JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation

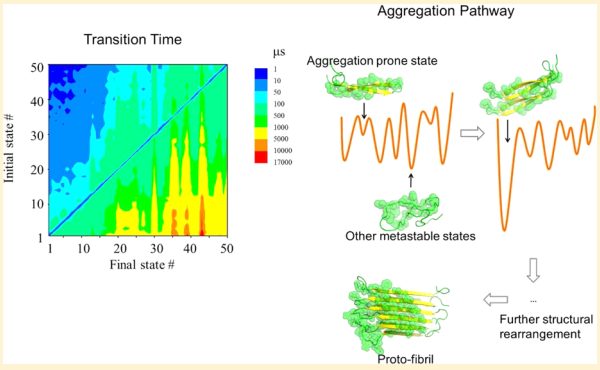
Qin Qiao,[†] Gregory R. Bowman,[∥] and Xuhui Huang*[,†,‡,§]

[†]Bioengineering Graduate Program, Division of Biomedical Engineering, [‡]Department of Chemistry, [§]Center of Systems Biology and Human Health, School of Science and Institute for Advance Study, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

[∥]Departments of Chemistry and Molecular and Cell Biology, University of California, Berkeley, California 94720, United States

Supporting Information

ABSTRACT: Amyloid fibril deposits of the intrinsically disordered hIAPP peptide are found in 95% of type II diabetes patients, and the aggregation of this peptide is suggested to induce apoptotic cell-death in insulin-producing β-cells. Understanding the structure and dynamics of the hIAPP monomer in solution is thus important for understanding the nucleation of aggregation and the formation of oligomers. In this study, we identify the metastable conformational states of the hIAPP monomer and the dynamics of transitioning between them using Markov state models constructed from extensive molecular dynamics simulations. We show that the overall structure of the hIAPP peptide is random coil-like and lacks a dominant folded structure. Despite this fact, our model reveals a large number of reasonably well-populated metastable conformational states (or local free energy minima) having populations of a few percent or less. The time scales for transitioning between these states range from several microseconds to milliseconds. In contrast to folded proteins, there is no kinetic hub. More strikingly, a few states contain significant amounts of β-hairpin secondary structure and extended hydrophobic surfaces that are exposed to the solvent. We propose that these states may facilitate the nucleation of hIAPP aggregation through a significant component of the conformational selection mechanism, because they may increase their populations upon aggregation by promoting hydrophobic interactions and at the same time provide a flat geometry to seed the ordered β-strand packing of the fibrils.



INTRODUCTION

Human islet amyloid polypeptide (hIAPP), also known as amylin, is a 37-residue hormone cosecreted by pancreatic β-cells along with insulin. hIAPP is associated with type II diabetes because its amyloid fibril deposit is found in 95% of type II diabetes patients. Experiments have shown that the formation of hIAPP deposits can induce apoptotic cell-death in insulin-producing β-cells.[1−3] Therefore, it has been proposed that fibril formation is responsible for the onset of type II diabetes. The fibril structure of hIAPP and its formation have been studied extensively, and various fibril structural models have been proposed with the common feature of a long-extended cross β-structure through residues 8−37.[4,5] However, recent experimental evidence suggests that small oligomers preceding the formation of fibrils may be the key cytotoxic elements, leading to a number of recent studies of oligomer formation.[6−8] The formation of oligomers may lead to the disruption of the cell membrane and thus becomes cytotoxic.[9−11] The presence of the cell membranes is found to stabilize the α-helical intermediate structures, which have been suggested as a critical component in the aggregation of hIAPP *in vivo*.[12−15] The subsequent fibrillation process can then be greatly accelerated upon α-helical to β-strand structural transitions.[12,15,16] Experimental observations further suggest that amyloid fibril formation is nucleation dependent,[17] and the nucleation phase contains a heterogeneity of pathways with different oligomers formed.[18−20]

Multiple mechanisms ranging from nucleated conformational conversion (NCC)[21] to conformational selection (CS)[22,23] have been proposed to understand the fibril formation for hIAPP and other intrinsically disordered peptides (IDPs).[21,24−26] The NCC[21] mechanism is an analogue to the induced fit model[27] developed for ligand binding systems, in which the monomers first collapse through either hydrophilic or nonspecific hydrophobic interactions and further rearrange to form the β-ordered proto-fibril.[28−32] For hIAPP, The NCC mechanism is supported by NMR experiments,[20] and the initial collapse is suggested to occur in the α-helix region in the N-terminus. Conformational selection[22,23] is another extreme of the fibril formation mechanism. In this mechanism, monomers containing pre-existing β-structures can selectively collapse and

Received: March 29, 2013
Published: September 10, 2013

   dx.doi.org/10.1021/ja403147m | J. Am. Chem. Soc. 2013, 135, 16092−16101

further grow to form fibrils. The dimerization process of hIAPP has been shown to be largely through the conformational selection by molecular dynamics (MD) simulations, and the cross sections of the dimer structures evolved from the stacking of β-rich monomers are consistent with the measurement of ion mobility mass spectroscopy.[33] Understanding the monomer structure is important to the mechanism since the structural preference of the monomers may determine the nucleation pathway as found in both all-atom and coarse-grained models of IDPs.[34−38] As for hIAPP, a large extent of random coil structures are observed in CD experiments.[39] However, both simulations and experiments revealed to some extent a secondary structure in the monomer structural ensemble, including an α helix in the N-terminal[40−44] as well as some extended β-hairpin structures,[42−45] which may serve as the precursor for fibril formation. The presence of aggregation prone states N* in the monomer ensemble of amyloid peptides has been suggested by Thirumalai, Straub, and co-workers.[46−48] Despite their high free energies, these N* states may still be sampled by thermal fluctuations, and the collapse of these N* states may facilitate aggregation. Both all-atom and coarse-grained simulations have identified the N* states in various amyloid systems.[30,34−38,49−51]

Dynamics play a crucial role in both mechanisms discussed above. If the collision rate for monomers is fast, the NCC mechanism will be dominant because monomers in different conformations may collapse frequently, and thus structural rearrangements are likely to be necessary for the system to reach the ordered fibril structure. On the other hand, the CS will be dominant when the collision rate is low. In this scenario, monomers have enough time to explore different conformations between two collisions, and thus only β-hairpin like monomers may form stable on-pathway oligomer and further grow to the fibrils. Lapidus and co-workers have previously shown that the kinetics of the aggregation process is optimized (or fastest) when these two time scales are comparable for different IDPs and also well-structured proteins.[52] A rough estimate of the collision interval for hIAPP monomers (at 1 μm concentration, with a diffusion coefficient equal to 3.6×10^{-10} m^2/s at 310 K reported in the diffusion NMR experiment)[53] is about tens of μs. Knowledge of the time scales for the monomers to transit between different metastable conformations is still required to shed light on the fibril nucleation mechanism.

The dynamics of the hIAPP monomer and other IDPs remain largely elusive both experimentally and theoretically. The free energy landscapes of IDPs are generally believed to contain numerous local minima but without a dominant global minimum.[43,54−60] Intuitively, one may speculate that transitions between local minima in a largely unstructured ensemble may be fast because of low-energy barriers. However, recent studies show that the dynamics for transitions between different unfolded states of a protein with a native structure can be surprisingly slow because of the enormity of conformational space.[61−65] Without a folded structure, it is still difficult to predict the time scales for the transitions between different states in IDPs. Several studies have applied straightforward MD simulations on microsecond time scales to study the dynamics of IDP monomers,[66,67] but it still remains challenging for these simulations to obtain a converged, statistical picture of the dynamics.

Here, we have constructed Markov state models (MSMs) from extensive all-atom MD simulations in explicit solvent (for a total of 70 μs of simulation) to investigate the dynamics of the hIAPP monomer. MSMs are kinetic network models that can predict long time scale dynamics from many short MD simulations. MSMs have been successfully applied to study protein folding and other conformational changes.[11,68−77] However, it has proven difficult to obtain statistically sound models of the dynamics of intrinsically disordered proteins, like the Aβ-peptide.[78] To overcome this limitation, we first ran extensive replica exchange molecular dynamics (REMD) simulations to obtain a converged picture of hIAPP's thermodynamics and then launched constant temperature simulations to infer the monomer's kinetics. Our model reveals a number of metastable states with populations of only a few percent or less rather than a single dominant free energy minimum. Surprisingly, the time scales for the transitions between these states are quite slow (ranging from several microseconds to milliseconds). Furthermore, we have found a few metastable states that share two common structural features: significant β-hairpin content and a large connected hydrophobic surface exposed to the solvent. We propose that these metastable states may serve as seeds for the nucleation of fibril formation.

METHODS

We constructed MSMs to study the thermodynamics and kinetics of the hIAPP monomer folding by the following procedure: We first performed two independent 200 ns REMD simulations to obtain initial sampling. These two simulations starting from totally different conformations are shown to converge by comparing the projection of the free energy landscape on the radius of gyration (R_g) and number of hydrogen bonds. Next, we selected representative conformations from REMD simulations to initiate around 1,400 20 ns or longer MD simulations (with an aggregation of 70 μs simulation time). Finally, we built MSMs from this set of extensive MD simulations.

REMD Simulations. REMD is an enhanced sampling algorithm that can help the system to escape kinetic traps in the free energy landscape by increasing the temperature.[79,80] In the REMD method, parallel replicas of the systems are simulated at different temperatures, and after a certain interval, the conformations at low temperatures can be swapped with those at high temperatures. The swapping probability is chosen to satisfy detailed balance. REMD simulations have been widely applied in enhancing the thermodynamic sampling for protein folding. However, despite several previous attempts,[81,82] it is still difficult to extract kinetic information at a particular temperature from REMD simulations due to the temperature swaps.

In this study, we have applied REMD simulations to obtain the initial thermodynamic sampling. Two independent REMD simulations were performed: one starting from a helical structure (extracted from a NMR structure for the hIAPP-micelle complex, PDB ID: 2KB8),[83] and the other one starting from a coil structure randomly selected from the first REMD simulation. The systems were equilibrated by energy minimization with steepest algorithm first and then a 200 ps position restraint in NVT ensemble. A 2 ns NPT simulation was also performed before the REMD production run in NVT ensemble. Each REMD simulation is 200 ns long with an exchange interval of 2 ps. We adopted the amber99sb force field[84] and the tip3p water model.[85] The simulation box contains 11 020 water molecules, and two Cl^- ions were added to make the system neutral. The electrostatic interactions were treated by the reaction field method[86] with a cutoff at 12 Å. We chose a dielectric constant of 78.5 for the solvent. The cutoff for the vdW interaction was set to be 11 Å, and the potential was smoothly switched off from 10 to 11 Å. The LINCS algorithm[87] was used to constrain all bonds. At each temperature, the system was coupled to a Nose Hoover thermostat[88,89] with a coupling frequency of 0.4 ps^{-1}.

Seeding MD Simulations. To select initial conformations for subsequent MD simulations, we first divided all the REMD conformations into 1200 clusters using the k-centers clustering

algorithm.[90] We then randomly chose two conformations from each cluster and initiated MD simulations from them. All the simulations were performed at the Folding@home distributed computing environment,[91] and the simulation length varied due to the heterogeneity of this platform. Finally, we have collected around 1400 simulations each of which is at least 20 ns long, giving 70 μs of simulation in total. All the seeding MD simulations were performed at 351 K with the same setup as described in the previous section except that the temperature coupling changed to velocity rescaling thermostat.[92]

MSM Construction. In a discrete-time and -state MSM, the conformational space is divided into a set of metastable states, and the fast protein motions are integrated out by coarse graining in time with a discrete unit of Δt. If Δt is longer than the relaxation time within each metastable state, the model is Markovian, i.e., the probability of a given state at time $t + \Delta t$ depends only on the state at time t. Under this condition, the long time scale dynamics can be modeled by a first-order master equation:

$$\mathbf{p}(n\Delta t) = \mathbf{T}^n(\Delta t)\mathbf{p}(0) \quad (1)$$

where $\mathbf{p}$ is the vector of state populations, $\mathbf{T}$ is the transition probability matrix, and Δt is the lag time of the model.

To construct MSMs, we have first applied the k-centers clustering algorithm[90] to divide all the MD conformations into 10 000 microstates (i.e., small clusters). We then constructed a 10 000 state MSM. To obtain the transition probability matrix ($\mathbf{T}$) for this MSM, i.e., probability of transitioning from states i to j after a certain lag time, we symmetrized and then normalized the transition count matrix by column. Elements of the transition count matrix (i.e., number of transitions between a pair of states at an interval of the lag time) were obtained directly from the MD simulations. This MSM was further validated and then used to compute all the quantitative properties reported in this study, such as equilibrium state populations and other kinetic properties. This microstate MSM is useful for obtaining quantitative properties but may contain too many states for human's appreciation. In order to better elucidate the folding mechanisms, we have further lumped the microstates that can interconvert quickly into the same metastable macrostate using the Perron cluster cluster algorithm (PCCA).[93] The MSM construction was performed using the MSMBuilder software.[94−96]

Validating MSMs. We first determine the lag time at which the model is Markovian[97] by examining the behavior of the implied time scales τ_k:

$$\tau_k = -\frac{\tau}{\ln \lambda_k} \quad (2)$$

where τ is lag time, λ_k is the k^{th} eigenvalue of the transition probability matrix $\mathbf{T}(\tau)$, and the Markovian time is when τ_k does not change with the lag time τ. If the model is Markovian at a certain lag time, the exponentiation of $\mathbf{T}(\tau)$ should be identical with MSMs built with longer lag times and the implied time scale plots will reach a plateau.

We further validated our MSMs using the residence probability test.[98,99] In this test, we compare the probability for the system to stay in a certain state (diagonal terms of the transition probability matrix) predicted by the propagation of the MSM with those directly obtained from MD trajectories.

Mean First Passage Time (MFPT). MFPT is defined as the average time taken from an initial state i to a final state f. We have followed the same procedure as in ref 95 to obtain the MFPT by solving the following set of equations:

$$\mathrm{MFPT}_{if} = \sum_j \mathbf{T}(t)_{ij}(t + \mathrm{MFPT})_{jf} \quad (3)$$

, where t is the lag time, and the boundary condition is: $\mathrm{MFPT}_{ff} = 0$.

RESULTS AND DISCUSSION

Model Validation. *Convergence of REMD simulation.* The two sets of REMD simulations are converged based on comparing projections of the free energy landscape for each onto R_g and number of hydrogen bonds within the protein. To obtain these projections, we used the weighted histogram analysis method (WHAM)[100] to combine data across all temperatures to obtain the thermodynamic properties at each temperature. As shown in Figure 1, even though the two sets of REMD simulations started from two totally different conformations, the ensemble average properties agree well with each other, indicating the simulations have reached certain convergence. The converged REMD data benefit the further seeding. We also observe that the R_g increases and the number of hydrogen-bond decreases with increasing temperature, indicating that the structure is more expanded and flexible at higher temperatures.

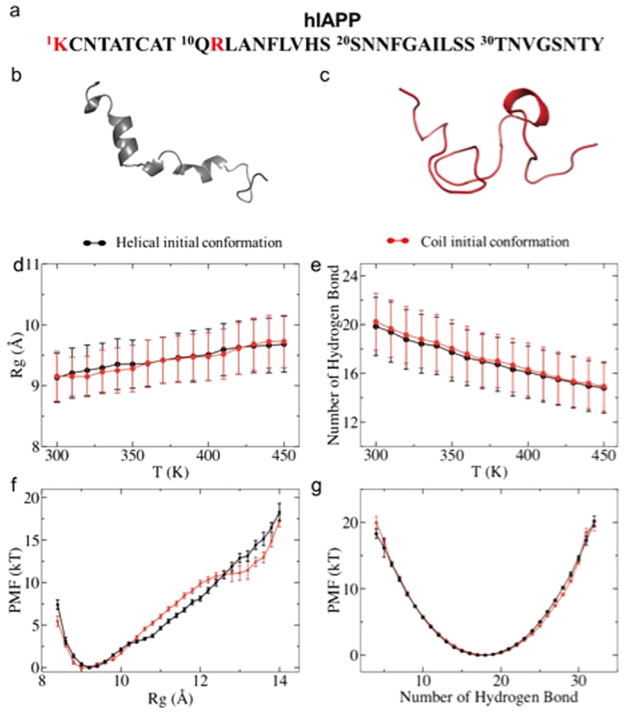


Figure 1. Convergence of REMD simulations. (a) The hIAPP sequence with positively charged residues shown in red. (b) One of two initial structures for REMD simulations, which comes from NMR studies in SDS micelle.[83] (c) The second initial structure used for REMD simulations, which is a random coil conformation selected form the first batch of REMD simulations. (d−g) Convergence check for the two sets of REMD simulations using WHAM. (d,e) Plots of the ensemble average R_g and the number of hydrogen bonds, respectively, as a function of temperature. (f,g) Projections of the potential of mean force onto the R_g and the number of hydrogen bonds, respectively, at 351 K. In each graph, the black curve comes from the first set of simulations initiated from the NMR structure in (b) and the red curve comes from simulations started from the random coil structure in (c). The two curves agree with each other well, indicating the simulations have converged. The increase in R_g and the decrease in the number of hydrogen bonds indicate the structure is more extended and flexible at higher temperatures.

Validation of MSM. We have first validated the 10 000 state MSM. As shown in Figure 2a, the implied time scale curves reach a plateau at around 5 ns, thus we choose this lag time to construct our MSMs. We further validated our model using the residence probability test.[98,99] The probabilities for the system to remain in a certain microstate predicted by the propagation of the MSM are compared with those directly obtained from

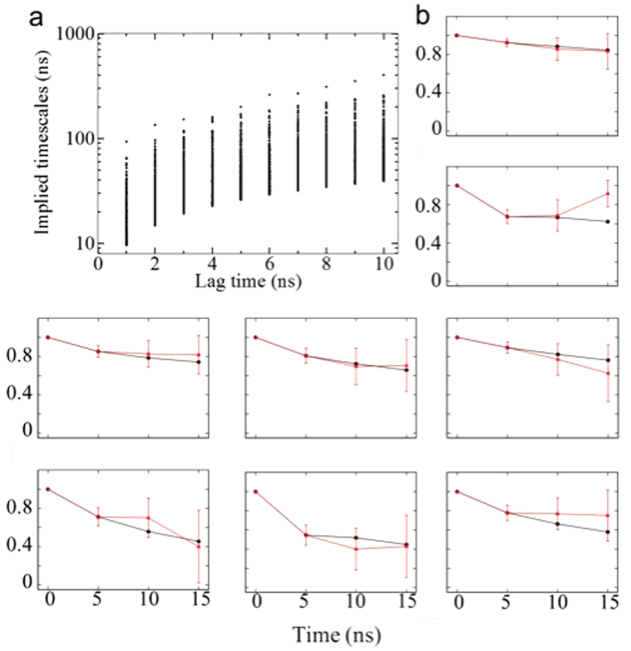


Figure 2. MSM validation. (a) Implied time scales plot as a function of the lag time. These plots are generated from the 10 000 state MSM. The implied time scales reach a plateau at 5 ns or longer, indicating the model is Markovian at these lag times. (b) 'Residence probability tests' for the eight most populated microstates, which show the probability to stay at a certain microstate as a function of the propagation time. The red curve comes directly from the MD simulations, while the black curve is calculated from a MSM with a lag time of 5 ns. The two curves agree with each other well, indicating the Markov model is a faithful representation of the original MD data.

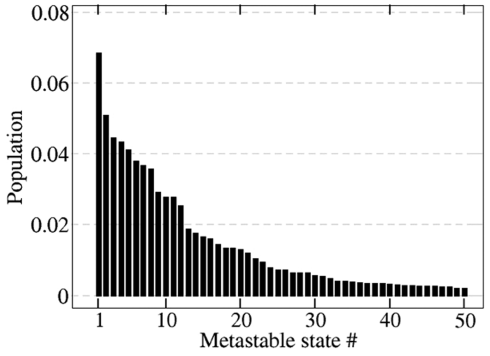


Figure 3. Equilibrium populations of the 50 most populated macrostates, from the most (6.9%) to the least (0.21%) populated.

MD simulations. As shown in Figure 2b, these two curves agree well with each other for various highly populated microstates, indicating that our model is able to reproduce the dynamics observed in MD simulations. We then employed this 10 000 state MSM to predict the long time scale dynamics for the folding of hIAPP.

The 10 000 state MSM discussed above is good for computing quantitative properties but contains too many states to clearly elucidate the features of the folding free energy landscape. We then further divided these 10 000 microstates into 200 macrostates by grouping microstates that can interconvert quickly into the same macrostate using the PCCA algorithm.[93] However, the residence probability test for this 200 macrostate MSM displayed a clear discrepancy

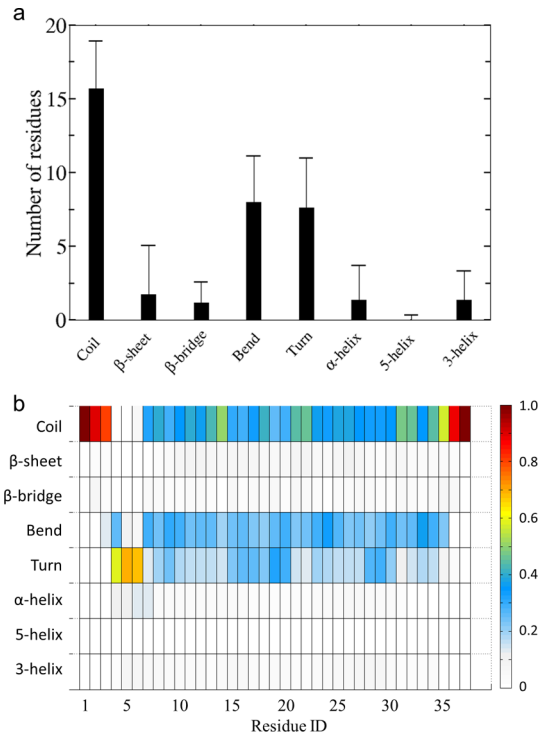


Figure 4. Secondary structure for hIAPP. (a) The average number of residues having each secondary structure type. The error bars represent the standard deviation. (b) Probability of each secondary structure type for each residue. The *x*-axis is the residue number, and the *y*-axis is the secondary structure type. The probability of each secondary structure type is color-coded according to the color bar shown on the right. The probability that each residue in a random coil structure is generally higher than 30%, except for residues 4−6, which tend to have turn and α-helical structures instead.

between the MSM prediction and MD simulation data (see SI Figure 1a). Further investigation revealed that this error of the MSM was caused by lumping small microstates (containing one or a few conformations) together with the large ones. By separating these small microstates and keeping them as independent states, the MSM showed great improvement in reproducing the dynamics observed in the original MD simulations (see SI Figure 1b−d and text). To elucidate the structure and dynamics of metastable states of hIAPP, we have constructed an MSM based on the 200 macrostate model but keeping microstates with populations <0.01% as independent states (see SI text for details).

We note that all the quantitative properties reported in this work are computed based on the well-validated 10 000 state MSM. For example, the population of each metastable macrostate is the sum of the equilibrium populations of the microstates it contains. The MFPTs are also computed based on the 10 000 state transition probability matrix by setting the initial and final states as a group of microstates.

The Free Energy Landscape Is Rugged. *Many Local Metastable States without a Global Minimum.* We have identified a number of metastable states in the folding free energy landscape of hIAPP, while none of these states dominate in population. As shown in Figure 3, the most populated metastable states all have an equilibrium population of only a few percent (with the largest one ~6.9%). Our model has thus confirmed the intrinsically disordered nature of the hIAPP peptide, where a global folded state is lacking. Furthermore, our

dx.doi.org/10.1021/ja403147m | J. Am. Chem. Soc. 2013, 135, 16092−16101

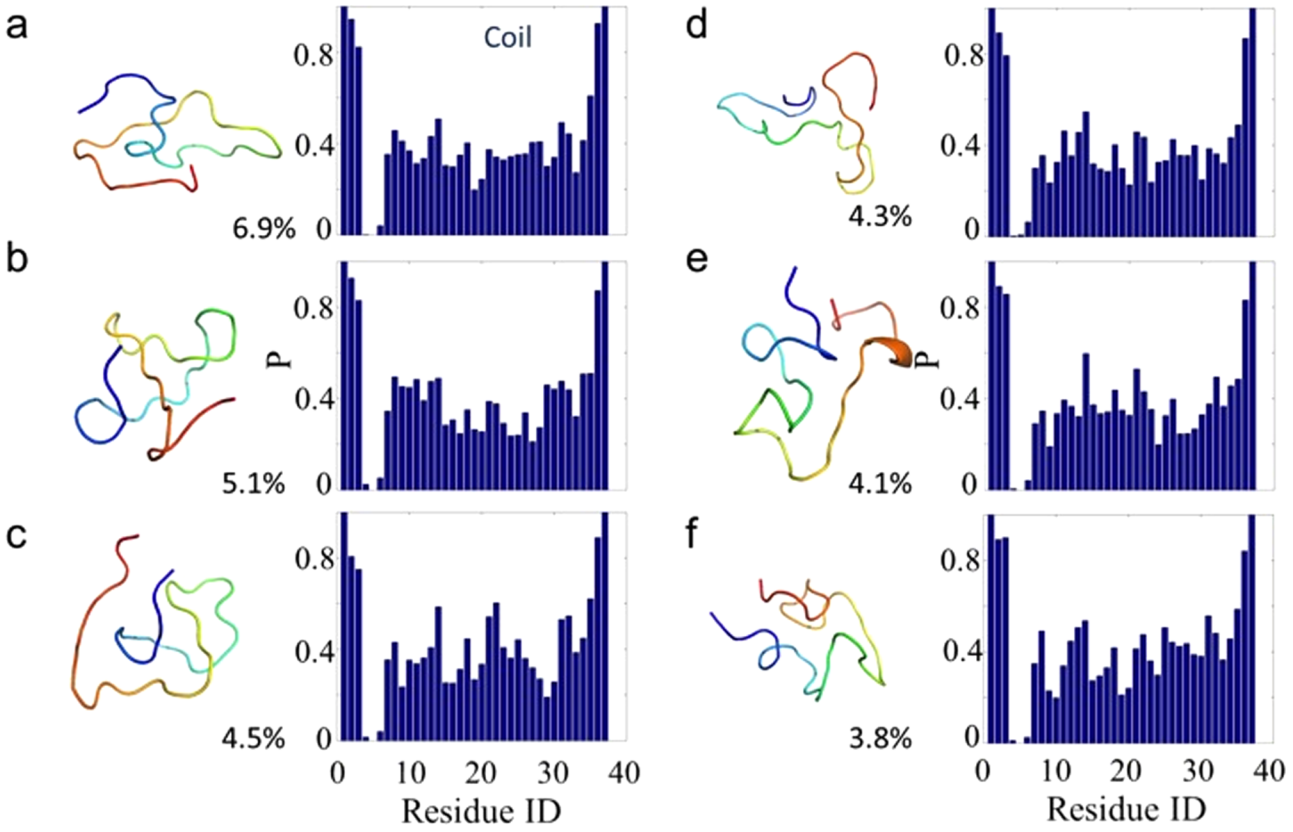


Figure 5. Representative structures and secondary structure preferences for the six most-populated metastable states. The protein structure is shown on the left in a cartoon representation with rainbow coloring (blue for the N-terminus and red for the C-terminus). The probability for each residue to adopt a random coil structure within its metastable state is shown on the right. Each conformation is also labeled with its equilibrium population. Most residues have a high probability of being random coil, except residues 4−6, which are mostly in turn and α-helical structures due to the disulfide bond between Cys2 and Cys7.

MSM has successfully elucidated the local free energy minima that are kinetically metastable. Alternatively, projection of the free energy landscapes on pairs of reaction coordinates has also been applied to provide structural information of local free energy minima for IDPs, but it is difficult to obtain kinetic information from thermodynamic simulations, such as REMD.[43,54,55,60] Using our MSM, we can also proceed to investigate kinetics as well as structural features of these metastable conformational states.

The Majority of Conformations Are Random Coil. To characterize the structural preferences of hIAPP, we first investigated the secondary structure distributions for the entire structural ensemble. As shown in Figure 4a, the majority of residues (16 out of 37) are in random coil conformations, which is consistent with the observations from CD experiments.[39] This result is also within expectation since IDPs generally lack a well-defined native state as well as distinguished secondary structures. Interestingly, there is some extent of turn and bend secondary structures (~8 residues each). Moreover, N-terminal residues 4−7 display a notable α-helical propensity (>10%, see Figure 4b), though the extent of this N-terminal α-helical segment may still be less than that observed in NMR experiments[40] and other MD simulations.[43] Some α-helix rich metastable states are also identified in our model (see SI Figure 4).

We further quantitatively compared the structural ensemble obtained from our MSM to two sets of experimental data: cross section from ion mobility mass spectrometry[42] and NMR chemical shift experiment.[40] The results show that the predictions from our structural ensemble can quantitatively reproduce the experimental observations on cross sections of the hIAPP peptide (see SI Figure 5) and residue-based chemical shift values (see SI Figure 6). The details of these comparisons are available in SI.

In addition to the overall structural ensemble, we have also investigated the structural features of the most populated metastable states. As shown in Figure 5 (right column), the majority of the residues in these states contain a significant fraction of random coil, especially residues at the N- and C-termini. Representative structures from these states also confirm the above observation and display random coil structures (see Figure 5, left column). However, these structures are not extended coil conformations but, instead, are rather compact. In particular, their R_g ranges from 8.3 to 13.9 Å, which is less than one-tenth of the length of an extended polypeptide chain with 37 residues. Previous fluorescence quenching experiments also indicated that hIAPP has a compact structure by estimating the end−end distance.[101] The hydrodynamic radius of hIAPP is 8.1 Å at 310 K as measured in diffusion NMR, two times smaller than that of a denatured protein with the same weight.[53] Compact coil structures have also been observed in other IDPs, such as poly-Q.[56]

Transitions between Different Metastable States Are Slow. To study the dynamics of hIAPP, we computed the MFPTs between the 50 most populated metastable states. As

dx.doi.org/10.1021/ja403147m | J. Am. Chem. Soc. 2013, 135, 16092−16101

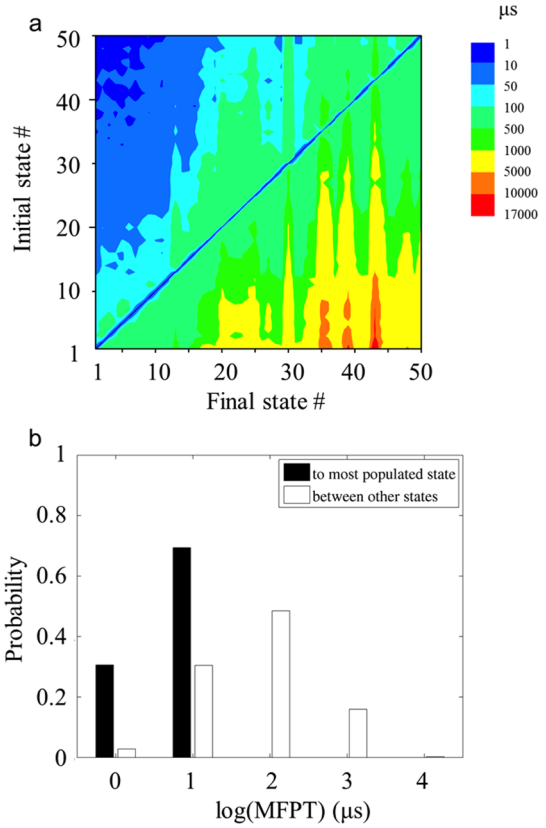


Figure 6. MFPTs between different states. (a) A contour plot of MFPTs between the 50 most populated states, with the initial state on the *y*-axis and the final state on the *x*-axis. The state indices are sorted by the state equilibrium population in descending order. The MFPTs range from several μs to tens of ms. (b) Distributions of MFPTs to the most populated state (black) and between all other pairs of states (white). See SI Figure 7 for the comparisons for the other large metastable states.

shown in Figure 6a, the time scales for these transitions span a wide range from several microseconds to milliseconds, indicating that different metastable free energy minima are well separated kinetically. These time scales are actually comparable or even slower than the folding time scales of typical peptides with similar length.[102] We speculate that transitions from one compact coil state to another in hIAPP are slow because they involve breaking existing contacts, reorganizing the conformation, and finally forming new contacts.

Unlike well-folded proteins, the most populated states of hIAPP do not appear to act as kinetic hubs. Interestingly, recent simulation studies and single molecule FRET experiments have shown that dynamics within the unfolded ensemble (or, more precisely, the non-native ensemble) may be slow compared to the time scale for folding.[62−65] The high kinetic accessibility of the native state is referred to as hub-like behavior because proteins are more likely to transition between different unfolded states by first folding and then unfolding rather than transitioning directly between the two unfolded states. While transitions from less well-populated states to more highly populated states are significantly faster than the reverse transitions, the most populated states in hIAPP do not appear to serve as kinetic hubs. As shown in Figure 6b and SI Figure 7, the distribution of MFPTs to the most populated states is indistinguishable from the distribution of MFPTs between less populated states.

Probable Aggregation Prone States: Flat β-Sheet Geometry with Extended Hydrophobic Surfaces. Interestingly, we have also identified a few metastable conformational states with populations of around 1% that contain a significant amount of β-sheet secondary structure. As shown in Figure 7 (left column), representative conformations from these states contain β-hairpin segments with around 10 residues or longer in length. In one of the states shown in Figure 7c, we even observed a single β-hairpin structure that almost spans the entire length of the peptide, which is consistent with the β-turn-β conformation found in the hIAPP fibril structure (see Figure 8). The β-hairpin segments can exist near both the N-terminus (see Figure 7a,b,d,f) and the C-terminus (see Figure 7e,f). Regardless of their position, all of these segments share one common striking structural feature: connected hydrophobic residues in proximity to the turn and exposed to the solvent. For example, in the conformational states shown in Figure 7a,d,f, five N-terminal hydrophobic residues (L12, A13, F15, L16, V17) form an extended hydrophobic surface that is exposed to the solvent. In the states shown in Figure 7a,b,f, an even larger hydrophobic cluster is observed near the C-terminus formed by residues: F23, G24, A25, I26, L27, V32, and G33. Additional MD simulations indicate that the above two distinct structural features of this set of metastable conformational states, β-hairpin segments and extended solvent-accessible hydrophobic surface areas, can be largely preserved under different force fields (see SI Figures 8 and 9 and text) and ionic strength (see SI Figures 10 and 11 and text).

The presence of β-hairpin segments and extended hydrophobic surfaces may make these metastable conformational states particularly aggregation prone. The extended hydrophobic surface can induce collapse of the hIAPP peptides through nonspecific hydrophobic interactions. These newly formed hydrophobic contacts are physically located in β-hairpin segments that are nearly flat and thus provide a template for further growth of the aggregate. We propose that both of these structural features are crucial for aggregation. Hydrophobic interactions may bring individual hIAPP peptides together, while the flat geometry facilitates further growth of the aggregates. Other conformational states, such as those shown in Figure 5, may also collapse into compact coils, but the subsequent conformational reorganization to a β-sheet geometry may require breaking and reforming many contacts and thus constitute a less favorable path for aggregation.

We further propose that conformational selection plays an important role in this aggregation mechanism. As discussed previously, the free energy landscape for the hIAPP monomer contains numerous metastable conformations (or local minima). When multiple hIAPPs collide to form an encounter complex, the aggregation prone states will be selectively stabilized via a conformational selection mechanism, as shown in Figure 8. The dynamics of our MSM are consistent with this picture as the rate of collision is similar to the time scale for reaching the aggregation prone states. In particular, the collision interval for hIAPP is estimated to be around tens of microseconds at a 1 μM concentration and a diffusion coefficient of 3.6×10^{-10} m^2/s.[53] Our MSM predicts that the time scales for transitioning to the aggregation prone states range from tens of microseconds to milliseconds (see Figure 6). Therefore, the system has sufficient time to reach various aggregation prone states between collision events, allowing conformational selection to occur. Our results cannot rule out

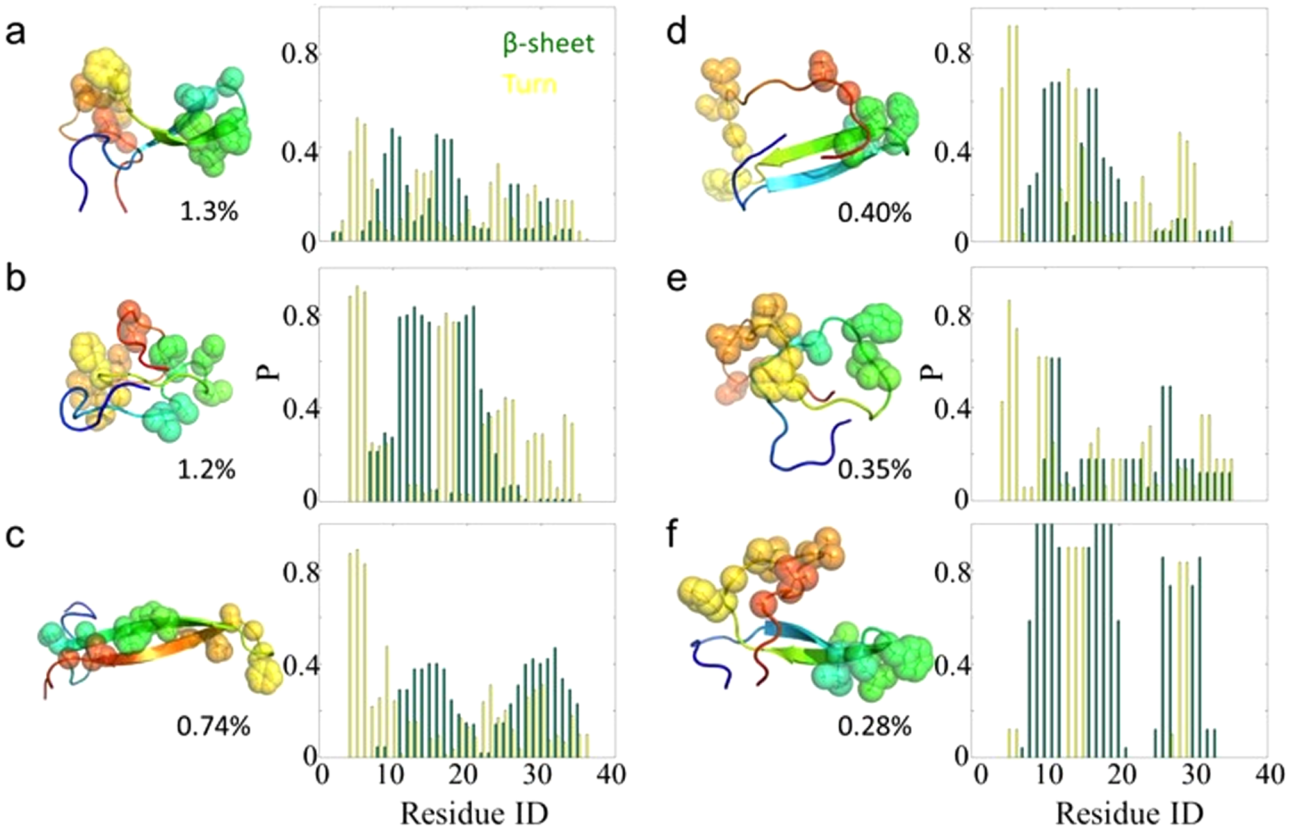


Figure 7. Representative structures and secondary structure analysis of the six metastable states whose β-sheet probabilities are >1 standard deviation above the mean. The protein is shown on the left in a cartoon representation with rainbow coloring (blue for the N-terminus and red for the C-terminus). The side chains of hydrophobic residues are shown as sphere with nonpolar hydrogen atoms hidden. The probability for each residue to adopt a β-sheet or turn structure within its metastable state is shown on the right. Adjacent hydrophobic side chains are connected to form clusters that are exposed to water. Each conformation is also labeled by its equilibrium population.

the possibility of a role for induced fit, however, we speculate that the conformational reorganization involved in such a mechanism may be associated with significant free energy barriers due to breaking and reforming many protein contacts. In the conformational selection mechanism, the stacking of aggregation prone states can already form a template similar to the fibril structure and thus requires a smaller free energy cost for fibrillation.

Our observations are consistent with previous simulation and experimental studies.[33,103] As in the N* states mechanisms proposed by Thirumalai, Straub, and co-workers,[46−48] our aggregation prone states are high free energy metastable states that consist of an ensemble of structures with lifetimes that are comparable or longer than the time scale for diffusion-driven encounter. Notably, in our model, the β-hairpin secondary structures in aggregation prone states can provide a perfect flat geometry for further growth of the aggregates. Earlier MD simulation studies[33,104] have shown that these relatively low-population β-hairpin structures in monomers can be greatly stabilized upon dimerization. Interestingly, we noticed that the β-sheet regions they reported[33,104] coincide with the hairpin motifs we observed in our aggregation prone states. These regions (residues 11−18 and 23−32) contain extensive hydrophobic residues and thus may promote aggregation through hydrophobic interactions. The role of the hydrophobic interactions in aggregation was also underlined by previous experimental studies.[105−107] For example, Doran et al.[105] showed that hydrophobicity is the essential factor in the aggregation of hIAPP (20−29) through their Phe23 mutants. More strikingly, another study showed that a single mutant I26P in the hydrophobic region (residue 23−27) can entirely inhibit fibril formation.[106] Moreover, Shim et al.[107] detected the order of aggregation for various residues using isotope labeling technique and 2DIR spectroscopy. They found that the hydrophobic regions in the middle of the chain (residues 15−17 and 23−27) aggregate first, followed by contact formation between other residues in the termini. In our proposed aggregation mechanism, conformational selection plays an important role. Recent 19F NMR experiments[90] support this conclusion because only fibril-like intermediate structures are found to exist during fibrillation. A role for the induced fit mechanism is not excluded in our model. A previous simulation study has also observed a transition from an α-helix or coil to a β-sheet during the dimerization of hIAPP, though the dominant pathway may still be conformational selection in their model.[33]

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

In this work, we constructed MSMs from extensive MD simulations of the hIAPP monomer. We demonstrated that the free energy landscape of the hIAPP monomer contains many local minima but no dominant global minimum. The transition time scales between different local minima range from several microseconds to milliseconds, which is comparable to or even slower than the folding time scales of proteins with a similar

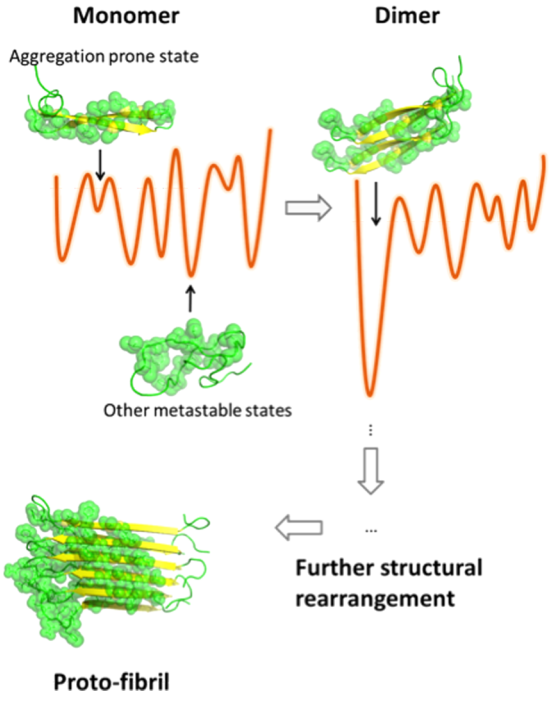


Figure 8. Schematic illustration of the hIAPP free energy landscape. The orange curves show the free energy landscape in one dimension. The β-hairpin structure is a minor minimum when hIAPP is a monomer and becomes the dominant minimum when more monomers come into contact. Further rearrangement is needed for fibril formation. We show a dimer for visual clarity, but the critical nucleus may contain additional peptide chains. Hydrophobic residues are shown in green spheres.

length. The structure of the hIAPP monomer is generally random coiled and occasionally has α-helix or β-sheet components. We have further identified a few potential aggregation prone states, all containing β-hairpin segments and extended hydrophobic surfaces that are exposed to the solvent. The hydrophobic surface can induce hydrophobic collapse, while the flat geometry may provide a template for further growth of the aggregate via a conformational selection mechanism.

Our model sheds light on the very initial stages of nucleation. In subsequent steps to form the fibrils, significant structural rearrangements are needed. For example, the β-hairpin is more extended in the fibril state (30 residues long) compared to our aggregation prone states (10−23 residues long). The intramolecular hydrogen bonds in the aggregation prone states also need to be converted to intermolecular hydrogen bonds in the fibril. Furthermore, additional interchain hydrophobic contacts must be formed in the fibril (with a corresponding reduction in intrachain hydrophobic contacts). Notably, these transitions from compact β-hairpins to extended β-strands have already been observed in the amyloid β-peptide in a recent MD study.[108] The NCC mechanism (or induced fit) may play an important role in these steps as suggested in previous simulation[28−30,108−110] as well as experimental studies.[21]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Details about model validation, quantitative comparisons between our structural ensemble and experiment (cross sections and NMR chemical shifts), and robustness tests of aggregation prone states at different force fields and under various ionic strengths. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

xuhuihuang@ust.hk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

X.H. acknowledges Hong Kong Research Grants Council (661011, SBI12SC01, HKUST6/CRF/10, and T13-607/12R), National Science Foundation of China (no. 21273188), and National Basic Research Program of China (973 program, 2013CB834703). Computing resources were provided by the National Supercomputing Center of China in Shenzhen and Folding@home distributed computing environment. GRB is supported by the Miller Institute.

■ REFERENCES

(1) Höppener, J. W. M.; Oosterwijk, C.; Nieuwenhuis, M. G.; Posthuma, G.; Thijssen, J. H. H.; Vroom, T. M.; Ahrén, B.; Lips, C. J. M. *Diabetologia* **1999**, *42*, 427.
(2) Lorenzo, A.; Razzaboni, B.; Weir, G. C.; Yankner, B. A. *Nature* **1994**, *368*, 756.
(3) Kahn, S. E.; Andrikopoulos, S.; Verchere, C. B. *Diabetes* **1999**, *48*, 241.
(4) Luca, S.; Yau, W.-M.; Leapman, R.; Tycko, R. *Biochemistry* **2007**, *46*, 13505.
(5) Bedrood, S.; Li, Y.; Isas, J. M.; Hegde, B. G.; Baxa, U.; Haworth, I. S.; Langen, R. *J. Biol. Chem.* **2011**, *287*, 5235.
(6) Haataja, L.; Gurlo, T.; Huang, C. J.; Butler, P. C. *Endocr. Rev.* **2008**, *29*, 303.
(7) Butterfield, S. M.; Lashuel, H. A. *Angew. Chem., Int. Ed.* **2010**, *49*, 5628.
(8) Meier, J. J.; Kayed, R.; Lin, C.-Y.; Gurlo, T.; Haataja, L.; Jayasinghe, S.; Langen, R.; Glabe, C. G.; Butler, P. C. *Am. J. Physiol.: Endocrinol. Metab.* **2006**, *291*, E1317.
(9) Mirzabekov, T. A.; Lin, M.-c.; Kagan, B. L. *J. Biol. Chem.* **1996**, *271*, 1988.
(10) Janson, J.; Ashley, R. H.; Harrison, D.; McIntyre, S.; Butler, P. C. *Diabetes* **1999**, *48*, 491.
(11) Anguiano, M.; Nowak, R. J.; Lansbury, P. T. *Biochemistry* **2002**, *41*, 11338.
(12) Jayasinghe, S. A.; Langen, R. *Biochemistry* **2005**, *44*, 12113.
(13) Knight, J. D.; Miranker, A. D. *J. Mol. Biol.* **2004**, *341*, 1175.
(14) Knight, J. D.; Hebda, J. A.; Miranker, A. D. *Biochemistry* **2006**, *45*, 9496.
(15) Hebda, J. A.; Miranker, A. D. *Annu. Rev. Biophys.* **2009**, *38*, 125.
(16) Fu, L.; Liu, J.; Yan, E. C. Y. *J. Am. Chem. Soc.* **2011**, *133*, 8094.
(17) Harper, J. D.; Lansbury, P. T. *Annu. Rev. Biochem.* **1997**, *66*, 385.
(18) Kirkitadze, M. D.; Condron, M. M.; Teplow, D. B. *J. Mol. Biol.* **2001**, *312*, 1103.
(19) Feige, M. J.; Groscurth, S.; Marcinowski, M.; Yew, Z. T.; Truffault, V.; Paci, E.; Kessler, H.; Buchner, J. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13373.
(20) Abedini, A.; Raleigh, D. P. *Phys. Biol.* **2009**, *6*, 015005.
(21) Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L. *Science* **2000**, *289*, 1317.
(22) Tsai, C.-J.; Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9970.
(23) Miller, Y.; Ma, B.; Nussinov, R. *Chem. Rev.* **2010**, *110*, 4820.
(24) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 111.

(25) Andrews, J. M.; Roberts, C. J. *J. Phys. Chem. B* **2007**, *111*, 7897.
(26) Xue, W.-F.; Homans, S. W.; Radford, S. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8926.
(27) Koshland, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98.
(28) Auer, S.; Meersman, F.; Dobson, C. M.; Vendruscolo, M. *PLoS Comput Biol* **2008**, *4*, e1000222.
(29) Nguyen, H. D.; Hall, C. K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16180.
(30) Li, M. S.; Klimov, D. K.; Straub, J. E.; Thirumalai, D. *J. Chem. Phys.* **2008**, *129*, 175101.
(31) Chong, S.-H.; Ham, S. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 7636.
(32) Chong, S.-H.; Ham, S. *Phys. Chem. Chem. Phys.* **2012**, *14*, 1573.
(33) Dupuis, N. F.; Wu, C.; Shea, J.-E.; Bowers, M. T. *J. Am. Chem. Soc.* **2011**, *133*, 7240.
(34) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G. *PLoS Comput Biol* **2007**, *3*, e173.
(35) Pellarin, R.; Caflisch, A. *J. Mol. Biol.* **2006**, *360*, 882.
(36) Pellarin, R.; Guarnera, E.; Caflisch, A. *J. Mol. Biol.* **2007**, *374*, 917.
(37) Bellesia, G.; Shea, J.-E. *J. Chem. Phys.* **2007**, *126*, 245104.
(38) Bellesia, G.; Shea, J.-E. *J. Chem. Phys.* **2009**, *130*, 145103.
(39) Higham, C. E.; Jaikaran, E. T. A. S.; Fraser, P. E.; Gross, M.; Clark, A. *FEBS Lett.* **2000**, *470*, 55.
(40) Yonemoto, I. T.; Kroon, G. J. A.; Dyson, H. J.; Balch, W. E.; Kelly, J. W. *Biochemistry* **2008**, *47*, 9900.
(41) Cort, J. R.; Liu, Z.; Lee, G. M.; Huggins, K. N. L.; Janes, S.; Prickett, K.; Andersen, N. H. *Protein Eng., Des. Sel.* **2009**, *22*, 497.
(42) Dupuis, N. F.; Wu, C.; Shea, J.-E.; Bowers, M. T. *J. Am. Chem. Soc.* **2009**, *131*, 18283.
(43) Laghaei, R.; Mousseau, N.; Wei, G. *J. Phys. Chem. B* **2010**, *114*, 7071.
(44) Reddy, A. S.; Wang, L.; Singh, S.; Ling, Y. L.; Buchanan, L.; Zanni, M. T.; Skinner, J. L.; Pablo, J. J. d. *Biophys. J.* **2010**, *99*, 2208.
(45) Xu, W.; Jiang, P.; Mu, Y. *J. Phys. Chem. B* **2009**, *113*, 7308.
(46) Massi, F.; Straub, J. E. *Proteins: Struct., Funct., Bioinf.* **2001**, *42*, 217.
(47) Thirumalai, D.; Klimov, D. K.; Dima, R. I. *Curr. Opin. Struct. Biol.* **2003**, *13*, 146.
(48) Straub, J. E.; Thirumalai, D. *Annu. Rev. Phys. Chem.* **2011**, *62*, 437.
(49) Tarus, B.; Straub, J. E.; Thirumalai, D. *J. Mol. Biol.* **2005**, *345*, 1141.
(50) Tarus, B.; Straub, J. E.; Thirumalai, D. *J. Am. Chem. Soc.* **2006**, *128*, 16159.
(51) Reddy, G.; Straub, J. E.; Thirumalai, D. *J. Phys. Chem. B* **2009**, *113*, 1162.
(52) Lapidus, L. J. *Mol. BioSyst.* **2013**, *9*, 29.
(53) Soong, R.; Brender, J. R.; Macdonald, P. M.; Ramamoorthy, A. *J. Am. Chem. Soc.* **2009**, *131*, 7079.
(54) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2007**, *368*, 1448.
(55) Baumketner, A.; Shea, J.-E. *J. Mol. Biol.* **2006**, *362*, 567.
(56) Crick, S. L.; Jayaraman, M.; Frieden, C.; Wetzel, R.; Pappu, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16764.
(57) Kayed, R.; Bernhagen, J. u.; Greenfield, N.; Sweimeh, K.; Brunner, H.; Voelter, W.; Kapurniotu, A. *J. Mol. Biol.* **1999**, *287*, 781.
(58) Jarvet, J.; Damberg, P.; Danielsson, J.; Johansson, I.; Eriksson, L. E. G.; Gräslund, A. *FEBS Lett.* **2003**, *555*, 371.
(59) Jarvet, J.; Damberg, P.; Bodell, K.; Göran Eriksson, L. E.; Gräslund, A. *J. Am. Chem. Soc.* **2000**, *122*, 4261.
(60) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2011**, *405*, 570.
(61) Singh, V. R.; Kopka, M.; Chen, Y.; Wedemeyer, W. J.; Lapidus, L. J. *Biochemistry* **2007**, *46*, 10046.
(62) Waldauer, S. A.; Bakajin, O.; Lapidus, L. J. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 13713.
(63) Bowman, G. R.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10890.
(64) Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 4702.
(65) Voelz, V. A.; Jäger, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; Weiss, S.; Pande, V. S. *J. Am. Chem. Soc.* **2012**, *134*, 12565.
(66) Gsponer, J.; Haberthür, U.; Caflisch, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5154.
(67) Han, W.; Wu, Y.-D. *J. Am. Chem. Soc.* **2005**, *127*, 15408.
(68) Vitalis, A.; Caflisch, A. *J. Chem. Theory Comput.* **2012**, *8*, 1108.
(69) Da, L.-T.; Wang, D.; Huang, X. *J. Am. Chem. Soc.* **2011**, *134*, 2399.
(70) Silva, D.-A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. *PLoS Comput. Biol.* **2011**, *7*, e1002054.
(71) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 4.
(72) Buchete, N. V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057.
(73) Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011.
(74) Huang, X.; Yao, Y.; Bowman, G. R.; Sun, J.; Guibas, L. J.; Carlsson, G.; Pande, V. S. *Pac. Symp. Biocomput.* **2010**, 228.
(75) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86.
(76) Morcos, F.; Chatterjee, S.; McClendon, C. L.; Brenner, P. R.; Lopez-Rendon, R.; Zintsmaster, J.; Ercsey-Ravasz, M.; Sweet, C. R.; Jacobson, M. P.; Peng, J. W.; Izaguirre, J. A. *PLoS Comput. Biol.* **2010**, *6*, e1001015.
(77) Zheng, W.; Andrec, M.; Gallicchio, E.; Levy, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15340.
(78) Lin, Y.-S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. *Biophys. J.* **2012**, *102*, 315.
(79) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140.
(80) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
(81) van der Spoel, D.; Seibert, M. M. *Phys. Rev. Lett.* **2006**, *96*, 238102.
(82) Chodera, J. D.; Swope, W. C.; Noe, F.; Prinz, J.-H.; Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2011**, *134*, 244107.
(83) Patil, S. M.; Xu, S.; Sheftic, S. R.; Alexandrescu, A. T. *J. Biol. Chem.* **2009**, *284*, 11982.
(84) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.
(85) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(86) Barker, J. A.; Watts, R. O. *Mol. Phys.* **1973**, *26*, 789.
(87) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.
(88) Nose, S. *Mol. Phys.* **1984**, *52*, 255.
(89) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(90) Gonzalez, T. F. *Theor. Comput. Sci.* **1985**, *38*, 293.
(91) Shirts, M.; Pande, V. S. *Science* **2000**, *290*, 1903.
(92) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(93) Deuflhard, P.; Huisinga, W.; Fischer, A.; Schütte, C. *Linear Algebra Appl.* **2000**, *315*, 39.
(94) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. *J. Chem. Theory Comput.* **2011**, *7*, 3412.
(95) Bowman, G. R.; Huang, X.; Pande, V. S. *Methods* **2009**, *49*, 197.
(96) Huang, X.; Bowman, G. R.; Bacallado, S.; Pande, V. S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19765.
(97) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571.
(98) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011.
(99) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schutte, C.; Noe, F. *J. Chem. Phys.* **2011**, *134*, 174105.
(100) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195.
(101) Vaiana, S. M.; Best, R. B.; Yau, W.-M.; Eaton, W. A.; Hofrichter, J. *Biophys. J.* **2009**, *97*, 2948.
(102) Ouyang, Z.; Liang, J. *Protein Sci.* **2008**, *17*, 1256.

(103) Suzuki, Y.; Brender, J. R.; Hartman, K.; Ramamoorthy, A.; Marsh, E. N. G. *Biochemistry* **2012**, *51*, 8154.
(104) Laghaei, R.; Mousseau, N.; Wei, G. *J. Phys. Chem. B* **2011**, *115*, 3146.
(105) Doran, T. M.; Kamens, A. J.; Byrnes, N. K.; Nilsson, B. L. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 1053.
(106) Abedini, A.; Meng, F.; Raleigh, D. P. *J. Am. Chem. Soc.* **2007**, *129*, 11300.
(107) Shim, S.-H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6614.
(108) Larini, L.; Shea, J.-E. *Biophys. J.* **2012**, *103*, 576.
(109) Ma, B.; Nussinov, R. *Biophys. J.* **2006**, *90*, 3365.
(110) Wu, C.; Shea, J.-E. *Curr. Opin. Struct. Biol.* **2011**, *21*, 209.