggett. Protein folding and unfolding at atomic resolution. *Cell*, 108:573–582, 2002.

[13] J. Kubelka, J. Hofrichter, and W. A. Eaton. The protein folding 'speed limit'. *Curr. Opinion Struct. Biol.*, 14:76–88, 2004.

[14] J.-E. Shea and C. L. Brooks III. From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.*, 52:499–535, 2001.

[15] M. Vieth, A. Kolinski, C. L. Brooks, and J. Skolnick. Prediction of the folding pathways and structure of the gcn4 leucine zipper. *J. Mol. Biol.*, 237:361–367, 1994.

[16] B. Ma and R. Nussinov. Stabilities and conformations of Alzheimer's-amyloid peptide oligomers (Aβ 16 22, Aβ 16–35, and Aβ 10 35): Sequence effects. *PNAS*, 99:14126–14131, 2002.

[17] Y. Levy, A. Caflisch, J.N. Onuchic, and P.G. Wolynes. The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simu lations. *J. Mol. Biol.*, 340(1):67–79, 2004.

[18] S. Yang, S.S. Cho, Y. Levy, M.S. Cheung, H. Levine, P.G. Wolynes, and J.N. Onuchic. Domain swapping is a consequence of minimal frustration. *Proc. Natl. Acad. Sci. U.S.A.*, 101(38):13786–13791, 2004.

[19] Y. Levy, G.A. Papoian, J.N. Onuchic, and P.G. Wolynes. Energy landscape analysis of protein dimers. *Isr. J. Chem.*, 44(1-3):281–297, 2004.

[20] S. Yang, H. Levine, J.N. Onuchic, and D.L. Cox. Structure of infectious prions: stabilization by domain swapping. *FASEB J.*, 19(13):1778–1782, 2005.

[21] M. Khalili, A. Liwo, F. Rakowski, P. Grochowski, and H. A. Scheraga. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. *J. Phys. Chem. B*, 109:13785–13797, 2005.

[22] M. Khalili, A. Liwo, A. Jagielska, and H. A. Scheraga. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model α-helical systems. *J. Phys. Chem. B*, 109:13798–13810, 2005.

[23] A. Liwo, M. Khalili, and H. A. Scheraga. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains; test of the approach on model proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 102:2362–2367, 2005.

[24] M. Khalili, A. Liwo, and H. A. Scheraga. Kinetic studies of folding of the B-domain of staphylococcal protein A with molecular dynamics and a united-residue (UNRES) model of polypeptide chains. *J. Mol. Biol.*, 355:536–547, 2006.

[25] A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.*, 18:849–873, 1997.

[26] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, S. Ołdziej, and H. A. Scheraga. A united-residue force field for off-lattice protein-structure simulations. ii: Parameterization of local interactions and determination of the weights of energy terms by z-score optimization. *J. Comput. Chem.*, 18:874–887, 1997.

[27] A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.*, 115:2323–2347, 2001.

[28] A. Liwo, P. Arłukowicz, C. Czaplewski, S. Ołdziej, J. Pillardy, and H.A. Scheraga. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: Application to the UNRES force field. *Proc. Natl. Acad. Sci. U.S.A.*, 99:1937–1942, 2002.

[29] S. Ołdziej, U. Kozłowska, A. Liwo, and H. A. Scheraga. Determination of the potentials of mean force for rotation about $c^{\alpha} \cdots c^{\alpha}$ virtual bonds in polypeptides from the *ab initio* energy surfaces of terminally-blocked glycine, alanine, and proline. *J. Phys. Chem. A*, 107:8035–8046, 2003.

[30] A. Liwo, S. Ołdziej, C. Czaplewski, U. Kozłowska, and H. A. Scheraga. Parameterization of backbone-electrostatic and multibody contributions to the UNRES force field for protein-structure prediction from *ab initio* energy surfaces of model systems. *J. Phys. Chem. B*, 108:9421–9438, 2004.

[31] S. Ołdziej, A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga. Optimization of the UNRES force field by hierarchical design of the potential-energy landscape: II. Off-lattice tests of the method with single proteins. *J. Phys. Chem. B*, 108:16934–16949, 2004.

[32] S. Ołdziej, J. Łągiewka, A. Liwo, C. Czaplewski, M. Chinchio, M. Nanias, and H. A. Scheraga. Optimization of the UNRES force field by hierarchical design of the potential-energy landscape: III. Use of many proteins in optimization. *J. Phys. Chem. B*, 108:16950–16959, 2004.

[33] C. Czaplewski, S. Ołdziej, A. Liwo, and H. A. Scheraga. Prediction of the structures of proteins with the UNRES force field, including dynamic formation and breaking of disulfide bonds. *Protein Eng. Des. Sel.*, 17:29–36, 2004.

[34] J. A. Saunders and H. A. Scheraga. Ab initio structure prediction of two α-helical oligomers with a multiple-chain united-residue force field and global search. *Biopolymers*, 68:300–317, 2003.

[35] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[36] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga. Prediction of protein conformation on the basis of a search for compact structures; test on avian pancreatic polypeptide. *Protein Sci.*, 2:1715–1731, 1993.

[37] T. Veitshans, D. Klimov, and D. Thirumalai. Protein kinetics: timescales, pathways, and energy landscapes in terms of sequence-dependent properties. *Fold. Des.*, 2:1–22, 1996.

[38] M. Cieplak, T. X. Hoang, and M. O. Robbins. Thermal folding and mechanical unfolding pathways of protein secondary structures. *Proteins: Struct., Funct., Genet.*, 49:104–113, 2002.

[39] P.-G. de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, Ithaca, 1979. Chapter VI.

[40] L. Verlet. Computer simulations of classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[41] S. Nosé. A molecular dynamics method for simulation in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.

[42] W. G. Hoover. Canonical dynamics: Equilibrum phase-space distribution. *Phys. Rev. A*, 31:1695–1697, 1985.

[43] H. C. Andersen. Molecular dynamics at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[44] A. Liwo, P. Arłukowicz, S. Ołdziej, C. Czaplewski, M. Makowski, and H. A. Scheraga. Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. I: Tests of the approach using simple lattice protein models. *J. Phys. Chem. B*, 108:16918–16933, 2004.

[45] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Oldziej, K. Wachucik, and H.A. Scheraga. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B*, 111:260–285, 2007.

[46] EK O'shea, JD Klemm, PS Kim, and T. Alber. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, 254(5031):539–544, 1991.

[47] P.R.E. Mittl, C. Deillon, D. Sargent, N. Liu, S. Klauser, R.M. Thomas, B. Gutte, M.G. Gr, et al. The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure. *Proc. Natl. Acad. Sci. U.S.A.*, 97(6):2562–2566, 2000.

[48] N.L. Ogihara, G. Ghirlanda, J.W. Bryson, M. Gingery, W.F. DeGrado, and D. Eisenberg. Design of three-dimensional domain-swapped dimers and fibrous oligomers. *Proc. Natl. Acad. Sci. U.S.A*, 98(4):1404–1409, 2001.

[49] PB Harbury, T. Zhang, PS Kim, and T. Alber. A switch between two-, three- and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science*, 262(5138):1401–1407, 1993.

[50] J. Lee, D. R. Ripoll, C. Czaplewski, J. Pillardy, W. J. Wedemeyer, and H. A. Scheraga. Optimization of parameters in macromolecular potential energy functions by conformational space annealing. *J. Phys. Chem. B*, 105:7291–7298, 2001.

[51] J. A. Saunders and H. A. Scheraga. Challenges in structure prediction of oligomeric proteins at the united-residue level: searching the multiple-chain energy landscape with csa and cfmc procedures. *Biopolymers*, 68:318–332, 2003.

[52] J. Lee, H. A. Scheraga, and S. Rackovsky. New optimization method for conformational energy calculations on polypeptides: Conformational space annealing. *J. Comput. Chem.*, 18:1222–1232, 1997.

[53] J. Lee, A. Liwo, and H. A. Scheraga. Energy-based *de novo* protein folding by conformational space annealing and an off-lattice united-residue force field: Application to the 10-55 fragment of staphylococcal protein a and to *apo*-calbindin d9k. *Proc. Natl. Acad. Sci. U. S. A.*, 96:2025–2030, 1999.

[54] B.A. Yankner and T. Lu. Amyloid β-protein toxicity and the pathogenesis of Alzheimer disease. *J. Biol. Chem.*, 284:4755, 2009.

[55] R. Nelson, M.R. Sawaya, M. Balbirnie, A.O. Madsen, C. Riekel, R. Grothe, and D. Eisenberg. Structure of the cross-β spine of amyloid-like fibrils. *Nature*, 435:773–778, 2005.

[56] C. Ritter, M.L. Maddelein, A.B. Siemer, T. Luhrs, M. Ernst, B.H. Meier, S.J. Saupe, and R. Riek. Correlation of structural elements and infectivity of the het-s prion. *Nature*, 435:844–848, 2005.

[57] O. S. Makin, E. Atkins, P. Sikorski, J. Johansson, and L. C. Serpell. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. USA*, 102:315–320, 2005.

[58] Robert Tycko. Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q. Rev. Biophys.*, 39:1–55, 2006.

[59] Aneta T. Petkova, Wai-Ming Yau, and Robert Tycko. Experimental Constraints on Quaternary Structure in Alzheimer's-Amyloid Fibrils. *Biochemistry*, 45:498–512, 2006.

[60] M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J.W. Wiltzius, H.T. McFarlane, A. Ø. Madsen, C. Riekel, and D. Eisenberg. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature*, 447:453–457, 2007.

[61] A. K. Paravastu, R. D. Leapman, W. M. Yau, and R. Tycko. Molecular structural basis for polymorphism in Alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci. USA*, 105:18349–18354, 2008.

[62] S. Chimon, M. A. Shaibat, C. R. Jones, D. C. Calero, B. Aizezi, and Y. Ishii. Evidence of fibril-like β-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's β-amyloid. *Nat. Struct. Mol. Biol.*, 14:1157–1164, 2007.

[63] Govardhan Reddy, John E. Straub, and D. Thirumalai. Dynamics of locking of peptides onto growing amyloid fibrils. *Proc. Natl. Acad. Sci. USA*, 106:11948–11953, July 2009.

[64] T. Ban, M. Hoshino, S. Takahashi, D. Hamada, K. Hasegawa, H. Naiki, and Y. Goto. Direct Observation of A β Amyloid Fibril Growth and Inhibition. *J. Mol. Biol.*, 344:757–767, 2004.

[65] K.L. Sciarretta, D.J. Gordon, A.T. Petkova, R. Tycko, S.C. Meredith, et al. Aβ40-Lactam(D23/K28) Models a Conformation Highly Favorable for Nucleation of Amyloid. *Biochemistry*, 44:6003–6014, 2005.

[66] E.P. O'Brien, Y. Okamoto, J.E. Straub, B.R. Brooks, and D. Thirumalai. Thermodynamic Perspective on the Dock-Lock Growth Mechanism of Amyloid Fibrils. *J. Phys. Chem. B*, 113:14421–14430, 2009.

[67] T. Takeda and D.K. Klimov. Replica exchange simulations of the thermodynamics of Aβ fibril growth. *Biophys. J.*, 96:442–452, 2009.

[68] E. Małolepsza, M. Boniecki, A. Kolinski, and L. Piela. Theoretical model of prion propagation: a misfolded protein induces misfolding. *Proc. Natl. Acad. Sci. USA*, 102(22):7835, 2005.

[69] N.L. Fawzi, Y. Okabe, E.H. Yap, and T. Head-Gordon. Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer's Aβ1-40 Peptide. *J. Mol. Biol.*, 365:535–550, 2007.

[70] S. Auer, C.M. Dobson, and M. Vendruscolo. Characterization of the nucleation barriers for protein aggregation and amyloid formation. *HFSP J.*, 1:137–146, 2007.

[71] A.V. Rojas, A. Liwo, and H.A. Scheraga. Molecular dynamics with the united-residue force field: Ab initio folding simulations of multichain proteins. *J. Phys. Chem. B*, 111:293–309, 2007.

[72] J. P. Lee, E. R. Stimson, J. R. Ghilardi, P. W. Mantyh, Y. A. Lu, A. M. Felix, W. Llanos, A. Behbin, and M. Cummings. 1H NMR of Aβ Amyloid Peptide Congeners in Water Solution. Conformational Changes Correlate with Plaque Competence. *Biochemistry*, 34:5191–5200, 1995.

[73] N. D. Lazo, M. A. Grant, M. C. Condron, A. C. Rigby, and D. B. Teplow. On the nucleation of amyloid β-protein monomer folding. *Protein Sci.*, 14:1581–1596, 2005.

[74] C. J. Barrow, A. Yasuda, P. T. M. Kenny, and M. G. Zagorski. Solution conformations and aggregational properties of synthetic amyloid β-peptides of Alzheimer's disease: analysis of circular dichroism spectra. *J. Mol. Biol.*, 225:1075–1093, 1992.

[75] H. Sticht, P. Bayer, D. Willbold, S. Dames, C. Hilbich, K. Beyreuther, R.W. Frank, and P. Rösch. Structure of amyloid A4-(1-40)-peptide of Alzheimer's disease. *Eur. J. Biochem.*, 233:293–298, 1995.

[76] O. Crescenzi, S. Tomaselli, R. Guerrini, S. Salvadori, A. M. D'Ursi, P. A. Temussi, and D. Picone. Solution structure of the Alzheimer amyloid β-peptide (1-42) in an apolar microenvironment - Similarity with a virus fusion domain. *Eur. J. Biochem.*, 269:5642–5648, 2002.

[77] S. Tomaselli, V. Esposito, P. Vangone, N.A.J. van Nuland, A.M.J.J. Bonvin, R. Guerrini, T. Tancredi, P.A. Temussi, and D. Picone. The α-to-β conformational transition of Alzheimer's Aβ-(1-42) peptide in aqueous media is reversible: A step by step conformational analysis suggests the location of β conformation seeding. *ChemBioChem*, 7:257–267, 2006.

[78] M. Coles, W. Bicknell, A. A. Watson, D. P. Fairlie, and D. J. Craik. Solution structure of amyloid β-peptide (1-40) in a water-micelle environment. Is the membrane-spanning domain where we think it is? *Biochemistry*, 37:11064–11077, 1998.

[79] H. Shao, S. Jao, K. Ma, and M.G. Zagorski. Solution structures of micelle-bound amyloid β-(1-40) and β-(1-42) peptides of Alzheimer's disease. *J. Mol. Biol.*, 285:755–773, 1999.

[80] M. Yang and D.B. Teplow. Amyloid β-protein monomer folding: free-energy surfaces reveal alloform-specific differences. *J. Mol. Biol.*, 384:450–464, 2008.

[81] P. Anand, FS Nandel, and U.H.E. Hansmann. The Alzheimer's β amyloid (Aβ_{1-39}) monomer in an implicit solvent. *J. Chem. Phys.*, 128:165102, 2008.

[82] M. D. Kirkitadze, M. M. Condron, and D. B. Teplow. Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis. *J. Mol. Biol.*, 312:1103–1119, 2001.

[83] Y. Fezoui and D. B. Teplow. Kinetic studies of amyloid β-protein fibril assembly. Differential effects of α-helix stabilization. *J. Biol. Chem.*, 277:36948–36954, 2002.

[84] H. Späth. *Cluster analysis algorithms for data reduction and classification of objects*. Halsted Press; New York, 1980.

[85] D. R. Ripoll, A. Liwo, and H. A. Scheraga. New developments of the electrostatically driven Monte Carlo method – Test on the membrane bound portion of melittin. *Biopolymers*, 46:117–126, 1998.

[86] R. Koradi, M. Billeter, and K. Wüthrich. Molmol: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, 14:51–55, 1996.

[87] K. Ono, M.M. Condron, and D.B. Teplow. Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proc. Natl. Acad. Sci. USA*, 106:14745–14750, 2009.

[88] Nicolae Viorel Buchete, Robert Tycko, and Gerhard Hummer. Molecular Dynamics Simulations of Alzheimer's β-Amyloid Protofilaments. *J. Mol. Biol.*, 353:804–821, 2005.

[89] JJ Dannenberg. The Importance of Cooperative Interactions and a Solid-State Paradigm to Proteins-What Peptide Chemists Can Learn from Molecular Crystals. *Adv. Protein Chem.*, 72:227–73, 2005.

[90] A. Lomakin, DS Chung, GB Benedek, DA Kirschner, and DB Teplow. On the nucleation and growth of amyloid β-protein fibrils: Detection of nuclei and quantitation of rate constants. *Proc. Natl. Acad. Sci. USA*, 93:1125–1129, 1996.

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

[92] H. Naiki and F. Gejyo. Section II. Characterization of in Vitro Protein Deposition-C. Monitoring Aggregate Growth by Dye Binding-20. Kinetic Analysis of Amyloid Fibril Formation. *Methods Enzymol.*, 309:305–317, 1999.

[93] K. Tsemekhman, L. Goldschmidt, D. Eisenberg, and D. Baker. Cooperative hydrogen bonding in amyloid formation. *Prot. Sci.*, 16:761–764, 2007.

[94] J.A. Plumley and JJ Dannenberg. The Importance of Hydrogen Bonding between the Glutamine Side Chains to the Formation of Amyloid VQIVYK Parallel β-Sheets: An ONIOM DFT/AM1 Study. *J. Am. Chem. Soc.*, 132:1758–1759, 2010.

[95] S.R. Collins, A. Douglass, R.D. Vale, and J.S. Weissman. Mechanism of prion propagation: Amyloid growth occurs by monomer addition. *PLOS Biol.*, 2:1582–1590, 2004.

[96] A.T. Petkova, R.D. Leapman, Z. Guo, W.M. Yau, M.P. Mattson, and R. Tycko. Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science*, 307:262–265, 2005.

[97] W.P. Esler, E.R. Stimson, J.M. Jennings, H.V. Vinters, J.R. Ghilardi, J.P. Lee, P.W. Mantyh, and J.E. Maggio. Alzheimer's Disease Amyloid Propagation by a Template-Dependent Dock-Lock Mechanism. *Biochemistry*, 39:6288–6295, 2000.

[98] M.J. Cannon, A.D. Williams, R. Wetzel, and D.G. Myszka. Kinetic analysis of beta-amyloid fibril elongation. *Anal. Biochem.*, 328:67–75, 2004.

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

[100] C. Czaplewski, S. Kalinowski, S. Ołdziej, A. Liwo, and H.A. Scheraga. Application of Multiplexed Replica Exchange Molecular Dynamics to the UNRES Force Field: Tests with α and $\alpha + \beta$ Proteins. *J. Chem. Theory Comput.*, 5:627–640, 2009.

Appendix: Copyright Permissions

Vita

Ana Rojas was born in Córdoba, Argentina. In Spetember of 2001 she obtained a Licenciado degree in physics from the Facultad de Matemática, Astronomía y Física (FaMAF) at the Universidad Nacional de Córdoba. In 2002 she moved to Baton Rouge, Louisiana, to pursue a doctoral degree in physics at Louisiana State University.