

Thermodynamic analysis of structural transitions during GNNQQNY aggregation

Kenneth L. Osborne,¹ Michael Bachmann,² and Birgit Strodel^{1,3*}

¹ Institute of Complex Systems: Structural Biochemistry, Research Centre Jülich, 52425 Jülich, Germany

² Center for Simulation Physics, The University of Georgia, Athens, Georgia 30602, USA

³ Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

ABSTRACT

Amyloid protein aggregation characterizes many neurodegenerative disorders, including Alzheimer's, Parkinson's, and Creutzfeldt-Jakob disease. Evidence suggests that amyloid aggregates may share similar aggregation pathways, implying simulation of full-length amyloid proteins is not necessary for understanding amyloid formation. In this study, we simulate GNNQQNY, the N-terminal prion-determining domain of the yeast protein Sup35 to investigate the thermodynamics of structural transitions during aggregation. Utilizing a coarse-grained model permits equilibration on relevant time scales. Replica-exchange molecular dynamics is used to gather simulation statistics at multiple temperatures and clear energy traps that would adversely impact results. Investigating the association of 3-, 6-, and 12-chain GNNQQNY systems by calculating thermodynamic quantities and orientational order parameters, we determine the aggregation pathway by studying aggregation states of GNNQQNY. We find that the aggregation of the hydrophilic GNNQQNY sequence is mainly driven by H-bond formation, leading to the formation of β -sheets from the very beginning of the assembly process. Condensation (aggregation) and ordering take place simultaneously, which is underpinned by the occurrence of a single heat capacity peak.

Proteins 2013; 81:1141–1155.

© 2013 Wiley Periodicals, Inc.

Key words: amyloid aggregation; coarse-grained simulations; replica-exchange molecular dynamics; statistical analysis; orientational order parameters; aggregation states.

INTRODUCTION

Diseases such as Alzheimer's disease, Parkinson's disease, type II diabetes, and others share a general characteristic: each of them is associated with the misfolding and subsequent aggregation of soluble peptides and proteins into soluble oligomeric assemblies, and later, into insoluble amyloid fibrils.^{1–3} The aggregated assemblies contain a similar cross- β structure with a steric zipper motif and may even share structural properties with soluble oligomers; evidence that amyloid aggregates may follow similar aggregation pathways despite lack of sequence similarity.^{4,5} During the past decade, compelling evidence has emerged that soluble, low molecular weight oligomers are more toxic than fully formed fibrils.^{6–9}

The study of protein fragments that retain the essential amyloid characteristics of the full-length sequences is attractive, because the short peptide length allows for a systematic computational investigation of aggregation kinetics and thermodynamics. One such example fragment is GNNQQNY, a polar heptapeptide from the N-terminal prion-determining region of the yeast prion

protein Sup35. It exhibits amyloidogenic properties similar to the prion-determining domain of Sup35, including an unbranched morphology, the binding of “flat” dyes such as Congo red and thioflavin T, cooperative aggregation kinetics with a concentration-dependent lag phase, the ability to seed aggregation, and unusual structural stability.¹⁰ All of these are diagnostic features of amyloid aggregates.

X-ray diffraction experiments of GNNQQNY microcrystals were able to provide a detailed atomic structure of fibrillar GNNQQNY aggregates, revealing a steric zipper motif based on in-register, parallel β -sheets.^{10–12} This finding was supported by magic angle spinning NMR,¹³ which has proved uniquely valuable in the

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NSF (to M.B.); Grant number: DMR1207437

*Correspondence to: Birgit Strodel, Institute of Complex Systems: Structural Biochemistry, Research Centre Jülich, 52425 Jülich, Germany.

E-mail: b.strodel@fz-juelich.de

Received 22 October 2012; Revised 16 January 2013; Accepted 17 January 2013

Published online 14 February 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24263

structural characterization of amyloid fibrils due to its ability to provide site-specific structural constraints on biological solids, without requiring crystalline samples.^{14,15} While it was recently established that toxic Alzheimer's amyloid- β (1-42) fibrillar oligomers share the cross- β structure with mature amyloid fibrils,¹⁶ the structural characterization of intermediate amyloid oligomers has been generally precluded experimentally due to the high complexity of the aggregation process, and the short-lived and meta-stable character of the early aggregates. Molecular simulations provide a valuable tool to address the first steps of amyloid aggregation and elucidate its structural and kinetic information.

The knowledge of the atomic structure of GNNQQNY fibrils, together with its small size, has led GNNQQNY to become a model amyloid test system for experimental and theoretical studies. One of the first computational studies of GNNQQNY examined the behavior of the trimer using molecular dynamics (MD) simulations at a constant temperature of 330 K and revealed that the in-register parallel β -sheet is stabilized by side-chain hydrogen bonding and π -stacking of the aromatic tyrosine residues.¹⁷ This study was followed by various other simulation studies, which characterized the structures and free energies of small GNNQQNY aggregates ranging from dimers to 20-mers starting from disordered states, or studied the stability of preformed GNNQQNY assemblies with cross- β or annular morphologies.^{18–31} In most of these computational works, MD simulations were performed either as constant-temperature MD runs or using the replica-exchange MD (REMD) technique,³² in connection with atomic or coarse-grained (CG) force fields. As a result of the breadth of time scales (from ns for the formation of early oligomers to days or months for the formation of mature fibrils) and length scales (from a nm-sized protein to several hundred nm-long aggregates) involved in aggregation, a hierarchy of models is used in the computational study of this process. Different levels of resolution allow probing of different elements of the aggregation process. Atomic models can provide invaluable information at a detailed level not accessible to experiment, but only allow probing of very early stages of aggregation. CG models, on the other hand, provide the possibility of extracting general characteristics of the thermodynamics and kinetics of aggregation.

CG models come in a number of resolutions, from models that represent the peptide as a single preformed unit, to single bead lattice and off-lattice models, to multibead off-lattice models.³³ Models utilized for studying amyloid aggregation include the model by Caflisch and coworkers, where each peptide consists of four spherical backbone beads and six spherical side-chain beads of hydrophobic and hydrophilic nature,^{34–36} the mid-resolution Shea model with two backbone and one side-chain bead per residue,^{37–40} discontinuous CG models used in connection with discrete MD,^{41–46} and the OPEP model

by Derreumaux and coworkers that uses a detailed representation of all backbone atoms and reduces each side chain to a single bead.^{47–50} The OPEP model was recently applied in a multiscale approach to characterize the formation of GNNQQNY 3-mers, 12-mers, and 20-mers. More details about CG models for protein aggregation can be found in a recent review by Wu and Shea.³³

In this article, we perform REMD simulations using the CG peptide model developed by Bereau and Deserno⁵¹ to investigate the thermodynamics of structural transitions during the aggregation of the 3-, 6-, and 12-chain GNNQQNY systems. Their intermediate resolution level CG force field (four beads per amino acid, implicit solvent) was shown to be able to fold proteins exhibiting both helical and β conformations with tertiary structures and amino acid sequences different from the one used for parameter tuning.^{51–54} Furthermore, this CG peptide model was also successful in predicting aggregation for the GNNQQNY peptide from a short REMD simulation.⁵¹ The aim of this study is to more thoroughly analyze the structures and thermodynamics during aggregation of the GNNQQNY peptide. The resulting aggregation dynamics in temperature space is characterized in terms of orientational order, secondary structure, and oligomer size distribution.

MODEL AND METHODS

CG peptide model

The primary motivation for using a CG model was to reach the greater than microsecond time scales associated with peptide aggregation. The C_β -type CG force field we use here was built to sample a balanced proportion of α -helical and β -extended configurations, with the aim of avoiding a bias toward any particular secondary structure.⁵¹ The backbone is represented by three beads per residue and one bead per side chain with the latter located in the position of the C_β atom. The force field was parametrized to reproduce local, secondary, and tertiary conformations. The nearly atomistic resolution of the backbone beads allows the force field to model physically relevant secondary structures, such as β -sheets and α -helices, without imposing a given secondary structure prior to simulation.⁵¹

Bonded interactions

The CG force field uses harmonic potentials to describe vibrations and bending:

$$E_{\text{bond}}(r) = \frac{1}{2} k_{\text{bond}} (r - r_0)^2, \quad (1)$$

$$E_{\text{angle}}(\theta) = \frac{1}{2} k_{\text{angle}} (\theta - \theta_0)^2 \quad (2)$$

with k_{bond} and k_{angle} as the spring constants for bonds and angles, respectively. The distance between two atoms

is r , and θ is the angle formed when one atom is connected to two others. The equilibrium positions of the bonds and angles, respectively, are r_0 and θ_0 . The dihedral potential is defined by

$$E_{\text{dih}}(\phi) = k_n[1 - \cos(n\phi - \phi_{n,0})], \quad (3)$$

with period $2\pi/n$, phase $\phi_{n,0}$, and k_n the strength of the torsional potential. The L chiral nature form of amino acids is accounted for by the inclusion of an improper dihedral angle between the beads $\text{NC}_\alpha\text{C}'\text{C}_\beta$, where N and C_α include the hydrogen atoms, and C' collectively represents the carbonyl group.

Nonbonded interactions

The CG model includes four nonbonded potentials: backbone, side chain, hydrogen bond, and dipole interactions. This model does not contain explicit electrostatics. The backbone interactions are modeled by a purely repulsive Weeks–Chandler–Anderson potential of the form

$$E_{\text{bb}}(r) = \begin{cases} 4\epsilon_{\text{bb}} \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - \left(\frac{\sigma_{ij}}{r} \right)^6 + \frac{1}{4} \right], & r \leq r_c, \\ 0 & r > r_c, \end{cases} \quad (4)$$

where σ_{ij} is the arithmetic mean between the two bead sizes involved, $r_c = 2^{1/6}\sigma_{ij}$ and ϵ_{bb} is a free parameter used for tuning the potential.

The parametrization of side-chain interactions is based on the statistical analysis of residue–residue contacts in crystallized proteins by Miyazawa and Jernigan (MJ).⁵⁵ They developed an inter-residue potential, which includes hydrophobic effects, charge attraction and repulsion, and the likelihood of forming side-chain hydrogen bonds. Bereau and Deserno used the 20×20 MJ interaction matrix to extract a relative attraction strength between residues to define their side-chain potential:

$$E_{\text{sc}}(r) = \begin{cases} 4\epsilon_{\text{sc}} \left[\left(\frac{\sigma_{\text{C}_\beta}}{r} \right)^{12} - \left(\frac{\sigma_{\text{C}_\beta}}{r} \right)^6 \right] + (\epsilon_{\text{sc}} - \epsilon'_{\text{sc}}\epsilon'_{ij}), & r \leq r_c, \\ 4\epsilon_{\text{sc}}\epsilon'_{ij} \left[\left(\frac{\sigma_{\text{C}_\beta}}{r} \right)^{12} - \left(\frac{\sigma_{\text{C}_\beta}}{r} \right)^6 \right], & r \leq r_c \leq r_{\text{sc,cut}}, \\ 0, & r > r_{\text{sc,cut}}. \end{cases} \quad (5)$$

Here, σ_{C_β} is the (same) size of all side-chain beads, the interaction energies ϵ'_{ij} are calculated from the MJ table, while ϵ_{sc} translates the MJ interactions to an absolute scale. The potentials in Eq. (5) at positions r_c and $r_{\text{sc,cut}}$ are joined such that the potentials and their first derivative are continuous.

Backbone hydrogen bonds (H-bonds) are modeled explicitly due to the lack of explicit electrostatic interactions. Thus, an H-bond interaction potential is introduced by a radial 12–10 Lennard-Jones potential combined with an angular term,

$$E_{\text{hb}}(r) = \epsilon_{\text{hb}} \left[5 \left(\frac{\sigma_{\text{hb}}}{r} \right)^{12} - 6 \left(\frac{\sigma_{\text{hb}}}{r} \right)^{10} \right] \times \begin{cases} \cos^2\theta_{\text{N}}\cos^2\theta_{\text{C}}, & |\theta_{\text{N}}|, |\theta_{\text{C}}| < 90^\circ \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Here, θ_{N} and θ_{C} are defined to be $\angle\text{NHC}'$ and $\angle\text{OC}'\text{N}$, respectively. The CG model does not explicitly contain hydrogen and the carbonyl O atoms, and thus the positions of these atoms are implied from the position of N and C. H-bonds will not form unless there are three amino acids between the interaction candidates. In Eq. (6), r represents the distance between the atoms involved in the H-bond, and σ_{hb} defines the equilibrium distance. While the H-bond potential is sufficient to stabilize α -helices, a dipole interaction was introduced helping to stabilize β -sheets:

$$E_{\text{dip}}(\Phi, \Psi) = k_{\text{dip}}[(1 - \cos \Phi) + (1 - \cos \Psi)], \quad (7)$$

where k_{dip} is the strength of dipole interactions, and Φ and Ψ are the backbone dihedral angles about the N– C_α and C_α – C' bonds, respectively.

Units

In this CG model, temperature is defined in a reduced temperature unit (RTU) corresponding to room temperature: $T = 1 \text{ RTU} \equiv 300 \text{ K}$. Correspondingly, energies are measured in units of $300 \text{ K} \cdot k_{\text{B}} \equiv 1 \text{ REU}$ (reduced energy unit), where k_{B} is the Boltzmann constant.

Although CG force fields retain the correct measure for temperature, lengths, and relative energies, the time scale does not usually translate directly to real world time. In general, the interpretation of the time scale in CG simulations is not straightforward. In comparison to atomistic models, the dynamics observed with CG models is faster. The main reason is that the underlying energy landscape is smoother than that of atomistic models, as the system has less degrees of freedom. The removal of the motion of individual atoms reduces the friction, which further accelerates the dynamics, leading to an effective speed up of simulation dynamics. As such, CG simulations are typically more efficient at sampling the conformational space of a molecular system than their atomistic counterparts, which allows for the study of processes currently inaccessible via atomistic approaches. Based on comparison of diffusion constants in the CG and atomistic models, the speed-up rate can be approximated. However, there is no universal speed-up factor for a given CG force field, as it varies depending on the system. For example, the MARTINI force field reports speed-up factors between 3 and 10.⁵⁶

REMD

In our simulations, we used the REMD method³² as implemented in the ESPRESSO simulation package.⁵⁷ In

an REMD simulation, multiple MD runs of the same system (replicas) are run simultaneously, but at different temperatures. After a specified number of time steps, replicas at neighboring temperatures can be exchanged, provided that a Metropolis criterion is satisfied.⁵⁸ This procedure allows high-energy structures to be accepted for the replicas at higher temperature. The associated configurational changes then migrate to the replicas at lower temperatures when exchanged with each other. The overall effect is that the system is able to overcome high energy barriers relatively quickly.

Each REMD simulation was started from 3, 6, or 12 chains of GNNQQNY peptides placed at random starting positions and in random conformations in a simulation box. Each REMD simulation was performed for 500 million time steps per replica, where the initial 100 million steps were used for system equilibration. The dynamics were propagated using the Langevin method, with frictional coefficient $\Gamma = \tau^{-1}$, where $\tau \sim 0.1$ ps is the intrinsic unit of time of the CG model in use and time step $\delta t = 0.01\tau$.⁵⁹ This results in about 500 ns simulation time per replica.

Studies of the 3-chain system used 10 temperature replicas (temperatures in RTU: 0.63, 0.68, 0.73, 0.78, 0.82, 0.86, 0.89, 0.95, 1.06, and 1.22), whereas the studies of the 6- and 12-chain systems used 16 replicas (temperatures in RTU: 0.63, 0.68, 0.73, 0.78, 0.82, 0.86, 0.89, 0.91, 0.93, 0.95, 0.95, 1.01, 1.06, 1.12, 1.17, and 1.22). The temperatures were chosen to surround the association/dissociation temperature, which was determined in a short preceding REMD simulation. Replica exchanges were attempted every 500 fs, and the average acceptance probability was $48.3 \pm 22.4\%$ throughout the considered temperature range. We performed the GNNQQNY simulations at a concentration of 80 mM. This concentration was chosen to compare our results with those obtained by Bellesia and Shea.³⁹

Analysis

Weighted histogram analysis method

Within each temperature thread of our REMD simulations resides a canonical (NVT) simulation. The weighted histogram analysis method (WHAM) combines the data from all replicas to extract the multicanonical ensemble averages.⁶⁰ It was used to calculate both several thermodynamic quantities, such as the heat capacity (C_v) and free energy (ΔG), and average behavior of impact parameters ($\langle P \rangle$) defined below as a function of T .

The statistical expectation value of an order parameter $\langle P \rangle$ is given by

$$\langle P \rangle = \frac{1}{Z} \sum_k P(k) e^{-E(k)/k_B T}, \quad Z = \sum_k e^{-E(k)/k_B T} \quad (8)$$

where Z is the canonical partition function, k is the system snapshot within a given temperature replica, and

$E(k)$ and $P(k)$ are the measured system energies and order parameters, respectively. For identification of transition points, we calculate the fluctuations of $\langle P \rangle$:

$$\frac{d}{dT} \langle P \rangle = \frac{1}{k_B T^2} (\langle PE \rangle - \langle P \rangle \langle E \rangle). \quad (9)$$

Order parameters

For the characterization of the aggregation behavior, we computed the liquid crystal order parameters P_1 and P_2 to determine the (anti)parallel structural order of the system.¹⁸ This is useful, because amyloidogenic sequences are expected to form β -sheets and, thus, be aligned in characteristic forms. The polar and nematic order parameters P_1 and P_2 are defined as follows:

$$P_1 = \frac{1}{N} \sum_{i=1}^N \hat{z}_i \cdot \hat{d}, \quad (10)$$

$$P_2 = \frac{1}{2N} \sum_{i=1}^N \left[3(\hat{z}_i \cdot \hat{d})^2 - 1 \right], \quad (11)$$

where i represents a given peptide, N is the number of peptide chains in the system, and \hat{z}_i is the unit vector pointing from the N bead of the first residue to the C' bead of the last residue of each peptide. The preferred direction of alignment is defined by the director \hat{d} , which is the eigenvector corresponding to the largest eigenvalue of the 3×3 ordering matrix \hat{Q} , with⁶¹

$$Q_{ab} = \frac{1}{2N} \sum_{i=1}^N \left[3z_i^{(a)} \cdot z_i^{(b)} - \delta_{ab} \right], \quad (12)$$

where $z_i^{(m)}$ is the m th component of vector \hat{z}_i . The nematic order parameter P_2 corresponds to the largest positive eigenvalue (λ_+) of \hat{Q} .⁶¹

The order parameters P_1 and P_2 describe different orientational properties of the system and provide complementary information. Both P_1 and P_2 have values ranging between 0 and 1.⁶² P_1 reveals information about the polarity of the system, that is, how much the molecular vectors \hat{z}_i point in the same direction. P_2 detects the orientational order of the system and discriminates between ordered ($P_2 = 1$) and disordered ($P_2 = 0$) states. A P_2 value of 1 means that the system is aligned in the direction of \hat{d} . $P_2 = P_1 = 1$ when the system is aligned parallel to \hat{d} , while $P_2 = 1$, $P_1 = 0$ when there is a 50:50 mix of \hat{z}_i vectors aligned parallel and antiparallel to \hat{d} .

The free energy ΔG associated with either order parameter P_i ($i = 1, 2$) is calculated using the following equation

Table I

Definition of the Unique Aggregation States of a System Containing Three GNNQQNY Peptides

State name	Oligomeric species	Population
0	Monomer	3
1	Monomer	1
	Dimer	1
2	Trimer	1

$$\Delta G(\tilde{P}_i) \sim -k_B T \ln \sum_k \delta(P_i(k) - \tilde{P}_i) e^{-E(k)/k_B T} \quad (13)$$

where k runs over all conformations sampled at T . The \tilde{P}_i histograms were accumulated in 30 equally spaced bins over the range of P_1 and P_2 , respectively.

Aggregation state

For the classification of the GNNQQNY oligomers in terms of their aggregation state, we analyzed the aggregate size for each structure sampled in our REMD simulations. Here, a single chain is said to belong to a given oligomer cluster, if it shares one or more hydrogen bonds with another chain in that oligomer. We determined all possible combinations of oligomer sizes for the 3- and 6-chain system and enumerated them in Tables I and II, respectively. Lower numbered aggregation states correspond to a less aggregated system, with 0 being the completely monomeric state. Conversely, the higher numbered aggregation states correspond to aggregated systems that include larger fragments.

The aggregation behavior for the 12-chain system was characterized in a different way by tracking only the largest β -sheet in the system. This different approach was chosen for two reasons. First, the study of the 3-chain and 6-chain system already allows us to follow the aggregation leading to smaller oligomers. Second, our objective for the 12-chain system was to identify if, and how, larger oligomers form into a dodecamer. The reason we did not include all possible aggregation states is that at this size scale, it is questionable how distinct some of the higher aggregation states are.

RESULTS

We simulated 3-, 6-, and 12-chain GNNQQNY systems. We will first present the results of the 3- and 6-chain systems, as we have performed 3 and 12 runs for these systems, respectively, and can thus provide an in-depth statistical analysis of the results. Conversely, because the computational cost of simulating the 12-chain system was high, we only performed one run for this system and separately present the results therefrom.

Table II

Definition of the Unique Aggregation States of a System Containing Six GNNQQNY Peptides

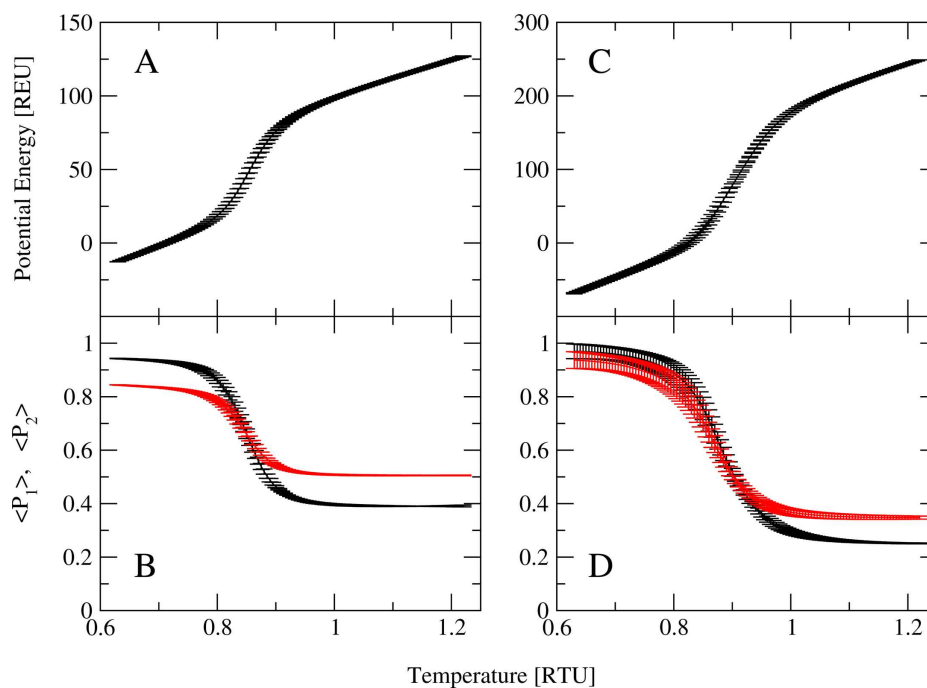
State name	Oligomeric species	Population
0	Monomer	6
1	Monomer	4
	Dimer	1
2	Monomer	2
	Dimer	2
3	Monomer	3
	Trimer	1
4	Monomer	1
	Dimer	1
	Trimer	1
5	Dimer	3
6	Monomer	2
	Tetramer	1
7	Dimer	1
	Tetramer	1
8	Trimer	2
9	Monomer	1
	Pentamer	1
10	Hexamer	1

Weighted histogram analysis

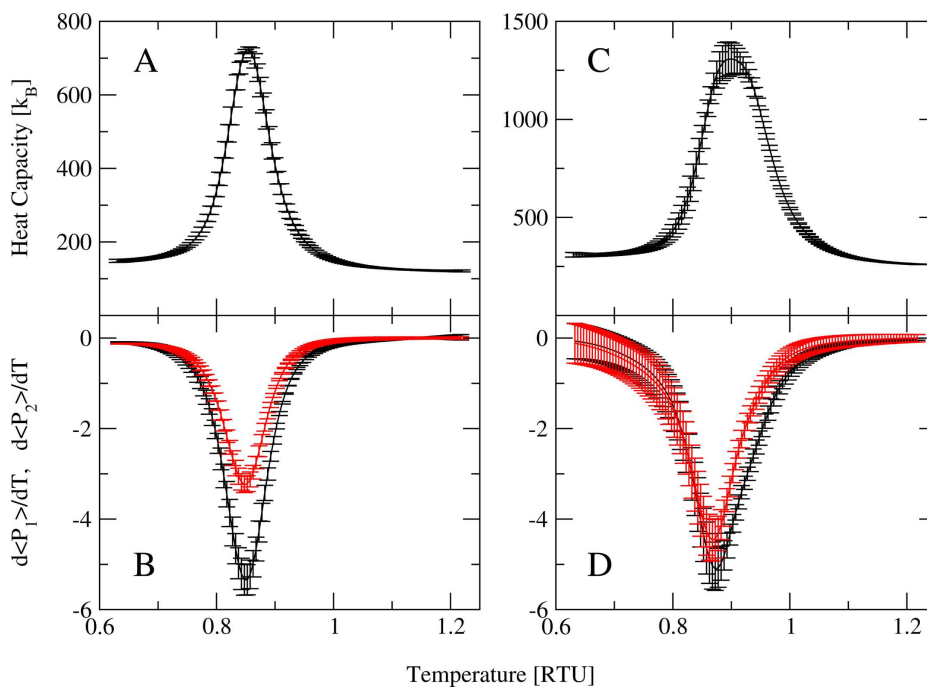
WHAM⁶⁰ was used to calculate various quantities important for understanding the behavior of the system as a function of temperature T . In our study, we examined the temperature dependence of the mean system energy, $\langle E \rangle$, the order parameters, $\langle P_1 \rangle$ and $\langle P_2 \rangle$, and the fluctuations of these quantities.

Figure 1 shows the temperature dependence of $\langle E \rangle$ and $\langle P_{1,2} \rangle$ versus T for 3- and 6-chain systems. In both systems, energy increases over the entire temperature range with a sharp increase in the middling temperatures. We see that $\langle P_1 \rangle$ and $\langle P_2 \rangle$ decrease over the entire temperature range with the sharpest decrease at the transition point. For ease of discussion, we introduce $T_{\text{low}} = 0.63$ RTU, the inflection point T_{trans} , and $T_{\text{high}} = 1.22$ RTU. The transition temperatures T_{trans} depend on the system size and are discussed later. $\langle E \rangle$ versus T behaves nearly the same for both system sizes, although T_{trans} slightly increases with increasing system size. Moreover, $\langle E \rangle$ increases dramatically at nearly the same T that $\langle P_1 \rangle$ and $\langle P_2 \rangle$ decrease. Figure 1 shows that at T_{low} , $\langle P_1 \rangle \neq \langle P_2 \rangle \neq 1$, and at T_{high} , $\langle P_1 \rangle \neq \langle P_2 \rangle \neq 0$. We investigate the matter in the section “Structural Results.”

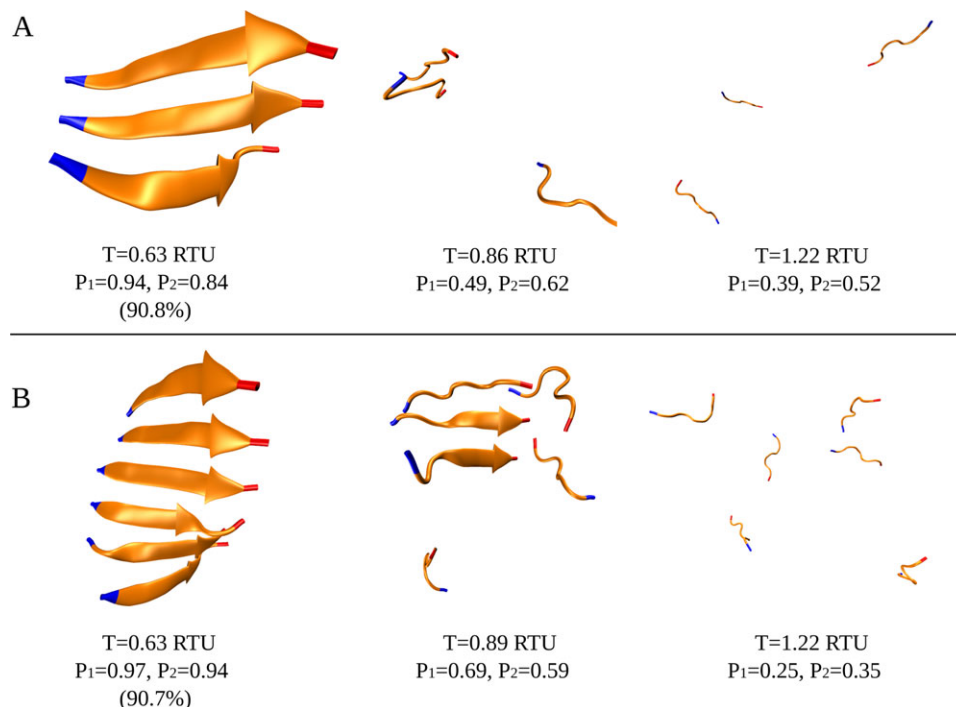
Figure 2 displays the T derivatives of $\langle E \rangle$, $\langle P_1 \rangle$, and $\langle P_2 \rangle$. Fluctuations in energy (A and C) appear maximal, whereas the fluctuations in order parameters (B and D) are minimal at the transition point. The peaks in C_v for the 3-chain system in (A) at $T_{\text{trans}} \approx 0.86$ RTU and for the 6-chain system in (C) at $T_{\text{trans}} \approx 0.89$ RTU occur at slightly higher temperatures than the peaks in $\frac{dP_1}{dT}$ and $\frac{dP_2}{dT}$: near 0.84 RTU (B) and 0.87 RTU (D) for the 3- and 6-chain system, respectively. The observed link between change in energy and change in system order suggests aggregation into ordered structures at T_{trans} .

**Figure 1**

3-chain system: $\langle E \rangle$ versus T (A) and $\langle P_{1,2} \rangle$ versus T (B). 6-chain system: $\langle E \rangle$ versus T (C) and $\langle P_{1,2} \rangle$ versus T (D). $\langle P_1 \rangle$ is represented in black, whereas $\langle P_2 \rangle$ is represented in red. Throughout the article, error bars for the 3-chain and 6-chain systems represent standard deviations calculated from independent samples obtained in different runs. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**Figure 2**

3-chain system: C_v versus T (A) and temperature fluctuation of $\langle P_1 \rangle$ (black) and $\langle P_2 \rangle$ (red) versus T (B). 6-chain system: C_v versus T (C) and temperature fluctuation of $\langle P_1 \rangle$ and $\langle P_2 \rangle$ versus T (D). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**Figure 3**

Representative structures at T_{low} (left), T_{trans} (mid), and T_{high} (right) are shown for (A) the 3-chain system and (B) the 6-chain system. For the structures at T_{low} , we included the Boltzmann populations (in parentheses) of the conformations at the given P_1 and P_2 values ± 0.05 .

The error bars for the 3- and 6-chain systems are relatively small, suggesting that the individual REMD runs were converged and are, therefore, representative of the system behavior in the simulated temperature range. However, error bars on $\langle P_1 \rangle$ and $\langle P_2 \rangle$ from the 6-chain system are larger than for the 3-chain system. This is because the system is not at one precise minimum for $\langle P_1 \rangle$ and $\langle P_2 \rangle$ at T_{low} and may be found within a range of similar, but slightly different, $\langle P_{1,2} \rangle$ values.

Structural results

We now discuss characteristic features of representative structures in the different phases (i.e., at T_{low} , T_{trans} , and T_{high}) to understand how aggregation proceeds. Figure 3 displays typical structures of the 3- and 6-chain systems for temperatures T_{low} , near the transition point T_{trans} , and at T_{high} . The values for P_1 and P_2 are given together with the structures and are representatives of the average $\langle P_1 \rangle$, $\langle P_2 \rangle$, and $\langle E \rangle$ values of the system at the given temperature in question.

The systems were all aggregated at T_{low} , partially associated at T_{trans} , and dissociated at T_{high} . At T_{low} the systems are completely aggregated into mostly in-register, parallel, twisted β -sheets. Because these β -sheets are twisted, $\langle P_1 \rangle \neq \langle P_2 \rangle \neq 1$. At T_{trans} , the system is in transition between completely dissociated and completely

associated, and between parallel order and disordered. At T_{high} , the systems are completely dissociated, and display very little ordering. The reason $\langle P_2 \rangle$ is not zero at T_{high} is because of finite size effects, which have also been observed by Caflisch and coworkers in their study of GNNQQNY trimer formation using an atomic force field.¹⁸ The ideal random case, where the average $\langle \cos^2(\theta_i) \rangle = 1/3$ and thus $3\langle \cos^2(\theta_i) \rangle/2 - 1/2 = 0$ with θ_i denoting the angle between vector \hat{z}_i and director \hat{d} [see Eq. (11)], is only true if the sample of z_i vectors is infinitely large. For the isotropic case, the N -dependence of the different eigenvalues of the ordering matrix \hat{Q} in Eq. (12) can be approximated, which is for λ_+ ⁶²

$$\lambda_+ \approx \frac{1}{\sqrt{N}} \left(\frac{\sqrt{3}}{2} + \frac{1}{6\sqrt{N}} \right). \quad (14)$$

Although this approximation was derived for N not too small ($N > 10$), it nonetheless predicts the finite P_2 values quite well, which we observe at T_{high} . For the 3-chain system, the disordered state corresponds to $\langle P_2 \rangle \approx 0.52$, whereas for the 6-chain system this value decreases to ≈ 0.34 . Equation (14) gives $\lambda_+ = 0.56$ for $N = 3$ and $\lambda_+ = 0.38$ for $N = 6$.

In the section “Weighted Histogram Analysis”, it was noted that the peaks in C_v very nearly match the peaks

in the fluctuation of $\langle P_1 \rangle$ and $\langle P_2 \rangle$. Here, we see that the change in internal energy is caused by aggregation. It was also mentioned that the peaks in $\frac{d\langle P_1 \rangle}{dT}$ and $\frac{d\langle P_2 \rangle}{dT}$ occur at slightly lower temperatures than the peaks in C_v . We posit that peak offsets are caused by the system first aggregating (change in E) and then reordering (change in $P_{1,2}$). Note that we do not suggest that the system first completely aggregates and then reorients itself, which would result in two or more peaks in C_v , only that chains come within a certain critical radius before reorienting.

Free energy profiles

To make certain that the average values $\langle P_1 \rangle$ and $\langle P_2 \rangle$ correspond to a certain aggregation state, we now look at the free energy ΔG along the order parameters P_1 and P_2 . We expect to see ΔG to have a minimum near high values of P_1 and P_2 at T_{low} and to have low values for small P_1 and P_2 at T_{high} . At T_{trans} , we anticipate the free energy barriers between all possible states to be low, allowing the system to easily shift between the aggregation states.

Figure 4 shows $\Delta G(\bar{P}_i)$ for the 3- and 6-chain GNNQQNY systems at T_{low} , T_{trans} , and T_{high} . This information is to be used with the $\langle P_1 \rangle$ and $\langle P_2 \rangle$ data in Figure 1 to make sure that we understand the full picture. From Figure 4(A), we see that at T_{low} the free energy of the 3-chain system has the global minimum near $P_1 \approx 0.94$ and a second minimum at $P_1 \approx 0.33$. These structures correspond to the parallel and antiparallel (i.e., two strands pointing in the same and one in the opposite direction) β -sheets, respectively.¹⁸ Both conformations yield a high P_2 value, giving rise to a broad ΔG minimum at $P_2 \approx 0.84$ at temperature T_{low} . At T_{low} the free energy barrier between the two (meta)stable β -sheet orientations is so large that only the global minimum is populated, giving rise to $\langle P_1 \rangle \approx 0.94$.

At T_{high} , the free energy of the 3-chain system has minima near $P_1 \approx 0.32$ and $P_2 \approx 0.48$. This is not exactly what we would expect from $\langle P_1 \rangle$ and $\langle P_2 \rangle$ in Figure 1, which suggests that ΔG should be minimal at $(P_1, P_2) \approx (0.39, 0.51)$ at T_{high} . However, the average values are not exactly equal to the global minima, because the barriers at T_{high} are small and easily overcome. Although many different P_1 and P_2 values are possible, the average values $\langle P_1 \rangle$ and $\langle P_2 \rangle$ do not change much for the individual REMD runs, as the small error bars for T_{high} in Figure 1 show. As expected at T_{trans} , we do not observe a pronounced free energy minimum for either P_1 or P_2 and, as also expected, the barriers between the possible conformations of the 3-chain system are small.

From Figure 4(B,D), we see that the free energy of the 6-chain system has a global minimum at $P_1 \approx 0.97$ and $P_2 \approx 0.94$ for T_{low} and at $P_1 = 0.00$ and $P_2 \approx 0.32$ for T_{high} . Like for the 3-chain system, a second minimum is observed in (B) at T_{low} when $P_1 \approx 0.36$, which corresponds to mixed parallel/antiparallel β -sheets, with four

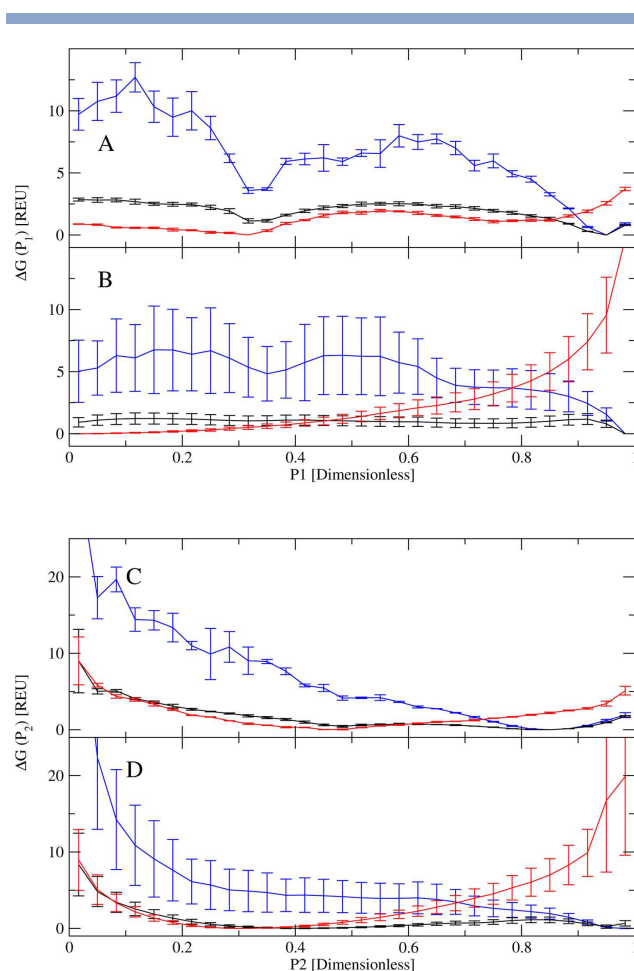


Figure 4

ΔG versus P_1 for the 3-chain (A) and 6-chain (B) systems. ΔG versus P_2 for the 3-chain (C) and 6-chain (D) systems. For each system, we visualize ΔG at T_{low} (blue), T_{trans} (black), and T_{high} (red). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

peptides pointing in the same direction and two pointing in the opposite direction. There is a free energy barrier in (B) between the minima at $P_1 = 0.36$ and $P_1 = 0.97$, representing the energy needed to transition between a mixed system and a purely parallel system. Judging by the average values, a mixed to parallel transition seems energetically plausible at T_{low} (barrier < 2 REU), whereas a parallel to mixed transition seems very unlikely (barrier > 7 REU). Although, (B) does have large error bars for ΔG at T_{low} , similarly large error bars are also present for $\langle P_1 \rangle$ and $\langle P_2 \rangle$ in Figure 1. Thus, for the 6-chain system statistically relevant conclusions regarding the free energies at T_{low} cannot be drawn.

As before, at T_{high} minima in ΔG do not exactly correspond to the average values $\langle P_1 \rangle$ and $\langle P_2 \rangle$ from Figure 1. The free energy barriers along P_1 are so low that the average $\langle P_1 \rangle$ is different from the minimum by a value of 0.25. The free energy profiles at T_{trans} are rather flat with

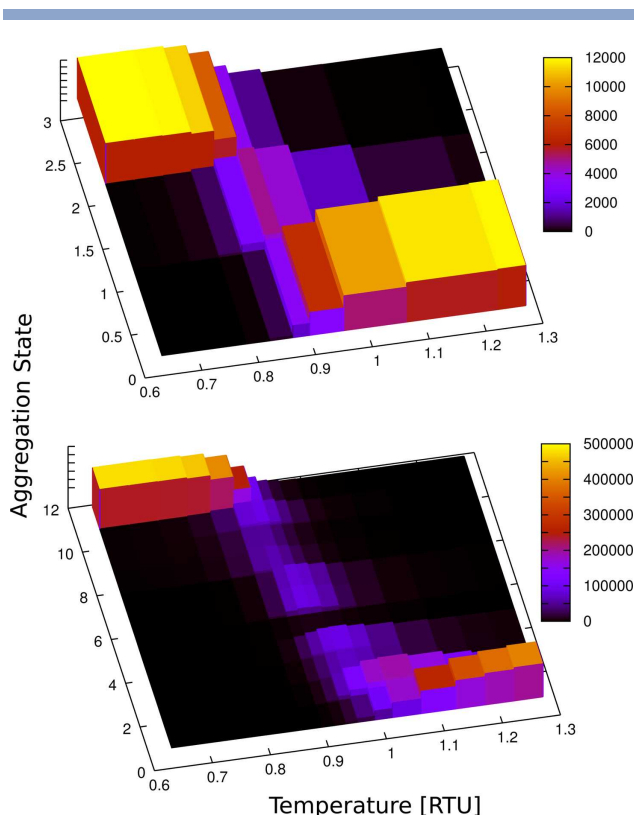


Figure 5

The aggregation pathway of GNNQQNY for the (A) 3- and (B) 6-chain systems. The depth axis is a measure of aggregation. The aggregation states are defined in Tables I and II, respectively. Along the wide axis is T , and along the tall axis is the population of the given state. The sum of the population across all states at any given T is constant. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

all values < 2 REU for $\Delta G(\tilde{P}_1)$ and $\Delta G(\tilde{P}_2)$ for $P_2 > 0.2$. As explained earlier, for small systems (small N) the nematic order parameter P_2 is not zero for the completely disordered state.

Aggregation pathway

The aggregation pathway was analyzed to determine the mechanism of aggregation. The oligomers occurring during the aggregation of the 3- and 6-chain systems were assigned to physically intuitive aggregation states defined in Tables I and II, respectively. Figure 5 shows the aggregation pathways as a function of temperature. For the 3-chain system, we see that at T_{low} the system is in a trimer conformation. Near T_{trans} the system visits all three aggregation states, each with similar probability. At T_{high} , the system is in a 3-monomers conformation. Because we see a smooth change from three monomers to one trimer, we conclude that the GNNQQNY trimer prefers to associate via monomer addition rather than a condensation-ordering transition. This is because an ini-

tial condensation process would be practically replete of the intermediate (1 dimer and 1 monomer) configuration. Furthermore, we would expect to see two peaks in the C_v curve in Figure 2, one for condensation and another for reordering.

For the 6-chain system, we see that at T_{low} the system is in a hexameric conformation. Near T_{trans} the system occupies every state; only the configuration consisting of a monomer, a dimer, and a trimer (aggregation state 4) is noticeably suppressed. At T_{high} , the system is either in a 6-monomers configuration or in a 4-monomers and 1-dimer configuration, with the probability of the completely monomeric configuration being much higher. As with the 3-chain system, the 6-chain system visits many intermediate states. This suggests that the system does not undergo an initial condensation and instead aggregates via a different mechanism. One possibility is monomer addition. In this process, a monomer attaches itself to a higher oligomeric species. But there is also the possibility of two small oligomers aggregating into a larger oligomer, which we call oligomer fusion.⁶³

12-Chain GNNQQNY

We performed a single 12-chain simulation for 500 ns per replica, and present the results separately from the 3- and 6-chain systems, because they are so striking.

Weighted histogram analysis and order parameters

Figure 6 shows the temperature dependence of $\langle E \rangle$ and C_v for the 12-chain system. As for the 3- and 6-chain systems, E increases over the entire temperature range with a sharp increase at T_{trans} . The peak in the heat capacity is at $T = T_{\text{trans}} = 0.95$ RTU. This supports our

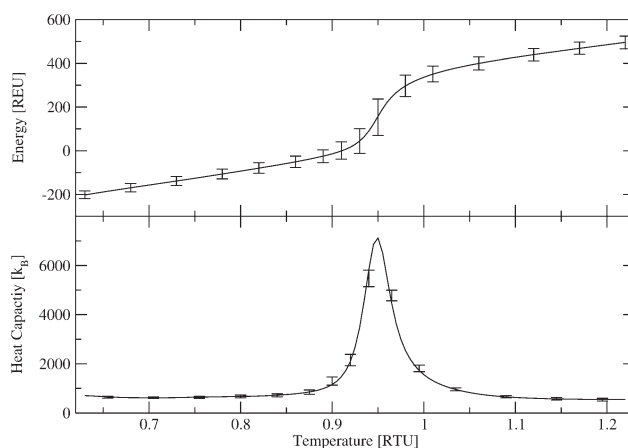


Figure 6

$\langle E \rangle$ versus T (top) and C_v versus T (bottom) for the 12-chain GNNQQNY system. Error bars were calculated from the standard deviation within each temperature thread after system equilibration.

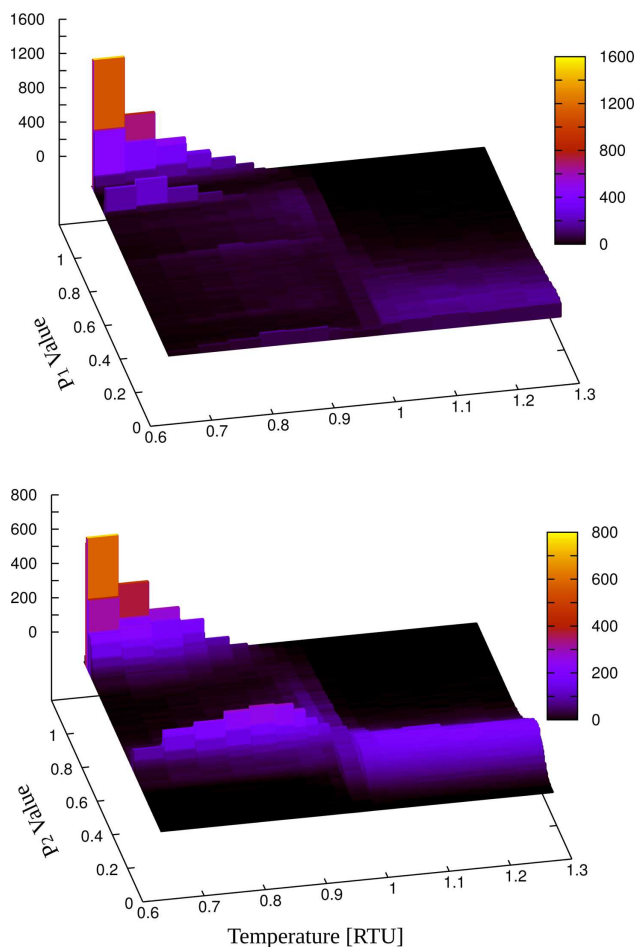


Figure 7

Histogram in dependence of P_1 and T (A) and P_2 and T (B) for the 12-chain GNNQQNY system. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

observation that with increasing system size aggregation occurs at higher temperatures. Although the 12-chain C_V is similar to 3- and 6-chains, the behavior of $P_{1,2}$ is dramatically different. To understand the aggregation behavior of the 12-chain system, we plotted a histogram in dependence of P_1 and P_2 versus T in Figure 7.

At low temperatures, we see a marked bimodal distribution of P_1 and P_2 . This indicates the system has two different behavioral phases that it exhibits at low temperature. For the 12-chain system, at $T < T_{\text{trans}}$ the P_2 parameter has two peaks located at $P_2 \approx 0.36$ and 0.99 , and the P_1 parameter has also two peaks at $P_1 \approx 0.84$ and 0.99 . Near T_{trans} , all P_1 and P_2 values are almost equally populated, witnessing a multitude of peptide conformations and orientations. However, P_2 has a slight peak near 0.45 at T_{trans} . The behavior for $T > T_{\text{trans}}$ is difficult to observe from Figure 7, but we can see that P_1 has decreasing population with increasing P_1 value, and the distribution of P_2 is centered around 0.3 .

Structural results

To understand the observed P_1 and P_2 distributions, we extracted typical structures, which are shown in Figure 8. Figure 8 reveals that at T_{low} the GNNQQNY dodecamer occurs in two distinct structures. The low P_2 value corresponds to two stacked β -sheets with an almost perpendicular relative orientation with respect to each other. Within each of these β -sheets, the peptides are aligned parallel. Although each of these sheets may individually have high P_2 values, the measured values of the system as a unit are low. The high P_2 value corresponds to 12 chains aligned in one parallel β -sheet. At T_{trans} , we see many small oligomers that have broken apart from the larger β -sheets aggregates, giving rise to the relatively flat P_1 and P_2 distributions. At T_{high} , we see that the 12-chain system is completely dissociated and disordered. The average value of $P_2 = 0.3$ corresponds well to the value predicted by Eq. (14), which is 0.26 .

Aggregation pathway

Figure 9 shows the aggregation pathway of the 12-chain GNNQQNY system. At T_{low} , the largest β -sheet in the sys-

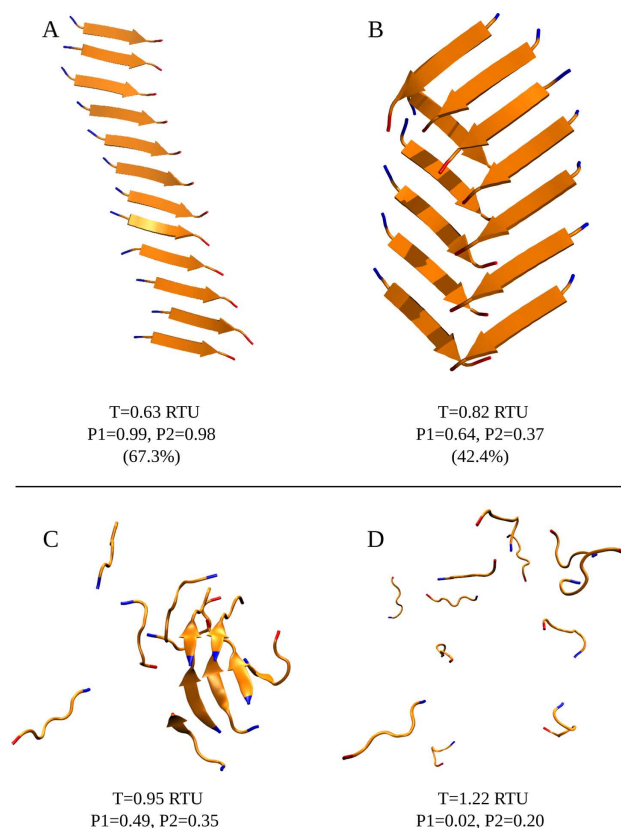


Figure 8

Typical structures of 12-chain GNNQQNY at T_{low} , T_{trans} , and T_{high} . For the structures at $T = 0.63$ RTU and $T = 0.82$ RTU, we included the Boltzmann populations (in parentheses) of the conformations at the given P_1 and P_2 values ± 0.05 . [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

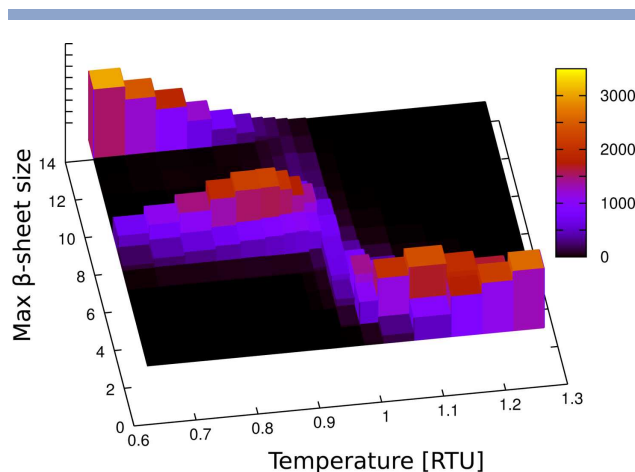


Figure 9

Aggregation states of 12-chain GNNQQNY in dependence of temperature. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

tem is either a dodecameric, heptameric, or hexameric conformation. The 7- and 6-chain β -sheets are paired with a 5- or 6-chain β -sheet, respectively. Near T_{trans} , all oligomers ranging from a dodecamer down to monomers only are populated, but with the highest populations being observed for the hexamer and heptamer. At T_{high} , the largest β -sheet species is either a monomer, dimer, or trimer, with the probability of 12 monomers being the highest.

Having understood how lower oligomeric species formed for the 6-chain system, the 12-chain system was used to study how larger aggregates form. As all β -sheet sizes are observed over the temperature range, we conclude that large GNNQQNY aggregates also do not form by hydrophobic collapse. However, unlike the 3- and 6-chain systems, the 12-chain system does not ramp up smoothly from a dimeric to a dodecameric β -sheet. At T_{high} , we saw monomers and dimers, as expected. At T_{low} , the 12-chain β -sheet is the dominant oligomeric species. But for $T_{\text{low}} < T < T_{\text{trans}}$, we observed that the system preferred to form double-layered β -sheets with a maximum sheet size of 6 or 7 GNNQQNY peptides (structure B in Fig. 8).

As the system cooled, most of the smaller β -sheets aggregated into a single 12-chain β -sheet. Furthermore, we conclude that the aggregation into a single sheet occurred via β -sheet fusion, where two smaller aggregates, for example, two 6-chain sheets or a 7- and a 5-chain sheet coming together. β -Sheet fusion is evidenced by the relative lack of abundance of 8–11 chain β -sheets, which would be expected to occur much more often if the aggregation from 7 chains to 12-chains was done via monomer addition.

DISCUSSION

The self-assembly of peptides and proteins into soluble oligomers that evolve to protofibrils and eventually to

fibrils is still challenging to study at atomic resolution using experimental approaches. Molecular simulations help reveal the molecular mechanism of the multistep process of amyloid aggregation, elucidating structural and kinetic information about this process. The aim of this work is to characterize structural transitions involved in oligomer formation for the model peptide GNNQQNY. To this end, we simulated the aggregation of 3, 6, and 12 GNNQQNY peptides via REMD using the CG force field of Bereau and Deserno.⁵¹ We used WHAM to study how various impact parameters vary with T . Because GNNQQNY is predicted to form parallel β -sheets in the amyloid fibril,^{10–13} we examined the parallel order of the systems by computing the orientational order parameters P_1 in Eq. (10) and P_2 in Eq. (11), and examined the change of free energy along these parameters. We determined the aggregation pathway of the 3-, 6-, and 12-chain systems by examining the populations of the oligomeric species in dependence of temperature.

GNNQQNY aggregates into parallel β -sheets

Our results clearly show that GNNQQNY has a strong propensity to form stable β -sheets independent of its oligomeric state. This observation is in agreement with the result of previous simulation studies investigating GNNQQNY oligomerization, ranging from dimer, trimer, and tetramer formation using atomic force fields^{17,18,22,23,26} to the aggregation into 12-mers and 20-mers using a CG protein model.³¹ Our REMD study leads to an almost exclusive population of parallel β -sheets independent of the oligomer size. This parallel order behavior was verified by examining the temperature and aggregation dependence of the order parameters $\langle P_1 \rangle$ and $\langle P_2 \rangle$, the free energy along these order parameters, and examination of the structures of the 3-, 6-, and 12-mers. In order not to miss antiparallel β -sheets, we examined the distributions of P_1 and P_2 values at T_{low} , T_{trans} , and T_{high} . The resulting histograms, shown in Figures S1–S3, Supporting Information, confirm that antiparallel structures also form, especially at T_{trans} . For both the 3- and 6-chain systems examples for mixed parallel/antiparallel structures are shown in Figure S4, Supporting Information. As the temperature decreases, parallel β -sheets become more favored, leading to very low population of antiparallel β -sheets at T_{low} . The almost exclusive preference for parallel β -sheets at low temperatures contrasts the findings of earlier simulation studies, which usually report both antiparallel and parallel β -sheets for GNNQQNY oligomers^{17,18,22,23,25–28,30,31} and other amyloidogenic peptides.^{64,65}

From atomic simulations, it could be reasoned that the stabilization of GNNQQNY parallel and antiparallel assemblies is based on different effects.^{23,26} Although side-chain interactions contribute to the stability of the

parallel states, antiparallel β -sheets form more H-bonds and are stabilized by electrostatic attraction between the charged peptide termini from neighbored strands in the sheet. The latter effect was evidenced by atomic simulations mimicking acidic pH conditions, with charged N-terminus and uncharged C-terminus, leading to a deficiency of antiparallel structures.²⁶ This finding was used to explain the parallel arrangement of β -strands in amyloid-like GNNQQNY microcrystals^{10–12} and fibrils,¹³ both of which have been grown in acidic conditions. Park et al.²⁷ have found that shape complementarity is the major factor for determining the sheet pattern, because it stabilizes the steric zipper motif. Parallel β -sheets are thus advantageous, as they enable better intra-sheet and intersheet side-chain contacts.^{17,20,23} Only the presence of charged side chains, like in VEALYL, KLVFFAE, and STVIIIE, favors antiparallel β -sheets, due to electrostatic interactions.^{27,66} Trovato et al.⁶⁷ derived sequence-specific interaction energies between pairs of protein fragments from a statistical analysis of the native folds of globular proteins, allowing them to predict the arrangement of ordered self-propagating cross- β structures. Based on this algorithm, they predict that the parallel arrangement for GNNQQNY is energetically favored. The gain in stabilization should increase with sheet size, which explains why simulation studies find an increased occurrence of parallel β -sheets the larger the GNNQQNY oligomers.^{26,31} This effect was also observed for the NNQQ sequence⁶⁴ and the amyloid tau fragment AceVQIVYK-NH₂,⁶⁵ which both display parallel β -strands in amyloid fibrils as was determined experimentally.

Overall, the observation of parallel β -sheets in our simulation is in agreement with experimental findings and simulation studies. The deficiency of antiparallel β -sheets can be attributed to the lack of charged termini in our study, whereas the CG study by Mousseau and coworkers³¹ and most of the atomic simulations^{17,18,22,23,25–28,30} used charged termini to study GNNQQNY aggregation.

GNNQQNY aggregation and β -sheet formation occur at the same temperature

The structural and aggregate-state analysis of our REMD data for 3-, 6-, and 12-chain GNNQQNY system revealed that changes in oligomer size and peptide order correspond to a single change in the heat capacity, C_v . GNNQQNY self-assembles into parallel β -sheets at the transition temperature T_{trans} of 225–285 K. The absence of further C_v peaks, along with the temperature dependence of the order parameters $\langle P_1 \rangle$ and $\langle P_2 \rangle$ further evidence that aggregation and β -sheet-formation occur at the same temperature, implying unordered GNNQQNY aggregates are thermodynamically unstable. Our REMD data do not allow us to conclude whether or not the intermediate aggregates at T_{trans} are kinetically stable. This

question needs to be answered with constant-temperature MD simulations. However, it is very likely that a high kinetic stability of unordered aggregates would manifest itself in a second heat capacity peak, as aggregation and β -sheet-formation would be expected to occur at different temperatures. The observation of only one heat capacity peak during the aggregation of GNNQQNY followed by REMD simulations is confirmed by other studies using different force fields and investigating other oligomers: for the dimer and trimer from atomic studies with implicit solvent^{18,23} and for the 3-mer, 12-mer, and 20-mer from a CG study using OPEP.³¹

The use of different force fields and concentration conditions ($c = 80$ mM in our study, 10 mM in Ref. 23, 12 mM in Ref. 18, and 4.15 mM in Ref. 31) leads to the observation of different transition temperatures: depending on system size between 250 and 280 K in our study and between 280 and 295 K in the study by Mousseau and coworkers,³¹ while temperatures of ≈ 400 K (Ref. 23) and 371 K (Ref. 18) were reported from simulations using CHARMM19.⁶⁸ Although the difference of more than 100 K between the atomic and CG T_{trans} results might be surprising, it should be noted that for the aggregation of the KFFE peptide it has been shown that the employment of different implicit solvent models in connection with the same atomic force field can result in different oligomeric structures and aggregation temperatures of more than 200 K difference.⁶⁹ In the light of this finding, the similar transition temperatures from our and the OPEP studies³¹ is quite remarkable. A possible explanation is that Bereau and Deserno followed a similar coarse-graining strategy⁵¹ to that used by Derreux and coworkers for the development of OPEP,^{47–50} using a detailed representation of the backbone atoms and reducing each side chain to one single bead. Furthermore, the OPEP and Bereau and Deserno energy functions are expressed in terms of similar types of interactions, and they treat solvent effects and electrostatic interactions originating from charged side chains implicitly.

By repeating the REMD simulation of the 6-chain system at a concentration of 10 mM, we verified that the lack of amorphous GNNQQNY aggregates is not due to the rather high concentration in our study. The results of this simulation (energy, C_v , P_1 , and P_2) are summarized in Figure S6 (Supporting Information). Like for $c = 80$ mM, we observe a single C_v peak that corresponds to the aggregation into ordered β -sheets. This is supported by the increase of P_1 and P_2 at the transition temperature. Expectedly, the lower concentration shifts the equilibrium between the dissociated and aggregated phases toward the dissociated state, leading to a slight decrease of T_{trans} . At T_{low} , P_1 and P_2 reach similar values as in the simulations with $c = 80$ mM, whereas at T_{high} the exact same values are found. Visual inspection of the final oligomers found at $c = 10$ mM further confirmed that the

GNNQQNY aggregation pathway is not very sensitive to concentration changes on the 10^{-5} molar scale.

GNNQQNY β -sheets grow via monomer addition and oligomer fusion

From Figures 5 and 9, we can observe that for the 3- and 6-chain systems, GNNQQNY aggregates into one long β -sheet as the system is cooled. In both cases, all oligomer sizes are visited during the formation of the trimeric and hexameric β -sheet, respectively. We assigned the combination of various oligomer sizes to specific aggregate states (Tables I and II), allowing us to follow the aggregation pathway in temperature space. The comparison with the order parameters $\langle P_1 \rangle$ and $\langle P_2 \rangle$ and the single C_v peak already revealed that aggregation and β -sheet formation occur at the same temperature. We can, therefore, conclude that the oligomerization into small β -sheets proceeds via monomer addition to existing β -sheets and fusion of smaller β -sheets. For the 12-chain system, GNNQQNY aggregates either into a single 12-mer β -sheet or into a double-layered β -sheet with the two β -sheets being of similar size. Figure 9 shows that the probability for the β -sheets consisting of 8–11 peptides is considerably reduced compared to the smaller and dodecameric β -sheets. Therefore, we can conclude that first β -sheets with a high preference for 5-mers, 6-mers, and 7-mers are formed, which then aggregate into a single- or double-layered 12-mer. The final stable structures, which we obtain for the 12-mer, are very similar to those found with OPEP, apart from the above-mentioned difference regarding the orientation of the strands within the sheets.³¹

The tendency of the GNNQQNY sequence to form small but stable β -sheets, which act as aggregation nuclei has already been noticed in previous simulation studies on this system.^{25,31} These studies reported tetrameric aggregation nuclei, while we also observe a high stability for pentameric and hexameric β -sheets. In agreement with previous simulation work²² and with entropic and energetic arguments,¹¹ we can confirm that short double layers do not form. Only for the 12-mer, we observe such structures corresponding to steric zipper formation. They are entropically favored over the single layer 12-mer for temperatures between 235 and 285 K, whereas below 235 K the latter is enthalpically preferred.

Using CG protein models, it has been demonstrated that β -sheet propensity, conformational flexibility of the monomeric peptide, and hydrophobicity determine the aggregation pathway.^{34,39,70} The importance of the monomeric structure for amyloid formation has been highlighted in work by Pellarin and Caflisch,³⁴ who used a CG model with varying stability of the β -prone state to investigate the kinetics and pathways of fibril formations. They found that different β -propensities give rise to different aggregation mechanisms, namely a disordered

aggregation pathway for the β -unstable state; fibril formation via oligomeric on-pathway intermediates for a more stable β -state; and fibril formation without intermediates in the case of a very β -stable state.³⁴ Similar conclusions were drawn by Bellesia and Shea, who used a different coarse-graining approach to compute a phase diagram for peptide aggregation as a function of temperature and β -sheet propensity.³⁹ Hall and coworkers simulated the aggregation of palindromic sequences into amyloid fibrils.⁷⁰ They found different aggregation pathways for the sequences AGAAAAGA and the more hydrophobic VAGAAAAGAV. The aggregation of the latter can be explained by the concept of the condensation-ordering transition,^{34,39,71} in which disordered oligomers formed by condensation at an early stage are transformed to ordered structures at a later stage. Here, the condensation step is driven by both hydrophobic interactions and hydrogen bonding, while for AGAAAAGA the initial aggregation is mostly driven by hydrogen bonding.⁷⁰ AGAAAAGA, like GNNQQNY, initially forms small β -sheets; these elongate by monomer addition and then associate into fibrils.⁷⁰

CONCLUSIONS

The main conclusion from our study is that the aggregation of the hydrophilic GNNQQNY sequence is mainly driven by H-bond formation, leading to the formation of β -sheets from the very beginning of the assembly process. The preceding condensation step, resulting in amorphous aggregates as observed for amyloidogenic hydrophobic peptides, does not occur for GNNQQNY. Instead, condensation (aggregation) and ordering take place simultaneously. This is evidenced by a single peak in the heat capacity curve for GNNQQNY assembly.^{23,31}

ACKNOWLEDGMENTS

The authors are grateful to Dr. Tristan Bereau for helpful discussions.

REFERENCES

1. Selkoe DJ. Folding proteins in fatal ways. *Nature* 2003;426:900–904.
2. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* 2006;75:333–366.
3. Uversky VN, Fink AL. Protein misfolding, aggregation and conformational disease. Heidelberg: Springer; 2006.
4. Kaye R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW, Glabe CG. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* 2003;300:486–489.
5. Thirumalai D, Klimov DK, Dima RI. Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr Opin Struct Biol* 2003;13:146–159.
6. Walsh DM, Klyubin I, Fadeeva JV, Cullen WK, Anwyl R, Wolfe MS, Rowan MJ, Selkoe DJ. Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature* 2002;416:535–539.

7. Kirkitadze MD, Bitan G, Teplow DB. Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: the emerging role of oligomeric assemblies. *J Neurosci Res* 2002;69:567–577.
8. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, Stefani M. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* 2002;416:507–511.
9. Klein WL, Stine WB, Teplow DB. Small assemblies of unmodified amyloid β -protein are the proximate neurotoxin in Alzheimer's disease. *Neurobiol Aging* 2004;25:569–580.
10. Balbirnie M, Grothe R, Eisenberg DS. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β -sheet structure for amyloid. *Proc Natl Acad Sci USA* 2001;98:2375–2380.
11. Nelson R, Sawaya MR, Balbirnie M, Madsen AØ, Riekel C, Grothe R, Eisenberg D. Structure of the cross- β spine of amyloid-like fibrils. *Nature* 2005;435:773–778.
12. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJ, McFarlane HT, Madsen AØ, Riekel C, Eisenberg D. Atomic structures of amyloid cross- β ; spines reveal varied steric zippers. *Nature* 2007;447:453–457.
13. van der Wel PCA, Lewandowski JR, Griffin RG. Structural characterization of GNNQQNY amyloid fibrils by magic angle spinning NMR. *Biochemistry* 2010;49:9457–9469.
14. Tycko R. Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q Rev Biophys* 2006;39:1–55.
15. Heise H. Solid-state NMR spectroscopy of amyloid proteins. *Chem-BioChem* 2008;9:179–189.
16. Stroud JC, Liu C, Teng PK, Eisenberg D. Toxic fibrillar oligomers of amyloid-beta have cross-beta structure. *Proc Natl Acad Sci USA* 2012;109:7717–7722.
17. Gsponer J, Haberthür U, Caflisch A. The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35. *Proc Natl Acad Sci USA* 2003;100:5154–5159.
18. Cecchini M, Rao F, Seeber M, Caflisch A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J Chem Phys* 2004;121:10748.
19. Lipfert J, Franklin J, Wu F, Doniach S. Protein misfolding and amyloid formation for the peptide GNNQQNY from yeast prion protein Sup35: simulation by reaction path annealing. *J Mol Biol* 2005;349:648–658.
20. Zheng J, Ma B, Tsai CJ, Nussinov R. Structural stability and dynamics of an amyloid-forming peptide GNNQQNY from the yeast prion Sup-35. *Biophys J* 2006;91:824–833.
21. Esposito L, Pedone C, Vitagliano L. Molecular dynamics analyses of cross- β -spine steric zipper models: β -sheet twisting and aggregation. *Proc Natl Acad Sci USA* 2006;103:11533–11538.
22. Zhang Z, Chen H, Bai H, Lai L. Molecular dynamics simulations on the oligomer-formation process of the GNNQQNY peptide from yeast prion protein Sup35. *Biophys J* 2007;93:1484–1492.
23. Strodel B, Whittleston CS, Wales DJ. Thermodynamics and kinetics of aggregation for the GNNQQNY peptide. *J Am Chem Soc* 2007;129:16005–16014.
24. Wang J, Tan C, Chen HF, Luo R. All-atom computer simulations of amyloid fibrils disaggregation. *Biophys J* 2008;95:5037–5047.
25. Meli M, Morra G, Colombo G. Investigating the mechanism of peptide aggregation: insights from mixed Monte Carlo-molecular dynamics simulations. *Biophys J* 2008;94:4414–4426.
26. Vitagliano L, Esposito L, Pedone C, de Simone A. Stability of single sheet GNNQQNY aggregates analyzed by replica exchange molecular dynamics: antiparallel versus parallel association. *Biochem Biophys Res Commun* 2008;377:1036–1041.
27. Park J, Kahng B, Hwang W. Thermodynamic selection of steric zipper patterns in the amyloid cross- β spine. *PLoS Comput Biol* 2009;5:e1000492.
28. Berryman JT, Radford SE, Harris SA. Thermodynamic description of polymorphism in Q- and N-rich peptide aggregates revealed by atomistic simulation. *Biophys J* 2009;97:1–11.
29. Periole X, Rampioni A, Vendruscolo M, Mark AE. Factors that affect the degree of twist in beta-sheet structures: a molecular dynamics simulation study of a cross-beta filament of the GNNQQNY peptide. *J Phys Chem B* 2009;113:1728–1737.
30. Reddy AS, Chopra M, de Pablo JJ. GNNQQNY-investigation of early steps during amyloid formation. *Biophys J* 2010;98:1038–1045.
31. Nasica-Labouze J, Meli M, Derreumaux P, Colombo G, Mousseau N. A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide GNNQQNY from the yeast prion sup-35. *PLoS Comput Biol* 2011;7:e1002051.
32. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 1999;314:141–151.
33. Wu C, Shea JE. Coarse-grained models for protein aggregation. *Curr Opin Struct Biol* 2011;21:209–220.
34. Pellarin R, Caflisch A. Interpreting the aggregation kinetics of amyloid peptides. *J Mol Biol* 2006;360:882–892.
35. Pellarin R, Guarnera E, Caflisch A. Pathways and intermediates of amyloid fibril formation. *J Mol Biol* 2007;374:917–924.
36. Magno A, Caflisch A, Pellarin R. Crowding effects on amyloid aggregation kinetics. *J Phys Chem Lett* 2010;1:3027–3032.
37. Bellesia G, Shea JE. Self-assembly of β -sheet forming peptides into chiral fibrillar clusters. *J Chem Phys* 2007;126:245104.
38. Bellesia G, Shea JE. Structure and stability of amyloid fibrils formed from synthetic β -peptides. *Front Biosci* 2008;13:6957–6965.
39. Bellesia G, Shea JE. Effect of β -sheet propensity on peptide aggregation. *J Chem Phys* 2009;130:145103.
40. Bellesia G, Shea JE. Diversity of kinetic pathways in amyloid fibril formation. *J Chem Phys* 2009;131:111102.
41. Nguyen HD, Hall CK. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc Natl Acad Sci USA* 2004;101:16180–16185.
42. Nguyen HD, Hall CK. Spontaneous fibril formation by polyalanines; discontinuous molecular dynamics simulations. *J Am Chem Soc* 2006;128:1890–1901.
43. Cheon M, Chang I, Hall CK. Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins: Struct Funct Bioinf* 2010;78:2950–2960.
44. Urbanc B, Cruz L, Ding F, Sammond D, Khare S, Buldyrev SV, Stanley HE, Dokholyan NV. Molecular dynamics simulation of amyloid β dimer formation. *Biophys J* 2004;87:2310–2321.
45. Urbanc B, Cruz L, Yun S., Buldyrev SV, Bitan G, Teplow DB, Stanley HE. In silico study of amyloid β -protein folding and oligomerization. *Proc Natl Acad Sci USA* 2004;101:17345–17350.
46. Urbanc B, Betnel M, Cruz L, Bitan G, Teplow DB. Elucidation of amyloid β -protein oligomerization mechanisms: discrete molecular dynamics study. *J Am Chem Soc* 2010;132:4266–4280.
47. Derreumaux P. From polypeptide sequences to structures using Monte-Carlo simulations and an optimized potential. *J Chem Phys* 1999;111:2301–2310.
48. Melquiond A, Boucher G, Mousseau N, Derreumaux P. Following the aggregation of amyloid-forming peptides by computer simulations. *J Chem Phys* 2005;122:174904.
49. Derreumaux P, Mousseau N. Coarse-grained protein molecular dynamics simulations. *J Chem Phys* 2007;126:025101.
50. Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. *Proteins: Struct Funct Bioinf* 2007;69:394–408.
51. Bereau T, Deserno M. Generic coarse-grained model for protein folding and aggregation. *J Chem Phys* 2009;130:235106.
52. Bereau T, Bachmann M, Deserno M. Interplay between secondary and tertiary structure formation in protein folding cooperativity. *J Am Chem Soc* 2010;132:13129–13131.

53. Bereau T, Deserno M, Bachmann M. Structural basis of folding cooperativity in model proteins: insights from a microcanonical perspective. *Biophys J* 2011;100:2764–2772.
54. Bereau T, Globisch C, Deserno M, Peter C. Coarse-grained and atomistic simulations of the salt-stable cowpea chlorotic mottle virus (ss-ccmv) subunit 2649: β -barrel stability of the hexamer and pentamer geometries. *J Chem Theory Comput* 2012;8:3750–3758.
55. Miyazawa S, Jernigan RL. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 1996;256:623–644.
56. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse grained force field: extension to proteins. *J Chem Theor Comput* 2008;4:819–834.
57. Limbach HJ, Arnold A, Mann BA, Holm C. ESPResSo—an extensible simulation package for research on soft matter systems. *Comput Phys Commun* 2006;174:704–727.
58. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087.
59. Bereau T, Swendsen RH. Optimized convergence for multiple histogram analysis. *J Comput Phys* 2009;228:6119–6129.
60. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules. 1: the method. *J Comput Chem* 1992;13:1011–1021.
61. Allen MP, Tildesley DJ. *Computer simulation of liquids*. Oxford: Clarendon Press; 1987.
62. Eppenga R, Frenkel D. Monte Carlo study of the isotropic and nematic phases of infinitely thin hard platelets. *Mol Phys* 1984;52:6:1303–1334.
63. Hill SE, Robinson J, Matthews G, Muschol M. Amyloid protofibrils of lysozyme nucleate and grow via oligomer fusion. *Biophys J* 2009;96:3781–3790.
64. Lu Y, Wei G, Derreumaux P. Structural, thermodynamical, and dynamical properties of oligomers formed by the amyloid NNQQ peptide: insights from coarse-grained simulations. *J Chem Phys* 2012;137:025101.
65. Li DW, Mohanty S, Irbäck A, Huo S. Formation and growth of oligomers: a Monte Carlo study of an amyloid tau fragment. *PLoS Comput Biol* 2008;4:e1000238.
66. Tsai HH, Gunasekaran K, Nussinov R. Sequence and structure analysis of parallel helices: implication for constructing amyloid structural models. *Structure* 2006;14:1059–1072.
67. Trovato A, Chiti F, Maritan A, Seno F. Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Comput Biol* 2006;2:e170.
68. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 1983;4:187–217.
69. Strodel B, Wales DJ. Implicit solvent models and the energy landscape for aggregation of the amyloidogenic KFFE peptide. *J Chem Theory Comput* 2008;4:657–672.
70. Wagoner VA, Cheon M, Chang I, Hall CK. Computer simulation study of amyloid fibril formation by palindromic sequences in prion peptides. *Proteins: Struct Funct Bioinf* 2011;79:2132–2145.
71. Auer S, Meersman F, Dobson CM, Vendruscolo M. A generic mechanism of emergence of amyloid protofilaments from disordered oligomeric aggregates. *PLoS Comput Biol* 2008;4:e1000222.