이학박사 학위논문

In Silico Investigation of the Structural Properties and Aggregation Processes of Amyloid Aggregates

아밀로이드 응집체의 구조적 성질과 응집 과정에 대한 계산 연구

2023 년 8 월

서울대학교 대학원

화학부 물리화학 전공

이 민 준

In Silico Investigation of the Structural Properties and Aggregation Processes of Amyloid Aggregates

지도교수 신 석 민

이 논문을 이학박사 학위논문으로 제출함
2023 년 8 월

서울대학교 대학원
화학부 물리화학 전공
이 민 준

이민준의 이학박사 학위논문을 인준함
2023 년 8 월

위 원 장 석 차 옥 (인)

부위원장 신 석 민 (인)

위 원 정 연 준 (인)

위 원 이 정 호 (인)

위 원 장 순 민 (인)

Abstract

In Silico Investigation of the Structural Properties and Aggregation Processes of Amyloid Aggregates

MinJun Lee
Department of Chemistry
The Graduate School
Seoul National University

Diseases arise from the disruption of precisely regulated networks of interactions among biomolecules in biological systems. Understanding disease origins involves deciphering disease-related regulatory networks, identifying malfunction causes, and restoring normal states. Recently, there has been a growing focus on targeting diseases at the protein level with a specific and systematic approach. The growing focus on age-related diseases, especially neurodegenerative diseases, is motivated by the extended human lifespan. Neurodegenerative diseases encompass various neurological disorders associated with age-related brain dysfunction and neuronal loss. These diseases are characterized by amyloid aggregation, a consequence of protein misfolding, which is an emerging and intriguing field of study. The complexity of amyloid aggregation, with its diverse

forms and intermediates, arises due to the characteristics of Intrinsically Disordered Proteins that possess energy landscapes with multiple local minima.

In this thesis, three computational discussions for the elucidation of neurodegenerative diseases are introduced: (1) a novel enhanced sampling method which can be used for systems where metastable states play important roles, (2) the structure and aggregation pathway of Aβ_{42} protofibril which is associated with Alzheimer's disease, and (3) the structural stability as the origin of the pathogenicity of α-synuclein protofibrils which is related to Parkinson's disease. These findings provide insights and information that can aid in understanding the structural and dynamical characteristics of the systems associated with neurodegenerative diseases.

Keyword: neurodegenerative diseases, amyloid fibril, molecular dynamics simulation, enhanced sampling methods, amyloid beta, alpha synuclein

Student Number: 2013-22931

Table of Contents

List of Figures and Tables

Chapter 1. Introduction

Understanding of the origins of the diseases is one of major objective of life science, and the development of effective remedies for each disease could ensure the long-term survival and prosperity of humanity. We have the perspective that diseases arise from the disruption of precisely regulated networks of interactions and mechanisms among biomolecules which perform specific roles in biological systems.[1] Therefore, understanding the origins of diseases involves deciphering disease-associated regulatory networks, identifying causes of malfunction, and seeking ways to restore them to normal states.[2] Due to the intricate regulation of biological processes, the investigation of disease origins should progress incrementally with consistently explaining multiple phenomena simultaneously.

In recent years, there has been a growing emphasis on approaching the treatment and prevention of diseases at the protein level, with a more specific and systematic approach. Proteomics, an interdisciplinary field encompassing chemistry, biology, medicine, and IT, plays a crucial role in biomedical research by studying the complete set of proteins produced by the genome, known as the proteome. Unlike the static nature of the genome, the proteome is highly dynamic. The dynamics can be influenced by the physiological states of cells or tissues, and affect protein generation and cellular localization. The functions of proteins within organisms are crucial and rely fundamentally on their structures. Therefore, the ultimate objective of studying protein folding and dynamics is to gain a fundamental understanding of how protein structures are formed under various interactions. This understanding allows not only the prediction of protein structures but also the design of proteins with desired functions.

The growing emphasis on diseases associated with aging is driven by the increasing average lifespan of humanity, and there is a rising interest in neurodegenerative diseases in recent times. Neurodegenerative diseases refer to a group of various neurological disorders which are closely associated with age-related brain dysfunction caused by the loss of neurons. They include Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), Amyotrophic lateral sclerosis (ALS), Multiple system atrophy (MSA) and others, and show several hallmarks: pathological protein aggregation, synaptic and neuronal network dysfunction, aberrant proteostasis, cytoskeletal abnormalities, altered energy metabolism, DNA and RNA defects, inflammation, and neuronal cell death.[3] The epidemiology, pathology, and genetics of neurodegenerative diseases have been investigated based on these features.[4]

The onset and progression of neurodegenerative diseases are known to be associated with amyloid aggregation.[5, 6, 7] The amyloid aggregation is one of the consequences associated with protein misfolding which is an emerging field of increasing interest. Understanding the structural transformations of specific proteins in response to changes in their surrounding conditions is a complex challenge that can be approached by leveraging the accumulated knowledge and investigative techniques from the study of protein structural diversity. Extensive research in proteomics has emphasized the importance of comprehending interactions between proteins-proteins, proteins-nucleic acids, and proteins-other bioactive molecules in order to gain insights into overall cellular functioning and activities.[8] The foundational knowledge derived from research on structural diversity of proteins should serve as a starting point.

However, the presence of fibril polymorphs and various aggregate

intermediates makes it challenging to explain within the “one sequence, one structure” paradigm that has been traditionally accepted in the protein folding problem.[9] This phenomenon occurs due to the characteristics of the Intrinsically Disordered Proteins (IDPs) which have energy landscapes with multiple local minima. Even with advancements in experimental techniques, it remains challenging to explain that the structural diversity of IDPs itself and the resulting various aggregate intermediates and polymorphic fibril structures. Recent research has been overcoming these challenges by utilizing Molecular Dynamics (MD) simulations to investigate the structures, dynamics, and energy landscape of IDPs.[10]

Even though MD simulations have the potential to reveal the characteristics of IDPs that cannot be directly observed through experiments, it is not necessarily an easy task. In the case of intricate molecular systems like biomolecules which have large degrees of freedom, conventional MD simulations often tend to confine molecular conformations to local minimum states at target temperature.[11] In particular, for systems such as IDPs where the structure of metastable states and their dynamics are crucial, achieving sufficient equilibrium using conventional MD simulations is challenging. To address the challenge of limited conformational sampling in traditional MD simulations, various “enhanced sampling” techniques have been suggested. These methods include umbrella sampling, metadynamics, and generalized ensemble approaches, such as Replica Exchange Molecular Dynamics (REMD). In **Chapter 2**, a novel enhanced sampling method named “*Tq*-REM” is introduced. In the proposed *Tq*-REM scheme which is one of Hamiltonian REMD, high temperature replicas in conventional REMD are substituted with *q*-replicas which have reduced barriers by using Tsallis effective potential.[12] This combined scheme is expected to exploit advantages of the conventional REMD and

Hamiltonian REMD resulting in improved sampling efficiency while minimizing the drawbacks of both approaches for simple protein systems. *Tq*-REM can provide useful tools to investigate systems where metastable states play important roles.

After that, the structural properties and aggregation process of IDP protein aggregates directly associated with neurodegenerative diseases are discussed by using the conventional MD simulations and several enhanced sampling methods. In **Chapter 3**, the structure and aggregation pathway of Aβ_{42} protofibril is presented. Amyloid accumulation consisting of Aβ protein within neuronal cells is a prominent characteristic of AD. Among various isoforms of Aβ, Aβ_{42} is known to exhibit relatively higher levels of toxicity, and this toxicity is believed to be attributed to its specific structural characteristics. Recently identified S-shaped triple-β fibril structure shows remarkable structural stability because of its complex residual interactions that form stable hydrophobic cores, and information about interactions between residues can be obtained during dock-and-lock process of fibril formation.[13]

In **Chapter 4**, the structural stability as the origin of the pathogenicity of α-synuclein protofibrils is discussed. α-Synuclein is a neuronal protein found in presynaptic regions, and its fibril form constitutes a significant part of Lewy bodies which are intraneuronal inclusions that serve as a hallmark of PD. Recent experiments on wild-type α-synuclein fibrils have revealed the presence of polymorphism, with the observation of two prominent polymorphic forms. The rod polymorph has stronger sidechain interactions and exhibits higher dissociation energy than the twister polymorph, which mean that the rod polymorph may exhibit high structural stability, comprising a major pathogenic strain of the wild-type α-synuclein fibril.[14]

In summary, this thesis introduces a novel enhanced sampling method (*Tq*-REM)

that can be applied to systems where sampling of metastable states is crucial, including IDPs involved in amyloid fibril formation. Additionally, it provides insights and information that can aid in understanding the structural characteristics and aggregation process of Aβ_{42} and α-synuclein, which form fibrils through aggregation and closely associated with neurodegenerative diseases.

Chapter 2. Conformational Sampling of Metastable States: *Tq*-REM as a Novel Replica Exchange Method

2.1. Introduction

Molecular dynamics (MD) simulations have been widely used to sample equilibrium ensembles for many systems.[15] However, for complex molecular systems such as biomolecules, conventional low-temperature MD schemes tend to trap conformations in local minimum states.[11] Conformational sampling by MD simulations for biomolecules such as proteins with large degrees of freedom has been recognized as an intrinsically difficult computational problem.[16] Several methods have been proposed[17, 18] to overcome this conformational sampling problem using conventional MD, such as umbrella sampling,[19] metadynamics,[20] and generalized ensemble approaches[21] (e.g., replica exchange molecular dynamics[22]).

Since first proposed for Monte Carlo simulation[23] in 1996 and applied to MD simulation[22] in 1999. the temperature replica exchange method (*T*-REM) has been a popular approach for overcoming conformational sampling problems for complex molecular systems. In *T*-REM, MD simulations of multiple replicas of the same system at different temperatures are performed independently and exchanges between neighboring replicas are attempted at certain intervals. Although it has been demonstrated that *T*-REM provides efficient conformational sampling for biomolecules such as proteins, *T*-REM has a major limitation: *T*-REM requires a large number of replicas proportional to the square root of the degrees of freedom of the system. In addition, the diffusion of all replicas in the temperature space is not efficient, which may make *T*-REM impractical for complex systems with limited computational resources. There have been attempts to overcome the limitation of *T*-

REM by developing a general form of REM, such as several Hamiltonian REMs in which selected factors of the Hamiltonian are modified instead of temperature. Fukunishi *et al.*[24] proposed hydrophobicity as a parameter for REM, and compared the performance with *T*-REM in small protein systems. Liu *et al.*[25] and Wang *et al.*[26] attempted replica exchange with solute tempering for efficient sampling of proteins in aqueous solutions (REST1 and REST2). Recently, Hamiltonian REM using dihedral-biasing potentials and a weighted histogram analysis method were also evaluated for oligosaccharides[27], and combined scheme of several enhanced sampling methods shows better convergences of conformational equilibria of N-glycan, chignolin, and adenyl kinase[28]. Examples and applications of enhanced sampling techniques including replica exchange methods can be found in recent reviews.[29]

There have been several attempts to utilize Tsallis statistics for overcoming the sampling problem.[30] For one of the attempts, *q*-REM, new Hamiltonian replica exchange approach, was proposed that uses the Tsallis-generalized effective potential as a coupling parameter.[31] Replicas of *q*-REM use different potential functions at the same temperature. It was also suggested that *q*-scaling using the Tsallis effective Hamiltonian may be applied only for selected components of potential energy. Applications of *q*-REM have demonstrated that *q*-REM can sample conformational spaces with a smaller number of replicas compared with conventional REM.[32] However, it was also noted that parameterization of the *q*-value is nontrivial because the rugged energy surface is highly sensitive to small changes in the *q*-value.[33] Another possible problem may arise since rapid equilibration of each replica, which is needed for efficient exchange attempts, is slow at relatively low temperatures.

In the present work, I propose a novel replica exchange scheme created by combining *T*-REM and *q*-REM, termed *Tq*-REM. In the implementation of *Tq*-REM, high-temperature *T*-replicas in *T*-REM are substituted with *q*-replicas. This scheme is expected to exploit the advantages of *T*-REM and *q*-REM that improves efficiency while minimizing the drawbacks of both approaches. I investigated the performance of *Tq*-REM, relative to *T*-REM, by performing all atom MD simulations on Met-enkephalin, $(AAQAA)_3$, and Trpzip2. The improvements in conformational sampling in using *Tq*-REM are illustrated and discussed.

2.2. Methods

Tq-REM *Tq*-REM combines the methodologies of conventional *T*-REM and *q*-REM. *T*-REM, introduced by Sugita *et al*.,[22] uses replicas of a target system in the canonical ensemble at different temperatures. Exchanges between *T*-replicas are determined by the following Metropolis criterion:

$$w\left(X_{T,i} \leftrightarrow X_{T,j}\right) = \begin{cases} 1, & \text{for } \Delta_T \leq 0, \\ \exp(-\Delta_T), & \text{for } \Delta_T > 0, \end{cases} \qquad \text{(eq. 2.1)}$$

where X represents states for replicas i and j, $\Delta_T = \left[\beta_i - \beta_j\right]\left(E_i - E_j\right)$, $\beta_{i,j} = 1/k_B T_{i,j}$, and *E* is the total potential energy of each *T*-replica.

In the original *q*-REM, introduced by Jang *et al*.,[31] replicas of a target system on different Tsallis effective Hamiltonians with the following generalized effective potential function are used:

$$U_q(\mathbf{r}^N, \epsilon) = \frac{q}{\beta(q-1)} \ln\{1 + \beta(q-1)[U(\mathbf{r}^N) + \epsilon]\}, \qquad \text{(eq. 2.2)}$$

where $U(\mathbf{r}^N)$ is the original potential in configurational space $\mathbf{r}^N$, q is a real number, and ϵ is an adjustable energy shift parameter. When the *q*-value approaches 1.0, U_q

becomes the original potential. As the q-value becomes larger than 1.0, the energy barriers of the transformed potential are reduced, providing much smoother potential energy surfaces. Note that the sampling of phase space by a q-value can be considered equivalent to the effective temperature β', as given by the following equation:

$$\frac{1}{\beta'} = \frac{1}{\beta} + (q-1)\epsilon. \qquad \text{(eq. 2.3)}$$

In our implementation of q-REM, the generalized effective potential is applied to dihedral, van der Waals, electrostatic, and solvation energy potentials. The resulting total effective potential is

$$E_q = U_{\text{bond}} + U_{\text{angle}} + \frac{q}{\beta(q-1)} \times$$
$$\ln\{1 + \beta(q-1)[U_{\text{dih}} + U_{\text{vdw}} + U_{\text{elec}} + U_{\text{solv}} + \epsilon]\}. \qquad \text{(eq. 2.4)}$$

Replicas in q-REM are at the same temperature and have the same ϵ, while their q-values are different. Exchanges between q-replicas are determined by the following Metropolis criterion:

$$w(X_{q,i} \leftrightarrow X_{q,j}) = \begin{cases} 1, & \text{for } \Delta_q \leq 0, \\ \exp(-\Delta_q), & \text{for } \Delta_q > 0, \end{cases} \qquad \text{(eq. 2.5)}$$

where X represents states for replicas i and j, $\Delta_q = \beta[\{E_{q,j}(X_{q,i}) - E_{q,j}(X_{q,j})\} - \{E_{q,i}(X_{q,i}) - E_{q,i}(X_{q,j})\}]$, and E_q is the total effective potential energy of each q-replica.

Combining T-REM and q-REM can be achieved in different ways. In a typical multiplex REM approach, each replica in the combined method is defined by two parameters for different quantities. One way of combining temperature (T) and q parameters is extending each T-replica by q-replicas. In other words, neighboring T-replicas are connected by placing several q-replicas between them. Preliminary

works using this approach showed some difficulties in its efficient implementation. In the present version of *Tq*-REM, I propose substituting high-temperature *T*-replicas with *q*-replicas. *T*-replicas are arranged from the 1st to the Nth replica, which corresponds to a temperature range varying from the lowest to some intermediate temperature. A quantity, M, of *q*-replicas follows the end of *T*-replica chain, from the (N+1)th to (N+M)th replica at the same temperature as the Nth temperature replica. I implemented this version of *Tq*-REM in the TINKER package.[34]

Multicanonical Simulation Since first proposed in MC version[35] in 1991, the multicanonical ensemble algorithm has been applied to various systems for efficient sampling including development of MD version.[36, 37] In multicanonical ensemble, flat probability distribution function is obtained by non-Boltzmann multicanonical weight factor:

$$P_{mc}(E) = \frac{1}{Z_{mc}} n(E) e^{-W(E)} = constant \qquad \text{(eq. 2.6)}$$

where $n(E)$ is the density of states, $W(E)$ is non-Boltzmann multicanonical weight factor, and $Z_{mc} = \sum_E n(E) e^{-W(E)}$. It facilitates a free random walk in potential energy space, therefore, the simulation can overcome the energy barriers. Practically, $W(E)$ is determined by iterations of preliminary runs.[37] Once $W(E)$is determined, product multicanonical simulations are performed. After the simulations, I can get expectation value of a physical quantity A at an arbitrary temperature T by re-weighting equation[38]:

$$\langle A \rangle_T = \frac{\sum_{i=1}^{n_t} A(x_i) e^{-\beta E(x_i) + W(E(x_i))}}{\sum_{i=1}^{n_t} e^{-\beta E(x_i) + W(E(x_i))}} \qquad \text{(eq. 2.7)}$$

where n_t is the total number of configurations, and x_i is the configurations at the ith MD step.

2.3. Simulation Protocols

I studied three peptide systems using Still's GB/SA implicit solvation model[39] to compare the conformational sampling performance of *T*-REM and *Tq*-REM: Met-enkephalin (Ace-YGGFM-Nme), a 15-residue helical peptide (Ace-$(AAQAA)_3$-Nme),[40] and Trpzip2 (SWTWENGKWTWK). Berendsen thermostat was used to maintain each replica at a constant temperature.[41] The velocity version of the Verlet integrator[42] was used with a 2 fs time step. All bonds containing hydrogen were made rigid by the RATTLE algorithm.[43] Simulation trajectories for all systems were saved every 0.5 ps and used for analysis.

Met-enkephalin (Ace-YGGFM-Nme) A minimized extended conformation of Met-enkephalin was prepared using the AMBER99 force field,[44] and equilibrated for 0.1 ns at eight different temperatures (268, 300, 336, 376, 419, 468, 512, and 580 K). Both *T*-REM and *Tq*-REM simulations used 8 replicas, and started from the same equilibrated conformation at each temperature. Replicas of *T*-REM were ensembles prepared at the 8 temperatures given above. The 4 low-temperature replicas (from 268 K to 376 K) of *Tq*-REM were the same as *T*-REM. *Tq*-REM substituted the 4 high-temperature *T*-replicas (from 419 K to 580 K) with 4 *q*-replicas at 376 K and ϵ=60 kcal/mol with different *q*-values: q=1.0025, 1.0045, 1.0065, and 1.0090, of which effective temperatures are 451, 512, 572, and 648 K, respectively. Thereafter, 100 ns product MD simulations for both REMs were performed. Exchanges of both REMs between adjacent replicas were attempted every 2 ps, and the overall exchange probabilities were ~35%. **Figure 2.6.1a** illustrates the energy distributions of each replica.

15-residue helical peptide (Ace-$(AAQAA)_3$-Nme) A minimized extended

conformation of the helical peptide was prepared using the AMBER99 force field.[44] This conformation was equilibrated for 0.1 ns at 16 different temperatures (244, 261, 280, 300, 321, 344, 368, 349, 421, 450, 481, 514, 549, 587, 627, and 669 K). For this helical peptide, three different REMs were performed: *T*-REM, *Tq*-REM(12,4), and *Tq*-REM(10,6). These three REM simulations used sixteen replicas, and started from the same equilibrated conformation at each temperature. Replicas of *T*-REM were ensembles at the 16 temperatures given above. *Tq*-REM(12,4) used the same 12 low-temperature *T*-replicas (from 244 K to 514 K) but the 4 high-temperature *T*-replicas were substituted with 4 *q*-replicas at 514 K and ϵ=250 kcal/mol with different *q*-values: *q*=1.0005, 1.0010, 1.0015, and 1.0020, of which effective temperatures are 577, 640, 703, and 766 K, respectively. *Tq*-REM(10,6) used the same 10 low-temperature *T*-replicas (from 244 K to 450 K) but the other 6 high-temperature *T*-replicas were substituted with 6 *q*-replicas at 450 K and ϵ=250 kcal/mol with different *q*-values: *q*=1.0005, 1.0010, 1.0014, 1.0018, 1.0020, and 1.0023, of which effective temperatures are 513, 576, 626, 676, 702, and 739 K, respectively. Thereafter, 200 ns product MD simulations of three REMs were performed. Exchanges of the three REMs between adjacent replicas were attempted every 2 ps and the overall exchange probabilities were ~40%. **Figure 2.6.1b** illustrates the energy distributions of each replica.

Trpzip2 (SWTWENGKWTWK) A minimized extended conformation of Trpzip2 was prepared using the AMBER96 force field.[45] These conformations were equilibrated for 0.1 ns at 16 different temperatures (249, 265, 282, 300, 319, 339, 360, 383, 407, 432, 459, 487, 517, 549, 582, and 618 K). For Trpzip2, three different REMs were performed: *T*-REM, *Tq*-REM(12,4), and *Tq*-REM(10,6). These three REM simulations used 16 replicas and started from the same equilibrated

conformation at each temperature. Replicas of *T*-REM were ensembles at the 16 temperatures given above. *Tq*-REM(12,4) used the same 12 low-temperature *T*-replicas (from 249 K to 487 K) but the other 4 high-temperature *T*-replicas were substituted with 4 *q*-replicas at 487 K and ϵ=200 kcal/mol with different *q*-values: *q*=1.00025, 1.00044, 1.00048, and 1.00051, of which effective temperatures are 512, 531, 535, and 538 K, respectively. *Tq*-REM (10,6) used the same 10 low-temperature *T*-replicas (from 249 K to 432 K) but the other 6 high-temperature *T*-replicas were substituted with 6 *q*-replicas at 432 K and ϵ=200 kcal/mol with different *q*-values: *q*=1.00022, 1.00037, 1.00041, 1.00043, 1.00047, and 1.00054, of which effective temperatures are 454, 469, 473, 475, 479, and 486 K, respectively. Thereafter, 200 ns product MD simulations of three REMs were performed. Exchanges of three REMs between adjacent replicas were attempted every 2 ps and the overall exchange probabilities were ~40%. **Figure 2.6.1c** illustrates the energy distributions of each replica.

Multicanonical simulation on Met-enkephalin It is shown that sampling for systems with first-order transitions, such as phase transition in spin systems or folding/unfolding transition of two-state folder proteins, can be enhanced by using multicanonical ensemble algorithm.[46] It was suggested by Okamoto *et al*. that random walk of multicanonical MD simulation could be less efficient than *T*-REM and MUCAREM in terms of transition frequency between highest and lowest energies.[22] Recently, modified iteration scheme was proposed to improve the folding/unfolding transition frequency.[47] Improvements of factor 30 over REM in the number of folding/unfolding events for systems with first-order-like transitions were suggested. Other than two state systems with first-order phase transition, the effectiveness of multicanonical sampling has not been tested for systems with

multiple metastable intermediate states in folding pathway or off-pathway of proteins/peptides. In order to compare multicanonical method with *Tq*-REM scheme in such regards, I performed multicanonical simulation on Met-enkephalin system.

2.4. Results

Met-enkephalin (Ace-YGGFM-Nme) Whole trajectories of the 100 ns simulations at the target temperature (2nd replica, 300 K) were used for the principal component analysis (PCA). **Figure 2.6.2** illustrates the free energy surfaces of *T*-REM and *Tq*-REM in PC space (two eigenvectors, PC1 and PC2). Four major local minimum structures were identified for both REM simulations. It is noted that sampling results for one of the local basins (denoted 3 in the **Figure 2.6.2**) showed significant differences between *T*-REM and *Tq*-REM. It can be argued that the *Tq*-REM samples conformations more efficiently than *T*-REM around this local minimum region. The evolution of free energy surfaces for different time windows during the simulations also illustrates this sampling difference (see **Figure 2.6.3a**). It was shown that the basin structures of local minimum points are developed more rapidly and more precisely during *Tq*-REM simulations compared with *T*-REM simulations.

In order to quantitatively measure the convergence behavior of sampling for different simulations, I introduce a quantity $< \delta F(t) >$ defined as follows:

$$< \delta F(t) > \equiv \sqrt{\frac{1}{N_{threshold}} \sum_{\{i,j \mid FE_{threshold}\}}^{n,n} \left(FE_{i,j}^{obj}(t) - FE_{i,j}^{ref} \right)^2} \quad \text{(eq. 2.8)}$$

where PC space is discretized by dividing the PC1 and PC2 axes into $n \times n$ uniform cells, $FE_{i,j}^{obj}(t)$ is the ijth component of the object free energy profile of a

particular simulation, calculated by the trajectory up to time t. Similarly, $FE_{i,j}^{ref}$ is the ijth component of the reference free energy profile of whole trajectory. We may consider convergence properties of free energy surfaces around local basin structures. $\{i,j \mid FE_{threshold}\}$ denotes the inclusion of cells which have their free energy values lower than $FE_{threshold}$. Similar 'ergodic measure' was used for measuring convergence efficiency of REST.[25] Ergodic measure measures how fast the free energy profiles of the two independent trajectories of the same system approach each other. In contrast, $< \delta F(t) >$ shows how rapidly the trajectory of the object system develops the basin topology of the free energy for the reference system. $< \delta F(t) >$ can be calculated for two independent trajectories of the same system or for two trajectories of different systems.

I calculated $< \delta F(t) >$for the simulations of *T*-REM and *Tq*-REM at 300 K for Met-enkephalin. Here I set $FE_{threshold}$ = –4 kcal/mol to consider only local minimum regions. For a benchmark system with small peptide, the intrinsic convergence properties of *T*-REM and *Tq*-REM are very similar as shown in **Figure 2.6.3c** where the reference for the calculation of $< \delta F(t) >$ is taken to be the free energy profile of whole trajectory for respective simulations. **Figure 2.6.3b** shows the behavior of $< \delta F(t) >$ for *T*-REM and *Tq*-REM when the reference is taken as the free energy profile of whole trajectory for *T*-REM. $< \delta F(t) >$ can be fit to a double exponential function: $y = a \times e^{-t/\tau_1} + (1-a) \times e^{-t/\tau_2}$. For *T*-REM, I obtained τ_1=3.39 and τ_2=11800, while τ_1=4.03 and τ_2=24742 for *Tq*-REM. It was observed that initial convergences are very fast up to t ~ 15ns. This early phase of sampling convergence can also be seen from **Figure 2.6.3a**. After t ~ 15ns, $< \delta F(t) >$ of *T*-REM continues to decrease while $< \delta F(t) >$ for *Tq*-REM seems to

be almost flat, as shown more clearly in the logarithmic scale curve (inset of **Figure 2.6.3b**). This difference at t > 15ns suggests that *Tq*-REM trajectory completes sampling of the basic topology of the converged free energy profile for *T*-REM faster than *T*-REM trajectory itself. In other words, it can be argued that *Tq*-REM samples conformations in the region of free energy minimum more efficiently than conventional *T*-REM.

The diffusion behavior of trajectories in PC space may illustrate the efficiencies of conformational samplings of the REM simulations. In order to examine the efficiency of sampling in conformational space, I calculated mean square displacement (MSD) in PC space, defined by $\mathrm{MSD}(\tau) = \langle (R(t+\tau) - R(t))^2 \rangle_t$ where $R(t)$ is coordinates of trajectory at time t in PC space. **Figure 2.6.3d** shows that the diffusion behavior in conformational space for *T*-REM and *Tq*-REM is similar, although *T*-REM seems to show faster diffusion in the intermediate time (region II). Increased MSD in region II can be attributed to faster diffusion between local basins, which may imply that *T*-REM's trajectory stays inside individual local basins for relatively shorter time. I calculated average residence times of the trajectories in local basin and non-basin regions of PC free energy surface. When I define the local basins as regions where the free energy is lower than −4 kcal/mol, residence times in local basin and non-basin regions are shorter (3.245 ps and 1.599 ps) for *T*-REM than *Tq*-REM (3.483 ps and 4.755 ps). When the energy criteria for local basin is changed to be −3 kcal/mol, basin residence time of *T*-REM is still shorter (33.21 ps) than *Tq*-REM (36.14 ps) while non-basin residence time of *T*-REM is comparable (0.9489 ps) to that of *Tq*-REM (0.9481 ps). Analysis of the present simulations suggests that the trajectories of *T*-REM wander about in higher

energy non-basin regions, while the trajectories of the *Tq*-REM sample the local minimum basin regions relatively efficiently. Since *Tq*-REM samples the low-energy region more efficiently, the free energy surface of the local basins in PC space converges better for *Tq*-REM than *T*-REM. One may conclude that *Tq*-REM samples local basin regions more intensively than *T*-REM, leading to improved basin sampling efficiency.

The trajectories of the replicas in the REM simulations are affected by the exchanges between neighboring replicas. In *Tq*-REM, substitution of high temperature replicas with *q*-scaling replicas influences the behavior of the lower temperature replicas. The 5th replica of the *T*-REM has a higher temperature than the 4th replica. In *Tq*-REM, the 5th replica has lower potential energy barriers than the 4th replica due to *q*-scaling, while the temperature is unchanged. In other words, the higher-temperature replica of *T*-REM increases the kinetic energy of the system, but *Tq*-REM reduces the energy barrier of the same replica. Additional 10 ns simulations for the 5th replica of *T*-REM and *Tq*-REM were performed without exchange to examine the differences in dynamics between the higher-temperature and the reduced-potential conditions. The time auto-correlation function of each dihedral angle was calculated by

$$C_I(\tau) = \langle P_2(\cos(\theta_{t+\tau} - \theta_t))\rangle_t \qquad \text{(eq. 2.9)}$$

where θ_t is a dihedral angle at time t, and P_2 is the second Legendre polynomial $P_2 = 1.5x^2 - 0.5$.[48] Relaxation of C_I quantitatively illustrates how rapidly a trajectory samples each dihedral angle (see **Figure 2.6.4a**). Except for the two terminal dihedral angles (φ of Tyr1 and ψ of Met5), C_I of the other eight dihedral angles relax faster in *Tq*-REM than *T*-REM. It can be argued that most dihedral angles are sampled more rapidly and efficiently under the reduced-barrier condition

than the higher-temperature condition. In particular, C_I for φ of Gly2 and Gly3 show large differences between *T*-REM and *Tq*-REM. I compared the evolutions of φ for Gly2 in the two REM trajectories (**Figure 2.6.4b**). It was shown that *Tq*-REM samples a diverse set of φ values for Gly2 more efficiently than *T*-REM. The trajectories of *Tq*-REM seem to sample uniformly the dihedral angles corresponding to all of the local minimum structures, while the φ angles for some of the local minima are inefficiently sampled in *T*-REM. Broadening of distribution of the radius of gyration also reflects more diverse basin sampling behavior of *q*-replica (see **Figure 2.6.4c**). These results illustrate that reducing the potential barrier is more efficient than raising the temperature for sampling diverse local minimum values of dihedral angles.

Sampling of dihedral angle space is examined by considering dihedral angle vector $\overrightarrow{d(t)}$ whose rth row components are $(\cos\varphi_r,\ \sin\varphi_r,\ \cos\psi_r,\ \sin\psi_r)$ with the is residue number r.[49] The sampling probability of local basin k can be calculated by $\text{Probability}_k(t) = \langle b_k(t)\rangle$ where $b_k(t)$ is 1 if $\cos^{-1}\left(\overrightarrow{d(t)}\cdot\overrightarrow{d_k}/\left|\overrightarrow{d(t)}\right|\left|\overrightarrow{d_k}\right|\right) \leqq \delta\overrightarrow{d_k}$, or otherwise 0. $\delta\overrightarrow{d_k} = \langle\cos^{-1}\left(\overrightarrow{d_l}\cdot\overrightarrow{d_k}/\left|\overrightarrow{d_l}\right|\left|\overrightarrow{d_k}\right|\right)\rangle_l$ where $\overrightarrow{d_k}$ is dihedral angle vector with respect to the deepest free energy region for each basin and l denotes grids in PC space with values of free energy lower than –4 kcal/mol. This probability function shows the sampling probability around certain local basin in dihedral angle space. To examine the effects of different dihedral angle sampling pattern on overall conformational sampling, I calculated the time evolution of sampling probabilities of four basins in dihedral space[49] corresponding to basin 1 to 4 in **Figure 2.6.2**. **Figure 2.6.4d** shows that sampling probability of each basin and the overall basin sampling probability (inset of **Figure 2.6.4d**) converge faster in *Tq*-

REM than *T*-REM. It was also shown that *Tq*-REM is capable of sampling metastable states (basin 1, 3) in addition to the global minimum state (basin 2). In contrast, *T*-REM predominantly samples global minimum state much more than metastable states. It can be argued that *q*-replica with reduced potential barrier is more efficient than *T*-replica with higher temperature for balanced sampling of various local minimum dihedral angles, resulting in improved sampling of metastable states.

Representative structures of the four local basins (denoted as 1, 2, 3, and 4 in **Figure 2.6.2**) are shown in **Figure 2.6.5**. Basin 4 shows extended structures with no backbone hydrogen bond, a large radius of gyration, and a large end-to-end distance. Basin 3 has a U-shaped bent structure with two backbone hydrogen bonds. (Carbonyl oxygen of Gly2 makes two hydrogen bonds with the N-H on Met5 and Nme) Basin 1 shows a helical structure with the carbonyl oxygen of Tyr1 making two hydrogen bonds with the N-H of Phe4 and Met5. For basin 2, Ace converts to a helical backbone and forms more hydrogen bonds. The free energy surface as a function of the radius of gyration and the end-to-end distance are shown in **Figure 2.6.6**. The region around basin 3 is much better represented in *Tq*-REM like the free energy surfaces in PC space. The number of hydrogen bonds increases in the order of basin $4 < 3 < 1 < 2$ (see **Figure 2.6.7**). It can be concluded that folding of Met-enkephalin from an extended conformation (basin 4) to a helical conformation (basin 2) follows the pathway of $4 \rightarrow 3 \rightarrow 1 \rightarrow 2$. The efficiency of conformational samplings around basins 1 and 2 are similar for both *Tq*-REM and *T*-REM. However, for the metastable folding intermediates (basin 3), *Tq*-REM provides much more efficient sampling than *T*-REM. It can be argued that our results of *Tq*-REM showed improved samplings for the global minima and local basins of metastable intermediates

compared with the previous Met-enkephalin simulations.[50]

Multicanonical Sampling of Met-enkephalin In order to compare multicanonical method with *Tq*-REM scheme, I performed multicanonical MD simulations on Met-enkephalin with the same force field and thermostat. I performed simulations at 580 K which is the highest temperature in *T*-REM of Met-enkephalin.[37] From the iterative runs from the simulation, I obtained a flat potential energy distribution (see **Figure 2.6.8a**). Using this muticanonical potential, I performed ten independent multicanonical simulations of 100 ns with different initial conformations at 300 K. Backbone RMSD (Root Mean Square Deviation) curves with respect to local basin structures identified from *T*-REM and *Tq*-REM indicate that multicanonical simulations successfully sample various local basin structures (see **Figure 2.6.8b**). I used all of these ten 100 ns trajectories to calculate free energy surface as a function of radius of gyration and the end-to-end distance (**Figure 2.6.6**).[38]

Multicanonical free energy surface shows different features in comparison with those for *T*-REM and *Tq*-REM. Multicanonical sampling captures additional local basins (denoted as 5, 6, 7, and 8 in **Figure 2.6.6**) which are not sampled as isolated local basins in *T*-REM and *Tq*-REM. Although multicanonical method also samples four local basins observed in *T*-REM and *Tq*-REM, two major local basins 1 and 2 are relatively poorly sampled with somewhat improved sampling of basin 3. It is noted again that backbone configurations are changed from extended (4) through bent (3) to helical (1→2) conformations with gradually increasing hydrogen bonds during typical folding pathway. On the other hand, structures 5 and 8 have highly distorted backbone torsional configurations in bent region and structures 6 and 7 have planar geometries. It was shown that these four conformations are energetically

unstable due to high backbone strain and weak hydrogen bond interactions (see **Figure 2.6.8c**).

Multicanonical method, with flat potential energy distribution, is capable of sampling various conformational states which are energetically very unstable and rarely sampled in the original potential energy space. However, major conformational states corresponding to global minima could be sampled relatively poorly because of the reduced stability under flat energy distribution. In the above example, *Tq*-REM successfully samples both of the major local basins 1 and 2 and also folding intermediate state of basin 3, while multicanonical simulation shows improved sampling of basin 3 but poor sampling of global minima corresponding to basins 1 and 2. It can be argued that *Tq*-REM is more efficient to sample most of the relevant local basins in comparison to multicanonical method.

15-residue helical peptide (Ace-(AAQAA)$_3$-Nme) Whole trajectories for the 200 ns simulation at the target temperature (4th replica, 300 K) were used for PCA. **Figure 2.6.9** illustrates the free energy surfaces of *T*-REM, *Tq*-REM(12,4), and *Tq*-REM(10,6) in PC space. Three major local minimum structures were identified for the REM simulations. It is noted that the sampling results for local basins 2 and 3 showed subtle differences between *T*-REM and *Tq*-REMs. *T*-REM seems to sample better in the region around basin 2 compared with *Tq*-REM(12,4), while the reverse is true for basin 3. *Tq*-REM(10,6) is shown to provide efficient sampling for basins 2 and 3. The evolution of the free energy surfaces at different times during the simulations also illustrates these differences (see **Figure 2.6.10**). The development of basin structures for all local minima are well represented by *Tq*-REM(10,6) compared with *T*-REM or *Tq*-REM(12,4).

Representative structures of the three local basins (denoted as 1, 2, and 3 in

Figure 2.6.9) are shown in **Figure 2.6.11**. Basin 1 has a full helical structure, in which all backbone hydrogen bonds are formed. Basins 2 and 3 have partial helical structures, where some backbone hydrogen bonds in the N-terminal region are lost. The increase in the number of hydrogen bonds, the radius of gyration and the end-to-end distance illustrates the development of the helicity of the peptide (see **Figure 2.6.12**). The free energy surface as a function of the radius of gyration and the end-to-end distance is shown in **Figure 2.6.13**. The region around basin 4 shows various collapsed structures (see **Figure 2.6.14**). The folding of $(AAQAA)_3$ follows the pathway of 4 → 3 → 2 → 1. After the collapse of the initial extended conformation, the C-terminal and the middle of $(AAQAA)_3$ form a helical conformation (4 → 3 → 2). The final stage of the folding is finished with formation of the last hydrogen bond at the N-terminal (2 → 1). It can be argued that *Tq*-REM (10,6) provides improved conformational sampling for the early folding intermediates (basins 4 and 3), compared with *T*-REM.

Trpzip2 (SWTWENGKWTWK) Whole trajectories of the 200 ns simulation at the target temperature (4^{th} replica, 300 K) were used to calculate the backbone RMSD relative to the native structure (PDB:1LE1). The free energy surfaces of the end-to-end distance and the backbone RMSD are shown for *T*-REM, *Tq*-REM(12,4), and *Tq*-REM(10,6) in **Figure 2.6.15a**. One can identify five local minimum basins for the REM simulations. It is noted that all of the local basins are well sampled by *Tq*-REM(10,6), while basin 2 is poorly represented by *T*-REM and *Tq*-REM(12,4). **Figure 2.6.15b** shows free energy as a function of the backbone RMSD at different times during the simulations. *T*-REM cannot sufficiently sample native-like global minimum structures. In addition, the second local minimum at the RMSD value of ~2.3 Å is not well represented by either *T*-REM or *Tq*-REM(12,4). *Tq*-REM(10,6)

seems to sample sufficiently all local minimum structures, and the resulting free energy surface from the 200 ns simulation is consistent with the free energy profile of the previous study with a 3.8 μs conventional MD simulation.[51]

It was suggested that the folded state is overpopulated at low temperatures in REMD with Berendsen (weak-coupling) thermostat due to distortion of configurational space distributions.[52] This may be attributed to the artificial bias toward replica exchanges between low energy conformations (global basin) at high temperature replicas and high energy conformations (partially folded states/local basins) at low temperature replicas. Such bias would consequently result in the underpopulation of local basins at low temperature. Interestingly, free energy profiles of *Tq*-REM simulations of Met-enkephalin and $(AAQAA)_3$ showed slightly reduced population of global basin region and clearly improved sampling of local basins compared to *T*-REM. It can be argued that possible artifacts of Berendsen thermostat are reduced in *Tq*-REM simulations. *Tq*-REM(10,6) simulation of Trpzip2 efficiently samples local minima (around the RMSD values of ~2.3 Å and ~2.7 Å), which are poorly sampled by *T*-REM. At the same time, *Tq*-REM(10,6) simulation was found to sample global minimum successfully which are not sufficiently sampled in *T*-REM with the same number of replicas. The structures of these local minima are essentially consistent with previous simulation performed using Nosé–Hoover thermostat,[51] which implies the reliability of our combined *Tq*-REM scheme. The behavior of our *Tq*-REM scheme under other thermostats such as strong-coupling bath will be the subject of future studies.

The representative structures of five local basins (denoted as 1, 2, 3, 4, and 5 in **Figure 2.6.15a**) are shown in **Figure 2.6.16**. Basin 1 has a folded structure aligned to the PDB native structure with a backbone RMSD of 1 Å. The folded structure

shows the two hydrophobic cores (Core1: Trp4 and Trp9, Core2: Trp2 and Trp11). Basins 2, 3, and 4 are partially folded structures in which four Tryptophan residues fail to form native hydrophobic cores. Basin 5 corresponds to the initially collapsed structure. The free energy surface as a function of the radius of gyration for the two hydrophobic cores is shown in **Figure 2.6.17**. This free energy surface suggests a possible folding pathway for Trpzip2 with the sequence of d → c → b → a. One can relate the conformations for a, b, c, and d with the local minima on the free energy surface as a function of the backbone RMSD (see **Figure 2.6.18**). The initial extended structure of Trpzip2 first collapses to conformation d, where Core1 is formed. In order to align correctly the native contacts, such as backbone hydrogen bonds, the initial formation of Core 1 needs to be broken (d → c). After the alignment, reformation of Core1 (c → b) and final formation of Core2 (b → a) can proceed sequentially. The sequence in which the outer Core2 is formed after the inner Core1 suggests the zip-out folding pathway for Trpzip2, consistent with the previous simulation study.[53] It is noted that step d → c seems to be the rate-determining step of the folding pathway because of its large free energy barrier. Structures for local basin 2 (**Figure 2.6.16**) are included for conformation c as an important metastable folding intermediate in the folding pathway. It can be concluded that *Tq*-REM(10,6) provides reliable descriptions of the folding pathways by efficiently sampling conformations of important metastable intermediates.

Application of Tq-REM to small peptide in explicit water model I performed simulations on trialanine in order to investigate the relative performances of *Tq*-REM and *T*-REM in explicit water. Trialanine is a model peptide used in several experimental[54] and theoretical[55] studies. Despite its small size, its diverse conformational dynamics make feasible the comparison between simulations and

experiments. It was shown that solvated trialanine has pPII structure predominantly and α, β, and Lα in small probabilities.[56, 57] I used OPLS-AA force field[58] with SPC/E water model[59] in order to compare *T*-REM and *Tq*-REM in explicit solvents. It was suggested that OPLS-AA force field with SPC/E water model accurately reproduces the conformational distributions of trialanine obtained by experimental data.[57] A minimized extended conformation of trialanine surrounded by 252 water molecules in a periodic cubic box of which length is 20 Å was prepared and equilibrated for 0.1 ns at 12 different temperatures (300, 310, 320, 330, 342, 355, 369, 383, 398, 414, 430, and 447 K). Three different REMs were performed: *Tq*-REM(6,2), *T*-REM(8) and *T*-REM(12). *T*-REM(8) used only 8 low temperatures from 300 K to 383 K, while *T*-REM(12) used all 12 temperatures. *Tq*-REM(6,2) used the same 6 low temperatures from 300 K to 355 K but the other 2 high-temperature *T*-replicas were substituted with 2 *q*-replicas at 355 K and ϵ=2700 kcal/mol with different *q*-values: *q*=1.000186944, and 1.000344448, of which effective temperatures are 609, and 823 K, respectively. These three REM simulations started from the same equilibrated conformation at each temperature. Exchanges between adjacent replicas were attempted every 2 ps and the overall exchange probabilities were ~30 % (see **Figure 2.6.20a**).

Ramachandran plots of dihedral angles for the central residue are calculated from the 20 ns trajectories at the target temperature 300 K. Ramachandran plots showed five distinct basins (see **Figure 2.6.19**). Compared to 50 ns conventional MD simulations with the same system in previous research,[57] 20 ns REM simulations successfully reproduce the experimental results. Solvated trialanine has predominantly pPII structure (basin 1) and small probabilities of β (basin 2), α (basin 3, 4) and Lα (basin 5). Though Lα conformation of basin 5 is rarely sampled by

conventional MD simulations, three REM simulations show distinct spots of the Lα conformation in Ramachandran plot. It was shown that *Tq*-REM(6,2) exhibits enhanced peaks of metastable α-helical conformations compared to two *T*-REMs. Time evolution of Ramachandran plots through different time windows during the simulations illustrates the different behavior for sampling local basin topology (see **Figure 2.6.20b**). It was suggested that *Tq*-REM(6,2) more rapidly captures the regions of five minima than *T*-REM(8) and *T*-REM(12) with smaller number of replicas. Notably, *Tq*-REM(6,2) samples Lα (basin 5) conformation from very early times of the simulation. These results are consistent with the behavior of *Tq*-REM and *T*-REM observed in simulations with implicit solvation system.

2.5. Discussion and Conclusions

Even though replica exchange methods have been developed for some time now for efficient conformational sampling methods in bio-molecular simulations and methods of choice for *ab initio* protein folding, their application to very large bio-systems are limited. Different versions of REMs have their own advantages and limitations, resulting in inefficient sampling performances.[60] One of the main obstacles for the conventional *T*-REM may be the large number of replicas needed for complex bio-molecular systems. Optimization of some of the Hamiltonian REM approaches, including the *q*-REM, may not be straightforward because of difficulties in parameterizations. As one of the attempts to address such issues, one can develop a novel REM by combining different REM versions.

In *T*-REM, the increased kinetic energy in the high-temperature replica can make the system overcome its energy barriers. However, accelerating all degrees of

freedom at an elevated temperature may be inefficient because the system stays mainly in a high-energy region rather than sufficiently sampling lower-energy local basins. *q*-REM is one of the Hamiltonian-REMs using the Tsallis effective potential, which reduces barriers of specific potentials. In *q*-REM, the reduced potential makes the system easily overcome energy barriers by accelerating selected degrees of freedom important for protein dynamics. I demonstrated that *q*-REM was able to sample conformations of biomolecules more efficiently than *T*-REM with a smaller number of replicas.[32] The folding dynamics of small peptides with a generalized simulated annealing algorithm using the Tsallis effective potential was compared with the conventional simulated annealing simulations with temperature scaling.[61] Even though these studies clearly illustrated advantages of using barrier-reduced effective potentials for enhanced conformational sampling, implementation of the *q*-REM to a wide variety of systems still needs further work to determine its optimum parameters. One of the difficulties that might arise is that the relatively low temperature (usually 300 K) adopted for *q*-REM does not guarantee efficient equilibrations for replicas with different *q*-values. Another disadvantage of *q*-REM is that one cannot obtain the temperature dependence of the system from a single simulation.

In the present work, I proposed a novel replica exchange scheme, termed as *Tq*-REM, by combining *T*-REM and *q*-REM. In the proposed *Tq*-REM scheme, high-temperature *T*-replicas in *T*-REM are substituted with *q*-replicas. The aim of this "combo" scheme is to exploit advantages of both *T*-REM and *q*-REM for improved efficiency while minimizing the drawbacks of both approaches. I applied the *Tq*-REM to all atom MD simulations on 3 different peptide systems: Met-enkephalin, $(AAQAA)_3$, and Trpzip2. The focus of the present study was to investigate the

behavior of *Tq*-REM during conformational sampling of these systems compared with the conventional *T*-REM. For a given number of replicas in *T*-REM, some of the replicas at the high-temperature end are replaced by replicas with different *q*-values. The performance of the two REMs in conformational sampling was examined by calculating the free energy surface for the folding dynamics of the peptides. It was found that convergence of the free energy surfaces is improved using *Tq*-REM compared with the conventional *T*-REM. In particular, the trajectories of *Tq*-REM were able to sample the relevant conformations for all of the metastable folding intermediates, while some of the local minimum structures were poorly represented by *T*-REM. In order to take full advantages of *Tq*-REM, one needs to include a sufficient number of *q*-replicas for a particular simulation.

The results of this study demonstrated that *Tq*-REM can provide very efficient sampling for conformations of biomolecules such as proteins, especially conformations of metastable structures important in understanding the dynamical behavior of a system. The reason for the improvements can be attributed to the fact that the trajectories of *q*-REM efficiently sample conformations near local minima. This *q*-REM sampling process occurs because the potential energy barriers are reduced. The trajectories of very high-temperature *T*-replicas of *T*-REM, on the other hand, wander about in high energy regions leading to inefficient sampling. Having a sufficient number of *T*-replicas through the intermediate temperature range for *Tq*-REM make sure that efficient equilibration can be achieved for all *T*-replicas and *q*-replicas, and the temperature-dependent behavior of the system is obtained from the simulation. Thanks to the improvements in sampling efficiency, it can be argued that *Tq*-REM may allow researchers to use a smaller number of replicas than *T*-REM. I also confirmed similar behavior of *Tq*-REM for a small peptide under explicit

solvent environment. However, further studies of applying *Tq*-REM to more complex systems are needed to systematically illustrate such advantages. The results of the present study suggest that *Tq*-REM can provide useful tools to investigate systems where metastable states play important roles.

2.6. Figures

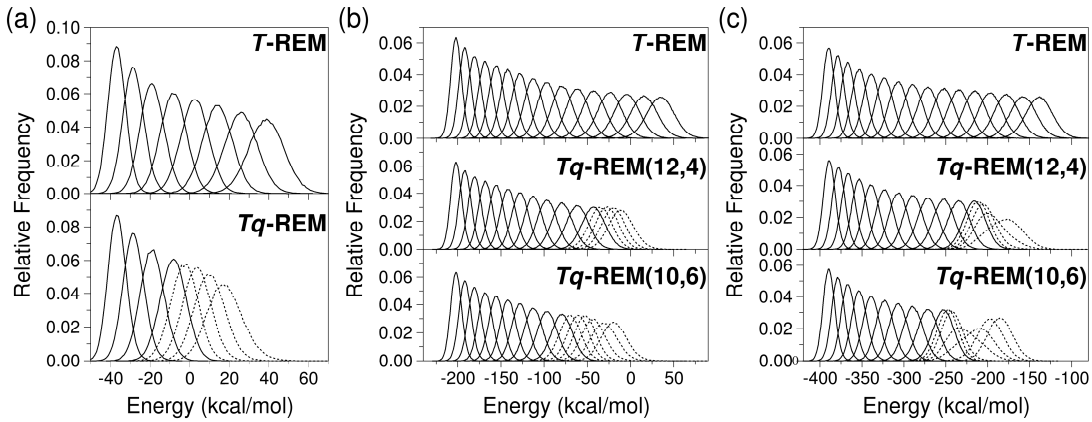


Figure 2.6.1. The energy distributions of *T*-replicas (solid lines) and *q*-replicas (dot lines) in *T*-REM and *Tq*-REM simulations for (a) Met-enkephalin, (b) $(AAQAA)_3$, and (c) Trpzip2.

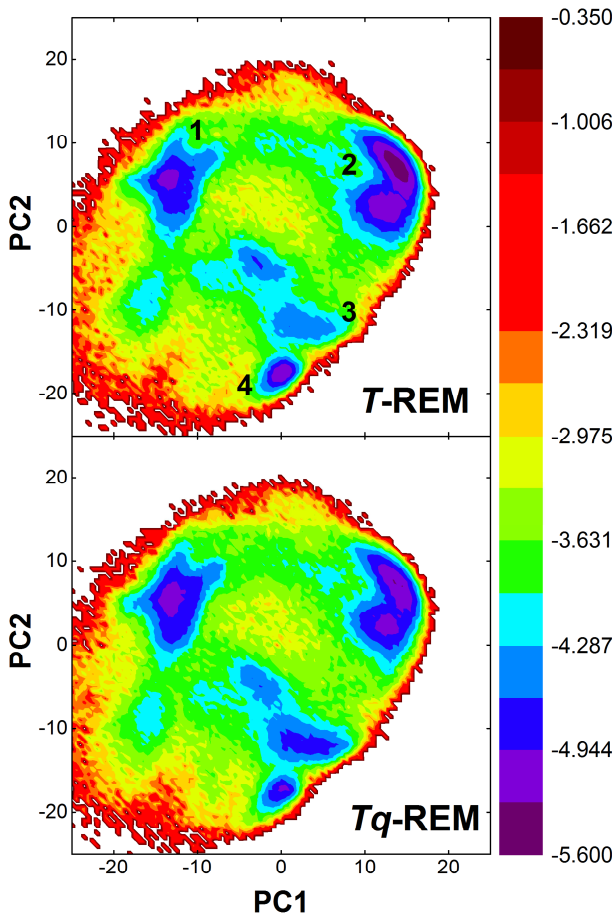


Figure 2.6.2. The free energy surfaces of *T*-REM and *Tq*-REM in PC space obtained from 100 ns trajectories at 300 K for Met-enkephalin. Basins of four local minimum structures are denoted 1 to 4.

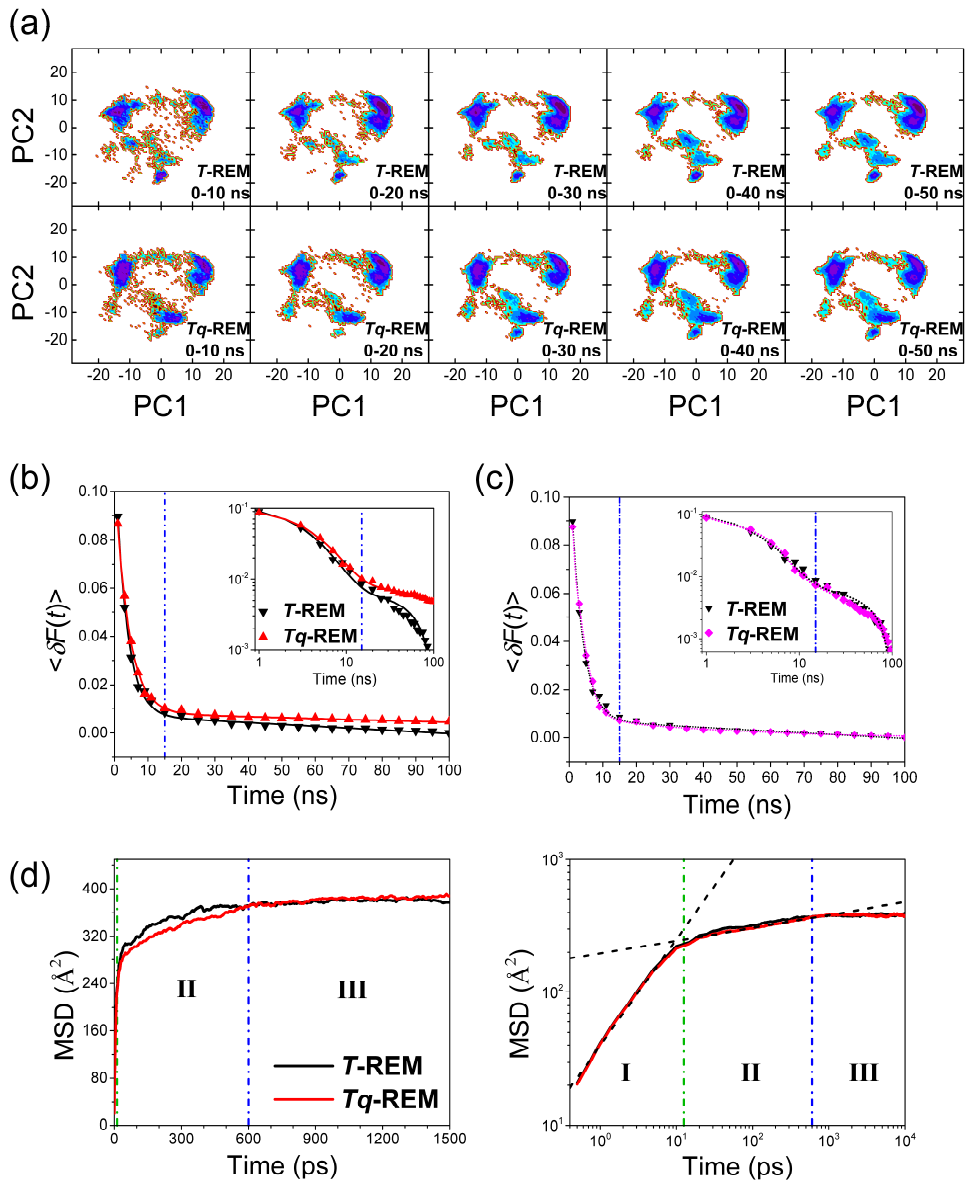


Figure 2.6.3. (a) Time evolutions of the free energy profiles of *T*-REM and *Tq*-REM in PC space at 300 K for Met-enkephalin. The free energy profile regions whose energies are lower than –4 kcal/mol are shown for clarity. (b) $< \delta F(t) >$ for *T*-REM and *Tq*-REM of which reference is free energy profile of whole trajectory for *T*-REM. The inset shows logarithm scale plots of the same graphs. (c) $< \delta F(t) >$ of *T*-REM and *Tq*-REM of which reference is free energy profile of whole

trajectories for the respective simulations. (d) Time evolution of mean square displacement as calculated by MSD(τ) = $\langle (R(t+\tau) - R(t))^2 \rangle_t$ where $R(t)$ is coordinates of trajectory at time t in PC space. Linear scale (left) and logarithm scale (right) plots are shown. MSD can be fit to a power law function of time t^α in region I (0–12.5 ps) and II (12.5–600 ps), respectively, as denoted by the black dashed lines.

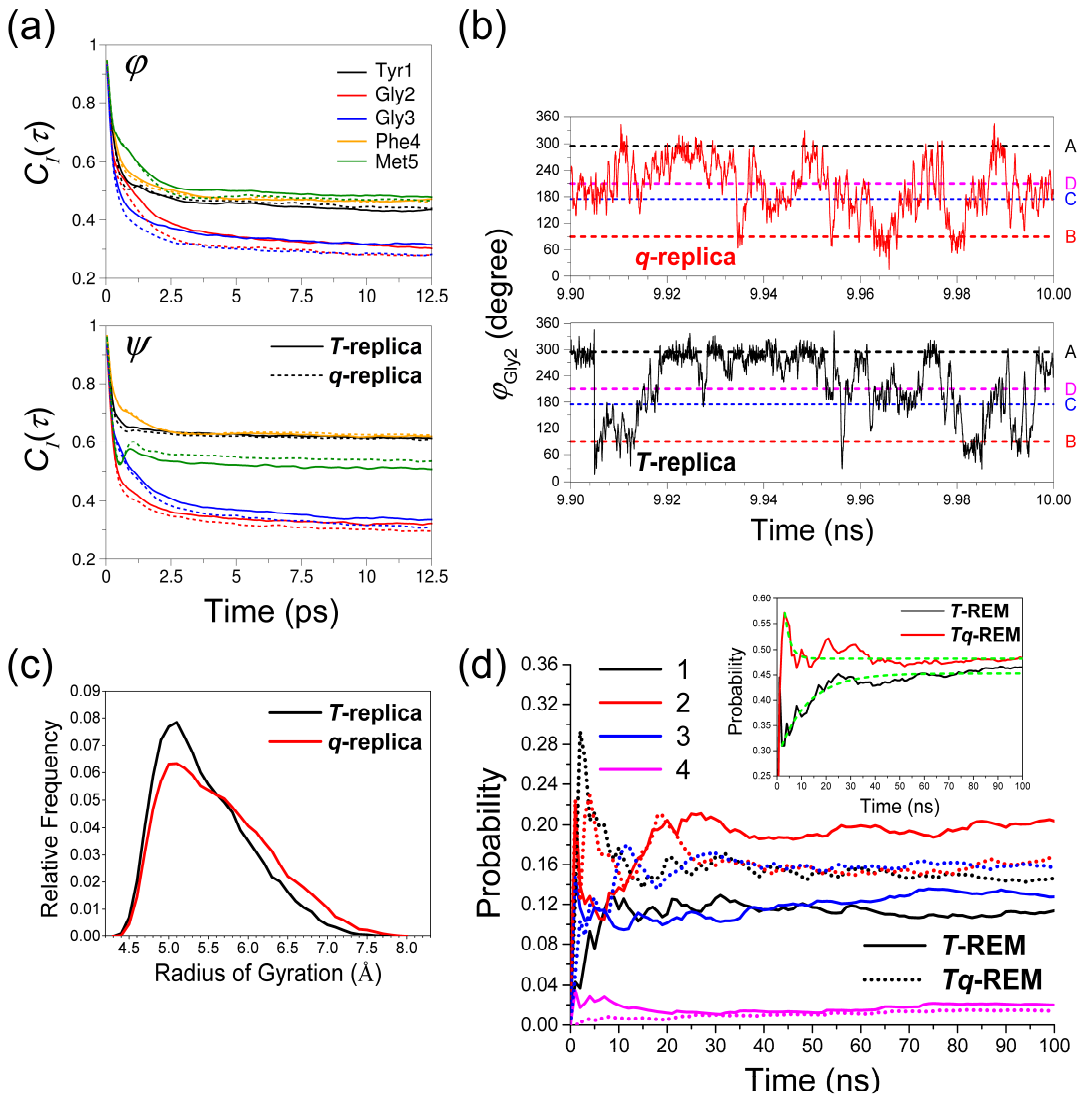


Figure 2.6.4. (a) Time auto-correlation functions of dihedral angles of Met-enkephalin, calculated for the trajectories corresponding to 5th replica of *T*-REM and *Tq*-REM. (b) Representative sampling behavior of φ_{Gly2} in *T*-replica and *q*-replica. Horizontal dash-dot lines indicate the φ values of local minima. (c) Distributions of radius of gyration of Met-enkephalin for *T*-replica and *q*-replica. (d) Time evolution of sampling probabilities of local basins.

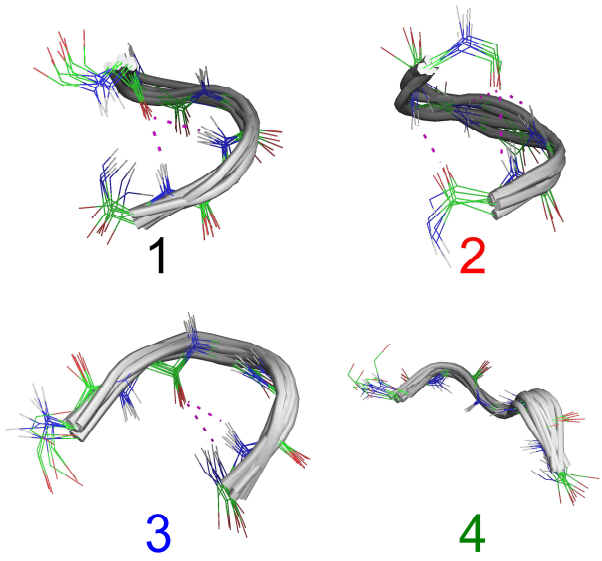


Figure 2.6.5. Representative structures of Met-enkephalin for the four local basins in the free energy surface as shown in **Figure 2.6.2**. Backbone hydrogen bonds are indicated by dashed lines.

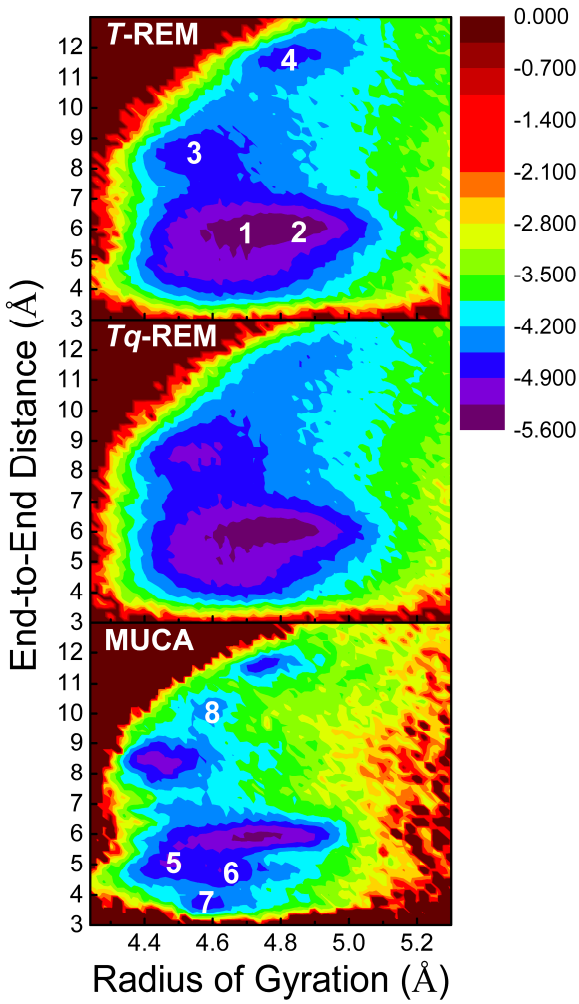


Figure 2.6.6. The free energy surfaces of *T*-REM, *Tq*-REM, and multicanonical simulation as a function of the radius of gyration and the end-to-end distance, obtained from whole trajectories at 300 K for Met-enkephalin. Basins of four local minimum structures are denoted 1 to 4 as given in **Figure 2.6.2**. Additional basins sampled in MUCA are also denoted 5 to 8.

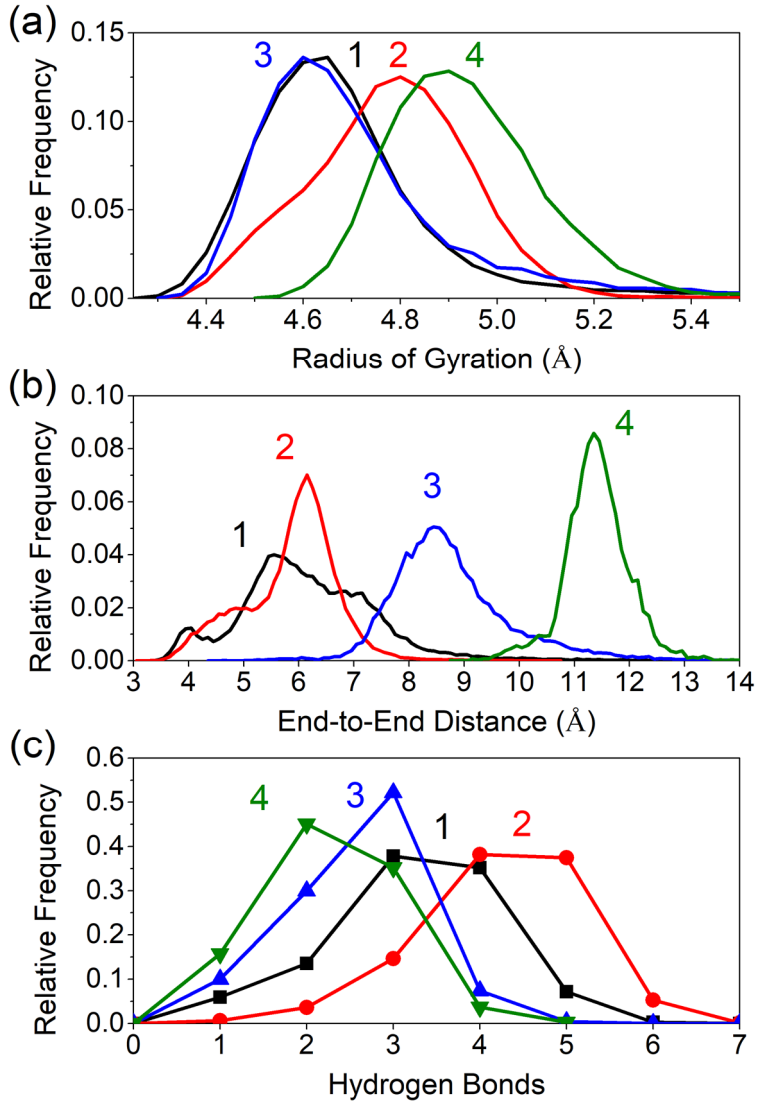


Figure 2.6.7. Distributions of (a) the radius of gyration, (b) the end-to-end distance, and (c) the number of hydrogen bonds for the four local basins in the free energy surfaces of *T*-REM and *Tq*-REM, as obtained from 100 ns trajectories at 300 K for Met-enkephalin.

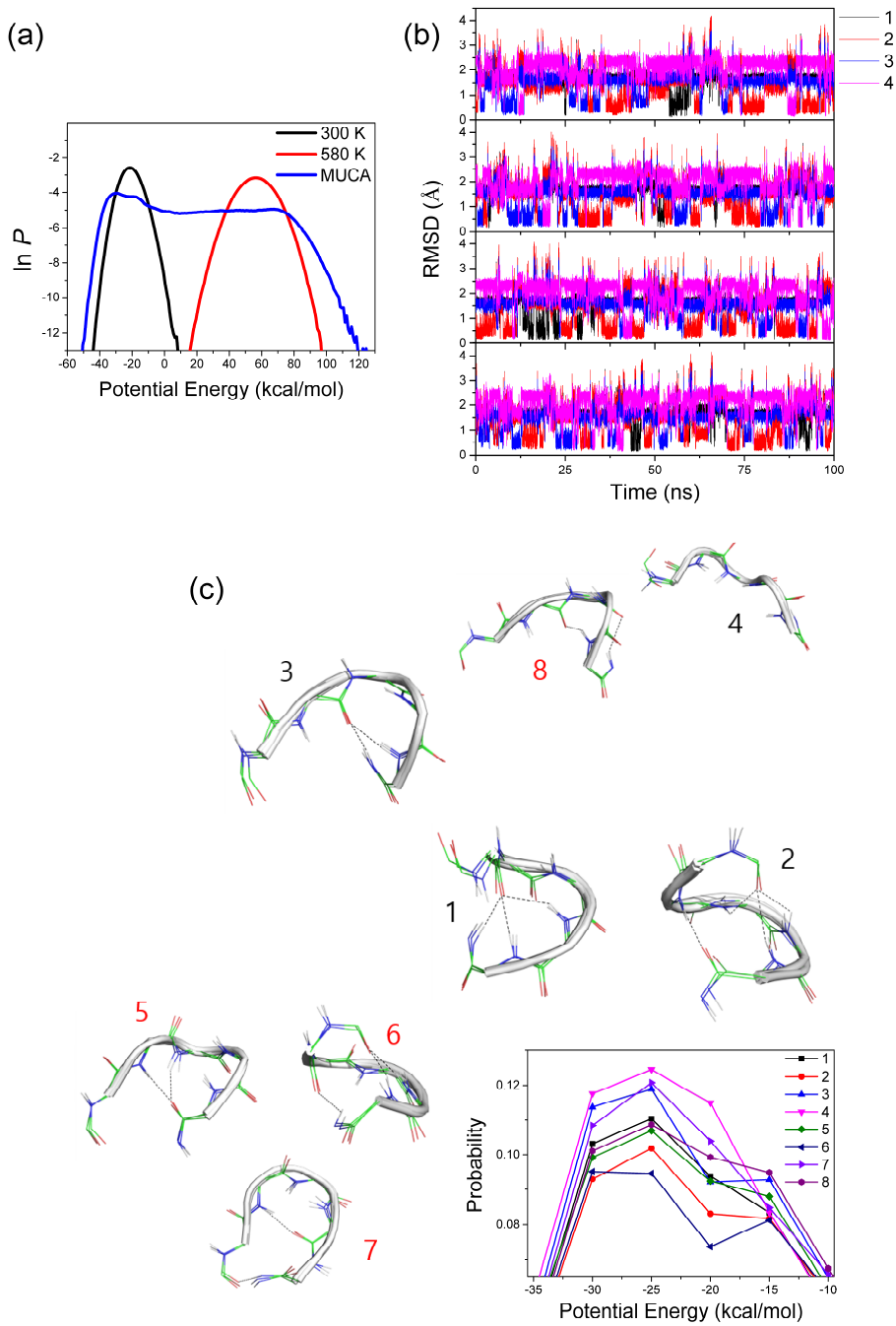


Figure 2.6.8. (a) Flat potential energy distribution obtained for multicanonical simulation. (b) Backbone RMSD plots with respect to local basin structures from *T*-REM and *Tq*-REM. (c) Structures and potential energy probabilities of various basin structures. Hydrogen bonds are indicated by dashed lines.

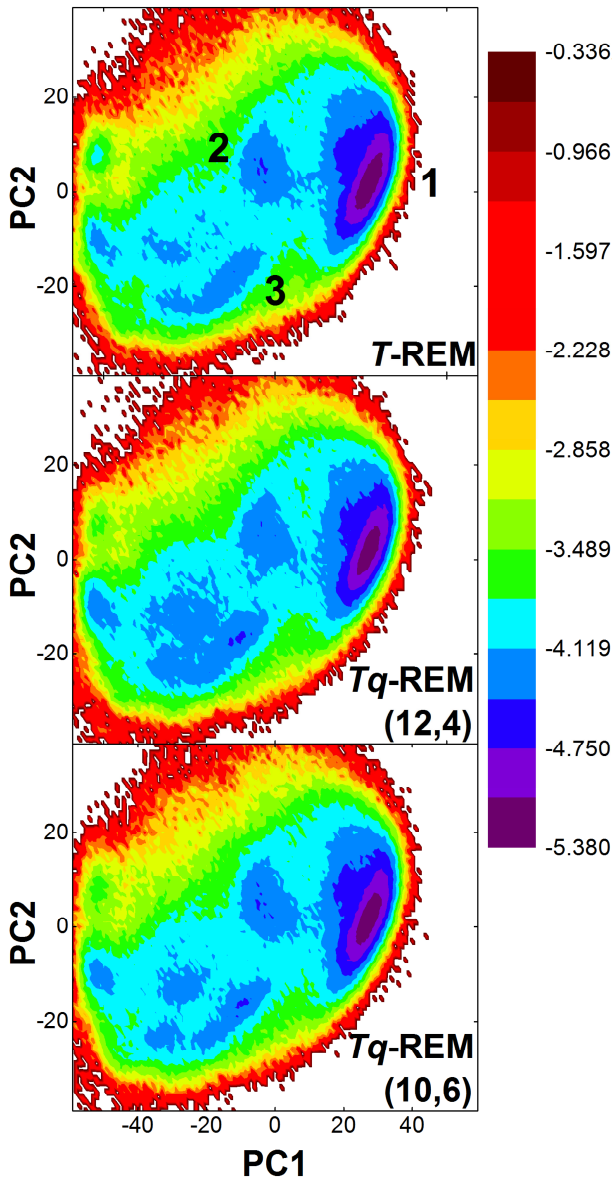


Figure 2.6.9. The free energy surfaces of *T*-REM and *Tq*-REM in PC space obtained from 200 ns trajectories at 300 K for $(AAQAA)_3$. Basins of three local minimum structures are denoted 1 to 3.

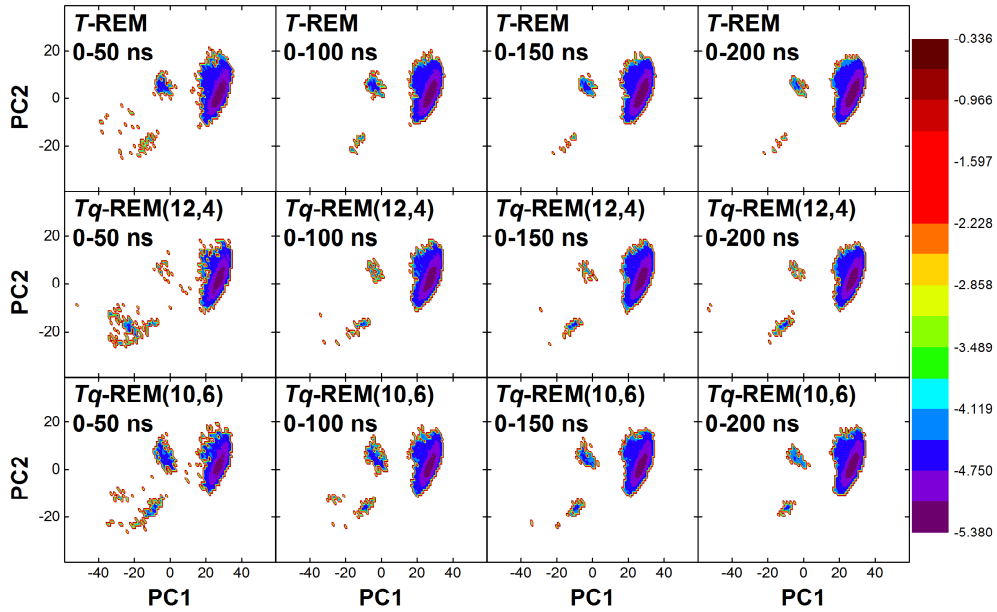


Figure 2.6.10. The evolutions of the free energy surfaces of *T*-REM and *Tq*-REM in PC space for different time windows during the 200 ns trajectories at 300 K for $(AAQAA)_3$. The free energy surface regions whose energies are lower than –4.35 kcal/mol are shown for clarity.

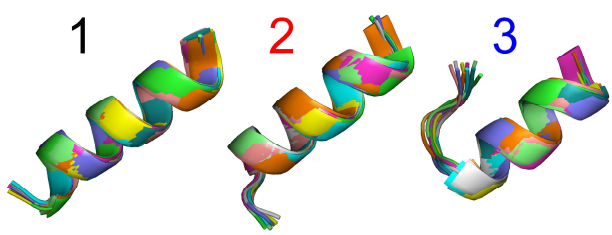


Figure 2.6.11. Representative structures of $(AAQAA)_3$ for the three local basins in the free energy surface as shown in **Figure 2.6.9**.

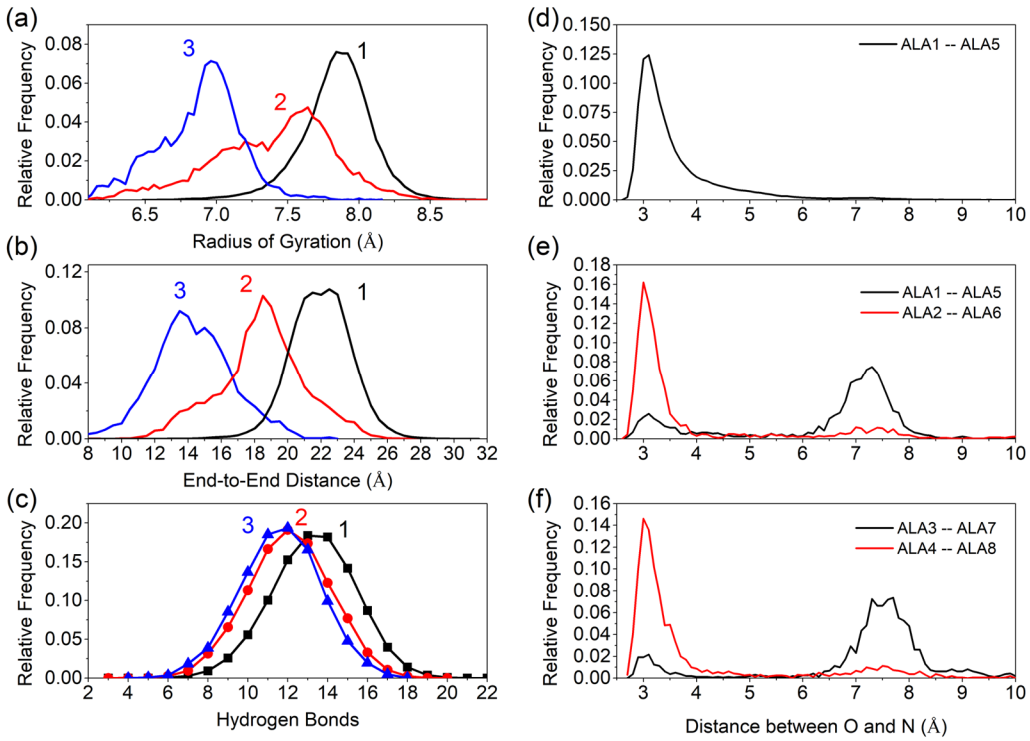


Figure 2.6.12. Distributions of (a) the radius of gyration, (b) the end-to-end distance, and (c) the number of hydrogen bonds for the three local basins in the free energy surfaces of *T*-REM and *Tq*-REM, as obtained from 200 ns trajectories at 300 K for $(AAQAA)_3$. Distributions of distances between backbone oxygen and nitrogen of the residues participating in hydrogen bonds, for the local minimum structures corresponding to (d) basin 1, (e) basin 2, and (f) basin 3. Backbone hydrogen bonds are sequentially formed from middle to N-terminal direction along the pathway of 3 → 2 → 1 indicating the folding of $(AAQAA)_3$ helix.

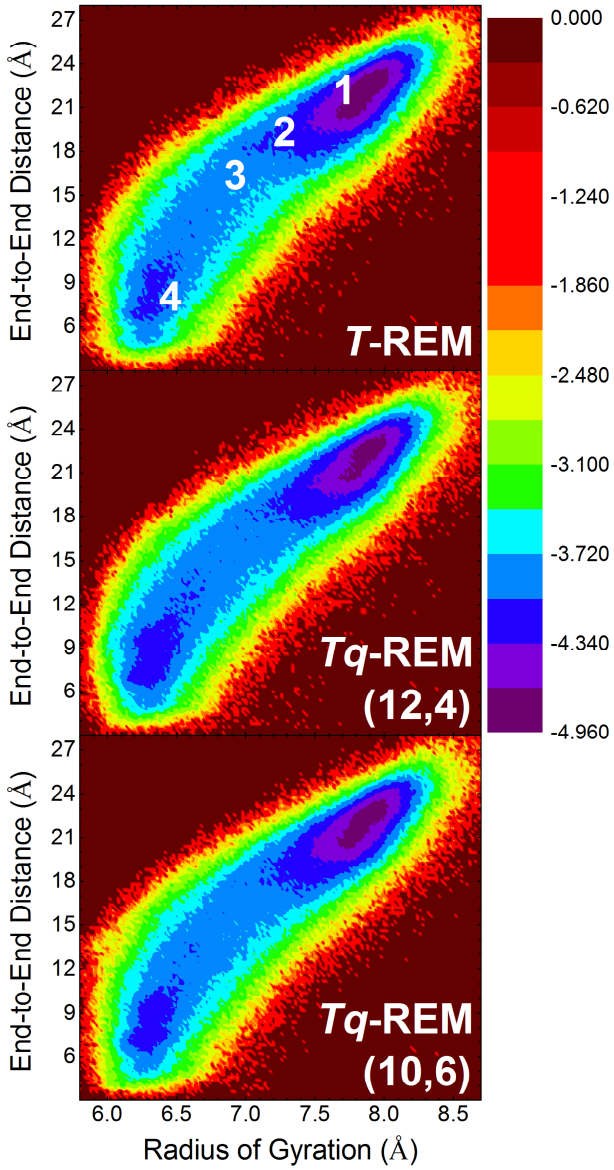


Figure 2.6.13. The free energy surfaces of *T*-REM and *Tq*-REM as a function of the radius of gyration and the end-to-end distance, obtained from 200 ns trajectories at 300 K for $(AAQAA)_3$. Basins of three local minimum structures are denoted 1 to 3 as given in **Figure 2.6.9**. Basin 4 corresponds to various collapsed structures.

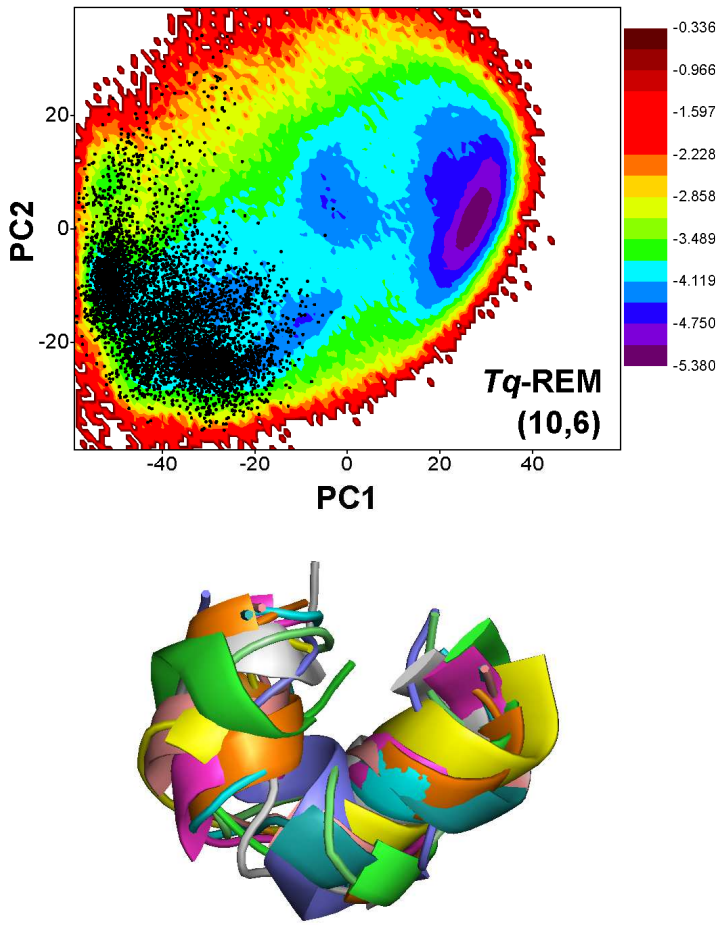


Figure 2.6.14. The free energy surfaces of *Tq*-REM(10,6) in PC space obtained from 200 ns trajectories at 300 K for $(AAQAA)_3$. Representative structures of $(AAQAA)_3$ for the local basin 4, as represented by black dots in the free energy surface.

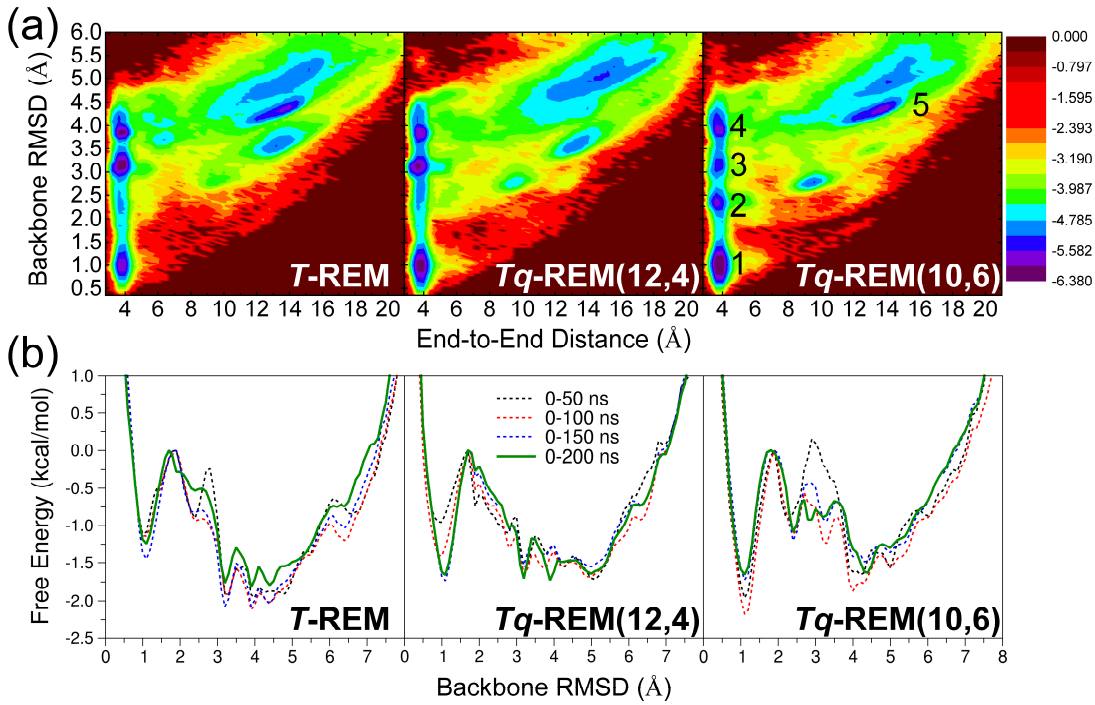


Figure 2.6.15. (a) The free energy surfaces of *T*-REM and *Tq*-REM as a function of the backbone RMSD and the end-to-end distance, obtained from 200 ns trajectories at 300 K for Trpzip2. Basins of five local minimum structures are denoted 1 to 5. (b) The free energy as a function of the backbone RMSD at different times during the simulations.

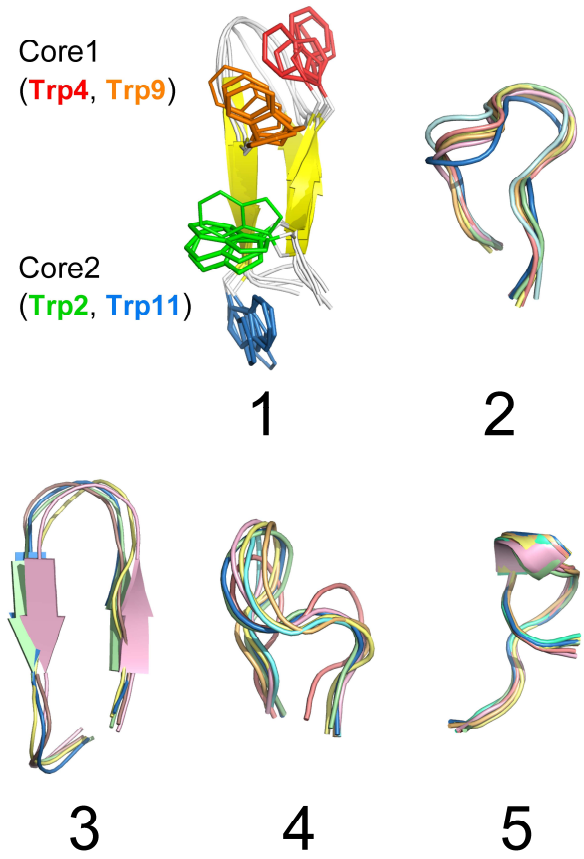


Figure 2.6.16. Representative structures of Trpzip2 for the five local basins in the free energy surface as shown in **Figure 2.6.15**.

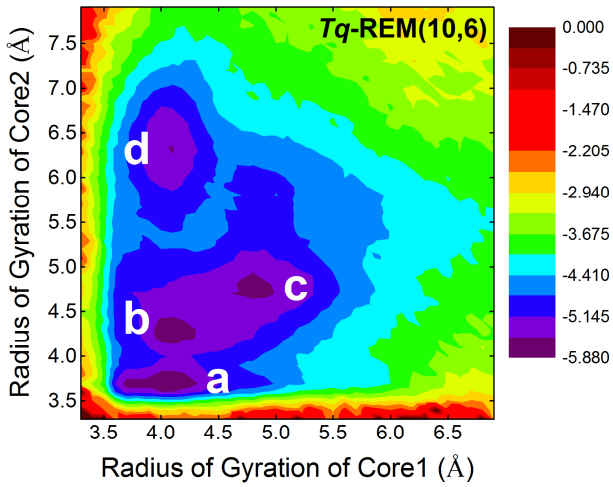


Figure 2.6.17. The free energy surface of *Tq*-REM(10,6) as a function of radius of gyration for the two hydrophobic cores (Core1: Trp4 and Trp9, Core2: Trp2 and Trp11), obtained from 200 ns trajectories at 300 K for Trpzip2. Basins of four local minimum structures are denoted a to d.

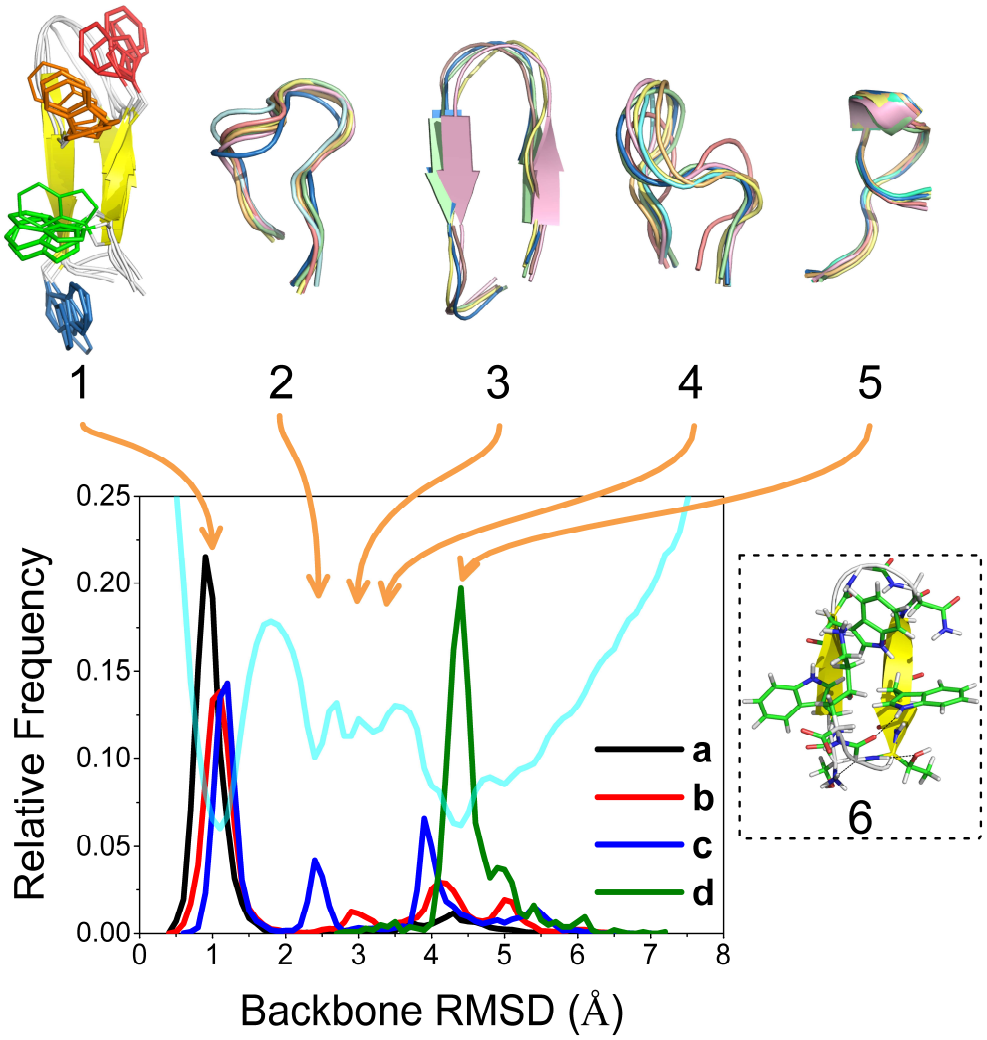


Figure 2.6.18. The distributions of the backbone RMSD of the structures corresponding to the four local basins (denoted a to d) in the free energy surface shown in **Figure 2.6.17**. The free energy as a function of the backbone RMSD, obtained from 200 ns trajectories at 300 K for Trpzip2, is also shown. The representative structures of Trpzip2, as shown in **Figure 2.6.16**, are related with the local minima in the free energy as a function of the backbone RMSD. The structure of off-folding pathway metastable state is denoted 6.

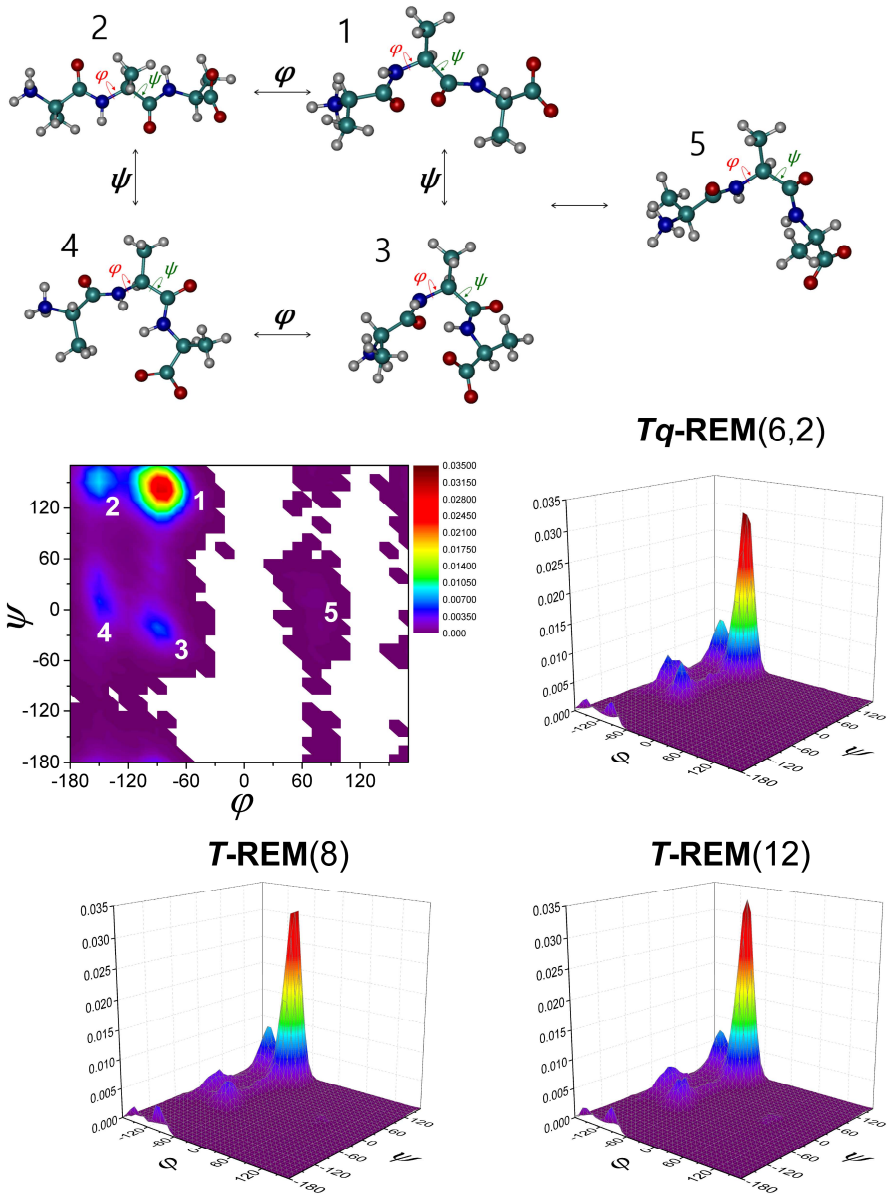


Figure 2.6.19. Representative structures of trialanine for the five local basins in the Ramachandran plots. Basins of five local minimum structures are denoted 1 to 5. Dihedral angles of central residue and transitions between basins are denoted. 3D graphs of Ramachandran distribution plots of each REMs show that *Tq*-REM(6,2) samples metastable α conformations more than two *T*-REMs.

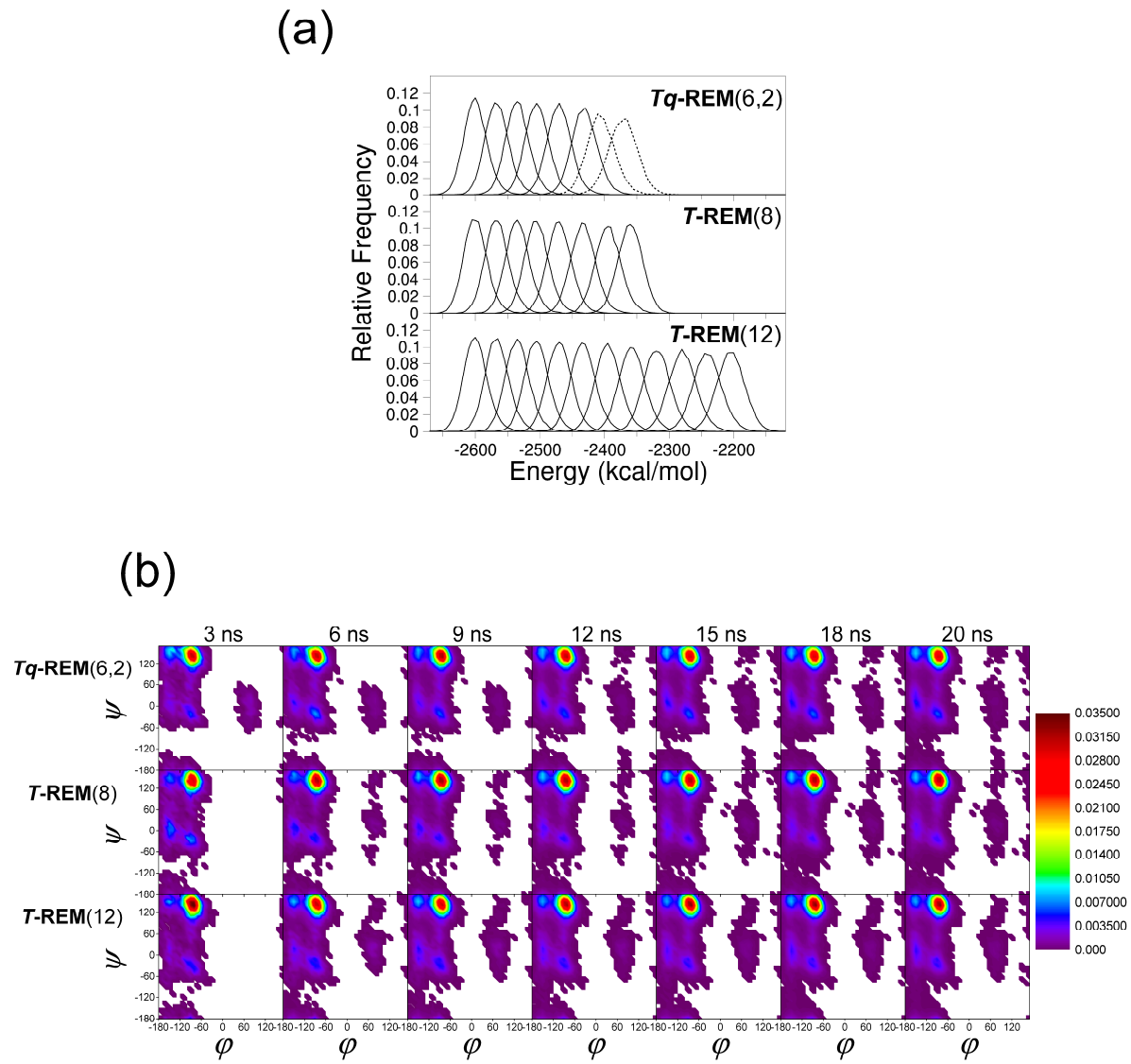


Figure 2.6.20. (a) The energy distributions of *Tq*-REM(6,2), *T*-REM(8), and *T*-REM(12). Solid lines are *T*-replicas and dot lines are *q*-replicas. (b) The time evolutions of the Ramachandran distributions of three REMs during 20 ns trajectories at 300 K for trialanine.

Chapter 3. Structure and Aggregation Pathway of Aβ_{42} Amyloid Protofibril

3.1. Introduction

Amyloid proteins are associated with neurodegenerative diseases; in particular, they are a major cause of amyloid fibril deposits in the neuronal cells of patients' brains.[62] Various neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, and type II diabetes, are related to specific amyloid proteins.[7, 63] Amyloid fibrils contain a characteristic β-spine structure in which antiparallel β-sheet structures are constructed by the formation of typical intramolecular β-turn-β motifs and intermolecular parallel β-sheets.[64]

Although there has been much research into the structure and function of amyloid fibrils, some issues, such as the role of these fibrils and the toxicity of oligomeric species, remain unclear.[65] It has been suggested through numerous in vitro experiments that off-pathway oligomeric species show cytotoxicity by disrupting membranes.[66] Conversely, it is possible that mature amyloid deposit-derived fibrils play a role as reservoirs of toxic amyloid oligomer species.[67] For this reason, amyloid oligomers are drawing strong interest as therapeutic targets for amyloid-related diseases.[68, 69] However, amyloid oligomers are structurally instable and heterogeneous, and there is minimal information about their structure, formation, and physiological effects.[70] Experimental and computational studies are highly useful to elucidate these topics.[18, 69, 71]

Aside from their toxicity, amyloid oligomers and fibrils are associated with amyloid fibril formation pathways either as intermediates or as final products. Solving the structure and determining the formation pathway of amyloid oligomers

and their fibril species are challenging issues owing to their high conformational plasticity and polymorphic nature based on their sensitive responses to environmental conditions.[72] Several recent studies have identified possible aggregation pathways, corresponding intermediate species, and principles surrounding the mechanisms of aggregation in monomeric, oligomeric, protofibrillar, and mature fibril species.[6, 73] In particular, several recent experiments successfully resolved the high-resolution fibril structures of several amyloid proteins.[74, 75, 76]

Amyloid beta peptide (Aβ) is known to be the major cause of Alzheimer's disease. Of the two Aβ isoforms most commonly associated with Alzheimer's disease, it has been suggested that Aβ_{42} is more toxic than Aβ_{40}, possibly because of its conformational features. Therefore, it is worthwhile to examine its structural features. Lührs (2005) suggested that Aβ_{42} has a U-shaped β-arch structure.[77] More recently, Xiao (2015) succeeded in resolving the high-resolution solid-state nuclear magnetic resonance (ssNMR) structure of Aβ_{42}, which unexpectedly showed S-shaped triple-β structure.[74] Subsequent ssNMR and cryogenic electron microscopy (cryo-EM) experiments showed similar, but not identical, S-shape structures with two attached protofibrils forming two-column fibril structures; this implies that Aβ_{42} amyloid fibrils may have polymorphic properties.[75]

Although these experimental results have provided valuable information on the fibrillar structure of Aβ_{42}, different aspects of its structural properties must be determined, especially the dynamic and thermodynamic properties of the fibril, as well as its interactions with water; this information can be used to better explain its conformational stability and mechanism of aggregation. In this paper, I performed molecular dynamics (MD) simulations to investigate the structural basis of the conformational stability and mechanisms of aggregation of Aβ_{42} at the molecular

level. To examine both the structural stability and aggregation pathway of S-shaped triple-β structure model of Aβ_{42}, I performed several different simulation methods, including straightforward extensive MD simulation, steered MD, and replica-exchange MD (REMD). By carefully analyzing the MD simulation trajectories, I clarified the structural features of the Aβ_{42} protofibril motif and related them to the fibril stability. Based on these results, I tried to interpret the trajectories of steered MD and REMD simulations and predict the aggregation pathway of Aβ_{42} fibril. These results could also provide a plausible explanation for the high toxicity of Aβ_{42} species.

3.2. Methods

MD simulations of Aβ₄₂ protofibril motif The ssNMR structure of the S-shaped triple-β structure published by Xiao (2015) was used, which includes residues 11–42 of the peptide (PDB ID: 2MXU).[74] Six Aβ_{42} fibril chains were taken from PDB and solvated in a dodecahedron box containing 11700 molecules of TIP3P water[78] with 47 Na^+ and 35 Cl^- counter ions to form a neutral 150 mM NaCl environment. CHARMM22/CMAP force field[79] was used for Aβ_{42}. After a short minimization, the system was gradually heated for 6.2 ns from 0 K to 310 K with position restraints on backbones (force constant k = 100 kJ/mol/nm^2). For pre-equilibration, several 2 ns runs were performed at 310 K with restraints of decreasing orders of strength (k = 100, 50, 30, 10 kJ/mol/nm^2). The system was further equilibrated at 310 K for 5 ns without position restraints to obtain the initial conformations for the product run. For all heating and equilibration steps throughout this paper, the isothermal–isobaric ensemble (NPT ensemble) with a velocity-

rescaling thermostat[80] and a Berendsen barostat[41] was used. After equilibration, a product run was performed for 10 μs, and the trajectory was saved every 2 ps for analysis. All product runs were performed using the NPT ensemble with a velocity-rescaling thermostat and a Parrinello-Rahman barostat,[81] and a 2 fs time step was used with the LINCS algorithm[82] for the constraints of bonds related with hydrogen atoms. GROMACS 5.1 package[83] was used for all simulations.

Steered MD simulations of S-shaped triple-β structure For steered MD simulations, an initial configuration with a trajectory of 10 μs was used for the S-shaped triple-β structure. Five adjacent chains of the protofibril motif were arranged in a 67 Å × 62 Å × 135 Å rectangular periodic box and solvated in 17544 molecules of TIP3P water with 61 Na^+, and 51 Cl^-. The fibril axis was aligned to the *z*-axis, and the protofibril motif was placed at a distance of 15 Å from one *xy*-plane of the box. After a short minimization and gradual heating to 310 K with position restraints on backbones (k = 10 kJ/mol/nm^2), a 2.5 ns equilibration was performed at 310 K with weaker restraints (k = 2 kJ/mol/nm^2), the final 2 ns of which were used to sample ten different initial configurations for ten independent runs. As described above, solvation, counter ion addition, energy minimization, and gradual heating were performed for ten independent systems and equilibrated for 200 ps at 310 K with restraints (k = 2 kJ/mol/nm^2). Steered MD simulations were performed with position restraints on the backbones of four neighboring chains (k = 10 kJ/mol/nm^2). The fifth chain was pulled along the z-axis with a constant pulling rate. For each of the ten initial configurations, steered MD simulations were performed with three different pulling rates each (k_{pulling} = 0.002, 0.001, 0.0005 nm/ps), producing thirty trajectories in total. For β-sheet rupture analysis, as described in detail below, I performed an additional 200 ps run using the same initial configuration of the steered

MD simulation, but without pulling, in order to obtain a stable trajectory. Trajectories were saved every 1 ps for analysis.

To analyze the β-sheet rupture pattern between adjacent chains, the changes in distance corresponding to backbone hydrogen bonds and sidechain interactions were calculated. At first, the distances of all interaction pairs were calculated from the 200 ps stable trajectory. For each backbone hydrogen bond, the distance between the participating N and O atoms was calculated. For each sidechain interaction, the distance between the heavy atomic centers of mass of two interacting sidechains was calculated. To obtain a clear distinction between bonded and nonbonded states, the logistic function $\varphi(r) = \frac{1}{1+e^{(r-r_0)/\sigma}}$ was used, where r is the distance between two interacting objects, and r_0 and σ were determined by considering the distance distribution of each bonded state: $\varphi(r) = 1$ for a bonded state, and $\varphi(r) = 0$ for a nonbonded state.

REMD simulation of S-shaped triple-β structure REMD[22] simulation was performed to examine the docking process of Aβ_{42} fibril formation. An initial configuration was designed in which the Aβ_{42} monomer was placed at a distance of 25 Å from the fibril motif structure. The system was gradually heated for 500 ps from 310 K to 500 K with position restraints on the backbones (k = 10 kJ/mol/nm^2). Equilibration was performed for 10 ns at 500 K with position restraints on the backbones of the protofibril motif (k = 10 kJ/mol/nm^2) and a fixed center of mass for the isolated monomer (k = 1000 kJ/mol/nm^2). Based on the equilibration trajectory, eighty different initial configurations for the REMD replica were prepared. The fibril axes of the initial configurations were aligned to the z-axis of a periodic 64 Å × 58 Å × 102.7 Å rectangular box containing 9642 molecules of TIP3P water, 40 Na^+, and

30 Cl^-. Temperature distribution was initially generated using a temperature generator (https://virtualchemistry.org/remd-temperature-generator/),[84] and then modified using several 200 ps REMD test runs to obtain the appropriate exchange ratio. The final temperature range was 303.13 K to 486.15 K with an average exchange ratio of 0.287. The initial configuration of each replica was minimized and gradually heated for 500 ps from 0 K to each target temperature and equilibrated for 500 ps with restraints (k = 10 kJ/mol/nm^2). The product run of the REMD simulation was performed for 250 ns with position restraints on the backbones of protofibril (k = 10 kJ/mol/nm^2). Exchanges between adjacent replicas were attempted every 2 ps. Trajectory was saved every 2 ps for analysis.

3.3. Results and Discussion

Recently, several high-resolution structures of the Aβ_{42} fibril motif were experimentally identified by ssNMR and cryo-EM.[74, 75, 76] In 2015, the first detailed Aβ_{42} fibril structure was published using ssNMR[74] (PDB ID: 2MXU), and it contained a triple-β motif (β-turn-β-turn-β) in contrast to its previously suggested structure.[77] In 2016, Hansmann *et al*. performed a simulation study using this structure and reported the stability of the triple-β motif of Aβ_{42}.[85] They used both Amber99sb-ildn and CHARMM22/CMAP force fields to run 500 ns and 200 ns simulations, respectively. Overall shape of triple-β was preserved within their simulation, although CHARMM22/CMAP showed a more ordered backbone arrangement. Although they mainly used their Amber99sb-ildn data, it was reported that CHARMM22/CMAP was the most accurate compared to experimental data for intrinsically disordered proteins.[86]

Stability and Structure of Triple-β Motif To investigate the structural stability and dynamics of the Aβ_{42} protofibril motif, I used the same structure (PDB ID: 2MXU) and performed an extensive 10 μs MD simulation using the CHARMM22/CMAP force field. Compared to the trajectory of the Amber99sb-ildn force field shown by Hansmann *et al.*,[85] the structure given by the CHARMM22/CMAP force field showed more ordered backbone arrangements and hydrophobic sidechain packings (**Figure 3.5.1a**). The Root Mean Square Deviation (RMSD) values at 500 ns were ~3.4 Å for CHARMM22/CMAP and ~5 Å for Amber99sb-ildn, respectively (**Figure 3.5.1c**). Furthermore, the structural features of the triple-β motif were highly preserved throughout the 10 μs simulation period (**Figure 3.5.1b**). In particular, the core residues (19–42) showed highly stable ordered arrangements of both the backbones and sidechains throughout the simulation (**Figure 3.5.1d**). By analyzing the simulation trajectory of equilibrium ensemble of this structure, I identified detailed sidechain interactions that stabilized the hydrophobic core structures of the triple-β shape Aβ_{42} fibril motif (**Figure 3.5.2** and **Table 3.5.1**). Interactions forming two hydrophobic cores – HP core I and HP core II – are designated by colors consistent with **Figure 3.5.2**. HP core I has two sub-cores named as HP core Ia and HP core Ib in the table. HP contact means hydrophobic interactions not included in HP core I and II. Two polar interactions are hydrogen bonding between Asn27-Asn27 sidechains and salt bridge between C-terminal carboxylic group and Lys28 sidechain. Ala30, Ile31, Ile32, and Met35 are pivotal residues in the two hydrophobic cores. Note that inter-chain interaction pairs are written in order of inner chain residue–outer chain residue. For example, Val39–Ile31 means that Val39 of the inner chain interacts with Ile31 of the outer chain.

Before describing these residual interactions in detail, it is important to note the

differences between the experimental and simulated structures. The locations of the three β-sheet sequences are slightly different, as shown in **Figure 3.5.3**, possibly because of the different force fields and solvation conditions used. **Figure 3.5.4** shows that the essential sidechain packing structures of the hydrophobic cores were almost identical. It was shown that the simulated structure had more compact hydrophobic interactions in the presence of water, which was absent in the ssNMR structure. Therefore, it can be concluded that our simulated structure was consistent with the experimentally determined structure.

As shown in **Figure 3.5.2** and **Table 3.5.1**, the triple-β structure of Aβ_{42} is primarily made up of three β-sheet regions (β_1, β_2, β_3) and two hydrophobic cores formed between β_1 and β_2, and between β_2 and β_3, designated HP core I and HP core II, respectively. HP core I is composed of two sub-cores, highlighted by the blue and yellow circles in **Figure 3.5.2**. Each sub-core has a structure of aromatic sidechains surrounding a small aliphatic sidechain at a pivotal position. In detail, His14, Leu17, and Phe19 from two adjacent chains formed one sub-core, with Ile32 as a pivotal residue; similarly, Phe19 and Phe20 from two adjacent chains formed the other sub-core, with Ala30 as a pivotal residue where two Phe19 residues are shared between the two sub-cores to form HP core I. In HP core II (highlighted by the light pink circle in **Figure 3.5.2**), Ile31 and Met35 are located between two Val39 sidechains to form a compact hydrophobic core structure. In addition to the two hydrophobic cores, a hydrogen bond between two Asn27 stabilizes HP core I, while a salt bridge formed between Lys28 and the C-terminal carboxylic group stabilizes HP core II. Accordingly, the remarkable structural stability of triple-β motif can be explained through this highly compact and complex structure of residual interactions.

N-terminal Fluctuation and Effect of Water in Association with

Polymorphism In addition to the major hydrophobic core structure described above, Val12 and Leu34 residues formed hydrophobic contact through both inter- and intra-chain interactions which stabilizes the N-terminal region in an aqueous environment. However, this contact was relatively unstable compared to HP core I and II, and showed flexible motion during the simulation (**Figure 3.5.2a** and **Figure 3.5.2d**). **Figure 3.5.5** and **Figure 3.5.6** show the overall conformational dynamics of our simulation trajectory over 10 μs. Although N-terminal hydrophobic contact is relatively weak compared to that of the major hydrophobic cores, it showed good stability in our simulation trajectory. The distance between Val12 and Leu34 slightly increased at 1170 ns while maintaining overall contact to optimize the conformation of HP core I in an aqueous environment. This structure was disrupted at 2500 ns because of the breaking of the backbone hydrogen bonds in the adjacent outermost chain (**Figure 3.5.6b**). Thus, it is expected that N-terminal hydrophobic contact in the triple-β protofilament would not be spontaneously broken in an aqueous environment.

Recently reported experimental structures showed slightly different structures for this N-terminal contact,[74, 75, 76] implying that this unstable contact provides flexibility to allow conformational changes in response to different environments: In two ssNMR structures, Met35 sidechain is flipped outside and Val36 is flipped inside which form a steric zipper interface with the adjacent protofilament, from which the linear β-sheet geometry of N-terminal sequence is naturally induced. On the other hand, the cryo-EM structure maintains the same sidechain orientations of Met35, and Val36 with the structure used in our simulation, and instead, Val39 and Ile41 sidechains are inverted outside to form hydrophobic zipper interface between two protofilaments, which is relatively weak compared to ssNMR structures. Because of

this, N-terminal tail sequence is folded toward steric zipper structure allowing Lys28 to form a salt bridge with Asp1 of adjacent protofilament, which stabilizes this polymorphic structure.

Thirumalai *et al*.[87] suggested that a two-step model which is postulated for amyloid protein crystallization (that is, protofilament formation) will hold for the formation of higher order amyloid structure. In this model, disordered oligomers are first formed which produce a protein-rich droplets. The aggregation-prone states, N* in these droplets could contain varying number of water molecules according to which conformation it has among ensemble of N* states. At second stage, when the size of droplet becomes large, the different N* oligomers will form different geometries of distinct fibril structures. These distinct fibril structures may contain embedded water or not, which are located at protofibril interface. The existence of embedded water depends on the interface formed between protofilaments which is determined by conformation of N* forming the protofibril. Once the fibril morphology containing embedded water is formed, it could be metastable or proceed to expel the water and form dry interface at slow rate. Whether N* with water-embedding conformation arises or not might depends on the extent of hydration when protein-rich droplet is formed.

It can be speculated that the differences in triple-β core structures and morphologies of fibrils of ssNMR and cryo-EM are explained through the two-step model described above. According to a certain hydration condition, N* conformation resembling one of these two triple-β core structures could be selected as a predominant oligomer species which drives the formation of protofilaments of corresponding geometry. When the triple-β structure corresponding to ssNMR data is selected, N-terminal stretch takes a linear β-sheet geometry in order to maximize

the hydrophobic interface between two protofilaments. On the other hand, when the triple-β structure corresponding to cryo-EM data is selected, N-terminal stretch is bent toward the steric interface in order to form a salt bridge between Asp1 and Lys28 and provide additional stabilization to protofilament interface. Interestingly, this salt bridge formation results in flat, highly hydrophilic surface (Asp7, Arg5, Glu3, Asp1, and Lys28) exposed to solvent environments, which may be favored in aqueous condition of cryo-EM. Therefore, it can be said that the conformational flexibility of N-terminal sequence is required in order to achieve polymorphism of Aβ_{42} amyloid fibrils.

In order to examine the flexibility of Aβ_{42} peptide in our simulation, I calculated average Root Mean Square Fluctuations (RMSF) of each residue for three different regions of 20 ns length in trajectory. At 800–820 ns, hydrophobic contact between Val12 and Leu34 exists and stabilizes N-terminal structure whereas at 3600–3620 ns, contact between Val12 and Leu34 is broken which makes β-sheet plane of N-terminal sequence more flexible. At 5650–5670 ns, hydrogen bonds forming β-sheet backbone structure in N-terminus are disrupted and allows a full conformational flexibility to N-terminal stretch. In **Figure 3.5.7**, I can first observe that Val12 and His13 residues show large increase of RMSF values as hydrophobic contact and backbone β-sheet hydrogen bonds are disrupted in order. In contrast, RMSF of His14 increases when Val12-Leu34 hydrophobic contact is disrupted while it does not increase any more when backbone β-sheet is broken, indicating that His14 is participating to HP core Ia hydrophobic core structure. Detailed conformational changes and concurrent changes of hydration states (radial distribution function of water) are shown in **Figure 3.5.8** and **Figure 3.5.9**. Although His14 manages to maintain the contact with Ile32 in HP core Ia in our trajectory, it is expected that this

contact can be easily broken which allows the flipping of sidechain orientations for Val12, His13, and His14. The conformational flexibility of this sequence might be important to the formation of the structures revealed by ssNMR and cryo-EM experiments: In ssNMR, sidechain orientations of these three residues are inverted with respect to original triple-β structure to form linear straight β-sheet backbone geometry of N-terminal tail, allowing maximum area of hydrophobic interface between two protofilaments. On the other hand, sidechain flipping of these three residues in cryo-EM structure makes N-terminal tail (residues 1 to 11) folded toward the C-terminus of the other protofilament, inducing the formation of Asp1-Lys28 salt bridge.

Gln15 and Lys16 show relatively high RMSF values in **Figure 3.5.7**, which is because the two sidechains are solvent-exposed by outward location with distorted backbone shape from standard β-sheet geometry (**Figure 3.5.2**). This structural feature provides additional instability to N-terminal stretch to make feasible the above-described conformational changes of N-terminal stretch. Similarly, Leu34 has also relatively high RMSF values because of its outward sidechain orientation and highly curved backbone geometry. I can expect that the flipping of Met35 and Val36 is facilitated by this structural instability of Leu34 and consequences the steric zipper interface in ssNMR structure.

Dynamics of Chains at Fibril Ends RMSD curves show that the two external chains were relatively unstable compared to the four internal chains (**Figure 3.5.5**). In the outermost chains, residues 21–29 and residues 38–42 were unstable and showed disordered conformations. Residues 38–42 make up a short sequence containing a C-terminal salt bridge that allows the molecule to restore its inter-chain interactions. Conversely, the backbone geometry of residues 21–29 was not adequate

to form a stable β-sheet hydrogen bond as well as stable sidechain interactions, making this region the least stable in the overall sequence. Therefore, the structure of this region is the first to be disrupted upon exposure to the fibril end, and is not easily restored (**Figure 3.5.6b**).

These local conformational fluctuations might be associated with the relative thermodynamic stability of local inter- and intra-chain interactions composed of backbone hydrogen bonds and sidechain packing. However, knowing the way in which a certain region of the hydrophobic core or backbone β-sheet is formed or disrupted would allow us to more precisely demonstrate its relative stability to other regions. In order to examine this, I performed pulling simulations, as described later in this paper.

Prediction of Lock Phase Process by Pulling Simulation As mentioned above, the structure of residues 21–29 is weak and relatively easily separated from interactions with adjacent chains and forms an unstructured or helical geometry, as shown in **Figure 3.5.1b**. In addition, the C-terminal of the peptide also underwent the disruption and reformation of hydrogen bonds along with the adjacent backbone (**Figure 3.5.6b**). These features of equilibrium conformational dynamics imply the lock phase mechanism in the dock-and-lock process of aggregation.[88] This lock phase mechanism would be shown by how the various residual interactions are grouped spatially, for example, by the two sub-cores of HP core I shown in **Figure 3.5.2**, and by how the relative stabilities between these interacting groups are determined.

In order to investigate this, I performed pulling simulation using steered MD, and obtained force extension curves for disaggregation of Aβ_{42} chains from the fibril motif. Ten independent pulling simulations were performed for each of three

different pulling rates: $k_{pulling}$ = 0.002 nm/ps, 0.001 nm/ps, and 0.0005 nm/ps. **Figure 3.5.10** shows the force curves of these pulling simulations. From these thirty trajectories, I calculated the residue-residue distances for both backbone hydrogen bonding and sidechain contact. By transforming the distance to a logistic function $\varphi(r)$, I obtained the rupture patterns for the thirty trajectories to clearly visualize the disruption of each backbone or sidechain interaction (**Figure 3.5.11**). These patterns were classified into several categories according to their shape. From **Table 3.5.2**, I identified that the major pattern appeared at $k_{pulling}$ = 0.0005 nm/ps where the deviation in peak size was the smallest. **Figure 3.5.11a** shows this major disruption pattern. I calculated both backbone and sidechain interaction distances to examine the order of breaking between them. Overall, it seems that backbone interactions were broken first, followed by the breaking of involved sidechains, although the difference was slight. The major pattern was composed of three interaction groups, shown as three boxes in **Figure 3.5.11a**. Compared to **Figure 3.5.2**, it was clear that the largest group corresponded to the disruption process of HP core I, while the other two groups corresponded to the disruption process of HP core II. This indicates that the disruption of HP core I occurred first in a collective manner in which residues 21–23 was more unstable and the polar Asn27–Asn27 hydrogen bond was the most persistent. Once this collective rupture of HP core I structure occurred, the remaining backbone interactions of the β_2-turn-β_3 were sequentially broken.

The force curve in **Figure 3.5.11b** clearly shows that the collective rupture of HP core I was the most difficult step. As I already mentioned, the disruption of this region drives conformational changes in the separated backbone sequence into an α-helix structure, which is the stable conformation in an isolated state. Therefore, the reverse process was assumed to be a part of the lock pathway. In other words, helix

in the β_2 region of the docked Aβ_{42} peptide were unfolded to construct the β-sheet backbone and form HP core I. In addition, the reverse process of the sequential rupture of the β_2-turn-β_3 is expected to occur before the unfolding of the β_2 helix. **Figure 3.5.12** shows second major pattern, which is different from **Figure 3.5.11** in that sequential disruption occurred at both sides of the salt bridge residues. Trajectories of the major and the second major rupture patterns are shown in **Figure 3.5.13**. **Table 3.5.2** indicates that this difference was because of different pulling rates.

Prediction of Docking Process by REMD Simulation Up to now, I performed equilibrium simulations and identified the stability and detailed structure of residue-residue interactions which form the backbone geometry and hydrophobic cores. Next, I performed pulling simulations to identify interacting groups and their disaggregation behaviors, from which I predicted the mechanism of the lock phase process. In this section, I performed REMD simulations to study the docking pathway of the Aβ_{42} peptide to the protofibril motif. **Figure 3.5.14** shows the overall behavior of the Aβ_{42} peptide monomer during the docking process. I calculated the RMSD of the monomer with respect to the triple-β conformation in order to monitor the occurrence of low-RMSD ("S-like") conformations; their positions are denoted by the numbers (1) to (4) and the green windows in **Figure 3.5.14**. I examined the configurational features of these four S-like states, summarized as follows.

(1) In an isolated state, S-like conformations are mainly in a monomeric conformational structure. As explained in the following analysis, this ensemble shares a common structural feature with an early-stage configuration of the lock phase. Thus, this monomeric ensemble could be considered as an aggregation-prone state. (2) When the monomer comes into contact with the hydrophilic surface of the

fibril, it forms stable interactions by maintaining its conformation in a manner similar to that of the isolated structure, which also has a hydrophilic surface. (3) When the monomer comes into contact with the surface of hydrophobic residues, β_2 forms a partial helix. This conformational transition allows close contact between the unfolded sequence of β_2 and the hydrophobic surface of the fibril, which plays an essential role as an anchor to initiate the lock phase. Additionally, it allows the approaching of the C-terminus of the monomer to the C-terminus of adjacent fibril peptides in order to form salt bridge interactions. (4) When two C-termini are within a certain distance because of the conformational changes described above, the formation of β-sheet structure may occur transiently in the β_3 region.

It should be noted that the conformational changes (1) to (4) do not exactly represent the docking pathway because only S-like conformations were extracted and monitored for this mechanism. For example, in the early stage of docking between (1) and (2), the RMSD curve showed a convex shape, while the distance curve showed a concave shape. This behavior can be more clearly seen in **Figure 3.5.15**. The RMSD curve in **Figure 3.5.14** does not explain this behavior at the early stage of the docking process. Interestingly, the dynamics shown in **Figure 3.5.15a** are analogous to the motion of soft spheres approaching each other. When they are apart from each other, attractive interactions accelerate docking, while a repulsive force acts when contact becomes too close; this eventually leads to an equilibrium distance. In addition, the RMSD of two Aβ_{42} chains (approaching chain + chain at fibril surface) with respect to the triple-β conformation showed that this distance change was correlated to the formation of fibril configurations (**Figure 3.5.15b**).

To more accurately explain this, I calculated the distances from the fibril surface and the secondary structure contents for β_1, β_2, and β_3 separately. The distance plots

in **Figure 3.5.16** show that the concave shape described above was definitely derived from the dynamics of the β_1 sequence. β_2 preserves a constant distance with the fibril surface from the beginning of the contact period. The distance between β_3 and the fibril surface consistently gradually decreased after the initiation of the contact. The time evolution of the α-helix content (**Figure 3.5.16**) explains the behaviors of the distance changes for β_1 and β_2. In β_1, α-helix content gradually developed until 50 ns. The concave shape of the distance change of β_1 during this time interval was therefore induced by α-helix formation in the β_1 sequence. That is, the transition of the secondary structure of the β_1 sequence from a β-sheet-like to an α-helix conformation induced an increase in the center of mass distance between the β_1 sequence and fibril surface. On the other hand, the α-helix content of β_2 sequence developed rapidly to reach a stable state at the beginning of the contact, which explains the shape of the distance curve of β_2. Finally, the α-helix content in the β_3 sequence remained low throughout the simulation, and the reason for the slope of its distance plot was not clear. Instead, **Figure 3.5.17** shows that the β-sheet content of the β_3 sequence dramatically increased after contact with the fibril surface. Therefore, the slope of distance plot in β_3 (**Figure 3.5.16**) was because of this conformational change.

After showing that the behaviors of the distance changes and secondary structures were correlated with fibril formation dynamics, the docking pathway can be accurately summarized using these observations. In an isolated state, the Aβ_{42} peptide prefers an S-like conformation to optimize hydrophobic interactions. When contact with the fibril surface occurs, the β_2 sequence undergoes a rapid conformational transition to adopt a partially folded conformation, as indicated by the changes in its secondary structure (**Figure 3.5.16** and **Figure 3.5.17**). This

partially folded conformation plays an important role in the initiation of the lock phase. **Figure 3.5.18** shows a more detailed depiction of this change. It can be seen in the secondary structure plot in **Figure 3.5.18a** that the β_2 sequence can be divided into three parts: Glu22–Asn27, Lys28–Ala30, and Ile31–Gly33. When contact occurs, helix begins to develop for Glu22–Asn27 and Ile31–Gly33, but not for Lys28–Ala30. This difference in secondary structures results in a specific geometry, as shown in **Figure 3.5.18b**. By adopting an approximate helix-turn-helix geometry, β_2 arranges the positions of β_1 and β_3 in the docked monomer to be in close proximity to their counterparts on the fibril surface. In addition, the unfolded stretch of β_2 sequence (Lys28–Ala30) can maintain close contact with the fibril and stabilize this arrangement. This situation is shown in the distance plot in **Figure 3.5.18c**. Finally, the displacement of two β_3 stretches in close proximity leads to the formation of a β-sheet structure by the formation of backbone hydrogen bonds. **Figure 3.5.19** shows that hydrogen bonding occurs at the C-terminus. This will also lead to the formation of a salt bridge between C-terminus and Lys28. It is clear that this final configuration of the docking pathway would be the initial stage of the lock phase.

3.4. Conclusions

The fibril structure of amyloids is an important topic in the study of amyloid-related disorders. In particular, their neuronal toxicity might be affected by their structure and stability. Recent experiments revealed the high-resolution structures of the Aβ_{42} fibril; in this study, I performed MD simulations to study the structure, stability, and equilibrium dynamics of the fibril in an aqueous environment. In addition, I applied two sampling methods, steered MD and REMD, to analyze the

trajectories of aggregation and disaggregation in the fibril motifs leading to the predicted dock-and-lock pathway of the Aβ_{42} fibril motif.

The structural features of the triple-β motif were strongly preserved throughout the 10 μs simulation. Particularly, core residues showed highly stable ordered arrangements in both the backbones and sidechains throughout the simulation. I identified the detailed structure of backbone and sidechain interactions stabilizing the hydrophobic cores of the triple-β shape Aβ_{42} fibril motif. Two hydrophobic cores, namely HP core I and HP core II, formed β_1-turn-β_2 and β_2-turn-β_3 structures, respectively, and HP core I was found to be composed of two sub-cores. Salt bridge interactions and polar sidechain interactions provided additional stabilization to these structures.

In addition to the hydrophobic core structures, the behavior of relatively unstructured N-terminal sequence was examined by calculation of RMSF and water radial distribution functions. Relatively large RMSF values of specific residues and corresponding sidechains and backbone conformational changes together with RDF changes could partially provide an explanation of the role of N-terminal flexibility in polymorphic experimental Aβ_{42} fibril structures.

Based on our analysis of pulling simulation trajectories, I found that the two hydrophobic cores each generate two residue interaction groups. The larger group corresponding to HP core I is disaggregated collectively, while the other group corresponding to HP core II is disaggregated sequentially. The disaggregated region of HP core I was observed to form a helix, consistent with the results of docking pathway analysis. In the docking pathway analysis, I observed that the β_2 sequence forms a partially folded conformation, comes into close contact with the fibril surface, and guides the two helices of β_1 and β_3 to their proper orientations. As this

arrangement induces the formation of β-sheet in the β_3 sequence, the lock process might be propagated from the β-sheet formation of β_3 to β_2, and finally β_1, by unfolding their helices. This is exactly the reverse process of disaggregation shown by the pulling trajectory analysis. To confirm this lock phase mechanism, further simulations or experimental studies will be needed.

3.5. Figures and Tables

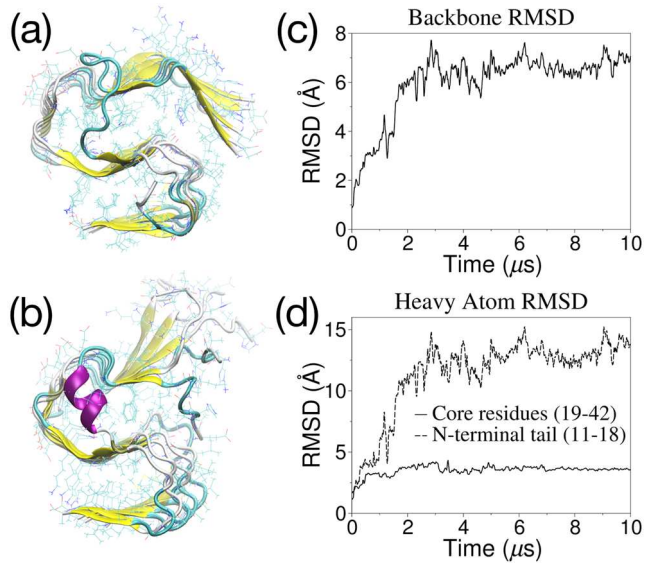


Figure 3.5.1. Simulation structures of triple-β motif of Aβ_{42} at (a) 500 ns and (b) 10 μs. (c) Backbone RMSD plot of 10 μs simulation. (d) Heavy atom RMSDs of core sequence (residues 19–42, solid line) and N-terminal tail sequence (residues 11–18, dashed line). RMSD of core sequence shows remarkably smooth plateau throughout the simulation at ~3.4 Å which reflects highly stable backbone and hydrophobic core structures compared to N-terminal tail region.

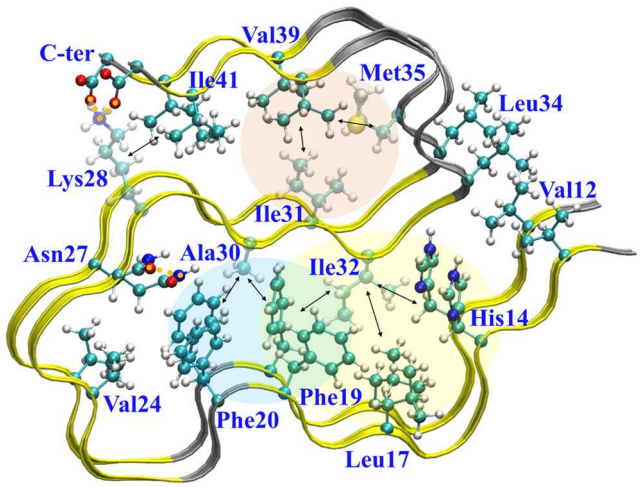


Figure 3.5.2. Structure of backbone and sidechain interactions of Aβ_{42} triple-β motif. β-sheet regions are indicated by yellow color. Ala30 and Ile32 play pivotal roles to form two sub-cores of HP core I (denoted by blue and yellow circles). Similarly, Ile31 and Met35 form HP core II (light pink circle). Sidechain interactions between residues are denoted by double-headed arrows. Hydrogen bond between Asn27 additionally stabilizes β_1-turn-β_2 structure (HP core I), and salt bridge of Lys28 and C-terminal carboxylic group stabilizes β_2-turn-β_3 structure (HP core II). Identification of β-sheet regions are calculated by STRIDE algorithm[89] using VMD[90].

Table 3.5.1. Sidechain interactions in the triple-β motif structure from simulation.

Intra-Chain Sidechain Interactions	
HP contact	Val12–Leu34
HP core Ia	His14–Ile32
	Leu17–Phe19
	Phe19–Ile32
HP core Ib	Phe19–Phe20
	Phe19–Ala30
	Phe20–Ala30
HP contact	Phe20–Val24
	Phe20–Asn27
HP core II	Ile31–Val39
	Met35–Val39
HP contact	Val39–Ile41
	Lys28–Ile41
Salt bridge	Lys28–C-ter

Inter-Chain Sidechain Interactions (inner chain–outer chain)		
HP contact	-	Val12–Val12
		Leu34–Leu34
HP core Ia	His14–Ile32	His14–His14
	Leu17–Ile32	Leu17–Leu17
	Phe19–Ile32	Phe19–Phe19
HP core Ib	Phe19–Ala30	
	Phe20–Ala30	Phe20–Phe20
HP contact	-	Val24–Val24
H-bond	-	Asn27–Asn27
HP core II	Val39–Ile31	-
	Met35–Val39	
HP contact	Ile41–Lys28	Ile41–Ile41
Salt bridge	C-ter–Lys28	-

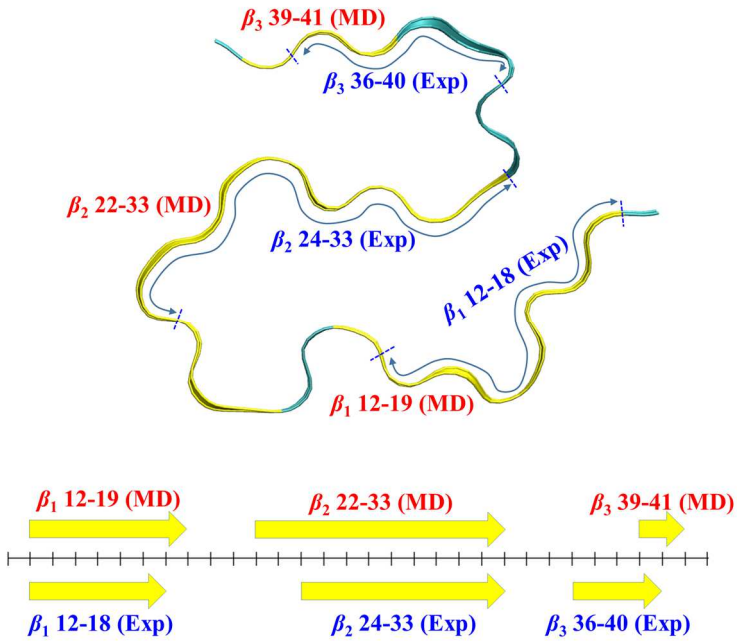


Figure 3.5.3. Comparison of triple-β motif between experimental and simulation structure. In experimental structure (PDB ID: 2MXU), residues 12 to 18, 24 to 33, and 36 to 40 form β-sheets (blue arrow). In simulations, residues 12 to 19, 22 to 33, and 39 to 41 form β-sheets (yellow ribbon).

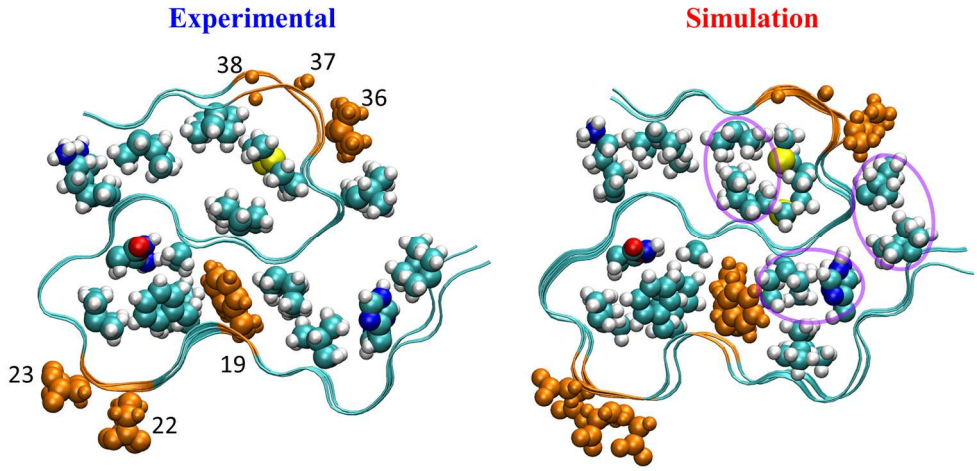


Figure 3.5.4. Comparison of triple-β motif between experimental (PDB ID: 2MXU) and simulation structure. Residues shown in orange color have β-sheet geometry in either one of experimental or simulation structure. Comparison of sidechain contact for these residues shows that simulation structure has more compact hydrophobic packing due to the presence of water. Additional sidechain contacts occur between residues 12–34, 14–32, and 32–39. (purple circles)

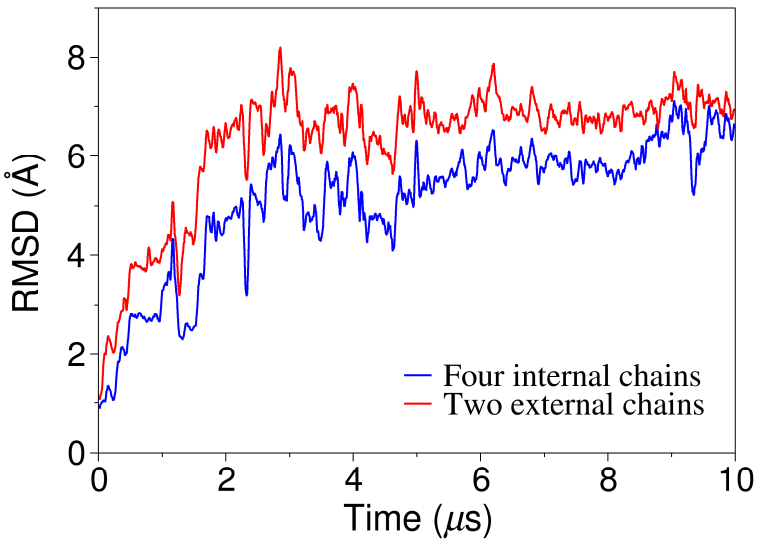


Figure 3.5.5. Backbone RMSD of the two external chains (red line) are larger than the four internal chains (blue line) of triple-β motif. Two external chains are relatively flexible compared to the four internal chains.

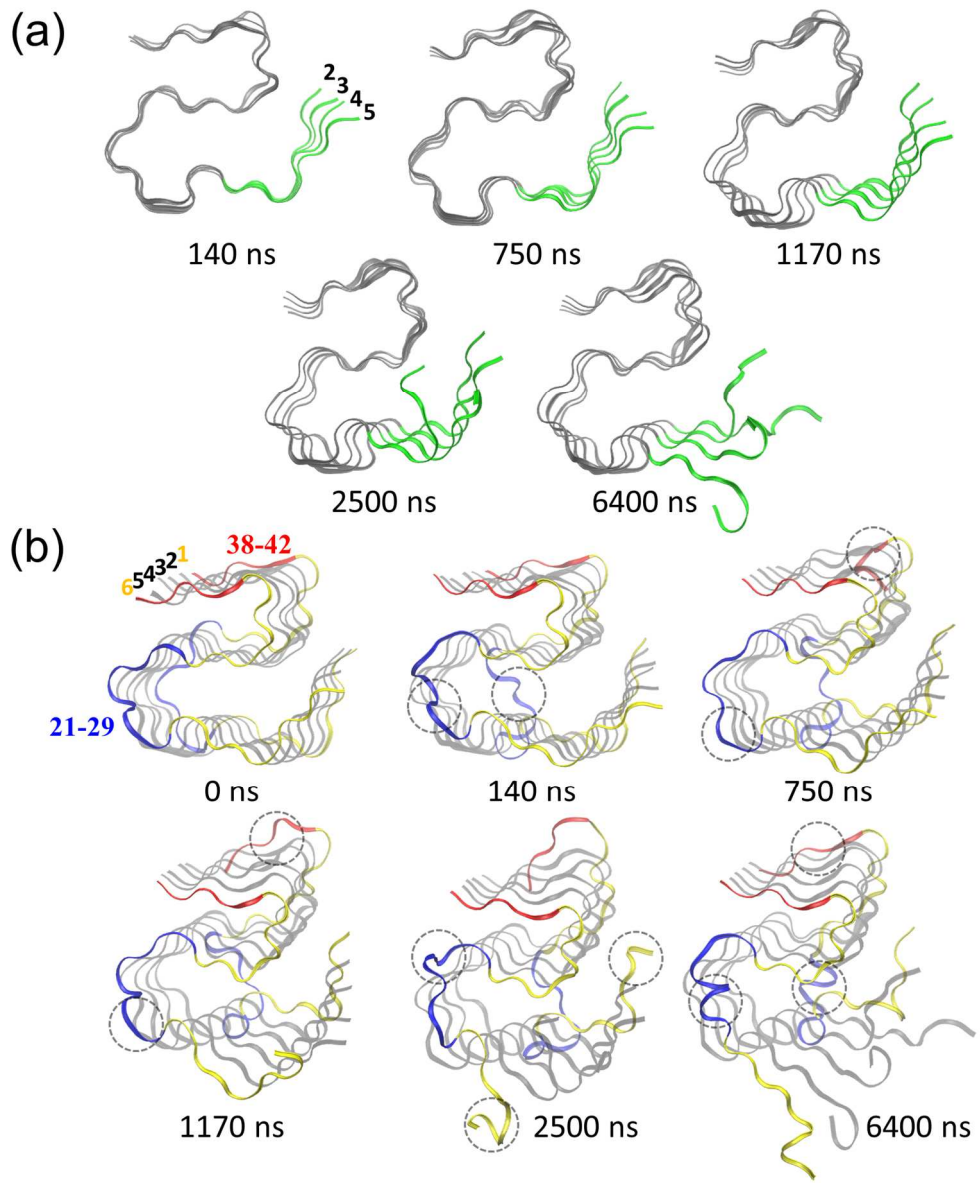


Figure 3.5.6. Conformational dynamics of triple-β motif through 10 μs MD simulation. (a) Conformational change of the four internal chains. Denoted numbers are chain numbers. Two external chains are not shown. Residues 19–42 form stable core of the structural motif for Aβ_{42} protofibril (gray ribbons). On the contrary, residues 11–18 form unstructured N-terminal tail (green ribbons). (b) Conformational change of the two external chains as represented by colored ribbons.

Residues 21–29 and 38–42 are colored in blue and red, respectively and the rest are in yellow. Four internal chains are represented by the transparent gray ribbons. Residues 21–29 and residues 38–42 are relatively unstable compared to other residues in core (19–42). Residues 11–15 form unstructured N-terminal tail similarly as in the four internal chains.

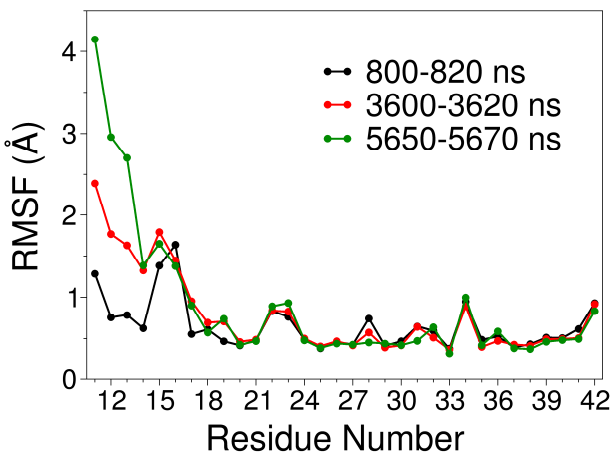


Figure 3.5.7. RMSF of N-terminal sequence for 800–820 ns, 3600–3620 ns, and 5650–5670 ns.

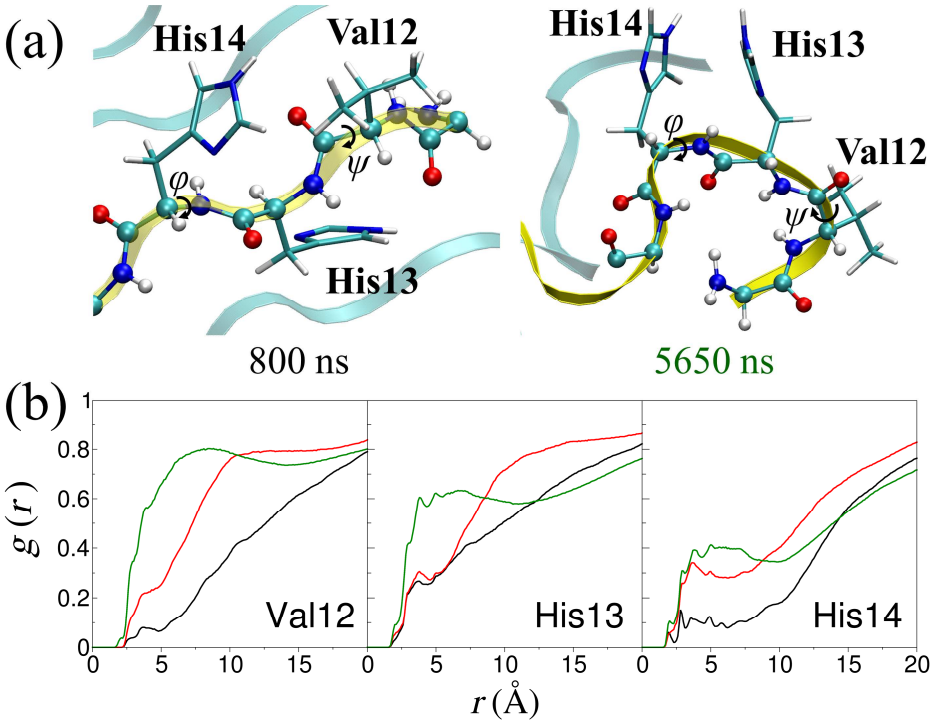


Figure 3.5.8. (a) Conformational changes of N-terminal residues, especially Val12, His13, and His14. (b) Changes of radial distribution functions (RDF) of water according to conformational changes in (a).

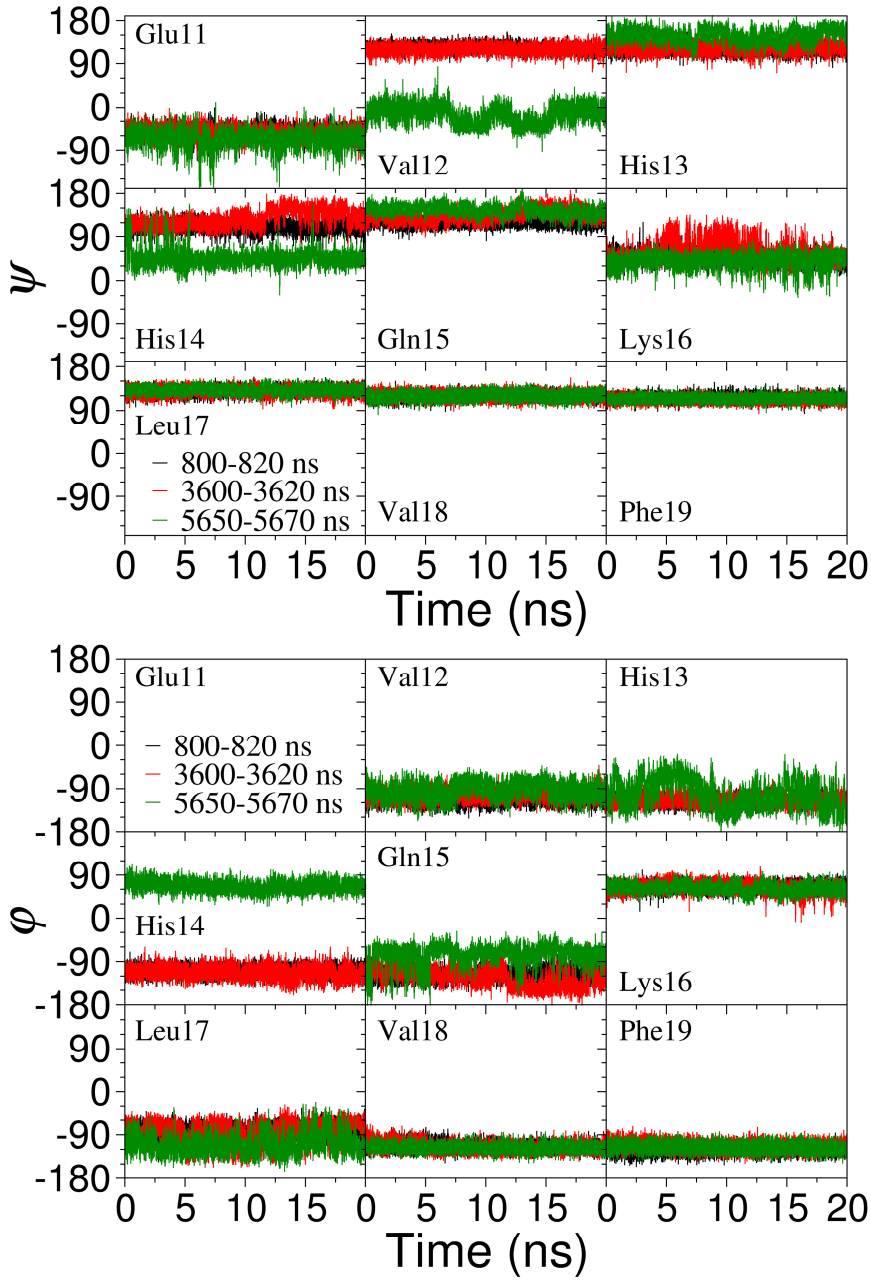


Figure 3.5.9. Dynamics of backbone torsion angles (φ, ψ) of N-terminal residues in triple-β fibril motif.

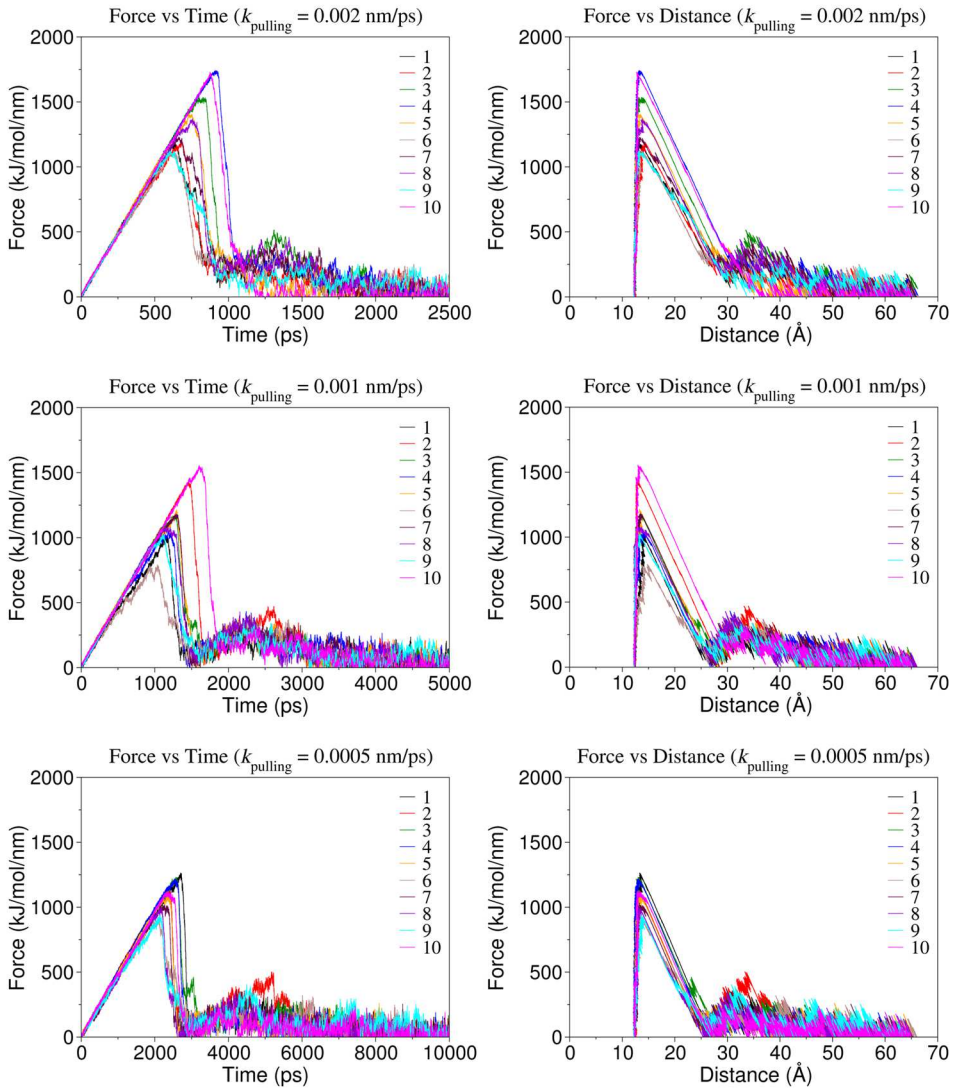


Figure 3.5.10. Force vs time curves (left column) and force vs distance curves (right column). Each row has two representations of force curves for the same set of ten independent simulation trajectories with constant pulling rate. As pulling rate is decreased, deviation of peak heights among different trajectories is reduced.

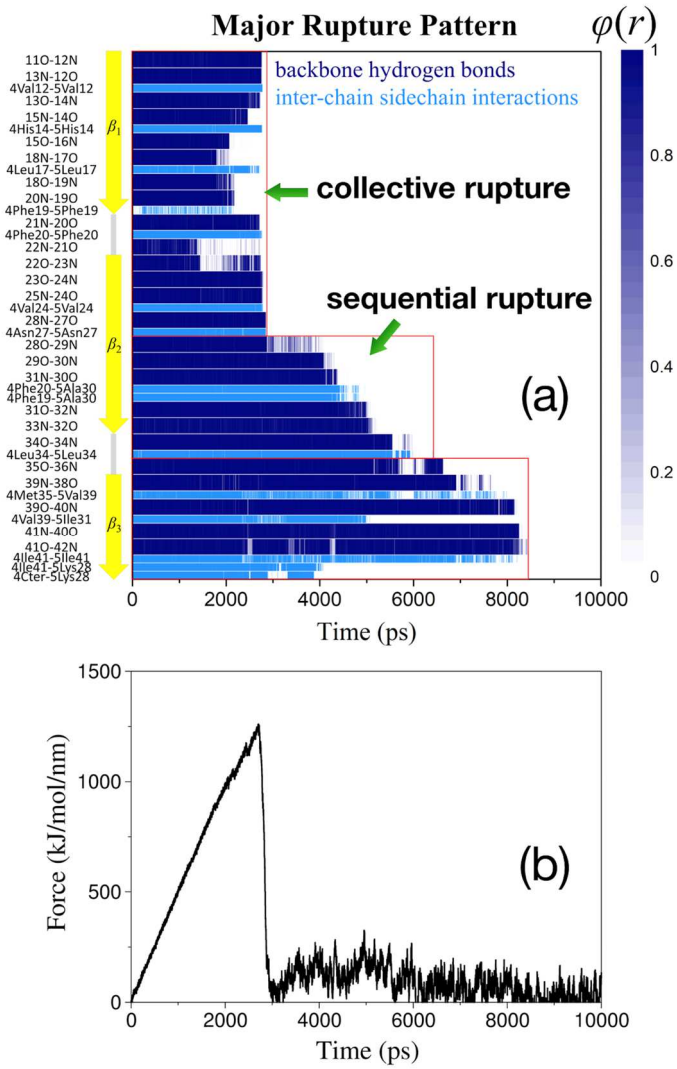


Figure 3.5.11. Structural change of Aβ_{42} triple-β motif during a pulling simulation. (a) Major rupture pattern of backbone hydrogen bonds and intermolecular sidechain interactions. $\varphi(r)$ = 1 when the interaction is formed, otherwise 0. Three boxes enclosed by red lines indicate three groups in which the residues are disrupted together. (b) Force extension curve corresponding to rupture pattern in (a). Large peak at 0–3000 ns is due to the collective rupture in (a). Broad peak at 3000–6000 ns corresponds to the sequential rupture in (a).

Table 3.5.2. Statistics of the logistic function profiles from steered MD simulations. Major pattern shown in **Figure 3.5.11** appears in 16 trajectories of the total 30 trajectories. Especially, most of the trajectories with $k_{pulling}$ = 0.0005 nm/ps (8 among 10 trajectories) show the major pattern. Second major rupture pattern denoted by '**a**' in the table appears in 7 trajectories. The rest of the patterns show minor contribution and corresponding plots are not shown here.

$k_{pulling}$	0.002 nm/ps	0.001 nm/ps	0.0005 nm/ps	total
major	2	6	8	16
a	3	3	1	7
a'	2	1	0	3
b	2	0	0	2
c	0	0	1	1
~ major	1	0	0	1

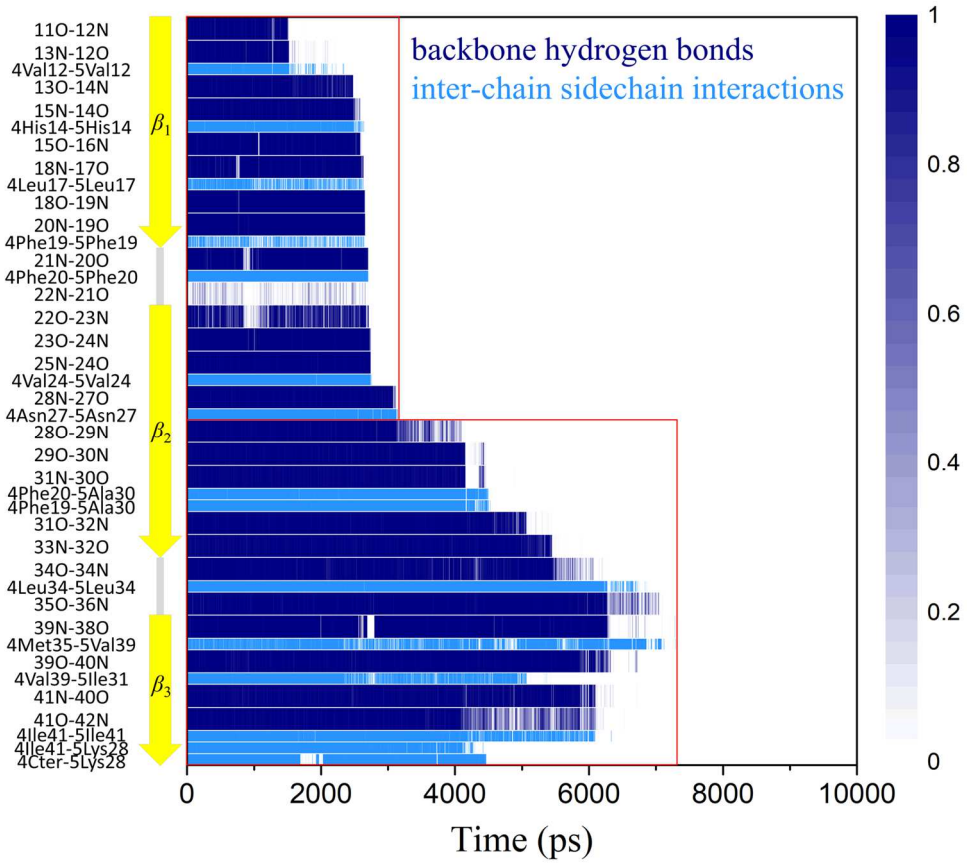


Figure 3.5.12. Second major rupture pattern (pattern a in **Table 3.5.2**) is observed in 7 trajectories. Collective rupture and sequential rupture in residues 11–34 are similar to the major pattern. On the other hand, interactions of residues 35–41 become disrupted much earlier compared to the major pattern. The shape of the plot in the second box (28O-29N to 4Cter-5Lys28) seems to be symmetric with respect to the 35O–36N, indicating that the disruption of HP core II occurs from the breaking of salt bridge interaction between Lys28 and C-terminus. It drives the propagation of simultaneous breaking in backbone hydrogen bonds from C-terminus to Met35 and from Lys28 to Met35.

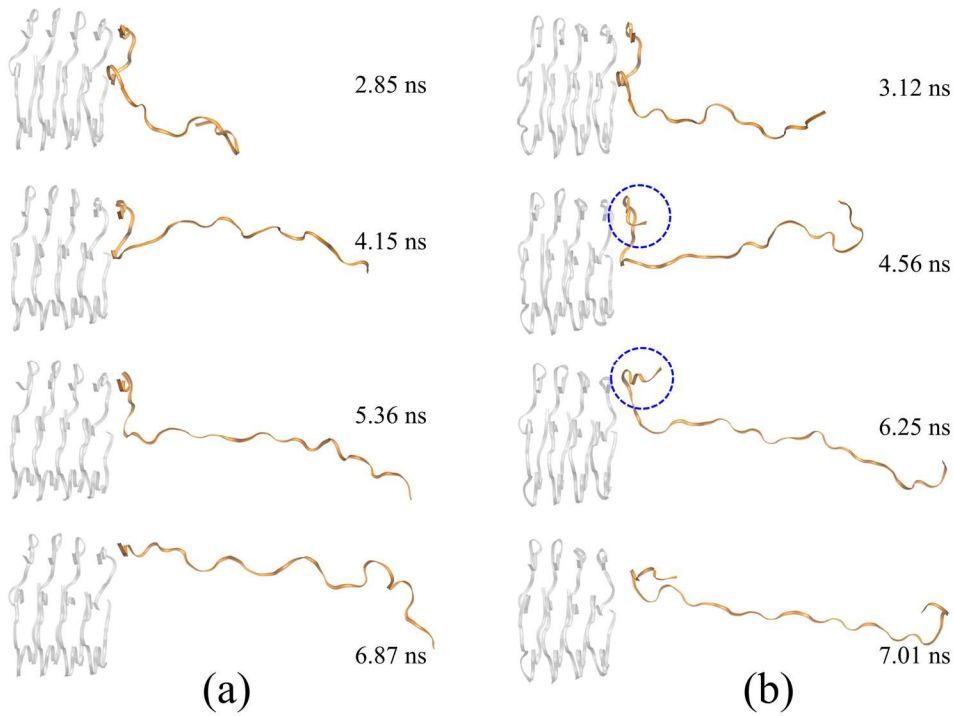


Figure 3.5.13. Configurations corresponding to the trajectories of (a) the major, and (b) the second major rupture pattern. In the major rupture pattern, C-terminus is the last part to be disrupted. In the second major rupture pattern, C-terminus is disrupted earlier than the major pattern as shown in blue dashed circles.

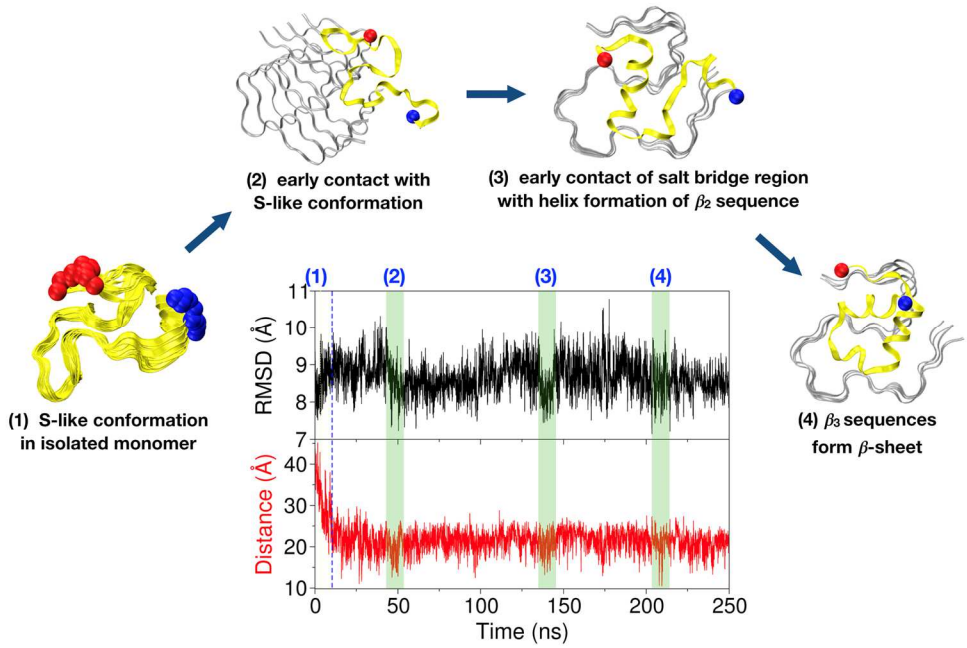


Figure 3.5.14. Dynamics of single Aβ_{42} peptide from isolated state to approaching the fibril structure. RMSD with respect to the triple-β conformation is plotted which shows the occurrence of S-shape conformations of single Aβ_{42} peptide during the REMD simulation. Red curve is the distance between C_α atoms of Ala30 in monomer and outermost chain of the fibril. The moments of low RMSD values are denoted by numbers (1) to (4) in order of time. Except for (1), they all happen for short monomer-fibril distances as highlighted by green windows. Representative configurations for these moments are also shown.

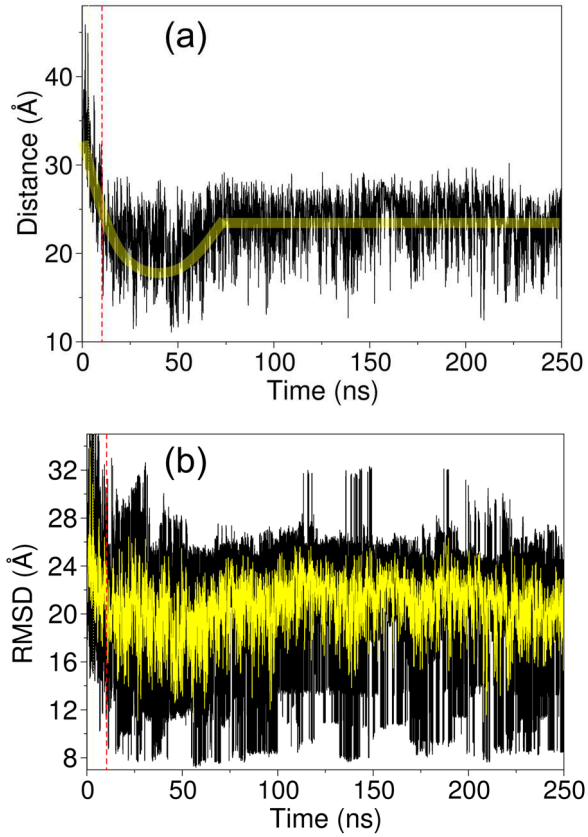


Figure 3.5.15. (a) Center of mass distance between the outermost chain of protofibril motif and Aβ_{42} monomer. (b) RMSD of the combined system (outermost chain in protofibril motif + monomeric Aβ_{42}) with respect to the triple-β configuration. Yellow curve shows the average plot of the original data (black curve).

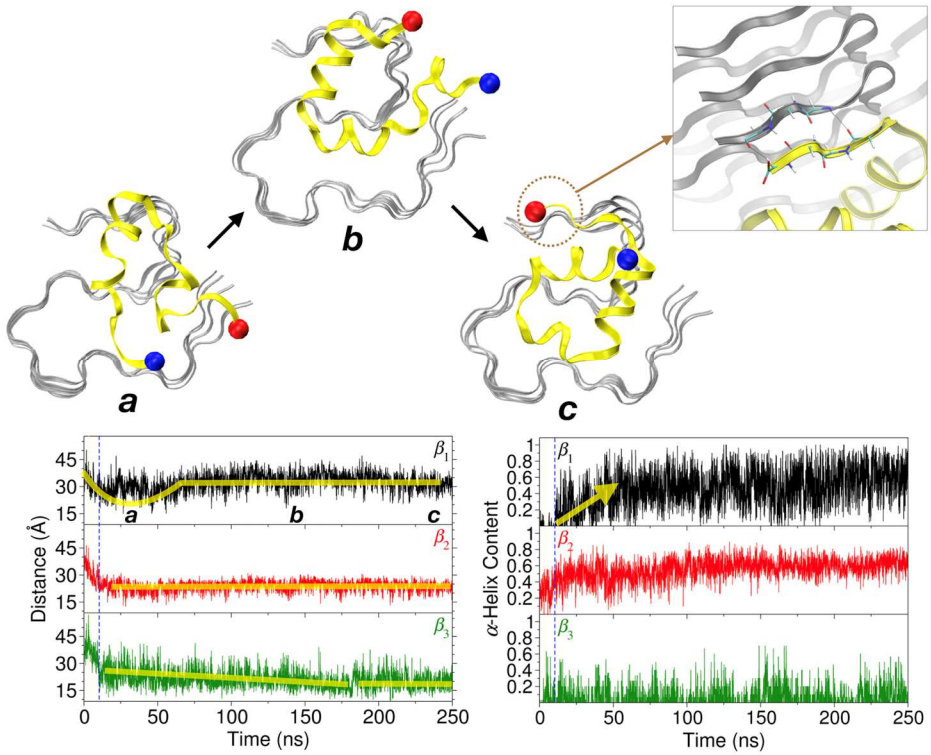


Figure 3.5.16. Docking pathway of Aβ_{42} monomer to preformed protofibril structure. (Up) Representative structures demonstrating the configurational changes in docking process. The zoomed-in view in square box shows backbone hydrogen bonds formation in β_3 region at the final stage of simulation. (Down Left) Center of mass distances between monomer and fibril surface for β_1, β_2, and β_3, respectively are plotted. Thick yellow transparent lines show overall features of the graphs. (Down Right) α-helix content of β_1, β_2, and β_3 sequences in Aβ_{42} monomer. Yellow arrow highlights the monotonic increase of α-helix content in β_1 sequence.

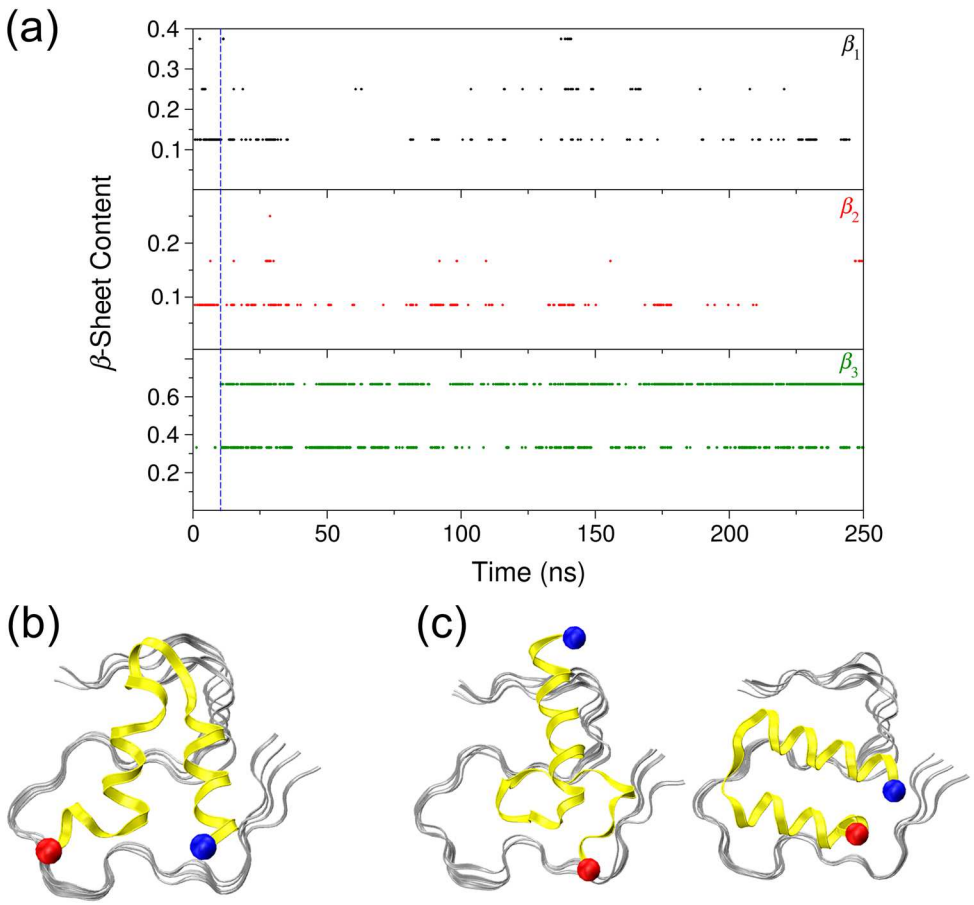


Figure 3.5.17. (a) Time evolution of β-sheet content of each β sequence. β-sheet character of β_1 and β_2 sequences is weakened as Aβ_{42} monomer forms a contact to protofibril. On the contrary, β_3 sequences shows increase of β-sheet content after contact between monomer and protofibril. (b) Additional conformation of simulation time range *a* in **Figure 3.5.16**. β_1 and β_2 sequences have partially helical conformations. (c) Additional conformations of simulation time range *b* in **Figure 3.5.16**. β_1 sequence shows structural change from partially helical to fully helical conformation. β_2 sequence maintains partially helical conformations.

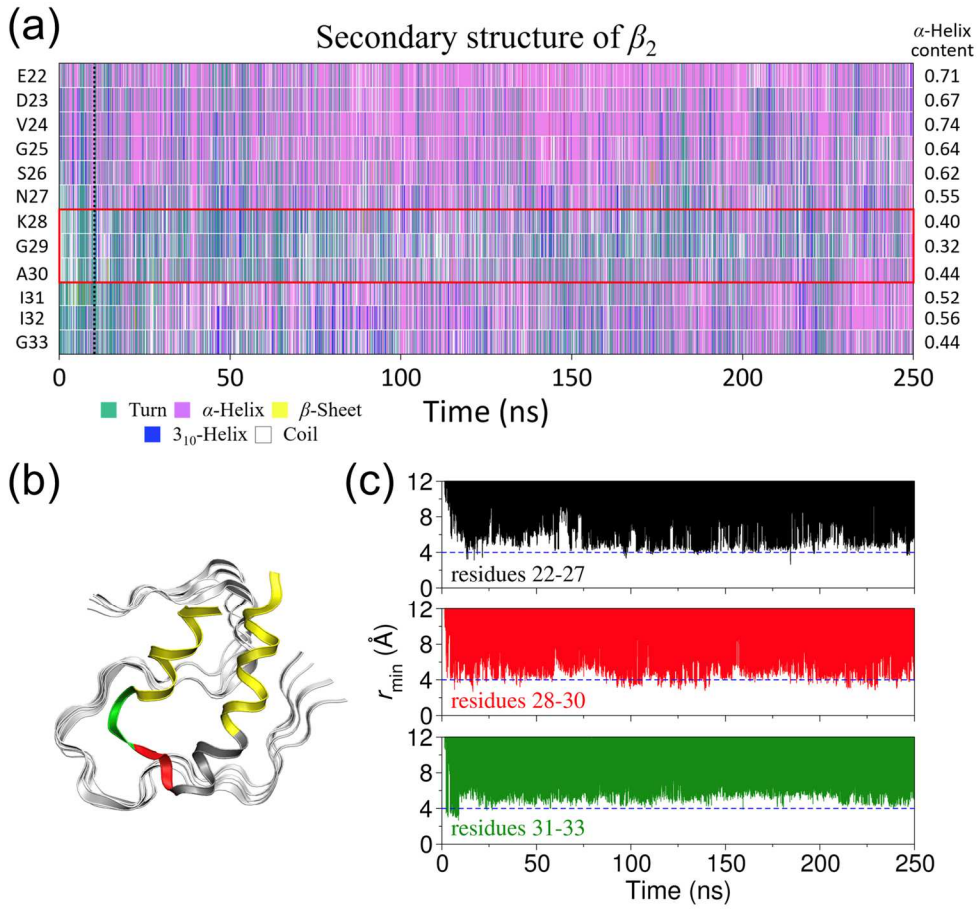


Figure 3.5.18. (a) Secondary structure of β_2 sequence of the monomer. It is shown that Lys28, Gly29, and Ala30 have low α-helix content compared to other residues which demonstrates partially unfolded geometry of β_2 sequence. (b) Representative structure for β_2 sequence of Aβ_{42} monomer. Unfolded geometry of residues 28–30 (red ribbon) allows close contact to fibril edge compared to residues 22–27 (black ribbon) and 31–33 (green ribbon). (c) Minimum heavy atom distance between the outermost chain in protofibril motif and the center of mass of each region in β_2. It is shown that the distance of residues 28–30 is around 4 Å whereas residues 22–27 and 31–33 show longer distances.

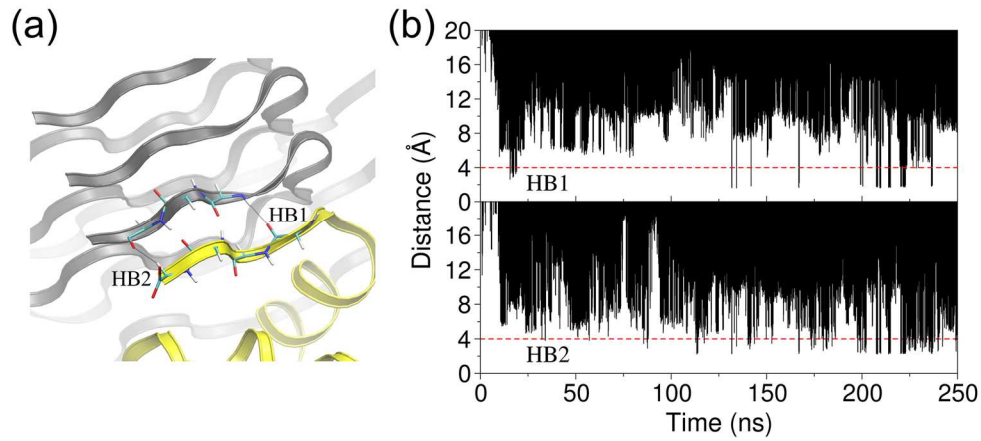


Figure 3.5.19. (a) β-sheet backbone hydrogen bonding formation in the C-terminal tail of β_3 sequence between Aβ_{42} monomer and fibril motif (b) N-O distances for two hydrogen bonds shown in (a) designated by dotted lines and denoted by HB1 and HB2. Formation of hydrogen bond can be identified by the N-O distance less than 4 Å. For HB2, N-O distance curve frequently visits the threshold distance throughout the simulation except for early stage. On the other hand, HB1 rarely visits 4 Å until 200 ns. From this, it is suggested that the development of β-sheet structure in β_3 sequence occurs in order of HB2 → HB1.

Chapter 4. In Silico Investigation of the Structural Stability as the Origin of the Pathogenicity of α-Synuclein Protofibrils

4.1. Introduction

α-Synuclein is a presynaptic neuronal protein that plays a crucial role in the etiology of Parkinson's disease (PD). Monomeric α-synuclein is known to be intrinsically disordered in aqueous phase and forms partial helix on lipid membrane.[91] Its aggregated form is a major constituent of intraneuronal inclusions called Lewy bodies (LBs) and Lewy neurites (LNs), the characteristic hallmarks of PD. LB is observed in the brains of PD patients, especially in the substantia nigra (SN). Various *in vitro* and *in vivo* studies have suggested that misfolding and aggregation of α-synuclein is a major pathogenic event in PD.[91] The familial form of PD is associated with specific mutations of α-synuclein, whereas sporadic PD is associated with elevated levels of α-synuclein in the brain tissue or post-translational modification.[92] Both induce misfolded states of α-synuclein and the formation of amyloid aggregates, including oligomers and fibrils. The increase in aggregates results in a decreased level of monomeric α-synuclein in nerve cells, especially in the synaptic terminal. As α-synuclein is known to play a chaperone-like role in SNARE-dependent trafficking of vesicles through its association with the presynaptic SNARE complex, the loss of this function by decreased levels of monomeric form, due to aggregate formation, causes the impairment of neuronal communication.[93] Therefore, understanding the principle of accelerating aggregation is an important issue for developing therapeutic strategies to prevent aggregation.

In principle, the entire aggregation process can be considered to proceed in two

stages. The first is the early stage of fibril formation, or in other words, the formation of oligomer species. This stage is a kinetically controlled reaction network that consists of more than a single step, including the formation of either on-pathway or off-pathway oligomers. While recent studies indicate that off-pathway oligomers are the most toxic species in PD pathogenesis,[94] it has been suggested that the early stage in the α-synuclein fibril formation process consists of two on-pathway oligomers.[95] The second stage is the transition from oligomers to fibrils, which is thermodynamically, rather than kinetically, controlled because the thermodynamic stability of the fibril structure is the major driving force of the process. Structural information on the amyloid protofibril motifs at the atomic level can provide insights into the thermodynamic stability of protofibril structures. They can also be considered as on-pathway intermediates and, therefore, can be used as the starting point for investigating the kinetics of the on-pathway amyloid formation process. Thus, elucidating the structural features of α-synuclein protofibrils is important for developing disease-modifying therapies for PD.

α-Synuclein fibrils show polymorphism where different strains have different protofibril structures, in terms of backbone geometries, sidechain interactions, and inter-protofilament interfaces.[96] These structural polymorphisms can result in different cytotoxicities, speeds of cell-to-cell spread, and seeding activities of the corresponding strains.[97] For example, in prion disease, it is referred to as “Prion strain phenomena”.[98] Experimental studies have shown that different α-synuclein fibrils prepared *in vitro* or *in vivo* may have different polymorph compositions.[99, 100] Because each polymorph or strain makes distinct contributions to the biological activities, different compositions may have different seeding activities and toxicities. Therefore, it is essential to characterize the biological function of each individual

polymorph to understand the pathological role of these complex polymorphic fibrils. Recent experiments using ssNMR, cryo-EM, and TEM have revealed various types of *α*-synuclein fibril structures, including wild-type, familial mutants, and truncated forms.[99, 101, 102, 103, 104] From this, it has been shown that two strains, called *rod* and *twister* polymorphs, are major constituents of *α*-synuclein fibrils.[99-101, 103] In addition, cytotoxicity tests and energy analysis suggested that *rod* polymorphs, having Greek-key like protofibril kernel, would be a major pathogenic strain.[99, 101, 103-105]

Although high-resolution structures from recent experiments provide valuable information on structural properties that may contribute to pathogenicity,[101] atomic-level information on how the structure behaves in an aqueous environment is required to explain the pathogenicity of a certain strain in association with its structural features. In this study, I performed molecular dynamics simulations of the two major polymorphs of wild-type *α*-synuclein fibrils. By analyzing and comparing their structural features, I investigated how the major strain obtains its structural stability in aqueous conditions, which might be crucial for explaining its pathogenicity. In particular, I aimed to describe how the residual packing of the protofibril kernel and of the interface between protofibrils are efficiently shaped to provide compact and energetically stable hydrophobic core structure while the relatively larger conformational fluctuations and the consequent solvent accessible surfaces are given to the peripheral parts of the fibril structure, and thereby these two structural counter aspects are combined to complementarily complete the structural stability of amyloid fibril in aqueous environment. The results of the present study would provide a basis for understanding the pathogenic behavior of diverse amyloid strains in terms of their structural properties.

4.2. System Preparation

ssNMR structure The ssNMR α-synuclein structure published by Tuttle *et al*. was used, which includes residues 1–140 of the peptide (PDB ID: 2N0A).[101] Six α-synuclein fibril chains were taken from PDB and unstructured residues (1–45 and 97–140) were removed (see **Figure 4.6.1**). After neutralizing terminal residues, a fibril consisting of six α-synuclein chains with residues 46–96 was prepared and solvated in a dodecahedron box containing 13875 molecules of TIP3P water[78] with 43 Na^+ and 43 Cl^- counter ions to form a neutral 150 mM NaCl environment. The CHARMM22/CMAP force field[79] was used for α-synuclein throughout this study.

Cryo-EM rod polymorph structure The cryo-EM structure of rod polymorph α-synuclein published by Li *et al*. was used, which includes residues 38–97 of the peptide (PDB ID: 6CU7).[99] One chain was copied and located parallel to the outermost chain for both protofilaments, and all terminal residues were neutralized. The paired protofilaments, which were composed of two hexameric α-synuclein chains with residues 38–97 (see **Figure 4.6.1**), were solvated in a dodecahedron box containing 48140 molecules of TIP3P water[78] with 145 Na^+ and 181 Cl^- counter ions to form a neutral 150 mM NaCl environment. A single protofilament system was constructed by removing the coordinates of one of the two protofilaments from the paired protofilament structure and solvated in a dodecahedron box containing 21529 molecules of TIP3P water[78] with 64 Na^+ and 82 Cl^- counter ions.

Cryo-EM twister polymorph structure The cryo-EM structure of twister polymorph α-synuclein published by Li *et al*. was used, which includes residues 43–83 of the peptide (PDB ID: 6CU8).[99] One chain was copied and located parallel to

the outermost chain for both protofilaments, and all terminal residues were neutralized. The paired protofilaments, which were composed of two hexameric α-synuclein chains with residues 43–83 (see **Figure 4.6.1**), were solvated in a dodecahedron box containing 48651 molecules of TIP3P water[78] with 141 Na^+ and 153 Cl^- counter ions to form a neutral 150 mM NaCl environment. A single protofilament system was constructed by removing the coordinates of one of the two protofilaments from the paired protofilament structure and was solvated in a dodecahedron box containing 21262 molecules of TIP3P water[78] with 62 Na^+ and 68 Cl^- counter ions.

4.3. Simulation Protocols

Conventional MD simulations First, all prepared systems were minimized using the steepest descent algorithm (10000 steps) and then the conjugated gradient algorithm (10000 steps). After minimization, the systems were gradually heated for 6.2 ns from 0 to 310 K with position restraints on the backbone heavy atoms (force constant k = 100 kJ/mol/nm^2). For pre-equilibration, several 2 ns runs were performed at 310 K with restraints of decreasing orders of strength (k = 100, 50, 30, and 10 kJ/mol/nm^2). Each system was further equilibrated at 310 K for 10 ns without position restraints to obtain the initial conformations for the product run. For all heating and equilibration steps throughout this work, the isothermal–isobaric ensemble (NPT ensemble) with a velocity-rescaling thermostat[80] and a Berendsen barostat[41] were used. After equilibration, product runs were performed for 1 μs for the ssNMR structure and the two single protofilament cryo-EM structures, and 500 ns for the two paired protofilament cryo-EM structures. The trajectories were saved

every 2 ps for analysis. All product runs were performed using the NPT ensemble with a velocity-rescaling thermostat,[80] a Parrinello-Rahman barostat,[81] and a 2 fs time step was used with the LINCS algorithm[82] for the constraints of bonds related to hydrogen atoms. The GROMACS 2020.4 package[106] was used for all simulations.

β-sheet contents were calculated using the DSSP algorithm[107] in GROMACS.[106] Electrostatic potential maps of each structure at the solvent accessible surface were calculated by solving the nonlinear Poisson-Boltzmann equation using the APBS Electrostatics plugin[108] in PyMOL[109]. I used the single Debye-Hückel boundary condition, cubic B-spline discretization of point charges on grid, and smoothed surface for dielectric and ion-accessibility coefficients at 310 K. Dielectric constants $\varepsilon_{protein}$=2.0 and ε_{water}=78.0 were used and ion concentrations were 150 mM for both Na^+ (radius=1.02 Å) and Cl^- (radius=1.81 Å).

Pulling Simulations Rod and twister structures are solvated with TIP3P water[78] boxes with dimensions of 13.6 nm × 10.9 nm × 30.0 nm and 11.6 nm × 9.0 nm × 30.0 nm, respectively. 150 mM of Na^+ and Cl^- counterions are added. Heating and equilibration are performed in the NPT ensemble using velocity-rescaling thermostat[80] and a Berendsen barostat[41]. Pulling simulations are performed in the NPT ensemble at 310 K and 1 bar using velocity-rescaling thermostat[80] and a Parrinello-Rahman barostat[81] with a pulling rate of 0.0005 nm/ps.

Binding Free Energy Calculations I calculated the potential of mean force (PMF) curves along with the pulling simulation trajectories of rod and twister polymorphs as the reaction coordinates by umbrella sampling[110] and weighted histogram analysis method (WHAM)[111]. From pulling simulation trajectories, starting configurations are taken for the umbrella sampling windows. An asymmetric distribution of sampling windows is used for increasing detail at smaller center of

mass (COM) distance. For rod polymorph, window spacing is 0.24 nm up to 4.63 nm COM separation, and 0.27 nm beyond 4.63 nm, which resulted in 39 windows. For twister polymorph, window spacing is 0.25 nm up to 3.95 nm COM separation, and 0.27 nm beyond 3.95 nm, which resulted in 23 windows. In each window, equilibration is performed in the NPT ensemble using velocity-rescaling thermostat[80] and a Berendsen barostat[41]. After the equilibration, MD simulations for umbrella sampling are performed in the NPT ensemble at 310 K and 1 bar using velocity-rescaling thermostat[80] and Parrinello-Rahman barostat[81]. For smaller COM distance windows, 5 ns simulations are performed, and 10 ns simulations are performed for larger COM distance windows. From these trajectories, weighted histogram analysis method (WHAM) calculations are performed to generate PMF curves.

4.4. Results and Discussion

Tuttle *et al*. identified a high-resolution structure of the wild-type α-synuclein with a single protofilament of the Greek-key β-sheet topology using ssNMR and X-ray spectroscopy.[101] They added the fibrils to primary hippocampal neurons and observed the induction of insoluble inclusions. They also showed that the same α-synuclein fibrils act as pathological seeds that may initiate a disease. Although the ssNMR structure of the fibril has a single protofilament morphology, it is possible that the fibril exists as a pair of two protofilaments inside the neuronal cells. Li *et al*. showed that the *in vitro* generated α-synuclein fibrils have a dose-dependent cytotoxicity and seeding activity.[99] Using cryo-EM, they found that the fibrils were composed of two major populations, referred to as “rod” and “twister”. The rod polymorph also has a Greek-key β-sheet topology, which is very similar to the

ssNMR structure identified by Tuttle *et al*. However, the rod polymorph was composed of a pair of two tightly matched protofilaments. Therefore, I postulated that the rod protofilament with the Greek-key β-sheet topology may exhibit high stability, composing a major pathogenic strain. To rationalize this in terms of structural stability, I performed systematic simulations of the experimental structures (ssNMR, cryo-EM rod, and cryo-EM twister), and made detailed comparisons of the structural features.

I performed simulations using two different initial structures with Greek-key β-sheet topologies. One was from the ssNMR structure determined by Tuttle *et al*.[101] and the other was from the cryo-EM structure of Li *et al*.[99], which are denoted as "rod_{ssNMR}" and "$rod_{cryo\text{-}EM}$", respectively. When compared with the paired structures, I will denote the single protofilament rod polymorph as "rod_{single}" instead of "$rod_{cryo\text{-}EM}$". **Figure 4.6.2** shows the two structures before and after 1 μs simulations and the changes in RMSD (Root Mean Square Deviation) values during the simulations. After 1 μs of simulations in an aqueous environment, the two converged structures had both common and different features. Both structures adopted compact interior sidechain packing with steric zipper geometries. In contrast, the sidechains at solvent-exposed surfaces adopted a more disordered arrangement compared to the initial configurations to maximize the solvent-accessible surface area. Consequently, the optimal backbone geometry in an aqueous environment was formed in compliance with simultaneous hydrophobic and hydrophilic sidechain rearrangements to achieve maximal structural stability through the most preferable interactions with water. In other words, various structural features of the α-synuclein fibril are formed in response to stabilizing interactions with water. Therefore, the differences and similarities in structural features between distinct fibril geometries

and their relative stabilities in aqueous conditions can be explained in the same way. The β-sheet content of the rod_{ssNMR} structure decreased after the simulation, as can be seen in **Figure 4.6.2**, while the $rod_{cryo\text{-}EM}$ structure exhibited a slightly increased area of the backbone with β-sheet content. Additionally, the $rod_{cryo\text{-}EM}$ structure had a significant increase in disorder in the arrangement of both C- and N-terminal residues after the simulation, whereas the rod_{ssNMR} structure maintained an ordered arrangement of terminal residues. These distinct differences between the two apparently highly similar structures demonstrated that the interactions of the fibril with water or other solvents delicately affect the fibril conformation to achieve maximal structural stability. It can be argued that the rod_{ssNMR} structure may not be favorable for aqueous environments because its sidechain arrangement and corresponding backbone geometry are adapted to dry conditions, while the $rod_{cryo\text{-}EM}$ structure, optimized for interactions with water, would show high structural stability in aqueous environments.

The experiments of Tuttle *et al.*[101] were successful in identifying the disordered region (residues 55–62) and the core region (residues 46–54 and 63–96). The RMSD curves in **Figure 4.6.2** show that the proposed disordered region of the rod_{ssNMR} structure exhibited large fluctuations. However, the same region in the $rod_{cryo\text{-}EM}$ structure had relatively low fluctuations, which was attributed to the formation of a well-ordered β-sheet geometry. In the core region for both rod structures, the well-arranged β-sheet backbone structure and the highly compact sidechain packing structure of the hydrophobic core resulted in relatively small fluctuations of the backbone and the inside sidechains, while the outside sidechains showed larger fluctuations. The fluctuations of the inside sidechains were even smaller than those of the backbone, especially for the $rod_{cryo\text{-}EM}$ structure, while the

fluctuations of the outside sidechains were significantly enhanced. These observations suggest that the $rod_{cryo\text{-}EM}$ fibril can maintain a highly stable structure in an aqueous environment because of the cooperative effects of the compact sidechain packing of the hydrophobic core, the backbone geometry of the maximal β-sheet contents wrapping the hydrophobic core, and the solvent-exposed sidechains with large fluctuations maximizing the solvation entropy.

Figure 4.6.3 shows the simulation results of the rod polymorph consisting of two interfaced $rod_{cryo\text{-}EM}$ protofilaments, denoted as "rod_{pair}". The most notable change in the RMSD curves of the rod_{pair} structure, compared to rod_{single} ($rod_{cryo\text{-}EM}$ in **Figure 4.6.2**), is that fluctuations of the disordered region (residues 55–62 in the ssNMR structure) were reduced, while the core region fluctuations slightly increased. This was attributed to the formation of an interface between the two protofilaments. An examination of the superimposed rod_{pair} and rod_{single} structures revealed that the backbone geometry of the region forming the interface became flat, and such conformational change propagated to adjacent sequences and induced the overall rearrangement of the backbone geometry. Consequently, the paired structure provides additional stability, especially at the interface region, by forming steric zipper interactions and hiding the hydrophobic residues from water exposure. This stable interface structure allowed a slight increase in fluctuations in the rest of the structure, which provided additional conformational and solvation entropy.

The simulation results of the twister polymorphs for a single protofilament "$twister_{single}$" and a pair of protofilaments "$twister_{pair}$" are shown in **Figure 4.6.4**. The twister polymorph had a less compact backbone geometry than the rod polymorph with a highly compact Greek-key β-sheet geometry. Therefore, $twister_{single}$ exhibited a more flexible and less stable backbone β-sheet and hydrophobic core structures.

Additionally, the twister polymorph had a different interface structure compared to the rod polymorph. The rod polymorph had a preNAC ($_{47}$GVVHGVATVA$_{56}$) interface, while the twister polymorph interface corresponded to a NACore ($_{68}$GAVVTGVTAVA$_{78}$). The NACore sequence was a part of the highly hydrophobic NAC (61–95) region, and both interfaces formed steric zippers for stability. However, the linear backbone geometry and the orientation of bulky hydrophobic sidechains with salt bridge interactions between E57 and K45 allowed the preNAC sequence to form the cross β-sheet interface structure of the rod polymorph, which is more rigid and stable than the bent shape NACore interface of the twister polymorph.

These differences in the interface structure, along with sidechain interactions in hydrophobic core regions, indicated that the twister polymorph, shown in **Figure 4.6.4a**, would be relatively less stable and more flexible in aqueous environments. In particular, the hydrophobic core and backbone conformation may not provide enough structural stability to counteract the instability of the hydrophobic interface region. Consequently, the backbone β-sheet planes, especially around the interface region, were twisted along the fibril axis so that a more stable geometry is adopted in an aqueous environment, which increases the hydrophilic surface and compacts the hydrophobic core sidechain packing (**Figure 4.6.4b**). For the twister polymorph with a pair of protofilaments, the formation of the hydrophobic steric zipper interface solved the problem of instability observed in the single protofilament, and the extent of backbone helical rotation was apparently reduced. The reduced rotation would be essential to form a helical structure of the twister polymorph by adjusting the pitch distance to be compatible with the persistent length of the backbone β-sheet structure.

The secondary structure contents were calculated using the DSSP algorithm to compare the geometries of the backbone structures in our simulations. The

backbone β-sheet content and structural patterns are shown in **Figure 4.6.5** and **Figure 4.6.6**. The secondary structure patterns in **Figure 4.6.6** reveal that all five structures of the different polymorphs mostly consist of β-sheets and turns. In particular, the rod_{pair} clearly exhibited distinct β-sheet/turn/β-sheet arrangements. The β-sheet content of the twister polymorph was found to be higher than that of the rod polymorph (**Figure 4.6.5a**), which suggested that the less compact backbone structure would be advantageous for large β-sheet contents. However, it can be argued that large β-sheet contents do not guarantee the stability of the fibril structure, even though the β-sheet backbone is generally preferred in aqueous environments to form a hydrophobic core. Effective formation of a highly compact folding structure for the hydrophobic core can benefit from the cooperative effects of the solvent-exposed surface of hydrophilic sidechains, the β-sheet hydrogen bonding, and the flexible β-turn-β motif, instead of the extended linear β-sheet motif (see **Figure 4.6.5b**). Additionally, the relatively short length of β-sheet stretches in the β-turn-β geometry is beneficial for forming stable backbone β-sheet hydrogen-bonding interactions. For both the rod and twister polymorphs, the β-sheet contents of the single protofilament structures were higher than those of the paired protofilament structures. This was attributed to the formation of inter-protofilament interfaces protecting the hydrophobic residues from exposure to water by forming hydrophobic steric zipper structures. The β-sheet content of the ssNMR structure was the lowest, indicating that this structure is relatively incompatible with the aqueous environment.

Previous analysis of the β-sheet contents and RMSD have shown the overall structural features and dynamic fluctuations of various experimentally identified α-synuclein protofibril structures. To analyze more detailed features of the

conformational and dynamic behavior of these structures, I calculated the RMSF (Root Mean Square Fluctuations) of individual amino acids (**Figure 4.6.7** and **Figure 4.6.8**). The backbone (C_α) RMSF plot exhibited an alternating peak-and-valley pattern, where the peaks and valleys correspond to turns and β-sheet sequences, respectively. In **Figure 4.6.7**, the locations of the peaks are denoted by thick bars labeled by letters (a, b, c, etc.). The large peaks of the sidechain RMSF plots occurred at the turn geometries (b and e of the rod polymorphs, and b and c of the twister polymorphs) because of the bulky sidechains exposed to the solvent. The low-sidechain RMSF regions (valleys) correspond to the sidechains participating in the hydrophobic core packing or inter-protofilament interface regions. **Figure 4.6.8** shows detailed local structural changes in the peak regions. By comparing these structures and the corresponding RMSF values, I investigated the structural features and fluctuations at the residue level, which can contribute to the overall structural stability of the corresponding polymorphs. For example, in the rod polymorph, region a corresponds to the interface and the RMSF values of the rod_{single} are larger than those of the rod_{pair} at this region. Region b contains solvent-exposed bulky residues such as E-K-K-Q and exhibits large peaks in the sidechain RMSF plot. The small RMSF peak of rod_{pair} in region c indicates a slight backbone rearrangement owing to interface formation. In the twister polymorph, the region a was located near the N-terminus and it had large RMSF values. Regions b and c correspond to the turns in which the solvent-exposed bulky sidechains exhibit large RMSF values. The low RMSF of the $twister_{pair}$ in region d may be attributed to the proximity of the corresponding position to the interface.

So far, I have discussed the influence of the aqueous environment on the stability of the fibril polymorph in terms of the formation of the hydrophobic core

and hydrophilic surface, as well as the residue-level conformational arrangements and fluctuations. It was also noted that solvent water can affect the stability of the fibril structure via direct interactions. One such direct interaction arises from the coordination of water molecules with the sidechains of the polar amino acid residues in the water channel. **Figure 4.6.9** shows the structures of the water channels in three different fibril polymorphs. The rod_{ssNMR} polymorph has a water channel consisting of three polar amino acid residues (E61, T59, and T72) that form hydrogen bonds with water molecules. This relatively small water channel allows only for the first and second hydration shells with only few isolated water molecules (**Figure 4.6.9a**). Similarly, the twister polymorph also has the first and second hydration shells, but with a narrower space (only two residues, E61 and T72, participate in hydrogen bonding with water), which is reflected by the lower peaks of the radial distribution functions (**Figure 4.6.9d**). In contrast, the $rod_{cryo\text{-}EM}$ polymorph has a water channel with a relatively large space and four hydrogen-bond participating residues (E61, T59, T54, and T75), which allows the formation of the third hydration shell (broad RDF peak at > 5.5Å in **Figure 4.6.9d**). The hydrogen bonding network of the water channel can provide additional structural stability by promoting coordination among the internal sidechains.

The other direct interaction of solvent water is with the solvent-exposed surface of the fibril polymorph. **Figure 4.6.10** shows the electrostatic potential energy maps of these solvent-exposed surfaces. In these maps, the surfaces can be categorized as unstable (high energy) and stable (low energy) in addition to water channels. The water channel of the rod_{ssNMR} polymorph has a narrow ellipsoidal shape, while the $rod_{cryo\text{-}EM}$ has a round channel with a larger area, allowing more water molecules to enter. The solvent-accessible surface at the N-terminal sequence, which is adjacent

to the interface sequence, had high electrostatic potential energy values in the rod_{single} polymorph. Therefore, the rod polymorph prefers the paired state to conceal this surface from water (**Figure 4.6.11**). In contrast, the large solvent-exposed pocket at the C-terminal sequence is stabilized by hydrogen bonding between water and S87 residue (see RDF in **Figure 4.6.12**). In the twister polymorph, the hydrophobic surface at the interface region can be stabilized by forming a paired polymorph (see **Figure 4.6.10** and **Figure 4.6.13**). The water RDF of V74 and V66 indicates that this region highly disfavors interaction with water (**Figure 4.6.14**).

To quantitatively compare the stabilities of the two polymorphs and support the results from the structural analyses, I performed two additional calculations. First, the interaction energies from the sidechain packing structures of the rod and twister polymorphs were computed and compared (**Figure 4.6.15**), demonstrating how the different sidechain interaction structures contribute to the thermodynamic stabilities of the two polymorphs. The overall sidechain interaction energies of the rod polymorphs were observed to be more favorable than those of the twister polymorph. The order of favorable contributions of the sidechain interaction energies was packing1 > interface > packing2 for the two polymorphs. The sidechain interaction energies can be decomposed into in- and inter-plane contributions (**Figure 4.6.15**).

For the rod polymorph, the in-plane contributions were significantly greater than the inter-plane contributions of the sidechain interactions of the interface. Electrostatic interactions provide dominant contributions to in-plane interactions, which can be mainly attributed to the two salt bridges (K45-E57) at the rod interface. Inter-plane van der Waals (vdW) interactions provide relatively weak contributions to the stabilization of the interface structure. For the packing1 region, both in- and inter-plane interactions provide stability to the structure. Electrostatic interactions

from the salt bridge (E46-K80) and hydrophilic residues in the water channel contribute significantly to the stability, whereas vdW interactions contribute weakly to stabilizing the packing1 structure. For the packing2 region of the rod polymorph, the in- and inter-plane vdW interactions slightly contribute to the stability.

For the twister polymorph, in- and inter-plane interaction energies provide relatively weak stabilization of the interface structure, mostly from vdW interactions. Both the in- and inter-plane electrostatic interactions were slightly repulsive. For the packing1 region, both in- and inter-plane interactions provide similar stability to the structure, with the major contributions coming from electrostatic interactions. For the packing2 region of the twister polymorph, both vdW and electrostatic interactions from the glutamine ladder contribute to the stabilization of the inter-plane region, whereas the in-plane energy consists of relatively small stabilization due to vdW interactions and rather large repulsion from electrostatic interactions. In brief, the Greek-key motif of the rod polymorph allowed more residues to participate in stabilizing the packing structures. The rod polymorph exhibited stronger vdW interactions than the twister polymorph. Electrostatic energies provided significantly greater stabilization in the rod polymorph, owing to the presence of salt bridges in the rod.

We also performed pulling simulations and examined the process of detaching a single α-synuclein chain from the fibril structure, which provided a comparison of the dissociation energies of the two polymorphs (**Figure 4.6.16**). Pulling simulations demonstrated that the single-chain dissociation energy (integrals of the force extension curves) of the rod polymorph (2405.5 kJ/mol) was greater than that of the twister polymorph (1481.9 kJ/mol). The dissociation of the sidechain packing structures in each of the two polymorphs occurred in three steps of collective rupture

through different patterns. The rod polymorph chain was observed to be dissociated in the following order: the interface and part of packing1 (step 1, 0–3.97 ns), the rest of packing1 (step 2, 3.97–6.35 ns), and packing2 (step 3, 6.35–14.47 ns). The dissociation energy of each step was 1506.3, 207.4, and 445.9 kJ/mol, respectively. Dissociation of the twister polymorph occurred in the following order: packing2 and part of packing1 (step 1, 0–3.11 ns), remainder of packing1 (step 2, 3.11–5.06 ns), and interface (step 3, 5.06–7.65 ns). The dissociation energy of each step was 964.7, 175.1, and 217.4 kJ/mol, respectively. The rod polymorph was observed to require a stronger pulling force and larger energy consumption for dissociation than the twister polymorph.

In step 1 of rod polymorph pulling (0–3.97 ns), the preNAC sequence (residues 45–57) participates in the formation of both the interface and part of packing1, which requires a significantly greater force to rupture this structure. In other words, the Greek-key structure of the rod polymorph provided a large energy barrier for dissociation. In particular, salt bridges at K45-E57 in the interface and E46-K80 in packing1 dominantly contributed to the stabilization of this structure. The twister polymorph exhibited collective rupture of packing2 and part of packing1 in step 1 (0–3.11 ns). The residues in packing1 participating in water channel formation were observed to be simultaneously dissociated with packing2.

We calculated the potential of mean force curves along with the pulling simulation trajectories of rod and twister polymorphs as the reaction coordinates, which is shown in **Figure 4.6.17**. The horizontal axis (ξ) is the center of mass (COM) distance between the chain being detached and the rest part of the protofilament. The dissociation process is finished at ξ = 7.6 nm for twister and ξ = 10.9 nm for rod respectively. Relatively long COM distance for complete dissociation for rod

polymorph indicates that more complex kernel structure of rod polymorph requires more sidechain interactions. Therefore, we expect that both dissociation and association process is more difficult for rod than twister polymorph. The binding free energy calculated from the PMF curves is 527.42 kJ/mol for rod polymorph and 265.14 kJ/mol for twister polymorph respectively. The colored circles below the PMF curves indicate the COM distance values where sidechain interactions are ruptured through the pulling process. That is, each circle represents the COM distance of a single sidechain-sidechain interaction dissociation. The colors of the circles correspond to the sidechains of the same colors shown in the polymorph structures in **Figure 4.6.17**. The circles of the same color indicate the group of spatially neighboring interactions. I can observe that some of these neighboring circles are clustered while other group of circles are sparsely distributed. The clustered circles indicate that the corresponding sidechain interactions are vulnerable and therefore the disruptions occur in a row. On the other hand, the sparsely distributed circles indicate that the corresponding sidechain interactions are more persistent and require more energy to rupture.

In the rod polymorph, the rupture pattern of the first cluster (red circles) shows that the entrance of kernel and interface structures are weak and disrupted quickly. The orange circles show more sparse rupture pattern which means the corresponding interface structure is strong. The distant interval between orange and green groups indicates strong salt bridge interaction at the interface. The distant interval before the cluster of purple circles indicates the compact and stable backbone β-sheet structure due to the turn geometry formed by A76 residue. It is noticeable that the PMF curve keeps rising for almost 2 nm of COM distance until the disruption of the last sidechain interaction. This is because backbone β-sheet at C-terminal region is

slightly tilted along the protofibril axis, which locates the sidechain residue of the last interaction buried from the outer surface and makes it hard to being detached. In fact, this slightly titled terminal backbone β-sheet geometry allows a long helical pitch of rod structure.

In the twister polymorph, the kernel structure is simple compared to the rod polymorph. The most contrast feature of the dissociation process of twister structure compared to rod structure is, the kernel structure is completed disrupted before the dissociation of interface structure begins. The half of the interface structure (green) is the most persistent region while the rest half (blue) is much weak. Therefore, I can expect that the formation of single protofibril of twister polymorph is energetically not preferable.

The results of the sidechain interaction analyses, pulling simulations, and binding free energy calculations clearly illustrated that the rod polymorph has stronger sidechain interactions and exhibits higher dissociation energy (higher kinetic barrier for dissociation) than the twister polymorph.

The Greek-key motif allows the formation of a compact steric zipper interface and the kernel of the compact hydrophobic core structure of the sidechains. In addition, the K45-E57 and E46-K80 salt bridges strongly anchor the interface and kernel structures, respectively. In contrast, the twister polymorph has a less compact steric zipper and kernel structure with no salt bridge. In APBS (Adaptive Poisson–Boltzmann Solver) determination of solvation energies of the 4-layer paired protofilament, rod_{pair} (-1.7467×10^4 kJ/mol) was observed to be more stable than $twister_{pair}$ (-1.6428×10^4 kJ/mol). In the case of a single protofilament, the solvation energy of rod_{single} was -8.9617×10^3 kJ/mol, which is lower than that of $twister_{single}$ (-8.5543×10^3 kJ/mol). Therefore, the solvation structures of the rod polymorphs

were more stable than those of the twister polymorphs in both the paired and single forms.

4.5. Conclusions

In this study, I performed molecular dynamics simulations using the experimentally identified structures of the two major fibril polymorphs of wild-type α-synuclein. The results of the simulations suggested that the predominance of the rod-type polymorphs with Greek-key kernel structure in various experimental studies can be rationalized in terms of their high structural stability in aqueous environments, which may be attributed to the cooperative contributions of various stabilizing features. RMSDs, RMSFs and β-sheet content analyses indicated that highly stable structures in aqueous environments could be maintained by the combination of compact sidechain packing in the hydrophobic core, backbone geometry of maximal β-sheet content wrapping the hydrophobic core, and solvent-exposed sidechains with large fluctuations maximizing the solvation entropy. Between the two major polymorphs, the twister polymorph exhibited more flexible and less stable backbone β-sheet and hydrophobic core structures, suggesting that the rod polymorph with a highly compact Greek-key β-sheet geometry would be preferred in aqueous environments.

The paired structure of the protofilaments provided additional stability, especially at the interface region, by forming steric zipper interactions and hiding the hydrophobic residues from exposure to water. The rod and twister polymorphs have different interface sequences: the preNAC ($_{46}$EGVVHGVATVA$_{56}$) sequence and the NACore ($_{68}$GAVVTGVTAVA$_{78}$) sequence, which is a part of the highly hydrophobic

NAC region (61–95), for the rod and twister polymorphs, respectively. Pulling simulation and binding free energy calculation results showed that the structure concerning the interface sequence of rod polymorph has relatively strong dissociation energy compared to the twister polymorph. The Greek-key structure in the rod polymorph allows the preNAC sequence participate in both interface and intramolecular sidechain packing (packing1), and the salt bridges thereby formed both in interface and packing1 provide strong electrostatic energy.

It should be noted that all six familial PD mutations (E46K, H50Q, G51D, A53E, A53T, and A53V) were associated with the preNAC sequence. Familial PD mutations might induce slight structural changes and modify the relative stabilities of the rod polymorphs of the wild-type α-synuclein, which results in different polymorphic fibril compositions and ultimately affects pathogenicity. There is experimental evidence that familial mutants have similar shapes to the wild-type rod polymorph.[112] According to these observations, familial mutations are not expected to induce major structural transformations, such as rod-to-twister conversion. Instead, subtle perturbations at the interface caused by familial mutations could lead to a new structure that is more prone to pathogenicity with enhanced stability, while maintaining the overall architecture of the predominant rod polymorph. A recent high-resolution cryo-EM fibril structure of the α-synuclein E46K mutant supports this argument. Eisenberg *et al.*[103] hypothesized that there is a deeper energy state than the wild-type rod polymorph structure, into which the folding pathway is kinetically trapped by the E46-K80 salt bridge interaction. The E46K mutation eliminates this trap and unlocks a more stable pathogenic structure. Although this structure is a significant rearrangement from the wild-type rod structure, it looks more like a rod than a twister structure. In fact, the protofilament interface forms at residues 45–57

for E46K, which is almost identical to the preNAC sequence but does not overlap with the NACore sequence. Eisenberg *et al*. reported that the twister polymorph of wild-type or mutant α-synuclein has not been observed in their cryo-EM study despite the predicted favorable interaction between K46 and E83 in E46K mutant.[103] They speculated that the difficulty in reproducing the twister structure for mutations or change in buffer conditions could be a stochastic nature of the kinetic process in the twister fibril formation.

From the calculations of sidechain interactions and solvation energies calculations, the energy difference between the paired and the single forms of the 4-layered protofilament can be evaluated as $\Delta E_{single\rightarrow paired} = E_{solv,paired} - 2 \times E_{solv,single} + 4 \times E_{interface}$ ($E_{solv,paired}$ and $E_{solv,single}$ are solvation energies of the paired and the single forms, respectively. $E_{interface}$ is the sidechain interaction energy of the interface for the single layer in paired protofilaments). In our simulations, $\Delta E_{single\rightarrow paired}$= –375.89 kJ/mol and 621.65 kJ/mol for the rod and the twister polymorphs, respectively. These results imply that the formation of the paired rod polymorph is energetically favorable process while the formation of the paired twister polymorph is energetically unfavorable process. Therefore, it is expected that the formation of the paired twister polymorph can occur in low probability through a stochastic diffusion, which is consistent with the speculation of Eisenberg *et al*.[103] In this case, the solvation entropy of water and specific solvent conditions (buffer condition, counter ions, etc.) will determine whether the paired twister polymorph can be stabilized or not.

It can be concluded that the rod polymorph may exhibit high structural stability, comprising a major pathogenic strain of the wild-type α-synuclein fibril. I anticipate

that the present study will provide insights for explaining the pathogenicity of various α-synuclein fibril structures and other amyloid peptides from the structural perspectives. To achieve an overall understanding of the mechanisms by which α-synuclein and other amyloid fibril polymorphisms play a role in the development of pathogenicity, further experimental and computational studies will be needed.

4.6. Figures

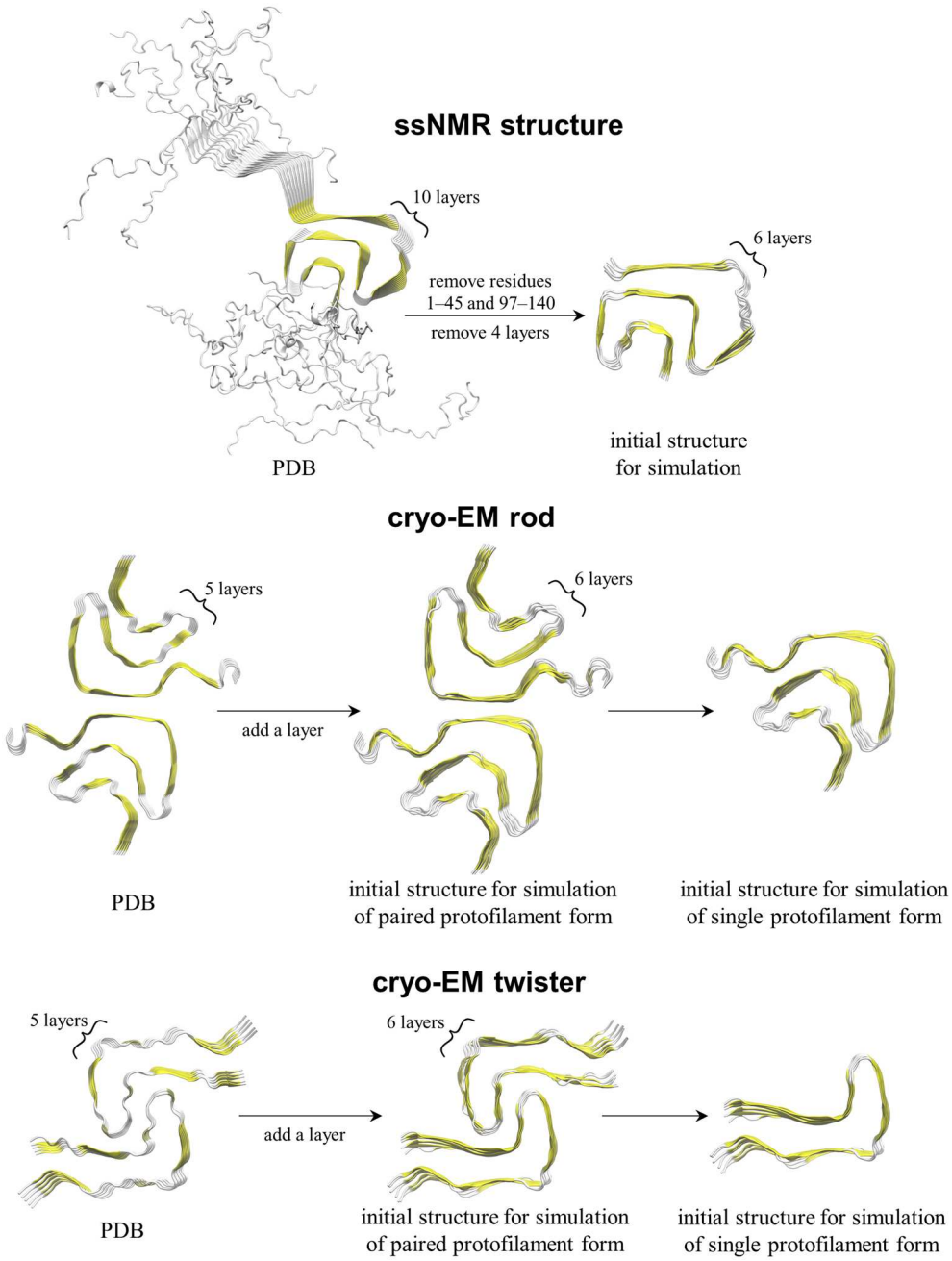


Figure 4.6.1. PDB structures of ssNMR (PDB ID: 2N0A) and cryo-EM (rod: 6CU7, twister: 6CU8) and the initial structures for simulations derived from PDB structures.

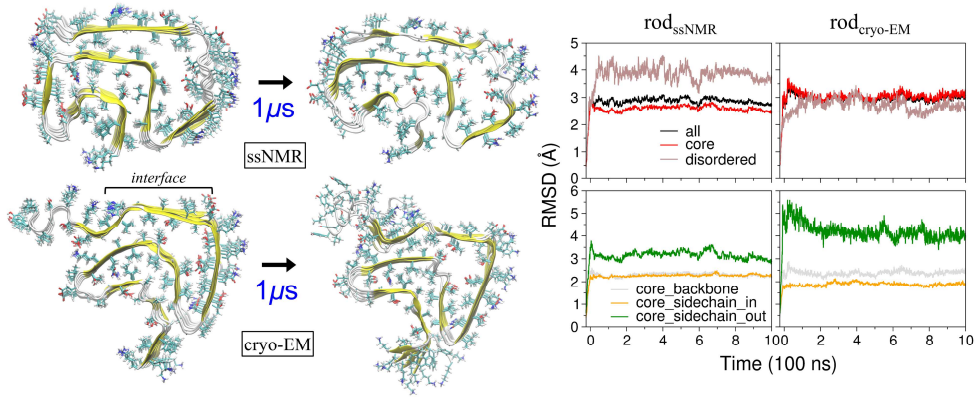


Figure 4.6.2. (left) Experimental rod structures of α-synuclein protofilament from ssNMR and cryo-EM experiments, and their converged structures after 1 μs simulation in an aqueous environment. (right) RMSD curves from the simulation. RMSD curves for whole sequence (residues 46–96), core region (residues 46–54 and 63–96), and disordered region (residues 55–62), shown in black, red, and brown colors, respectively. Core and disordered regions are defined from ssNMR data[101]. For the core region, RMSD curves for the backbone (grey), sidechains buried inside (orange), and sidechains exposed outward (green) are separately plotted.

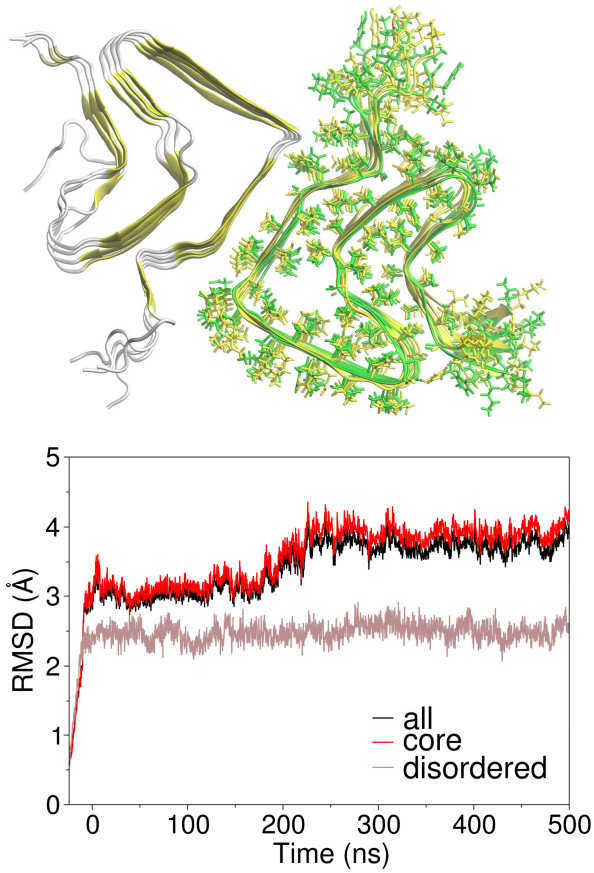


Figure 4.6.3. Structure of the rod polymorph consisting of paired and interfaced protofilaments after 500 ns simulation (yellow) and the corresponding RMSD curve. For comparison, a simulation structure of single protofilament (green) is superimposed.

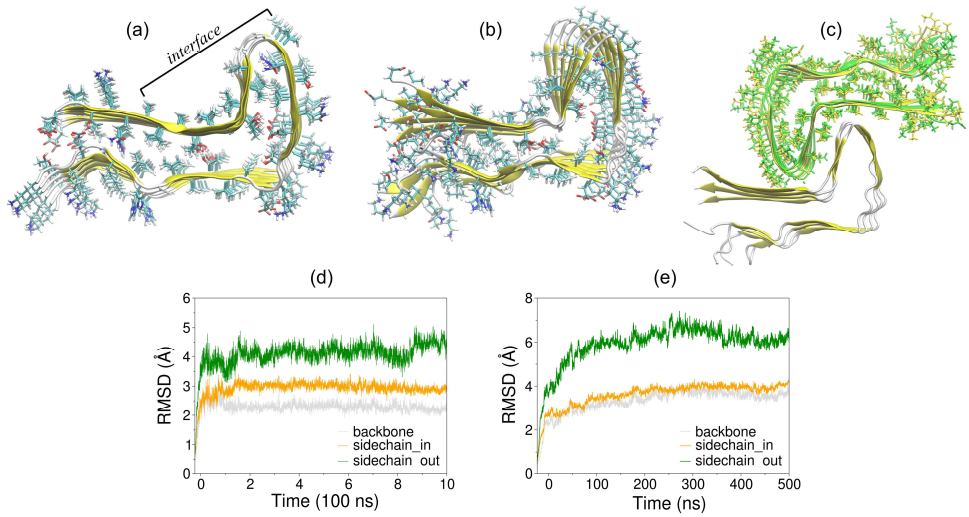


Figure 4.6.4. Simulation results of twister polymorphs. (a) Experimental cryo-EM structure. (b) Structure after 1 μs simulation. (c) Structure of twister polymorph consisting of paired interfaced protofilaments after 500 ns simulation (yellow), on which a single protofilament simulation structure (green) is superimposed for comparison. (d) RMSD curves from 1 μs simulation of a single protofilament structure. (e) RMSD curves from 500 ns simulation of paired protofilaments structure. RMSD curves for backbone (grey), sidechains buried inside (orange), and sidechains exposed outward (green) are separately plotted.

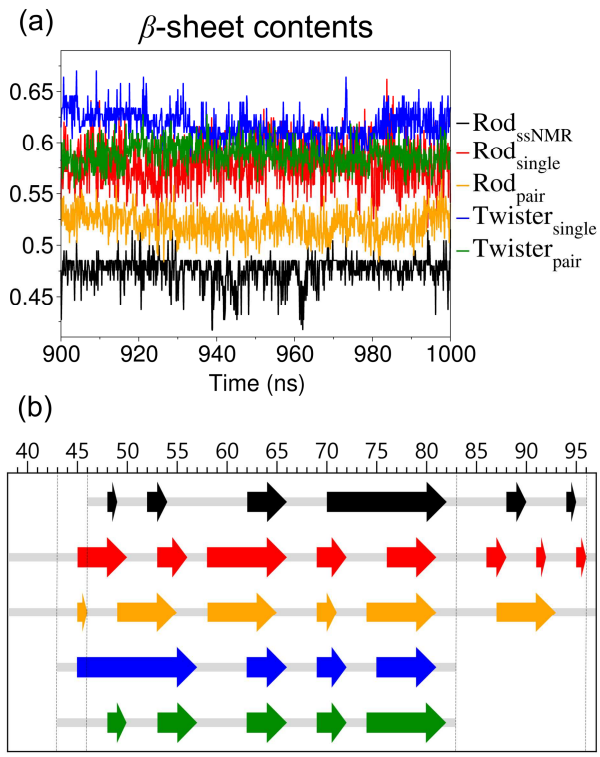


Figure 4.6.5. (a) β-sheet contents during the last 100 ns of each simulation using the ssNMR rod structure (Rod_{ssNMR}), rod polymorph with single protofilament (Rod_{single}), rod polymorph with a pair of interfaced protofilaments (Rod_{pair}), twister polymorph with a single protofilament ($Twister_{single}$), and twister polymorph with a pair of interfaced protofilaments ($Twister_{pair}$). (b) Backbone β-sheet geometries of five structures obtained from (a).

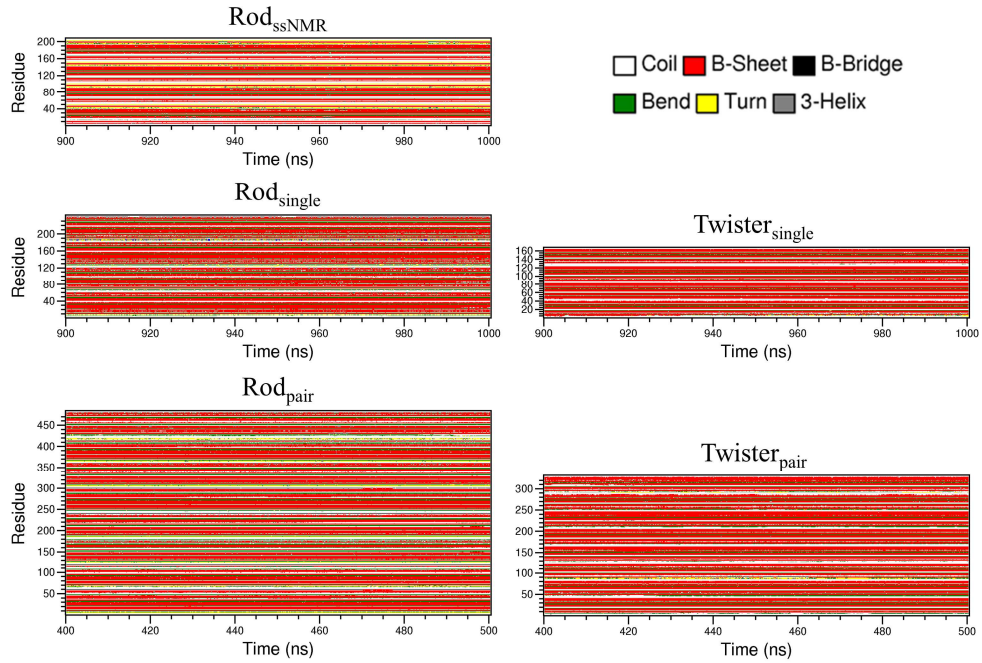


Figure 4.6.6. Secondary structure maps calculated by DSSP algorithm for the last 100 ns of each trajectory of five different structures.

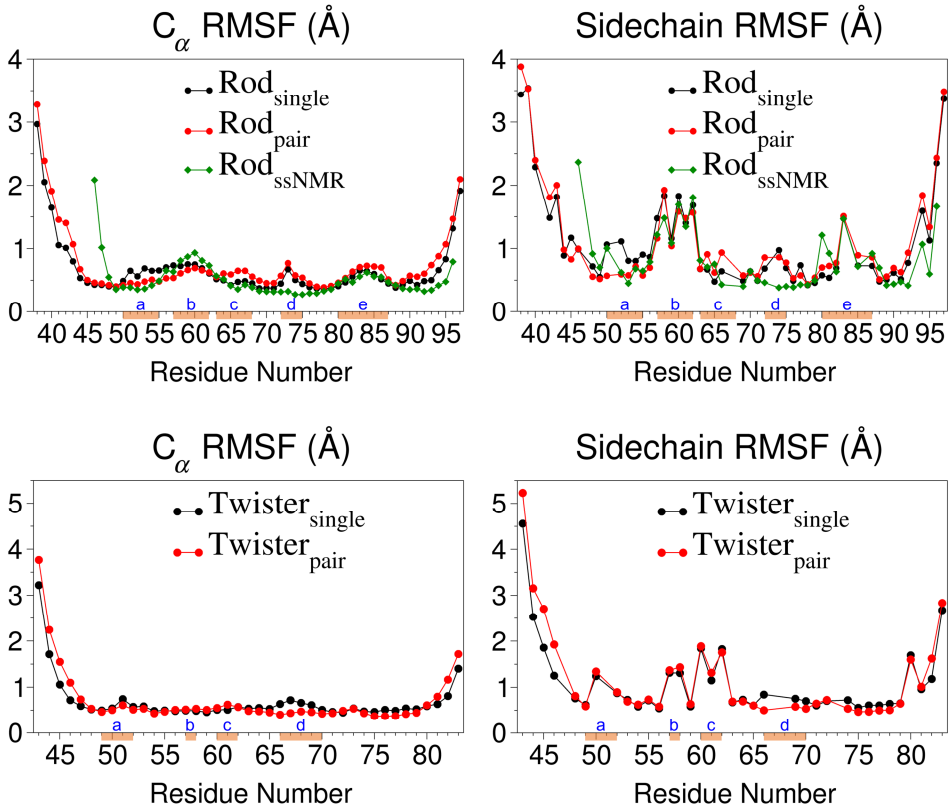


Figure 4.6.7. RMSF of C_α (left column) and sidechain (right column) for each residue of rod polymorphs (top row) and twister polymorphs (bottom row). The thick brown bars with the letters a, b, c, and so on, upon them denote the characteristic positions, mainly the peaks, where corresponding structures are shown in **Figure 4.6.8**.

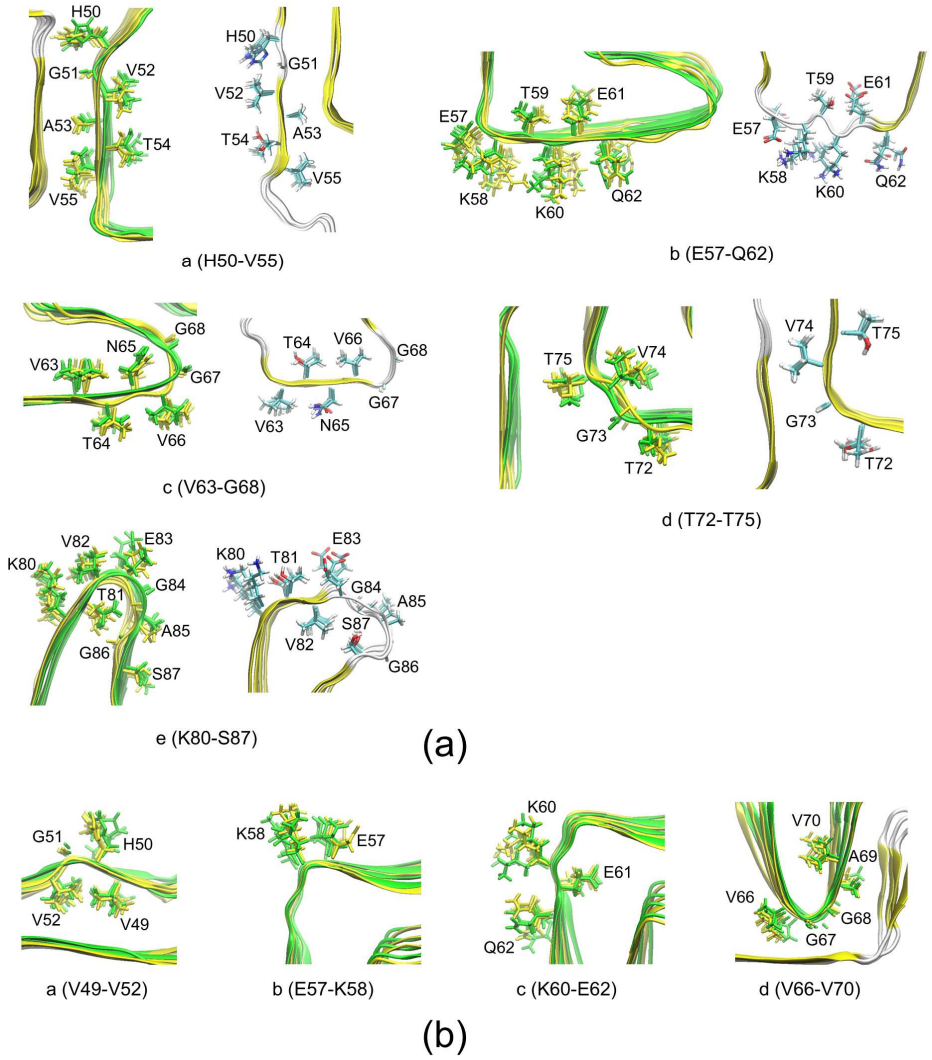


Figure 4.6.8. (a) Structures of sequences denoted by a, b, c, d, e in the RMSF of rod polymorphs in **Figure 4.6.7**. Shown on the left side, single protofilament structure (Rod_{single}, green) is superimposed on paired protofilament structure (Rod_{pair}, yellow). On the right side, Rod_{ssNMR} structure is drawn in yellow backbone and colored sidechains. (b) Structures of sequences denoted by a, b, c, d in the RMSF of twister polymorphs in **Figure 4.6.7**. Single protofilament structure ($Twister_{single}$, green) is superimposed on paired protofilament structure ($Twister_{pair}$, yellow).

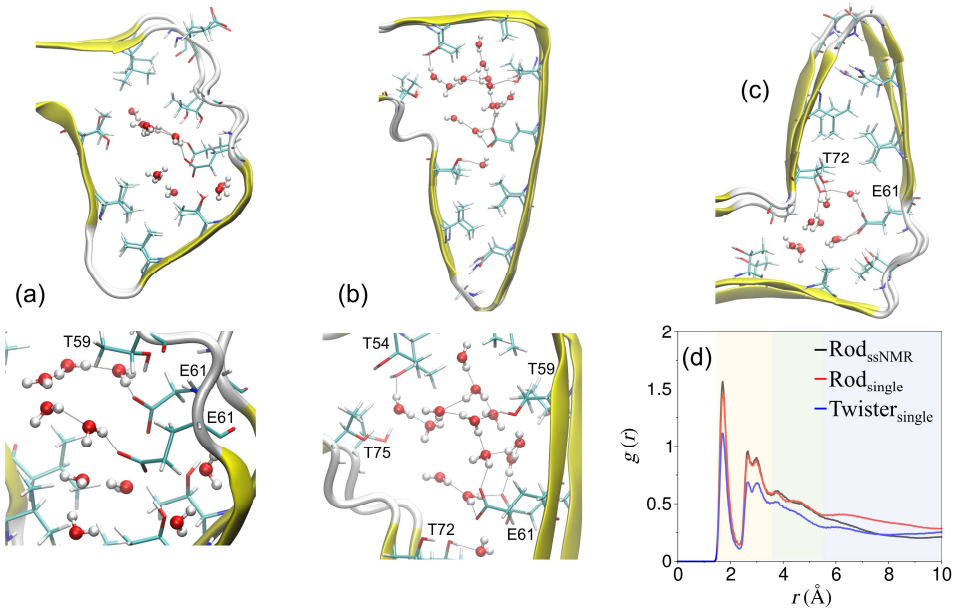


Figure 4.6.9. Structures of water channels in the simulations of (a) ssNMR structure, (b) cryo-EM rod structure, (c) cryo-EM twister structure, and (d) radial distribution functions of water around residue E61. Yellow, green, and blue background colors denote the areas of different hydration shells.

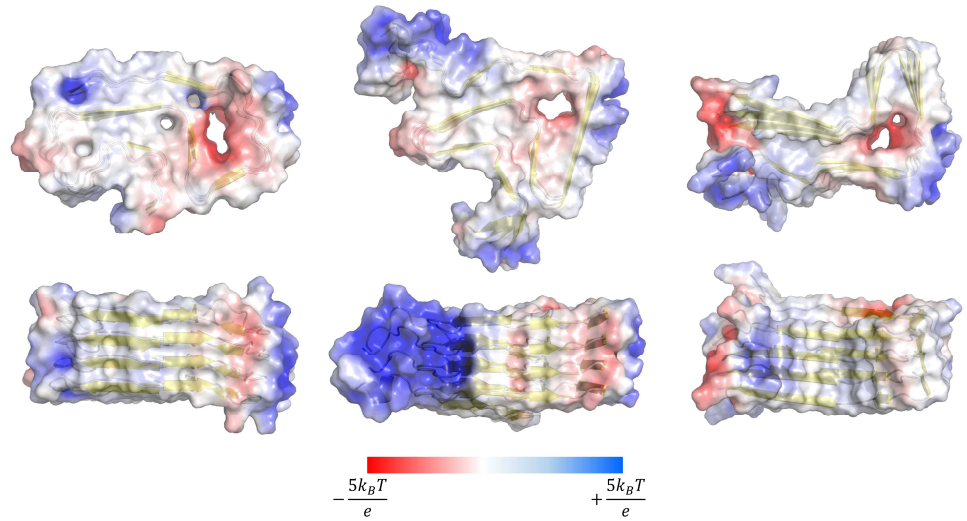


Figure 4.6.10. Electrostatic potential maps at solvent-accessible surfaces of the structures in an aqueous environment derived from the simulations of ssNMR (left), cryo-EM rod (middle), and cryo-EM twister (right) structures.

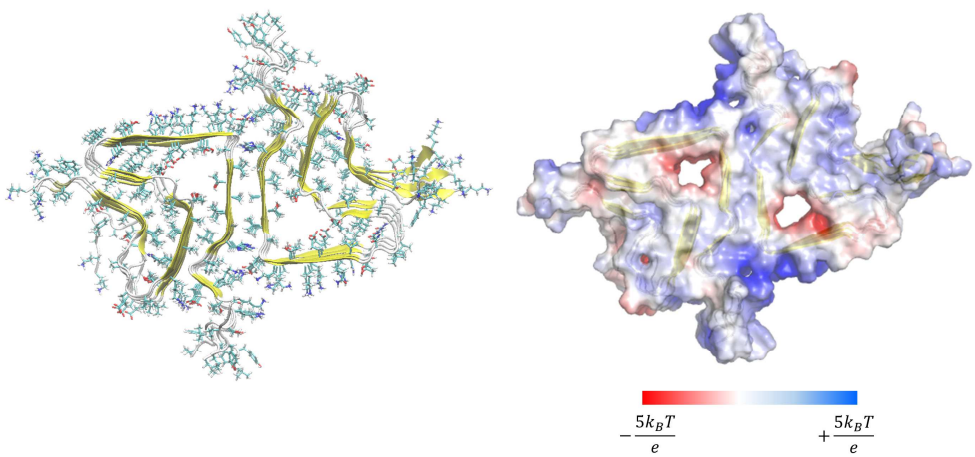


Figure 4.6.11. Electrostatic potential map at the solvent-accessible surface from the simulation of the rod polymorph with paired interfaced protofilaments.

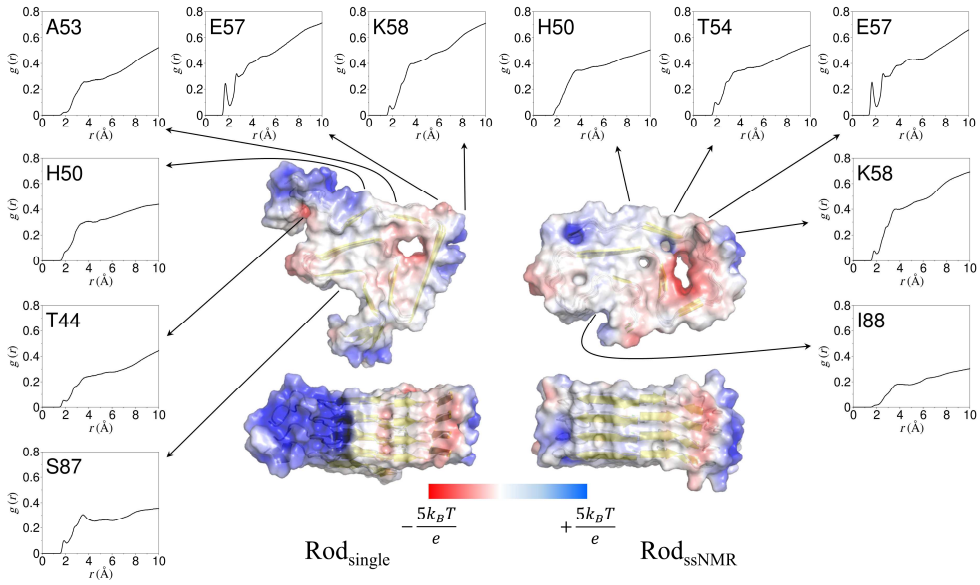


Figure 4.6.12. Electrostatic potential maps at the solvent-accessible surface from the simulations of single rod protofilament structures (left: cryo-EM, right: ssNMR), with radial distribution functions of water around the residues at the interface region (Rod$_{single}$: H50, A53, E57, K58 / Rod$_{ssNMR}$: H50, T54, E57, K58) and the residues at the pocket shape surfaces (Rod$_{single}$: T44, S87 / Rod$_{ssNMR}$: I88).

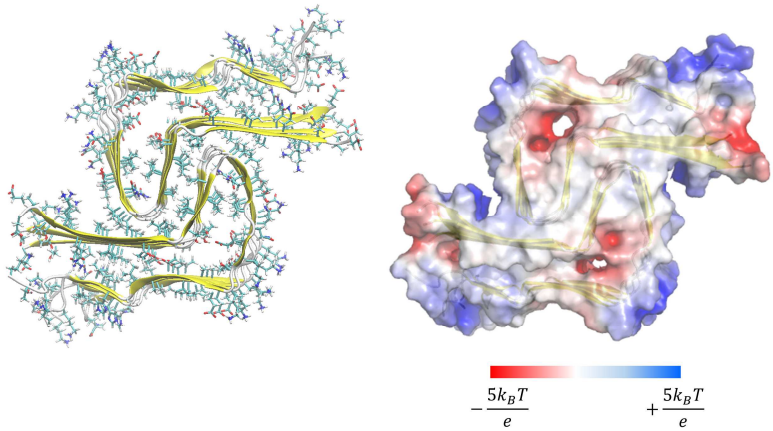


Figure 4.6.13. Electrostatic potential map at the solvent-accessible surface from the simulation of the twister polymorph with paired interfaced protofilaments.

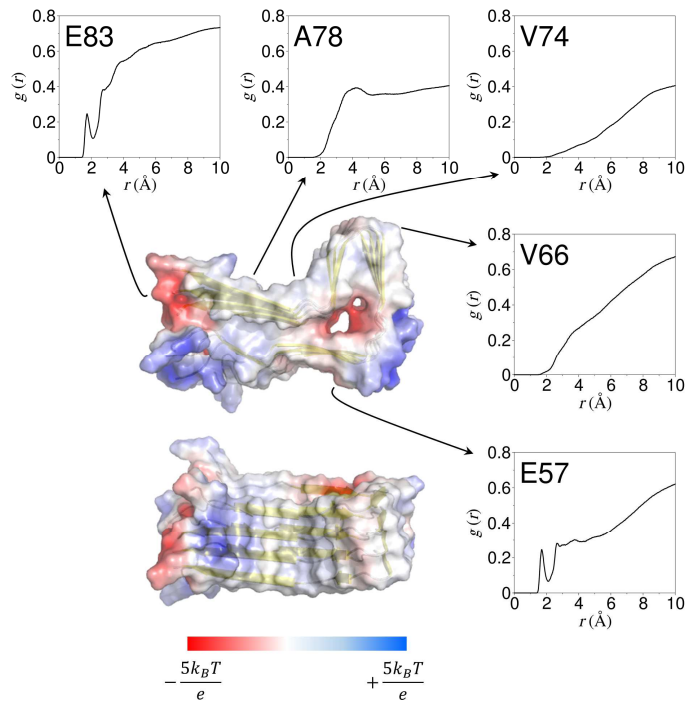
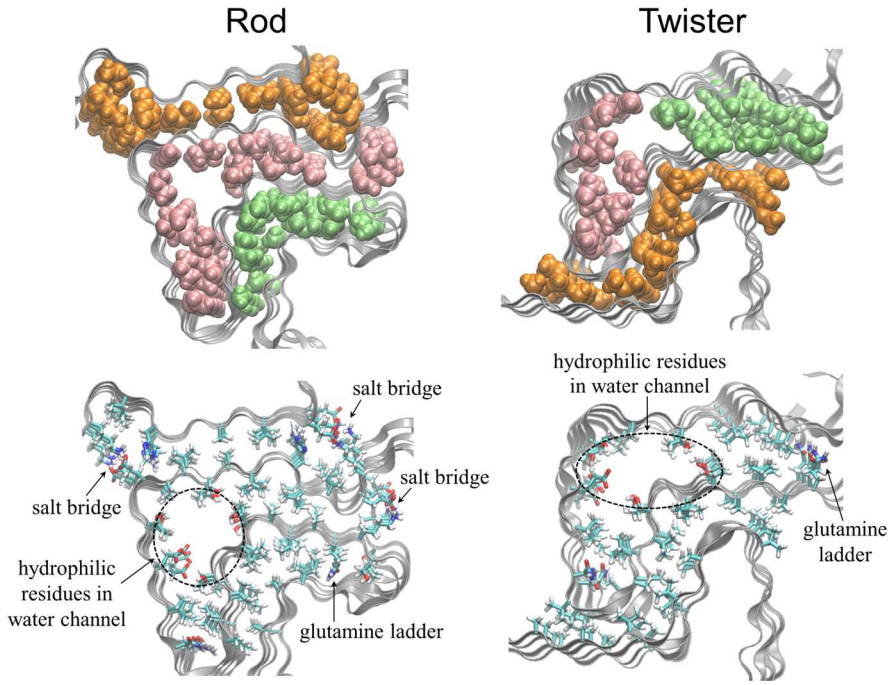


Figure 4.6.14. Electrostatic potential maps at the solvent-accessible surface from the simulation of the twister polymorph with a single protofilament, with radial distribution functions of water around A78, V74, V66, E83, and E57.



Unit : kJ/mol

Rod	in-plane sidechain interactions			inter-plane sidechain interactions		
region	total	electrostatic	vdW	total	electrostatic	vdW
interface	-219.235	-220.197	0.962289	11.08531	36.0795	-24.9942
packing1	-218.010	-191.140	-26.8693	-144.640	-104.972	-39.6673
packing2	182.614	202.155	-19.5412	-29.2253	3.20132	-32.4267

Twister	in-plane sidechain interactions			inter-plane sidechain interactions		
region	total	electrostatic	vdW	total	electrostatic	vdW
interface	-7.44678	2.77933	-10.2261	-7.26053	3.40480	-10.6653
packing1	-69.6141	-55.9193	-13.6948	-56.7332	-31.8854	-24.8478
packing2	264.059	270.224	-6.16534	-36.4295	-7.22127	-29.2082

Figure 4.6.15. Interaction energies of sidechain packing structures of rod and twister different polymorphs. (Up) Sidechain structures of van der Waals packing and polar interactions for rod and twister polymorphs. (Down) In-plane and inter-plane electrostatic, van der Waals, and total sidechain interaction energy values of packing1 (pink spheres), packing2 (green spheres), and interface (orange spheres) regions for rod and twister polymorphs.

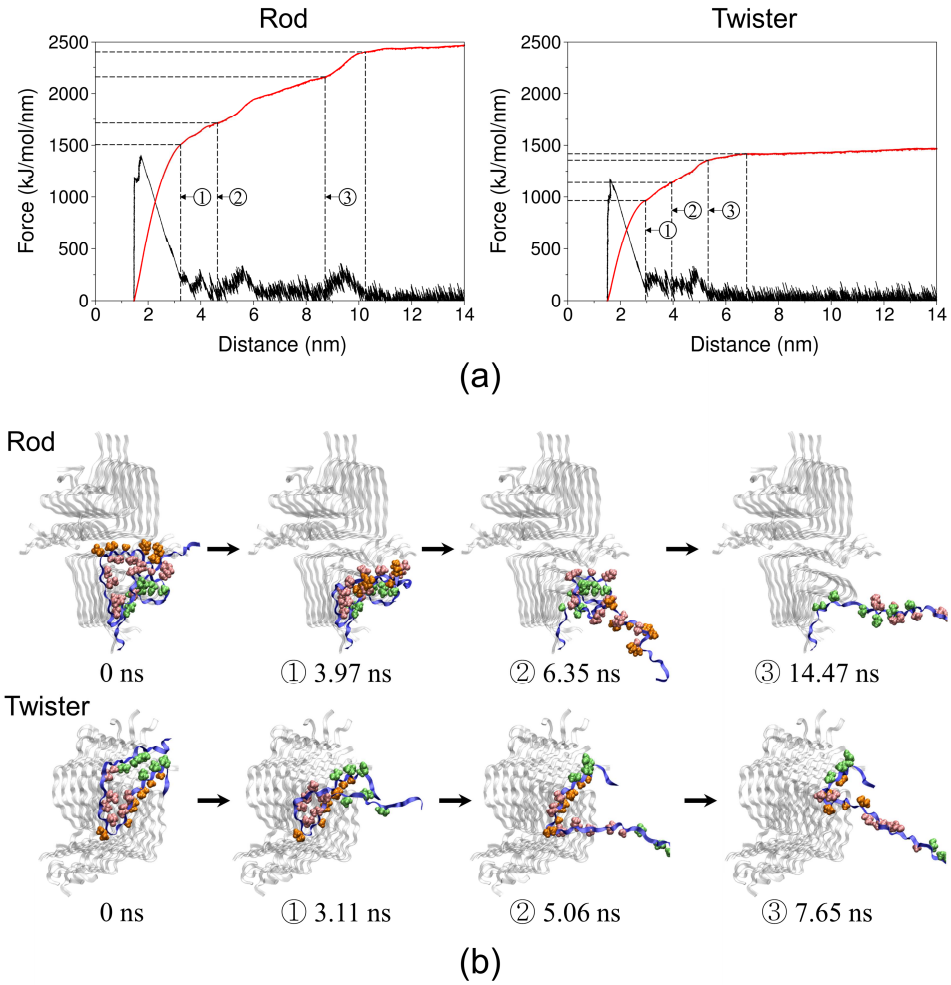


Figure 4.6.16. (a) Force extension curves from pulling simulations of rod and twister fibrils. The horizontal axis represents the distance between the center of mass of the chain being separated by pulling and the center of mass of the adjacent chain in the fibril structure. Red curves represent the integrals of force curves indicating the energies consumed for pulling up to the distance. Dashed vertical lines indicate the moments of collective ruptures of sidechain packing structures. (b) Trajectory snapshots corresponding to the dashed lines in (a) that show the different patterns of collective rupture between rod and twister structures.

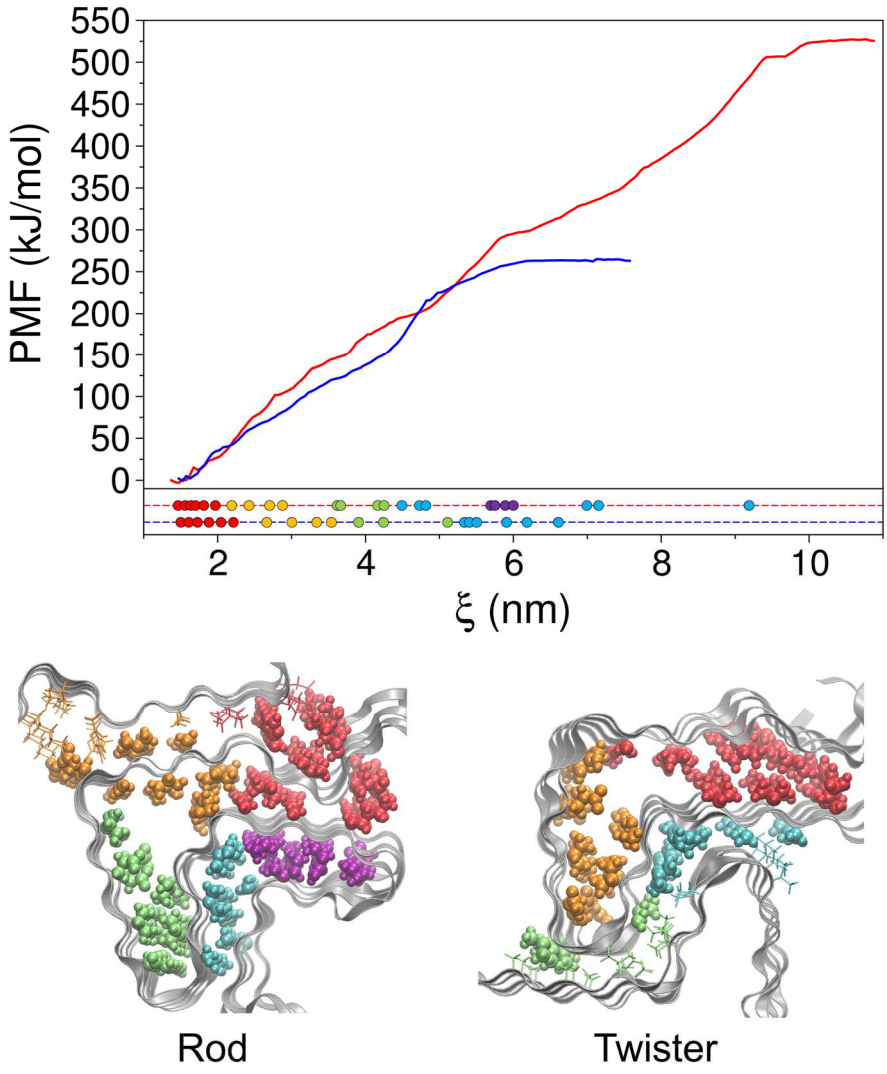


Figure 4.6.17. Potential of mean force (PMF) curves along with the pulling simulation trajectories of rod (red) and twister (blue) polymorphs as the reaction coordinates. ξ is the COM distance between the chain being detached and the rest part of the protofilament. The circles on the dashed horizontal lines show the COM distances where residual interactions of kernel structures are ruptured.

Chapter 5. Conclusion

Neurodegenerative diseases, including Alzheimer's disease (AD) and Parkinson's disease (PD), encompass a range of neurological disorders characterized by neuronal loss. The progression of neurodegenerative diseases is associated with the aggregation of amyloid proteins, specifically those belonging to the category of Intrinsically Disordered Proteins (IDPs). Despite advancements in experimental techniques, explaining the structural diversity of IDPs, as well as the diverse intermediate aggregates and polymorphic fibril structures they give rise to, continues to present challenges. In this thesis, theoretical approaches using Molecular Dynamics (MD) simulations are employed to overcome such challenges.

In **Chapter 2**, a novel replica exchange scheme, termed as *Tq*-REM, was proposed combining *T*-REM and *q*-REM. The scheme of *Tq*-REM substitutes the high-temperature *T*-replicas used in the *T*-REM with *q*-replicas using the effective potential with reduced barriers. The goal of this “combo” scheme is to optimize efficiency by combining the advantages of both *T*-REM and *q*-REM methods while mitigating their individual drawbacks. The conformational sampling abilities of the two REM methods (*T*-REM, and *Tq*-REM) were evaluated by determining the free energy surface for the folding dynamics of three small peptides. The results indicated that the convergence of the free energy surfaces was enhanced when utilizing the *Tq*-REM approach in comparison to the conventional *T*-REM method. Notably, the trajectories of *Tq*-REM effectively explored and sampled the relevant conformations associated with all of the metastable folding intermediates, while the *T*-REM method exhibited poor sampling results for certain local minimum structures.

In **Chapter 3**, the structure and aggregation pathway of Aβ_{42} protofibril were

discussed. The structural characteristics of the triple-β motif remained highly conserved throughout the duration of the 10 μs simulation. Through analysis, I determined the specific arrangement of backbone and sidechain interactions that contribute to stabilizing the hydrophobic cores within the triple-β shape of the Aβ_{42} fibril motif. In addition to investigating the hydrophobic core structures, I also explored the behavior of the relatively unstructured N-terminal sequence. This analysis shed light on the potential role of N-terminal flexibility in understanding the polymorphic nature observed in experimental Aβ_{42} fibril structures. By analyzing the trajectories from pulling and REMD simulations, we obtained valuable insights that provide potential explanations for the dock-and-lock mechanism of Aβ_{42} fibrillization.

In **Chapter 4**, the structural stability as the origin of the pathogenicity of α-synuclein protofibrils was discussed. It was shown that the dominance of rod polymorphs with a Greek-key kernel structure in multiple experimental studies can be justified by their exceptional structural stability in aqueous environments. This stability was likely a result of the synergistic contribution of various stabilizing elements: the combination of closely packed sidechains within the hydrophobic core, the structural arrangement of the backbone that maximizes the β-sheet content surrounding the hydrophobic core, and sidechains that are exposed to the solvent and exhibit significant fluctuations, thereby maximizing the entropy associated with solvation. In addition, the outcomes of pulling simulations and calculations of binding free energy revealed that the interface sequence in the rod polymorph exhibited a relatively higher dissociation energy compared to the twister polymorph. Based on these findings, it could be inferred that the rod polymorph, which displays high structural stability, represents a prominent pathogenic variant of the wild-type

α-synuclein fibril.

All the discussions presented in this thesis are based on theoretical approaches, which allowed for the examination of structural and dynamical detailed features that are difficult to observe in experiments. Even at this moment, ongoing efforts to elucidate neurodegenerative diseases persist through not only *in silico* but also *in vivo* and *in vitro* experiments. The development of techniques capable of elucidating the structure, such as ssNMR and cryo-EM, as well as advancements in imaging technologies like PET and MRI, which allow direct observations of the patient's brain, have greatly contributed to understanding the causes and progression of neurodegenerative diseases. I hope that the new sampling method proposed in this thesis and the discussed characteristics of protofibril structures related to AD and PD, will be integrated with the results of future studies and will be contribute to the prevention and treatment of neurodegenerative diseases.

Multiple research groups have consistently engaged in protein structure prediction studies, aiming to computationally determine the three-dimensional structure of proteins or peptides based on their sequence information.[113] While a deep learning technique called AlphaFold has seemingly resolved the challenge of structure prediction for conventional proteins with known native structures, it should be noted that AlphaFold may not possess the capability to accurately predict the specific folding pathway of proteins.[114] Concurrently, endeavors are being made to employ AlphaFold for predicting the structure of IDPs or proteins that encompass Intrinsically Disordered Regions (IDRs).[115] There is an ongoing effort to apply AlphaFold to the study of amyloid aggregation, however, the presence of amyloid polymorphs and diverse aggregation intermediates presents a significant challenge that hinders accurate prediction using this method.[116] It can be said that the

development of machine learning techniques capable of predicting the structural and dynamic characteristics of IDPs and amyloid aggregates is also a meaningful research field.

Current research indicates that mature aggregates, specifically fibrils, are not inherently harmful to the nervous system and might even possess a protective function. On the other hand, it is widely believed that oligomers are responsible for the toxic effects observed in various pathological conditions.[117] Consequently, there has been a significant surge in research focused on investigating the structure and mechanisms of amyloid oligomers, which are considered a key toxic entity implicated in the initiation of neurodegenerative diseases.[118] Recently, the atomic structures of complete Aβ_{42} tetramers and octamers were successfully unveiled in a membrane-like setting through the use of NMR and MS techniques.[119] The results demonstrated that the oligomeric structure has the ability to form a pore stabilized by lipids, which can ultimately lead to the development of toxicity. Characterizing small prefibrillar aggregates, which are transient structures during the conversion from toxic oligomers to fibrils, is still challenging experimentally. Therefore, structural and dynamic analysis using computer simulations can be a valuable approach in the study of those aggregates in the future works.

Bibliography

(1) Jalili, M.; Gebhardt, T.; Wolkenhauer, O.; Salehzadeh-Yazdi, A. Unveiling Network-Based Functional Features Through Integration of Gene Expression into Protein Networks. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2018**, *1864* (6, Part B), 2349-2359. DOI: 10.1016/j.bbadis.2018.02.010. del Sol, A.; Balling, R.; Hood, L.; Galas, D. Diseases as Network Perturbations. *Current Opinion in Biotechnology* **2010**, *21* (4), 566-571. DOI: 10.1016/j.copbio.2010.07.010.

(2) Somvanshi, P. R.; Venkatesh, K. V. A Conceptual Review on Systems Biology in Health and Diseases: From Biological Networks to Modern Therapeutics. *Systems and Synthetic Biology* **2014**, *8* (1), 99-116. DOI: 10.1007/s11693-013-9125-3. Csermely, P.; Korcsmáros, T.; Kiss, H. J. M.; London, G.; Nussinov, R. Structure and Dynamics of Molecular Networks: A Novel Paradigm of Drug Discovery: A Comprehensive Review. *Pharmacology & Therapeutics* **2013**, *138* (3), 333-408. DOI: 10.1016/j.pharmthera.2013.01.016.

(3) Wilson, D. M.; Cookson, M. R.; Van Den Bosch, L.; Zetterberg, H.; Holtzman, D. M.; Dewachter, I. Hallmarks of Neurodegenerative Diseases. *Cell* **2023**, *186* (4), 693-714. DOI: 10.1016/j.cell.2022.12.032.

(4) Gan, L.; Cookson, M. R.; Petrucelli, L.; La Spada, A. R. Converging Pathways in Neurodegeneration, from Genetics to Mechanisms. *Nature Neuroscience* **2018**, *21* (10), 1300-1309. DOI: 10.1038/s41593-018-0237-7. Erkkinen, M. G.; Kim, M.-O.; Geschwind, M. D. Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology* **2018**, *10* (4), a033118. DOI: 10.1101/cshperspect.a033118.

Dugger, B. N.; Dickson, D. W. Pathology of Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology* **2017**, *9* (7), a028035. DOI: 10.1101/cshperspect.a028035. Hinz, F. I.; Geschwind, D. H. Molecular Genetics of Neurodegenerative Dementias. *Cold Spring Harbor Perspectives in Biology* **2017**, *9* (4), a023705. DOI: 10.1101/cshperspect.a023705.

(5) Sengupta, U.; Kayed, R. Amyloid β, Tau, and α-Synuclein Aggregates in the Pathogenesis, Prognosis, and Therapeutics for Neurodegenerative Diseases. *Progress in Neurobiology* **2022**, *214*, 102270. DOI: 10.1016/j.pneurobio.2022.102270. Hansson, O. Biomarkers for Neurodegenerative Diseases. *Nature Medicine* **2021**, *27* (6), 954-963. DOI: 10.1038/s41591-021-01382-x. Soto, C.; Pritzkow, S. Protein Misfolding, Aggregation, and Conformational Strains in Neurodegenerative Diseases. *Nature Neuroscience* **2018**, *21* (10), 1332-1340. DOI: 10.1038/s41593-018-0235-9. Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nature Reviews Molecular Cell Biology* **2014**, *15* (6), 384-396. DOI: 10.1038/nrm3810.

(6) Chiti, F.; Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry* **2017**, *86* (1), 27-68. DOI: 10.1146/annurev-biochem-061516-045115.

(7) Sipe, J. D.; Benson, M. D.; Buxbaum, J. N.; Ikeda, S.-i.; Merlini, G.; Saraiva, M. J. M.; Westermark, P. Amyloid Fibril Proteins and Amyloidosis: Chemical Identification and Clinical Classification International Society of Amyloidosis 2016 Nomenclature Guidelines. *Amyloid* **2016**, *23* (4), 209-213. DOI: 10.1080/13506129.2016.1257986.

(8) Elhabashy, H.; Merino, F.; Alva, V.; Kohlbacher, O.; Lupas, A. N. Exploring

Protein-Protein Interactions at the Proteome Level. *Structure* **2022**, *30* (4), 462-475. DOI: 10.1016/j.str.2022.02.004.

(9) Lutter, L.; Aubrey, L. D.; Xue, W.-F. On the Structural Diversity and Individuality of Polymorphic Amyloid Protein Assemblies. *Journal of Molecular Biology* **2021**, *433* (20), 167124. DOI: 10.1016/j.jmb.2021.167124. Almeida, Z. L.; Brito, R. M. M. Structure and Aggregation Mechanisms in Amyloids. *Molecules* **2020**, *25* (5), 1195. DOI: 10.3390/molecules25051195. Adamcik, J.; Mezzenga, R. Amyloid Polymorphism in the Protein Folding and Aggregation Energy Landscape. *Angewandte Chemie International Edition* **2018**, *57* (28), 8370-8382. DOI: 10.1002/anie.201713416.

(10) Kulkarni, P.; Leite, V. B. P.; Roy, S.; Bhattacharyya, S.; Mohanty, A.; Achuthan, S.; Singh, D.; Appadurai, R.; Rangarajan, G.; Weninger, K.; et al. Intrinsically Disordered Proteins: Ensembles at the Limits of Anfinsen's Dogma. *Biophysics Reviews* **2022**, *3* (1), 011306. DOI: 10.1063/5.0080512. Strodel, B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. *Journal of Molecular Biology* **2021**, *433* (20), 167182. DOI: 10.1016/j.jmb.2021.167182. Fatafta, H.; Samantray, S.; Sayyed-Ahmad, A.; Coskuner-Weber, O.; Strodel, B. Chapter Five - Molecular Simulations of IDPs: From Ensemble Generation to IDP Interactions Leading to Disorder-to-Order Transitions. In *Progress in Molecular Biology and Translational Science*, Uversky, V. N. Ed.; Vol. 183; Academic Press, 2021; pp 135-185.

(11) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. *Annual Review of Biophysics* **2011**, *40* (1), 41-62. DOI: 10.1146/annurev-biophys-042910-155255.

(12) Lee, M.; Yoon, J.; Jang, S.; Shin, S. Conformational Sampling of

Metastable States: Tq-REM as a Novel Replica Exchange Method. *Physical Chemistry Chemical Physics* **2017**, *19* (7), 5454-5464. DOI: 10.1039/C6CP05322J.

(13) Lee, M.; Yoon, J.; Shin, S. Computational Study on Structure and Aggregation Pathway of Aβ42 Amyloid Protofibril. *The Journal of Physical Chemistry B* **2019**, *123* (37), 7859-7868. DOI: 10.1021/acs.jpcb.9b07195.

(14) Yoon, J.; Lee, M.; Park, Y.; Lee, K.; Shin, S. In Silico Investigation of the Structural Stability as the Origin of the Pathogenicity of α-Synuclein Protofibrils. *Journal of Biomolecular Structure and Dynamics* **2023**, 1-13. DOI: 10.1080/07391102.2023.2199077 *(Published Online)*.

(15) Adcock, S. A.; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chemical Reviews* **2006**, *106* (5), 1589-1615. DOI: 10.1021/cr040426m. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press, 2017. DOI: 10.1093/oso/9780198803195.001.0001. Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press, 2002. DOI: 10.1016/B978-0-12-267351-1.X5000-7.

(16) Onuchic, J. N.; and, Z. L.-S.; Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annual Review of Physical Chemistry* **1997**, *48* (1), 545-600. DOI: 10.1146/annurev.physchem.48.1.545. Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. Challenges in Protein-Folding Simulations. *Nature Physics* **2010**, *6* (10), 751-758. DOI: 10.1038/nphys1713. Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding. *Current Opinion in Structural Biology* **2013**, *23* (1), 58-65. DOI: 10.1016/j.sbi.2012.11.002. Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science*

2012, *338* (6110), 1042-1046. DOI: 10.1126/science.1219021.

(17) Christen, M.; van Gunsteren, W. F. On Searching in, Sampling of, and Dynamically Moving Through Conformational Space of Biomolecular Systems: A Review. *Journal of Computational Chemistry* **2008**, *29* (2), 157-166. DOI: 10.1002/jcc.20725.

(18) Morriss-Andrews, A.; Shea, J.-E. Computational Studies of Protein Aggregation: Methods and Applications. *Annual Review of Physical Chemistry* **2015**, *66* (1), 643-666. DOI: 10.1146/annurev-physchem-040513-103738.

(19) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *Journal of Computational Physics* **1977**, *23* (2), 187-199. DOI: 10.1016/0021-9991(77)90121-8.

(20) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proceedings of the National Academy of Sciences* **2002**, *99* (20), 12562-12566. DOI: 10.1073/pnas.202427399.

(21) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers. *Peptide Science* **2001**, *60* (2), 96-123. DOI: 10.1002/1097-0282(2001)60:2<96::AID-BIP1007>3.0.CO;2-F.

(22) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314* (1), 141-151. DOI: 10.1016/S0009-2614(99)01123-9.

(23) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *Journal of the Physical Society of Japan* **1996**, *65* (6), 1604-1608. DOI: 10.1143/JPSJ.65.1604.

(24) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to

Protein Structure Prediction. *Journal of Chemical Physics* **2002**, *116*, 9058-9067. DOI: 10.1063/1.1472510.

(25) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proceedings of the National Academy of Sciences* **2005**, *102* (39), 13749-13754. DOI: 10.1073/pnas.0506346102.

(26) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **2011**, *115* (30), 9431-9438. DOI: 10.1021/jp204407d.

(27) Yang, M.; MacKerell, A. D., Jr. Conformational Sampling of Oligosaccharides Using Hamiltonian Replica Exchange with Two-Dimensional Dihedral Biasing Potentials and the Weighted Histogram Analysis Method (WHAM). *Journal of Chemical Theory and Computation* **2015**, *11* (2), 788-799. DOI: 10.1021/ct500993h.

(28) Oshima, H.; Re, S.; Sugita, Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *Journal of Chemical Theory and Computation* **2019**, *15* (10), 5199-5208. DOI: 10.1021/acs.jctc.9b00761.

(29) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced Sampling in Molecular Dynamics. *The Journal of Chemical Physics* **2019**, *151* (7), 070902. DOI: 10.1063/1.5109531. Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *International Journal of Molecular Sciences* **2020**, *21* (17), 6339. DOI: 10.3390/ijms21176339.

(30) Andricioaei, I.; Straub, J. E. Generalized Simulated Annealing Algorithms Using Tsallis Statistics: Application to Conformational Optimization of a Tetrapeptide. *Physical Review E* **1996**, *53* (4), R3055-R3058. DOI: 10.1103/PhysRevE.53.R3055. Andricioaei, I.; Straub, J. E. On Monte Carlo and Molecular Dynamics Methods Inspired by Tsallis Statistics: Methodology, Optimization, and Application to Atomic Clusters. *The Journal of Chemical Physics* **1997**, *107* (21), 9117-9124. DOI: 10.1063/1.475203. Whitfield, T. W.; Bu, L.; Straub, J. E. Generalized Parallel Sampling. *Physica A: Statistical Mechanics and its Applications* **2002**, *305* (1), 157-171. DOI: 10.1016/S0378-4371(01)00656-2.

(31) Jang, S.; Shin, S.; Pak, Y. Replica-Exchange Method Using the Generalized Effective Potential. *Physical Review Letters* **2003**, *91* (5), 058305. DOI: 10.1103/PhysRevLett.91.058305.

(32) Son, W.-J.; Jang, S.; Pak, Y.; Shin, S. Folding Simulations with Novel Conformational Search Method. *The Journal of Chemical Physics* **2007**, *126* (10), 104906. DOI: 10.1063/1.2538966.

(33) Kim, J.; Straub, J. E. Optimal Replica Exchange Method Combined with Tsallis Weight Sampling. *The Journal of Chemical Physics* **2009**, *130* (14), 144114. DOI: 10.1063/1.3108523.

(34) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *Journal of Chemical Theory and Computation* **2018**, *14* (10), 5273-5289. DOI: 10.1021/acs.jctc.8b00529.

(35) Berg, B. A.; Neuhaus, T. Multicanonical Algorithms for First Order Phase Transitions. *Physics Letters B* **1991**, *267* (2), 249-253. DOI: 10.1016/0370-2693(91)91256-U.

(36) Hansmann, U. H. E.; Okamoto, Y.; Eisenmenger, F. Molecular Dynamics, Langevin and Hydrid Monte Carlo Simulations in a Multicanonical Ensemble. *Chemical Physics Letters* **1996**, *259* (3), 321-330. DOI: 10.1016/0009-2614(96)00761-0.

(37) Nakajima, N.; Nakamura, H.; Kidera, A. Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides. *The Journal of Physical Chemistry B* **1997**, *101* (5), 817-824. DOI: 10.1021/jp962142e.

(38) Ferrenberg, A. M.; Swendsen, R. H. New Monte Carlo Technique for Studying Phase Transitions. *Physical Review Letters* **1988**, *61* (23), 2635-2638. DOI: 10.1103/PhysRevLett.61.2635.

(39) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *The Journal of Physical Chemistry A* **1997**, *101* (16), 3005-3014. DOI: 10.1021/jp961992r.

(40) Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. A Neutral, Water-Soluble, α-Helical Peptide: the Effect of Ionic Strength on the Helix-Coil Equilibrium. *Journal of the American Chemical Society* **1991**, *113* (13), 5102-5104. DOI: 10.1021/ja00013a079. Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of Helicity within the Model Peptide Acetyl(AAQAA)3amide. *Journal of the American Chemical Society* **1994**, *116* (18), 8288-8293. DOI: 10.1021/ja00097a039.

(41) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics* **1984**, *81* (8), 3684-3690. DOI: 10.1063/1.448118.

(42) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer

Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *The Journal of Chemical Physics* **1982**, *76* (1), 637-649. DOI: 10.1063/1.442716.

(43) Andersen, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **1983**, *52* (1), 24-34. DOI: 10.1016/0021-9991(83)90014-1.

(44) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *Journal of Computational Chemistry* **2000**, *21* (12), 1049-1074. DOI: 10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F.

(45) Kollman, P. A. Advances and Continuing Challenges in Achieving Realistic and Predictive Simulations of the Properties of Organic and Biological Molecules. *Accounts of Chemical Research* **1996**, *29* (10), 461-469. DOI: 10.1021/ar9500675.

(46) Berg, B. A.; Neuhaus, T. Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions. *Physical Review Letters* **1992**, *68* (1), 9-12. DOI: 10.1103/PhysRevLett.68.9. Berg, B. A.; Celik, T. New Approach to Spin-Glass Simulations. *Physical Review Letters* **1992**, *69* (15), 2292-2295. DOI: 10.1103/PhysRevLett.69.2292. Hao, M.-H.; Scheraga, H. A. Monte Carlo Simulation of a First-Order Transition for Protein Folding. *The Journal of Physical Chemistry* **1994**, *98* (18), 4940-4948. DOI: 10.1021/j100069a028. Okamoto, Y.; Hansmann, U. H. E. Thermodynamics of Helix-Coil Transitions Studied by Multicanonical Algorithms. *The Journal of Physical Chemistry* **1995**, *99* (28), 11276-11287. DOI: 10.1021/j100028a031. Kidera, A. Enhanced Conformational

Sampling in Monte Carlo Simulations of Proteins: Application to a Constrained Peptide. *Proceedings of the National Academy of Sciences* **1995**, *92* (21), 9886-9889. DOI: 10.1073/pnas.92.21.9886.

(47) Jiang, P.; Yaşar, F.; Hansmann, U. H. E. Sampling of Protein Folding Transitions: Multicanonical Versus Replica Exchange Molecular Dynamics. *Journal of Chemical Theory and Computation* **2013**, *9* (8), 3816-3825. DOI: 10.1021/ct400312d.

(48) Jarymowycz, V. A.; Stone, M. J. Fast Time Scale Dynamics of Protein Backbones: NMR Relaxation Methods, Applications, and Functional Consequences. *Chemical Reviews* **2006**, *106* (5), 1624-1671. DOI: 10.1021/cr040421p.

(49) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins: Structure, Function, and Bioinformatics* **2005**, *58* (1), 45-52. DOI: 10.1002/prot.20310.

(50) Sanbonmatsu, K. Y.; García, A. E. Structure of Met-Enkephalin in Explicit Aqueous Solution Using Replica Exchange Molecular Dynamics. *Proteins: Structure, Function, and Bioinformatics* **2002**, *46* (2), 225-234. DOI: 10.1002/prot.1167. Su, L.; Cukier, R. I. Hamiltonian and Distance Replica Exchange Method Studies of Met-Enkephalin. *The Journal of Physical Chemistry B* **2007**, *111* (42), 12310-12321. DOI: 10.1021/jp073314i.

(51) Chen, C.; Xiao, Y. Observation of Multiple Folding Pathways of β-Hairpin Trpzip2 from Independent Continuous Folding Trajectories. *Bioinformatics* **2008**, *24* (5), 659-665. DOI: 10.1093/bioinformatics/btn029.

(52) Rosta, E.; Buchete, N.-V.; Hummer, G. Thermostat Artifacts in Replica Exchange Molecular Dynamics Simulations. *Journal of Chemical Theory and*

Computation **2009**, *5* (5), 1393-1399. DOI: 10.1021/ct800557h.

(53) Yang, L.; Shao, Q.; Gao, Y. Q. Thermodynamics and Folding Pathways of Trpzip2: An Accelerated Molecular Dynamics Simulation Study. *The Journal of Physical Chemistry B* **2009**, *113* (3), 803-808. DOI: 10.1021/jp803160f.

(54) Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P. Subpicosecond Conformational Dynamics of Small Peptides Probed by Two-Dimensional Vibrational Spectroscopy. *Proceedings of the National Academy of Sciences* **2001**, *98* (20), 11254-11258. DOI: 10.1073/pnas.201169498. Woutersen, S.; Hamm, P. Isotope-Edited Two-Dimensional Vibrational Spectroscopy of Trialanine in Aqueous Solution. *The Journal of Chemical Physics* **2001**, *114* (6), 2727-2737. DOI: 10.1063/1.1336807. Woutersen, S.; Hamm, P. Structure Determination of Trialanine in Water Using Polarization Sensitive Two-Dimensional Vibrational Spectroscopy. *The Journal of Physical Chemistry B* **2000**, *104* (47), 11316-11320. DOI: 10.1021/jp001546a. Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. Dihedral Angles of Trialanine in D2O Determined by Combining FTIR and Polarized Visible Raman Spectroscopy. *Journal of the American Chemical Society* **2001**, *123* (39), 9628-9633. DOI: 10.1021/ja016202s. Schweitzer-Stenner, R. Dihedral Angles of Tripeptides in Solution Directly Determined by Polarized Raman and FTIR Spectroscopy. *Biophysical Journal* **2002**, *83* (1), 523-532. DOI: 10.1016/S0006-3495(02)75188-6. Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. Tripeptides Adopt Stable Structures in Water. A Combined Polarized Visible Raman, FTIR, and VCD Spectroscopy Study. *Journal of the American Chemical Society* **2002**, *124* (48), 14330-14341. DOI: 10.1021/ja027381w. Woutersen, S.; Pfister, R.; Hamm, P.; Mu, Y.; Kosov, D. S.; Stock, G. Peptide Conformational Heterogeneity Revealed from Nonlinear Vibrational Spectroscopy and Molecular-

Dynamics Simulations. *The Journal of Chemical Physics* **2002**, *117* (14), 6833-6840. DOI: 10.1063/1.1506151. Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. Stable Conformations of Tripeptides in Aqueous Solution Studied by UV Circular Dichroism Spectroscopy. *Journal of the American Chemical Society* **2003**, *125* (27), 8178-8185. DOI: 10.1021/ja034625j.

(55) Gorbunov, R. D.; Nguyen, P. H.; Kobus, M.; Stock, G. Quantum-Classical Description of the Amide I Vibrational Spectrum of Trialanine. *The Journal of Chemical Physics* **2007**, *126* (5), 054509. DOI: 10.1063/1.2431803. Cai, K.; Han, C.; Wang, J. Molecular Mechanics Force Field-Based Map for Peptide Amide-I Mode in Solution and Its Application to Alanine Di- and Tripeptides. *Physical Chemistry Chemical Physics* **2009**, *11* (40), 9149-9159, 10.1039/B910269H. DOI: 10.1039/B910269H.

(56) Mu, Y.; Stock, G. Conformational Dynamics of Trialanine in Water: A Molecular Dynamics Study. *The Journal of Physical Chemistry B* **2002**, *106* (20), 5294-5301. DOI: 10.1021/jp013977e. Mu; Kosov, D. S.; Stock, G. Conformational Dynamics of Trialanine in Water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS Force Fields to NMR and Infrared Experiments. *The Journal of Physical Chemistry B* **2003**, *107* (21), 5064-5073. DOI: 10.1021/jp022445a. Bastida, A.; Zúñiga, J.; Requena, A.; Miguel, B.; Candela, M. E.; Soler, M. A. Conformational Changes of Trialanine in Water Induced by Vibrational Relaxation of the Amide I Mode. *The Journal of Physical Chemistry B* **2016**, *120* (2), 348-357. DOI: 10.1021/acs.jpcb.5b09753.

(57) Toal, S.; Meral, D.; Verbaro, D.; Urbanc, B.; Schweitzer-Stenner, R. pH-Independence of Trialanine and the Effects of Termini Blocking in Short Peptides: A Combined Vibrational, NMR, UVCD, and Molecular Dynamics Study. *The Journal*

of Physical Chemistry B **2013**, *117* (14), 3689-3706. DOI: 10.1021/jp310466b.

(58) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B* **2001**, *105* (28), 6474-6487. DOI: 10.1021/jp003919d.

(59) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *The Journal of Physical Chemistry* **1987**, *91* (24), 6269-6271. DOI: 10.1021/j100308a038.

(60) Chodera, J. D.; Shirts, M. R. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *The Journal of Chemical Physics* **2011**, *135* (19), 194110. DOI: 10.1063/1.3660669.

(61) Son, W.-j.; Jang, S.; Shin, S. Simulated Q-annealing: Conformational Search with an Effective Potential. *Journal of Molecular Modeling* **2012**, *18* (1), 213-220. DOI: 10.1007/s00894-011-1072-x.

(62) Sipe, J. D.; Cohen, A. S. Review: History of the Amyloid Fibril. *Journal of Structural Biology* **2000**, *130* (2), 88-98. DOI: 10.1006/jsbi.2000.4221.

(63) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry* **2006**, *75* (1), 333-366. DOI: 10.1146/annurev.biochem.75.101304.123901.

(64) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the Cross-β Spine of Amyloid-Like Fibrils. *Nature* **2005**, *435* (7043), 773-778. DOI: 10.1038/nature03680.

(65) Stefani, M. Structural Features and Cytotoxicity of Amyloid Oligomers: Implications in Alzheimer's Disease and Other Diseases with Amyloid Deposits.

Progress in Neurobiology **2012**, *99* (3), 226-245. DOI: 10.1016/j.pneurobio.2012.03.002.

(66) Pfefferkorn, C. M.; Jiang, Z.; Lee, J. C. Biophysics of α-Synuclein Membrane Interactions. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818* (2), 162-171. DOI: 10.1016/j.bbamem.2011.07.032. Serra-Batiste, M.; Ninot-Pedrosa, M.; Bayoumi, M.; Gairí, M.; Maglia, G.; Carulla, N. Aβ42 Assembles into Specific β-Barrel Pore-Forming Oligomers in Membrane-Mimicking Environments. *Proceedings of the National Academy of Sciences* **2016**, *113* (39), 10866-10871. DOI: 10.1073/pnas.1605104113. Evangelisti, E.; Cascella, R.; Becatti, M.; Marrazza, G.; Dobson, C. M.; Chiti, F.; Stefani, M.; Cecchi, C. Binding Affinity of Amyloid Oligomers to Cellular Membranes is a Generic Indicator of Cellular Dysfunction in Protein Misfolding Diseases. *Scientific Reports* **2016**, *6* (1), 32721. DOI: 10.1038/srep32721.

(67) Brody, D. L.; Jiang, H.; Wildburger, N.; Esparza, T. J. Non-Canonical Soluble Amyloid-Beta Aggregates and Plaque Buffering: Controversies and Future Directions for Target Discovery in Alzheimer's Disease. *Alzheimer's Research & Therapy* **2017**, *9* (1), 62. DOI: 10.1186/s13195-017-0293-3.

(68) Salahuddin, P.; Fatima, M. T.; Abdelhameed, A. S.; Nusrat, S.; Khan, R. H. Structure of Amyloid Oligomers and Their Mechanisms of Toxicities: Targeting Amyloid Oligomers Using Novel Therapeutic Approaches. *European Journal of Medicinal Chemistry* **2016**, *114*, 41-58. DOI: 10.1016/j.ejmech.2016.02.065.

(69) Cline, E. N.; Bicca, M. A.; Viola, K. L.; Klein, W. L. The Amyloid-β Oligomer Hypothesis: Beginning of the Third Decade. *Journal of Alzheimer's Disease* **2018**, *64*, S567-S610. DOI: 10.3233/JAD-179941.

(70) Breydo, L.; Uversky, V. N. Structural, Morphological, and Functional

Diversity of Amyloid Oligomers. *FEBS Letters* **2015**, *589* (19PartA), 2640-2648. DOI: 10.1016/j.febslet.2015.07.013.

(71) Galzitskaya, O. V.; Selivanova, O. M. Rosetta Stone for Amyloid Fibrils: The Key Role of Ring-Like Oligomers in Amyloidogenesis. *Journal of Alzheimer's Disease* **2017**, *59*, 785-795. DOI: 10.3233/JAD-170230. Xi, W.; Hansmann, U. H. E. Ring-like N-fold Models of Aβ42 fibrils. *Scientific Reports* **2017**, *7* (1), 6588. DOI: 10.1038/s41598-017-06846-0. Nagel-Steger, L.; Owen, M. C.; Strodel, B. An Account of Amyloid Oligomers: Facts and Figures Obtained from Experiments and Simulations. *ChemBioChem* **2016**, *17* (8), 657-676. DOI: 10.1002/cbic.201500623. Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N.-V.; Coté, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; et al. Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chemical Reviews* **2015**, *115* (9), 3518-3563. DOI: 10.1021/cr500638n. Best, R. B. Computational and Theoretical Advances in Studies of Intrinsically Disordered Proteins. *Current Opinion in Structural Biology* **2017**, *42*, 147-154. DOI: 10.1016/j.sbi.2017.01.006. Huang, J.; MacKerell, A. D. Force Field Development and Simulations of Intrinsically Disordered Proteins. *Current Opinion in Structural Biology* **2018**, *48*, 40-48. DOI: 10.1016/j.sbi.2017.10.008. Levine, Z. A.; Shea, J.-E. Simulations of Disordered Proteins and Systems with Conformational Heterogeneity. *Current Opinion in Structural Biology* **2017**, *43*, 95-103. DOI: 10.1016/j.sbi.2016.11.006. Matthes, D.; Gapsys, V.; Brennecke, J. T.; de Groot, B. L. An Atomistic View of Amyloidogenic Self-assembly: Structure and Dynamics of Heterogeneous Conformational States in the Pre-nucleation Phase. *Scientific Reports* **2016**, *6* (1), 33156. DOI: 10.1038/srep33156. Chiricotto, M.; Tran, T. T.; Nguyen, P. H.; Melchionna, S.; Sterpone, F.; Derreumaux, P. Coarse-Grained and All-Atom

Simulations Towards the Early and Late Steps of Amyloid Fibril Formation. *Israel Journal of Chemistry* **2017**, *57* (7-8), 564-573. DOI: 10.1002/ijch.201600048.

(72) Stefani, M. Structural Polymorphism of Amyloid Oligomers and Fibrils Underlies Different Fibrillization Pathways: Immunogenicity and Cytotoxicity. *Current Protein & Peptide Science* **2010**, *11* (5), 343-354. DOI: 10.2174/138920310791330631.

(73) Fändrich, M.; Nyström, S.; Nilsson, K. P. R.; Böckmann, A.; LeVine III, H.; Hammarström, P. Amyloid Fibril Polymorphism: A Challenge for Molecular Imaging and Therapy. *Journal of Internal Medicine* **2018**, *283* (3), 218-237. DOI: 10.1111/joim.12732. Iadanza, M. G.; Jackson, M. P.; Hewitt, E. W.; Ranson, N. A.; Radford, S. E. A New Era for Understanding Amyloid Structures and Disease. *Nature Reviews Molecular Cell Biology* **2018**, *19* (12), 755-773. DOI: 10.1038/s41580-018-0060-8.

(74) Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y. Aβ(1–42) Fibril Structure Illuminates Self-Recognition and Replication of Amyloid in Alzheimer's Disease. *Nature Structural & Molecular Biology* **2015**, *22* (6), 499-505. DOI: 10.1038/nsmb.2991.

(75) Wälti, M. A.; Ravotti, F.; Arai, H.; Glabe, C. G.; Wall, J. S.; Böckmann, A.; Güntert, P.; Meier, B. H.; Riek, R. Atomic-Resolution Structure of a Disease-Relevant Aβ(1–42) Amyloid Fibril. *Proceedings of the National Academy of Sciences* **2016**, *113* (34), E4976-E4984. DOI: 10.1073/pnas.1600749113. Gremer, L.; Schölzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R. B. G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; et al. Fibril Structure of Amyloid-β(1–42) by Cryo–Electron Microscopy. *Science* **2017**, *358* (6359), 116-119. DOI: 10.1126/science.aao2825.

(76) Colvin, M. T.; Silvers, R.; Ni, Q. Z.; Can, T. V.; Sergeyev, I.; Rosay, M.; Donovan, K. J.; Michael, B.; Wall, J.; Linse, S.; et al. Atomic Resolution Structure of Monomorphic Aβ42 Amyloid Fibrils. *Journal of the American Chemical Society* **2016**, *138* (30), 9663-9674. DOI: 10.1021/jacs.6b05129.

(77) Lührs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Döbeli, H.; Schubert, D.; Riek, R. 3D Structure of Alzheimer's Amyloid-β(1–42) Fibrils. *Proceedings of the National Academy of Sciences* **2005**, *102* (48), 17342-17347. DOI: 10.1073/pnas.0506723102.

(78) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926-935. DOI: 10.1063/1.445869.

(79) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* **2010**, *6* (2), 459-466. DOI: 10.1021/ct900549r.

(80) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *The Journal of Chemical Physics* **2007**, *126* (1), 014101. DOI: 10.1063/1.2408420.

(81) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics* **1981**, *52* (12), 7182-7190. DOI: 10.1063/1.328693.

(82) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *Journal of Computational Chemistry* **1997**, *18* (12), 1463-1472. DOI: 10.1002/(SICI)1096-

987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(83) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1), 43-56. DOI: 10.1016/0010-4655(95)00042-E.

(84) Patriksson, A.; van der Spoel, D. A Temperature Predictor for Parallel Tempering Simulations. *Physical Chemistry Chemical Physics* **2008**, *10* (15), 2073-2077, 10.1039/B716554D. DOI: 10.1039/B716554D.

(85) Xi, W.; Wang, W.; Abbott, G.; Hansmann, U. H. E. Stability of a Recently Found Triple-β-Stranded Aβ1–42 Fibril Motif. *The Journal of Physical Chemistry B* **2016**, *120* (20), 4548-4557. DOI: 10.1021/acs.jpcb.6b01724.

(86) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *Journal of Chemical Theory and Computation* **2015**, *11* (11), 5513-5524. DOI: 10.1021/acs.jctc.5b00736.

(87) Thirumalai, D.; Reddy, G.; Straub, J. E. Role of Water in Protein Aggregation and Amyloid Polymorphism. *Accounts of Chemical Research* **2012**, *45* (1), 83-92. DOI: 10.1021/ar2000869.

(88) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. Monomer Adds to Preformed Structured Oligomers of Aβ-Peptides by a Two-Stage Dock–Lock Mechanism. *Proceedings of the National Academy of Sciences* **2007**, *104* (1), 111-116. DOI: 10.1073/pnas.0607440104.

(89) Frishman, D.; Argos, P. Knowledge-Based Protein Secondary Structure Assignment. *Proteins: Structure, Function, and Bioinformatics* **1995**, *23* (4), 566-579. DOI: 10.1002/prot.340230412.

(90) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics.

Journal of Molecular Graphics **1996**, *14* (1), 33-38. DOI: 10.1016/0263-7855(96)00018-5.

(91) Mehra, S.; Sahay, S.; Maji, S. K. α-Synuclein Misfolding and Aggregation: Implications in Parkinson's Disease Pathogenesis. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2019**, *1867* (10), 890-908. DOI: 10.1016/j.bbapap.2019.03.001.

(92) Singleton, A. B.; Farrer, M.; Johnson, J.; Singleton, A.; Hague, S.; Kachergus, J.; Hulihan, M.; Peuralinna, T.; Dutra, A.; Nussbaum, R.; et al. α-Synuclein Locus Triplication Causes Parkinson's Disease. *Science* **2003**, *302* (5646), 841-841. DOI: 10.1126/science.1090278. Chartier-Harlin, M.-C.; Kachergus, J.; Roumier, C.; Mouroux, V.; Douay, X.; Lincoln, S.; Levecque, C.; Larvor, L.; Andrieux, J.; Hulihan, M.; et al. α-Synuclein Locus Duplication as a Cause of Familial Parkinson's Disease. *The Lancet* **2004**, *364* (9440), 1167-1169. DOI: 10.1016/S0140-6736(04)17103-1. Ross, O. A.; Braithwaite, A. T.; Skipper, L. M.; Kachergus, J.; Hulihan, M. M.; Middleton, F. A.; Nishioka, K.; Fuchs, J.; Gasser, T.; Maraganore, D. M.; et al. Genomic Investigation of α-Synuclein Multiplication and Parkinsonism. *Annals of Neurology* **2008**, *63* (6), 743-750. DOI: 10.1002/ana.21380. Farrer, M.; Kachergus, J.; Forno, L.; Lincoln, S.; Wang, D.-S.; Hulihan, M.; Maraganore, D.; Gwinn-Hardy, K.; Wszolek, Z.; Dickson, D.; et al. Comparison of Kindreds with Parkinsonism and α-Synuclein Genomic Multiplications. *Annals of Neurology* **2004**, *55* (2), 174-179. DOI: 10.1002/ana.10846. Sekine, T.; Kagaya, H.; Funayama, M.; Li, Y.; Yoshino, H.; Tomiyama, H.; Hattori, N. Clinical Course of the First Asian Family with Parkinsonism Related to SNCA Triplication. *Movement Disorders* **2010**, *25* (16), 2871-2875. DOI: 10.1002/mds.23313. Ahn, T.-B.; Kim, S. Y.; Kim, J. Y.; Park, S.-S.; Lee, D. S.; Min, H. J.; Kim, Y. K.; Kim, S. E.; Kim, J.-M.;

Kim, H.-J.; et al. α-Synuclein Gene Duplication is Present in Sporadic Parkinson Disease. *Neurology* **2008**, *70* (1), 43-49. DOI: 10.1212/01.wnl.0000271080.53272.c7.

(93) Burré, J.; Sharma, M.; Tsetsenis, T.; Buchman, V.; Etherton, M. R.; Südhof, T. C. α-Synuclein Promotes SNARE-Complex Assembly in Vivo and in Vitro. *Science* **2010**, *329* (5999), 1663-1667. DOI: 10.1126/science.1195227.

(94) Winner, B.; Jappelli, R.; Maji, S. K.; Desplats, P. A.; Boyer, L.; Aigner, S.; Hetzer, C.; Loher, T.; Vilar, M.; Campioni, S.; et al. In Vivo Demonstration That α-Synuclein Oligomers are Toxic. *Proceedings of the National Academy of Sciences* **2011**, *108* (10), 4194-4199. DOI: 10.1073/pnas.1100976108. Karpinar, D. P.; Balija, M. B. G.; Kügler, S.; Opazo, F.; Rezaei-Ghaleh, N.; Wender, N.; Kim, H.-Y.; Taschenberger, G.; Falkenburger, B. H.; Heise, H.; et al. Pre-Fibrillar α-Synuclein Variants with Impaired β-Structure Increase Neurotoxicity in Parkinson's Disease Models. *The EMBO Journal* **2009**, *28* (20), 3256-3268. DOI: 10.1038/emboj.2009.257.

(95) Dear, A. J.; Meisl, G.; Šarić, A.; Michaels, T. C. T.; Kjaergaard, M.; Linse, S.; Knowles, T. P. J. Identification of On- and Off-Pathway Oligomers in Amyloid Fibril Formation. *Chemical Science* **2020**, *11* (24), 6236-6247. DOI: 10.1039/C9SC06501F.

(96) Hoyer, W.; Antony, T.; Cherny, D.; Heim, G.; Jovin, T. M.; Subramaniam, V. Dependence of α-Synuclein Aggregate Morphology on Solution Conditions. *Journal of Molecular Biology* **2002**, *322* (2), 383-393. DOI: 10.1016/S0022-2836(02)00775-1.

(97) Jucker, M.; Walker, L. C. Self-Propagation of Pathogenic Protein Aggregates in Neurodegenerative Diseases. *Nature* **2013**, *501* (7465), 45-51. DOI:

10.1038/nature12481. Brundin, P.; Melki, R.; Kopito, R. Prion-Like Transmission of Protein Aggregates in Neurodegenerative Diseases. *Nature Reviews Molecular Cell Biology* **2010**, *11* (4), 301-307. DOI: 10.1038/nrm2873. Aguzzi, A.; Rajendran, L. The Transcellular Spread of Cytosolic Amyloids, Prions, and Prionoids. *Neuron* **2009**, *64* (6), 783-790. DOI: 10.1016/j.neuron.2009.12.016. Mougenot, A.-L.; Nicot, S.; Bencsik, A.; Morignat, E.; Verchère, J.; Lakhdar, L.; Legastelois, S.; Baron, T. Prion-Like Acceleration of a Synucleinopathy in a Transgenic Mouse Model. *Neurobiology of Aging* **2012**, *33* (9), 2225-2228. DOI: 10.1016/j.neurobiolaging.2011.06.022. Luk, K. C.; Kehm, V. M.; Zhang, B.; O'Brien, P.; Trojanowski, J. Q.; Lee, V. M. Y. Intracerebral Inoculation of Pathological α-Synuclein Initiates a Rapidly Progressive Neurodegenerative α-Synucleinopathy in Mice. *Journal of Experimental Medicine* **2012**, *209* (5), 975-986. DOI: 10.1084/jem.20112457.

(98) Morales, R.; Abid, K.; Soto, C. The Prion Strain Phenomenon: Molecular Basis and Unprecedented Features. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2007**, *1772* (6), 681-691. DOI: 10.1016/j.bbadis.2006.12.006. Collinge, J.; Clarke, A. R. A General Model of Prion Strains and Their Pathogenicity. *Science* **2007**, *318* (5852), 930-936. DOI: 10.1126/science.1138718. Aguzzi, A.; Heikenwalder, M.; Polymenidou, M. Insights into Prion Strains and Neurotoxicity. *Nature Reviews Molecular Cell Biology* **2007**, *8* (7), 552-561. DOI: 10.1038/nrm2204.

(99) Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S.; et al. Cryo-EM of Full-Length α-Synuclein Reveals Fibril Polymorphs with a Common Structural Kernel. *Nature Communications* **2018**, *9* (1), 3609. DOI: 10.1038/s41467-018-05971-2.

(100) Spillantini, M. G.; Crowther, R. A.; Jakes, R.; Hasegawa, M.; Goedert, M. α-Synuclein in Filamentous Inclusions of Lewy Bodies from Parkinson's Disease and Dementia with Lewy Bodies. *Proceedings of the National Academy of Sciences* **1998**, *95* (11), 6469-6473. DOI: 10.1073/pnas.95.11.6469.

(101) Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; et al. Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α-Synuclein. *Nature Structural & Molecular Biology* **2016**, *23* (5), 409-415. DOI: 10.1038/nsmb.3194.

(102) Guerrero-Ferreira, R.; Taylor, N. M. I.; Mona, D.; Ringler, P.; Lauer, M. E.; Riek, R.; Britschgi, M.; Stahlberg, H. Cryo-EM Structure of Alpha-Synuclein Fibrils. *eLife* **2018**, *7*, e36402. DOI: 10.7554/eLife.36402. Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X. Amyloid Fibril Structure of α-Synuclein Determined by Cryo-Electron Microscopy. *Cell Research* **2018**, *28* (9), 897-903. DOI: 10.1038/s41422-018-0075-x. Ni, X.; McGlinchey, R. P.; Jiang, J.; Lee, J. C. Structural Insights into α-Synuclein Fibril Polymorphism: Effects of Parkinson's Disease-Related C-Terminal Truncations. *Journal of Molecular Biology* **2019**, *431* (19), 3913-3919. DOI: 10.1016/j.jmb.2019.07.001. Boyer, D. R.; Li, B.; Sun, C.; Fan, W.; Sawaya, M. R.; Jiang, L.; Eisenberg, D. S. Structures of Fibrils Formed by α-Synuclein Hereditary Disease Mutant H50Q Reveal New Polymorphs. *Nature Structural & Molecular Biology* **2019**, *26* (11), 1044-1052. DOI: 10.1038/s41594-019-0322-y. Zhao, K.; Li, Y.; Liu, Z.; Long, H.; Zhao, C.; Luo, F.; Sun, Y.; Tao, Y.; Su, X.-d.; Li, D.; et al. Parkinson's Disease Associated Mutation E46K of α-Synuclein Triggers the Formation of a Distinct Fibril Structure. *Nature Communications* **2020**, *11* (1), 2643. DOI: 10.1038/s41467-020-16386-3. Sun, Y.;

Hou, S.; Zhao, K.; Long, H.; Liu, Z.; Gao, J.; Zhang, Y.; Su, X.-D.; Li, D.; Liu, C. Cryo-EM Structure of Full-Length α-Synuclein Amyloid Fibril with Parkinson's Disease Familial A53T Mutation. *Cell Research* **2020**, *30* (4), 360-362. DOI: 10.1038/s41422-020-0299-4. Watson, M. D.; Lee, J. C. N-Terminal Acetylation Affects α-Synuclein Fibril Polymorphism. *Biochemistry* **2019**, *58* (35), 3630-3633. DOI: 10.1021/acs.biochem.9b00629. Guerrero-Ferreira, R.; Taylor, N. M. I.; Arteni, A.-A.; Kumari, P.; Mona, D.; Ringler, P.; Britschgi, M.; Lauer, M. E.; Makky, A.; Verasdonck, J.; et al. Two New Polymorphic Structures of Human Full-Length Alpha-Synuclein Fibrils Solved by Cryo-Electron Microscopy. *eLife* **2019**, *8*, e48907. DOI: 10.7554/eLife.48907.

(103) Boyer, D. R.; Li, B.; Sun, C.; Fan, W.; Zhou, K.; Hughes, M. P.; Sawaya, M. R.; Jiang, L.; Eisenberg, D. S. The α-Synuclein Hereditary Mutation E46K Unlocks a More Stable, Pathogenic Fibril Structure. *Proceedings of the National Academy of Sciences* **2020**, *117* (7), 3592-3602. DOI: 10.1073/pnas.1917914117.

(104) Schweighauser, M.; Shi, Y.; Tarutani, A.; Kametani, F.; Murzin, A. G.; Ghetti, B.; Matsubara, T.; Tomita, T.; Ando, T.; Hasegawa, K.; et al. Structures of α-Synuclein Filaments from Multiple System Atrophy. *Nature* **2020**, *585* (7825), 464-469. DOI: 10.1038/s41586-020-2317-6.

(105) Yang, Y.; Shi, Y.; Schweighauser, M.; Zhang, X.; Kotecha, A.; Murzin, A. G.; Garringer, H. J.; Cullinane, P. W.; Saito, Y.; Foroud, T.; et al. Structures of α-Synuclein Filaments from Human Brains with Lewy Pathology. *Nature* **2022**, *610* (7933), 791-795. DOI: 10.1038/s41586-022-05319-3.

(106) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, 19-25. DOI:

10.1016/j.softx.2015.06.001.

(107) Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G. A Series of PDB-Related Databanks for Everyday Needs. *Nucleic Acids Research* **2014**, *43* (D1), D364-D368. DOI: 10.1093/nar/gku1028. Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577-2637. DOI: 10.1002/bip.360221211.

(108) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; et al. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Science* **2018**, *27* (1), 112-128. DOI: 10.1002/pro.3280.

(109) Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. 2015.

(110) Lemkul, J. A.; Bevan, D. R. Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics. *The Journal of Physical Chemistry B* **2010**, *114* (4), 1652-1660. DOI: 10.1021/jp9110794.

(111) Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of Chemical Theory and Computation* **2010**, *6* (12), 3713-3720. DOI: 10.1021/ct100494z.

(112) Lemkau, L. R.; Comellas, G.; Kloepper, K. D.; Woods, W. S.; George, J. M.; Rienstra, C. M. Mutant Protein A30P α-Synuclein Adopts Wild-Type Fibril Structure, Despite Slower Fibrillation Kinetics. *Journal of Biological Chemistry* **2012**, *287* (14), 11526-11532. DOI: 10.1074/jbc.M111.306902. Lemkau, L. R.; Comellas, G.; Lee, S. W.; Rikardsen, L. K.; Woods, W. S.; George, J. M.; Rienstra,

C. M. Site-Specific Perturbations of Alpha-Synuclein Fibril Structure by the Parkinson's Disease Associated Mutations A53T and E46K. *PLOS ONE* **2013**, *8* (3), e49750. DOI: 10.1371/journal.pone.0049750. Comellas, G.; Lemkau, L. R.; Zhou, D. H.; George, J. M.; Rienstra, C. M. Structural Intermediates during α-Synuclein Fibrillogenesis on Phospholipid Vesicles. *Journal of the American Chemical Society* **2012**, *134* (11), 5090-5099. DOI: 10.1021/ja209019s.

(113) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. Critical Assessment of Methods of Protein Structure Prediction (CASP)—Round XIV. *Proteins: Structure, Function, and Bioinformatics* **2021**, *89* (12), 1607-1617. DOI: 10.1002/prot.26237.

(114) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; et al. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706-710. DOI: 10.1038/s41586-019-1923-7.

(115) Wilson, C. J.; Choy, W.-Y.; Karttunen, M. AlphaFold2: A Role for Disordered Protein/Region Prediction? *International Journal of Molecular Sciences* **2022**, *23* (9), 4591. DOI: 10.3390/ijms23094591. Chakravarty, D.; Porter, L. L. AlphaFold2 Fails to Predict Protein Fold Switching. *Protein Science* **2022**, *31* (6), e4353. DOI: 10.1002/pro.4353. Piovesan, D.; Monzon, A. M.; Tosatto, S. C. E. Intrinsic Protein Disorder and Conditional Folding in AlphaFoldDB. *Protein Science* **2022**, *31* (11), e4466. DOI: 10.1002/pro.4466.

(116) Pinheiro, F.; Santos, J.; Ventura, S. AlphaFold and the Amyloid Landscape. *Journal of Molecular Biology* **2021**, *433* (20), 167059. DOI: 10.1016/j.jmb.2021.167059.

(117) Siddiqi, M. K.; Malik, S.; Majid, N.; Alam, P.; Khan, R. H. Chapter Ten -

Cytotoxic Species in Amyloid-Associated Diseases: Oligomers or Mature Fibrils. In *Advances in Protein Chemistry and Structural Biology*, Donev, R. Ed.; Vol. 118; Academic Press, 2019; pp 333-369. Hubin, E.; van Nuland, N. A. J.; Broersen, K.; Pauwels, K. Transient Dynamics of Aβ Contribute to Toxicity in Alzheimer's Disease. *Cellular and Molecular Life Sciences* **2014**, *71* (18), 3507-3521. DOI: 10.1007/s00018-014-1634-z. Treusch, S.; Cyr, D. M.; Lindquist, S. Amyloid Deposits: Protection Against Toxic Protein Species? *Cell Cycle* **2009**, *8* (11), 1668-1674. DOI: 10.4161/cc.8.11.8503. Bucciantini, M.; Giannoni, E.; Chiti, F.; Baroni, F.; Formigli, L.; Zurdo, J.; Taddei, N.; Ramponi, G.; Dobson, C. M.; Stefani, M. Inherent Toxicity of Aggregates Implies a Common Mechanism for Protein Misfolding Diseases. *Nature* **2002**, *416* (6880), 507-511. DOI: 10.1038/416507a.

(118) Wells, C.; Brennan, S.; Keon, M.; Ooi, L. The Role of Amyloid Oligomers in Neurodegenerative Pathologies. *International Journal of Biological Macromolecules* **2021**, *181*, 582-604. DOI: 10.1016/j.ijbiomac.2021.03.113. Nguyen, P. H.; Ramamoorthy, A.; Sahoo, B. R.; Zheng, J.; Faller, P.; Straub, J. E.; Dominguez, L.; Shea, J.-E.; Dokholyan, N. V.; De Simone, A.; et al. Amyloid Oligomers: A Joint Experimental/Computational Perspective on Alzheimer's Disease, Parkinson's Disease, Type II Diabetes, and Amyotrophic Lateral Sclerosis. *Chemical Reviews* **2021**, *121* (4), 2545-2647. DOI: 10.1021/acs.chemrev.0c01122. Diociaiuti, M.; Bonanni, R.; Cariati, I.; Frank, C.; D'Arcangelo, G. Amyloid Prefibrillar Oligomers: The Surprising Commonalities in Their Structure and Activity. *International Journal of Molecular Sciences* **2021**, *22* (12), 6435. DOI: 10.3390/ijms22126435. Madhu, P.; Mukhopadhyay, S. Distinct Types of Amyloid-β Oligomers Displaying Diverse Neurotoxicity Mechanisms in Alzheimer's Disease. *Journal of Cellular Biochemistry* **2021**, *122* (11), 1594-1608. DOI:

10.1002/jcb.30141. Kayed, R.; Lasagna-Reeves, C. A. Molecular Mechanisms of Amyloid Oligomers Toxicity. *Journal of Alzheimer's Disease* **2013**, *33*, S67-S78. DOI: 10.3233/JAD-2012-129001. Bemporad, F.; Chiti, F. Protein Misfolded Oligomers: Experimental Approaches, Mechanism of Formation, and Structure-Toxicity Relationships. *Chemistry & Biology* **2012**, *19* (3), 315-327. DOI: 10.1016/j.chembiol.2012.02.003. Sakono, M.; Zako, T. Amyloid Oligomers: Formation and Toxicity of Aβ Oligomers. *The FEBS Journal* **2010**, *277* (6), 1348-1358. DOI: 10.1111/j.1742-4658.2010.07568.x.

(119) Ciudad, S.; Puig, E.; Botzanowski, T.; Meigooni, M.; Arango, A. S.; Do, J.; Mayzel, M.; Bayoumi, M.; Chaignepain, S.; Maglia, G.; et al. Aβ(1-42) Tetramer and Octamer Structures Reveal Edge Conductivity Pores as a Mechanism for Membrane Damage. *Nature Communications* **2020**, *11* (1), 3014. DOI: 10.1038/s41467-020-16566-1.

국문초록

질병은 생물학적 시스템 내에서 생체 분자 간에 정확하게 규제되고 있던 상호작용 네트워크에 교란이 일어나서 발생한다. 질병의 기원에 대한 연구한다는 것은 관련된 네트워크에 대한 이해, 그 네트워크에 발생한 장애의 원인 식별, 그리고 그것의 정상 상태 회복을 뜻한다. 최근 과학 및 기술의 발달에 따라 단백질 수준에서 질병에 대해 연구하기 시작했다. 특히, 인간의 수명이 늘어나면서 발생하고 있는 연령 관련 질병에 대한 관심이 증가하고 있다. 신경퇴행성 질환은 연령 관련 뇌 기능 장애와 신경세포의 손실과 관련된 다양한 신경학적 장애를 포괄한다. 이러한 질병은 단백질 변형의 결과인 아밀로이드 응집체로 특징 지어지며, 이는 흥미로운 연구 분야인 단백질 변형의 결과로 이해되고 있다. 다양한 형태와 중간체를 가지는 아밀로이드 응집체의 복잡성은 에너지 랜드스케이프에 다중 지역 최소값을 가지는 본질적으로 무질서한 단백질의 특성에 기인한다.

이 학위 논문은 신경퇴행성 질환의 규명을 위해 수행된 세 가지의 계산 연구 결과를 담고 있다. 첫번째는 새로 개발한 강화 샘플링 방법으로 준안정상태들이 중요한 시스템의 샘플링에 사용할 수 있다. 두번째, 알츠하이머병과 관련된 아밀로이드 베타 프로토 파이브릴의 구조적 안정성을 밝혔고 또한 응집 과정을 유추할 수 있는 실마리에 대해 논하였다. 마지막으로 파킨슨병과 관련된 알파 시뉴클린 프로토 파이브릴 구조들의 구조적 안정성을 비교하여 어떤 구조가 병원성

원인으로서 작용할 것인지를 알아보았다. 이 결과들 및 여기에서 사용한 방법론들이 신경퇴행성 질환과 관련된 시스템의 구조적 및 동적 특성을 이해하는 것에 대한 정보와 통찰을 제공할 것으로 기대된다.

주요어: 신경퇴행성 질환, 아밀로이드 파이브릴, 분자동역학 시뮬레이션, 강화 샘플링 방법, 아밀로이드 베타, 알파 시뉴클린

학 번: 2013-22931

감사의 글

여러 가지로 부족한 제가 많은 분의 도움 속에 비로소 열매를 맺을 수 있게 되었습니다. 이 자리를 빌려 짧게나마 그분들께 감사함을 남겨보려 합니다.

학부생 때부터 지금까지 긴 시간 동안 지도편달을 아끼지 않으신 신석민 지도교수님께 깊은 감사를 전합니다. 항상 더 넓은 시야로 연구의 방향을 잡아 주신 것이 저의 학위 과정에 있어서 큰 도움이 되었습니다. 제가 보지 못하는 것을 짚어 주시는 교수님의 지도 덕분에 더 많은 논의를 포함하는 연구를 할 수 있었습니다.

또한, 바쁘신 가운데에도 제 학위논문의 심사위원을 맡아주시고 좋은 조언들 해 주신 석차옥, 정연준, 이정호, 그리고 장순민 교수님께도 큰 감사를 드립니다.

더불어 저의 연구의 진행에 있어 가장 많이 힘써 주신 윤제성 선배님께 감사를 표합니다. 연구 진행의 모든 단계에서 해 주셨던 세심한 조언들과 격려가 없었다면 저는 이 과정을 마칠 수 없었을 것입니다.

연구실 처음 들어왔을 때 계셨던 모든 분께도 감사드립니다. 장보영 박사님, 박성병 형님, 박희수 형님, 김태래 형님, 그리고 우민우 형님의 배려가 없었다면 저의 연구실 적응이 쉽지 않았을 것으로 생각합니다. 또한, 저보다 늦게 들어온 영범이, 채수, 영재, 명근이, 윤수, 민우, 동건이에게 별로 도와준 것이 없어 미안하다는 말을 전합니다.

마지막으로, 저를 믿고 응원해준 가족들에게 너무나 고맙다고 말하고 싶습니다. 인생의 모범을 꾸준하게 보여주시는 아버지, 가족 구성원들의 뒷바라지에 헌신하시는 어머니, 그리고 지금은 멀리 떨어져 있지만 언제나 우리 가족에게 활력을 불어넣어 주는 내 동생 강빈이, 모두 정말 고맙습니다. 그리고 사랑합니다.