THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B XXXX, XXX, XXX−XXX

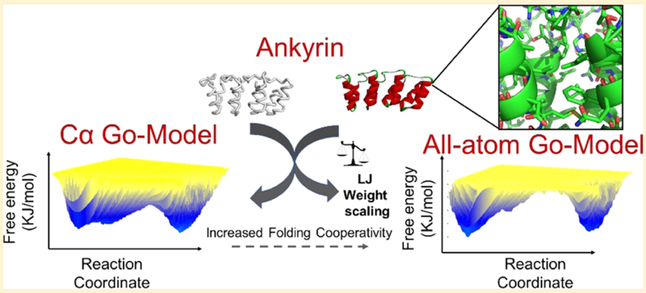
pubs.acs.org/JPCB

Pushing the Limits of Structure-Based Models: Prediction of Nonglobular Protein Folding and Fibrils Formation with Go-Model Simulations

Lokesh Baweja and Julien Roche*

Department of Biochemistry, Molecular Biology and Biophysics, Iowa State University, Ames, Iowa 50011, United States

Supporting Information



ABSTRACT: The development of computational efficient models is essential to obtain a detailed characterization of the mechanisms underlying the folding of proteins and the formation of amyloid fibrils. Structure-based computational models (Go-model) with Cα or all-atom resolutions have been able to successfully delineate the mechanisms of folding of several globular proteins and offer an interesting alternative to computationally intensive simulations with explicit solvent description. Here, we explore the limits of Go-model predictions by analyzing the folding of the nonglobular repeat domain proteins Notch Ankyrin and p16^{INK4} and the formation of human islet amyloid polypeptide (hIAPP) fibrils. Folding trajectories of the repeat domain proteins revealed that an all-atom resolution is required to capture the folding pathways and cooperativity reported in experimental studies. The all-atom Go-model was also successful in predicting the free-energy landscape of hIAPP fibrillation, suggesting a "dock and lock" mechanism of fibril elongation. We finally explored how mutations can affect the co-assembly of hIAPP fibrils by simulating a heterogeneous system composed of wild-type and mutated hIAPP peptides. Overall, this study shows that all-atom Go-model-based simulations have the potential of discerning the effects of mutations and post-translational modifications in protein folding and association and may help in resolving the dichotomy between experimental and theoretical studies on protein folding and amyloid fibrillation.

INTRODUCTION

A fundamental understanding of protein folding is critical to predict the potential effects of mutations and post-translational modifications on protein structure and stability.[1] Despite decades of intense research, the question of whether the folding of proteins occurs through multiple routes or through a single dominant pathway remains yet the subject of intense debates.[2−4] Experimental folding studies suggest that most proteins fold along a single dominant pathway[5,6] in contrast to theoretical studies or computational simulations, which often show the presence of parallel folding pathways.[7,8] From a theoretical point of view, protein folding is regarded as the process by which a polymeric chain diffuses on a multidimensional free-energy landscape to find and occupy a minimum energy state. The energy landscape theory posits that evolution has selected protein sequences for which the energy gap between the native state minimum and the ensemble of non-native minima is maximized.[9] This results in minimally frustrated, funnel-shaped free-energy landscapes that favor the rapid and efficient diffusion of proteins into their folded states.[10−12] This theory led to the development of coarse-grained computational models, called Go-models, that allow the simulation of proteins on an ideal, perfectly unfrustrated, free-energy landscape by attributing an attractive potential to native contacts and a simple sphere repulsion term to any non-native contacts.[13−15] Go-models are the so-called structure-based models because they require a reference structure to establish the list of native contacts. Previous studies using Go-models with Cα resolution have provided insights into the folding mechanisms of two-state folders[16] but are known to fail in some instances, particularly for proteins that exhibit high degree of symmetry,[17] a problem that can potentially be resolved with all-atom Go-model simulations.[18−23] The phenomenon of amyloid fibrillation related to diseases such as Alzheimer's and type 2 diabetes[24] is another challenging process for computational simulations. Understanding the mechanisms of fibril formation is yet crucial for the design of small molecules and peptide-based inhibitors.[25,26] Conventional all-atom molecular dynamics simulations of amyloid fibrils can potentially provide atomistic level information but are highly computational-intensive, especially for systems with a significant number of peptides.[27]

In the present study, we used Go-model simulations to characterize the folding mechanism of the repeat domain proteins Notch Ankyrin (Notch AR)[28] and p16^{INK4}[29] and the fibrillation landscape of the human islet amyloid polypeptide

Received: December 8, 2017
Revised: February 3, 2018
Published: February 9, 2018

DOI: 10.1021/acs.jpcb.7b12129

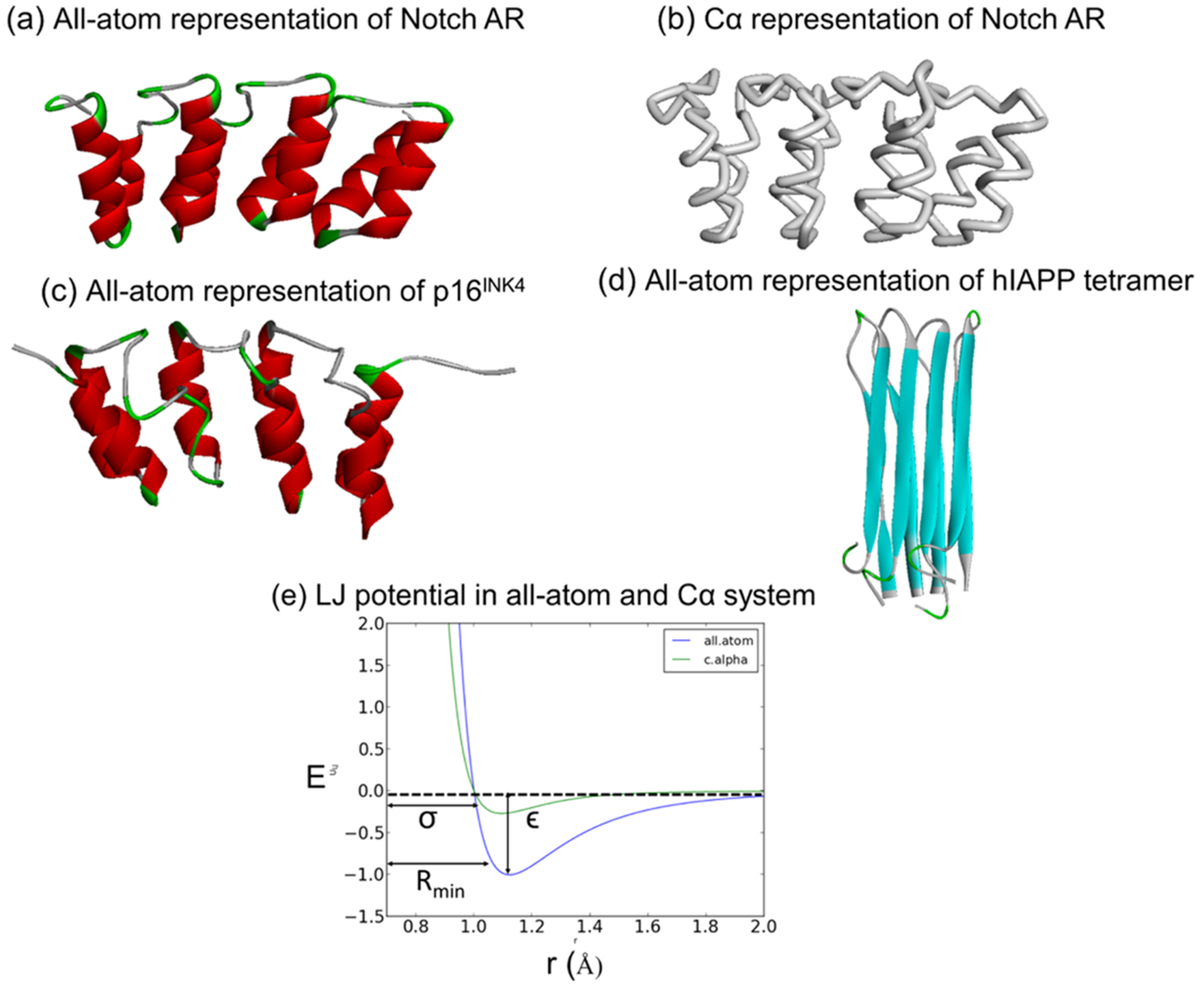

(a) All-atom representation of Notch AR
(b) Cα representation of Notch AR
(c) All-atom representation of p16INK4
(d) All-atom representation of hIAPP tetramer
(e) LJ potential in all-atom and Cα system
all.atom
c.alpha
E
σ
ϵ
Rmin
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.8
1.0
1.2
1.4
1.6
1.8
2.0
r (Å)

Figure 1. Reference structures used in this study with four repeats of Notch AR (PDB ID: 1OT8) in (a) all-atom and (b) Cα representation, (c) four repeats of p16$^{\mathrm{INK4}}$ (PDB ID: 1BI7) in all-atom representation, and (d) a tetramer extracted from the reference structure of the hIAPP fibril.[35] (e) Lennard-Jones (LJ) potential with respect to distance is presented for the all-atom and Cα models. The potential function is described in eqs 1 and 2.

(hIAPP), a disordered peptide whose fibrillation is linked to type 2 diabetes.[30] We chose these model proteins to test the limit of Go-model predictions and ask whether structure-based models can accurately reproduce the association and/or folding of nonglobular, non-two-state systems. We first addressed this question by comparing the folding trajectories of Notch AR obtained with the different level of coarse graining, i.e., Cα and all-atom resolutions. We showed that compared to the simple Cα model, the introduction of all-atom resolution increases the folding cooperativity of Notch AR and offers a significantly better agreement with the experimental studies. We also compared the folding trajectories obtained for Notch AR to that of p16$^{\mathrm{INK4}}$, a repeat protein with a very similar topology, and observed that the all-atom Go-model is able to distinguish the differences in the folding mechanisms of these two repeat domain proteins. When applied to the problem of fibril formation, all-atom Go-model simulations revealed that partially folded dimers were the smallest nucleus formed along the fibrillation pathway of hIAPP and suggest a general "lock and dock" mechanism similar to that of the amyloid β fibrils. We finally took advantage of the all-atom resolution offered by our Go-model to design an in silico mutant of the hIAPP peptide and study the mechanisms of association in the context of a heterogeneous system composed of wild-type and mutated peptides. On the basis of these findings, we concluded that all-atom Go-models have a high potential for predicting the effect of mutations and post-translational modifications on protein folding and fibril formation and may help in developing efficient peptide-based inhibitors for targeting pathogenic amyloid assembly.

■ MATERIALS AND METHODS

We used PDB IDs 1OT8,[31] 1BI7,[32] and 2A5E[33] (the latter being the NMR structure of p16$^{\mathrm{INK4}}$ used for comparison purpose) as reference structures for Notch AR and p16$^{\mathrm{INK4}}$, from which the coordinates of 2, 3, or 4 repeat domains were extracted for the purpose of the simulations. The missing residues in the PDB file were replaced using PyMOL, followed by energy minimization.[34] The coordinates of the hIAPP amyloid fibril were obtained from the solid-state NMR structure reported by Luca et al.[35] The lists of native contacts were determined for each reference structure using the Shadow Contact Map-SMOG algorithm,[36,37] and only heavy atoms were considered for establishing native contacts. The following energy functions were used in the Cα and all-atom models:

For Cα Representation.

$$E = \sum_{\text{bonds}} K_{\text{r}}(r - r_0)^2 + \sum_{\text{angles}} K_{\theta}(\theta - \theta_0)^2 + \sum_{\text{dihedrals}}^{n=1,3} K_{\varphi}(1 - \cos(n(\varphi - \varphi_0))) + \sum_{i<j-3}^{\text{native}} \varepsilon_1 \left[5\left(\frac{r'_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{r'_{ij}}{r_{ij}}\right)^{10} \right] + \sum_{i<j-3}^{\text{non-native}} \varepsilon_2 \left(\frac{r_{\text{rep}}}{r_{ij}}\right)^{12} \tag{1}$$

where r_{ij} is the distance between atom i and j and r'_{ij} represents the distance between i and j at which the interaction energy is a minimum (Figure 1e). The instantaneous bond lengths, bond angles, and dihedral angles are given by r, θ, and φ, respectively, and r_0, θ_0, and φ_0 are the corresponding values in the reference PDB structure. The ratios between interactions parameters for this model are: $K_{\text{r}} = 100\varepsilon$, $K_{\theta} = 20\varepsilon$, and $K_{\varphi} = \varepsilon_1 = \varepsilon_2$.

For All-Atom Representation.

$$E = \sum_{\text{bonds}} K_{\text{r}}(r - r_0)^2 + \sum_{\text{angles}} K_{\theta}(\theta - \theta_0)^2 + \sum_{\text{impropers/planar}} K_{\chi}(\chi - \chi_0)^2 + \sum_{\text{backbone}} \varepsilon_{\text{BB}} F_{\text{D}}(\varphi) + \sum_{\text{sidechains}} \varepsilon_{\text{sc}} F_{\text{D}}(\varphi) \sum_{i<j-3}^{\text{native}} \varepsilon_1 \left[\left(\frac{r'_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r'_{ij}}{r_{ij}}\right)^{6} \right] + \sum_{i<j-3}^{\text{non-native}} \varepsilon_2 \left(\frac{r_{\text{rep}}}{r_{ij}}\right)^{12} \tag{2}$$

where

$$F_{\text{D}}(\varphi) = [1 - \cos(\varphi - \varphi_0)] + \frac{1}{2}[1 - \cos(3(\varphi - \varphi_0))]$$

The dihedral interaction weights for the backbone and side chains are given by ε_{BB} and ε_{sc}, respectively. The ratios between interactions parameters for this model are $K_{\text{r}} = 50\varepsilon$, $K_{\theta} = 40\varepsilon$, and $K_{\chi} = 10\varepsilon$

Renormalization of Native Contact Weights. For all-atom models, SMOG normalizes the weight of native contacts ε_1, ε_{BB}, and ε_{SC} by the ratio of the number of atoms/number of native contacts in the systems such that the total stabilizing energy is roughly the same between molecular systems of different sizes.[36,37] However, because of their specific elongated architecture, the number of native contacts found in repeat proteins does not scale linearly with the total number of atoms as the number of repeats increases. For example, 488 native contacts are formed by the 492 atoms of Notch AR with two repeats, but Notch AR with four repeats form 1232 native contacts for 981 atoms resulting in low value of normalizing factor. We noted that the normalization conditions imposed by SMOG for all-atom models tend to excessively decrease the weight of native contacts for longer repeat proteins or fibrils compared to the shorter constructs. We therefore rescaled the native contact weights of our all-atom model for the constructs with three and four repeat domains based on the SMOG output obtained for our shorter construct (with two repeats). For example, the native contact (ε_1) and dihedral (ε_{BB} and ε_{SC}) weights of Notch AR with four repeats were uniformly renormalized by factors of 1.27 and 1.02, respectively, to match the weights obtained for Notch AR with two repeats. This renormalization of the all-atom models allowed us to monitor the folding transitions of the longer constructs on a relevant time scale, which was otherwise not possible under the initial normalization conditions obtained from SMOG (see Figure S1). Reversing the weight normalization imposed by SMOG is potentially useful for any system in which the number of atoms is much lower than the number of contacts and helps in deriving the folding trajectories in reasonable computational time. The same renormalization methodology was applied for hIAPP. Lists of native contacts for the hIAPP dimer, trimer, and tetramer were generated in a manner that individual peptides may swap their position during the simulations, which is accomplished by swapping coordinates of the peptides when establishing the list of native contacts to account for all combinatorial possibilities. Repeated intramolecular and intermolecular interactions were deleted to prevent redundancy.

Details of Simulations. All simulations were performed using GROMACS 5.0.7[38] program with the leapfrog stochastic dynamic integrator and a time step of 0.0005. For each system, independent simulations were performed over a wide range of temperatures and the folding temperature (T_{f}) was determined using the weighted histogram analysis method (WHAM).[39] In the case of Notch AR and p16$^{\text{INK4}}$, T_{f} represents the temperature at which both the unfolded and folded ensembles are equally populated. We define T_{a} in the case of the hIAPP system as the temperature at which both dissociation and association events can be equally observed.

Analysis. The fraction of native contacts (Q) was used as a reaction coordinate to monitor the folding of Notch AR and p16$^{\text{INK4}}$ and hIAPP assembly. A contact is considered formed when the distance between two atoms, r_{ij}, is less than 1.2 σ_{ij}. Free-energy landscapes were obtained using root-mean square deviation (RMSD) and Q as two coordinates using g_sham module in the GROMACS program.[38] The heat capacity (C_{v}) is derived as the partial derivative of the total energy E with respect to temperature

$$C_{\text{v}}(T) = \frac{\partial \langle E \rangle}{\partial T} = \frac{1}{k_{\text{B}}T^2}\langle (\Delta E)^2 \rangle$$

where k_{B} is the Boltzmann constant and T is the temperature in reduced units. The resulting graphs were further processed using in-house python scripts. The intermediate conformations were analyzed using cluster algorithm as implemented in g_cluster module of GROMACS,[38] using a cutoff of 1.35 nm, which was selected to find the top three clusters representing the majority of conformations (ca. 90−95%).[40] This seemingly high cutoff is due to the partially folded nature of the conformational ensembles selected for Notch AR and hIAPP from the free-energy landscape analysis.

DOI: 10.1021/acs.jpcb.7b12129

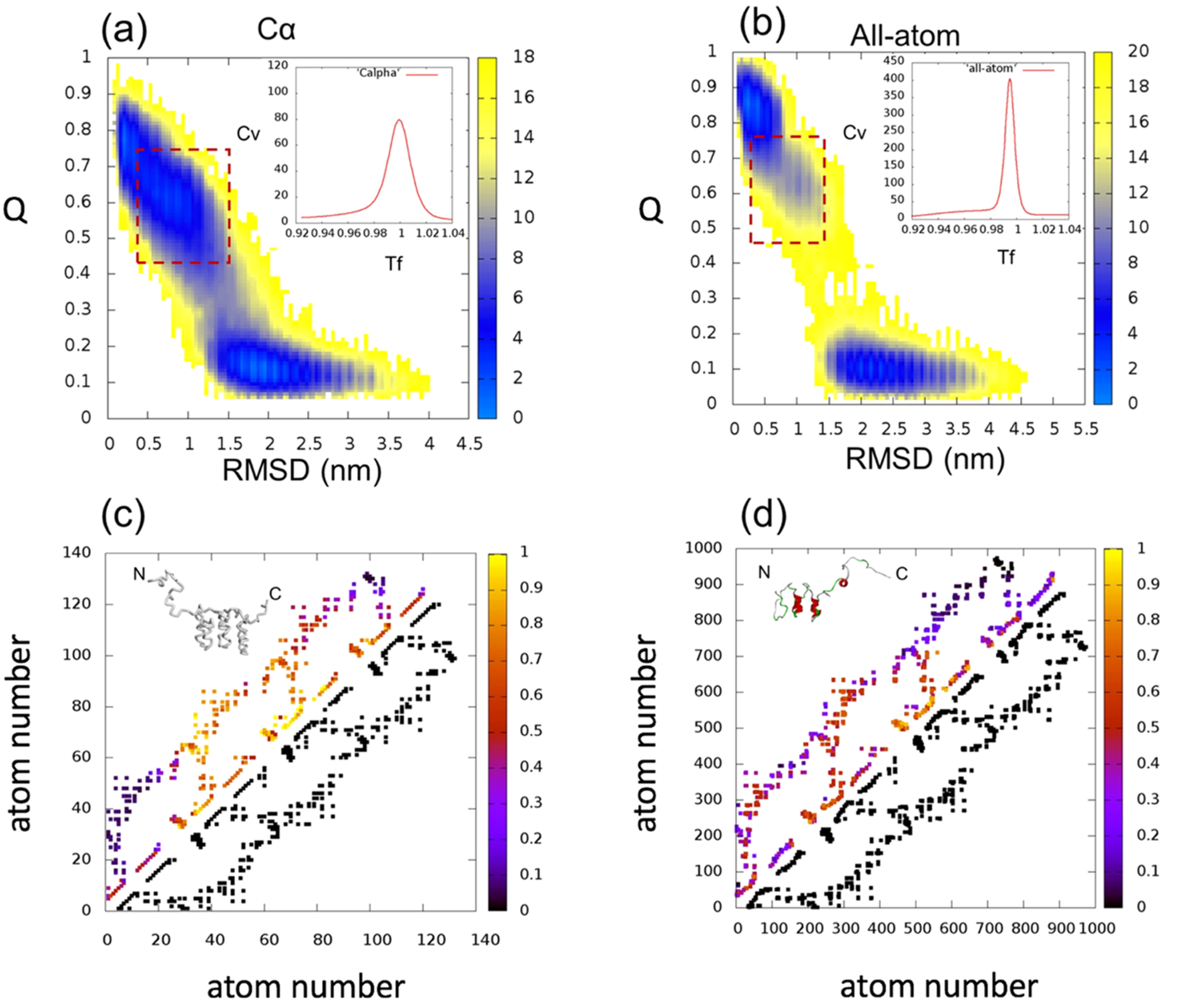


Figure 2. Folding free-energy profiles at T_f as a function of native fraction (Q) and root-mean-square deviation (RMSD) calculated from the Go-model simulations of Notch AR with Cα (a) and all-atom representations (b). The scale bar represents the free-energy scale in kJ/mol. The inset shows the respective change of specific heat over a wide range of temperatures around T_f. Contact map analysis of intermediate states populations identified from (a) and (b) (Q = 0.45−0.75 and RMSD = 0.5−1.5 nm, highlighted with red dots in the free-energy maps) for Notch AR in Cα (c) and all-atom representations (d). The color code in the upper half of the map corresponds to the probability, ranging from 0 to 1, to find each atom in its native contact conformation, whereas the lower half of the map represents all of the native contacts present in the native state. The insets show the central conformation of the dominant cluster (Figure S2).

■ RESULTS

Folding Behavior of Notch AR in All-Atom and Cα Representation. We first explored the effect of coarse graining on the cooperativity of folding by simulating Notch AR with four domain repeats in Cα and all-atom representations (Figure 1a,b). The folding temperature (T_f) was independently determined for both systems using the WHAM[39] algorithm, and both systems were then simulated at $T = T_f$ for 10^9 time steps. The change in specific heat over the wide range of temperatures sampled shows a sharper folding transition, indicative of a more cooperative folding mechanism for the all-atom representation compared to the Cα-model (insets in Figure 2a,b). Two-dimensional free-energy landscape analysis, based on the fraction of native contacts (Q) and the root-mean-square deviation (RMSD) from the reference PDB structure, revealed a clear difference in the folding cooperativity between the Cα and all-atom Go-model simulations (Figure 2a,b), suggesting that the incorporation of side-chain information in all-atom Go-model increases the cooperativity of Notch AR folding. The results from these all-atom Go-model simulations are therefore in closer agreement with the experimental studies, which showed that Notch AR folding kinetics can be fitted to the classical two-state mechanism.[41−43] In contrast, the free-energy landscape obtained from the Cα Go-model simulations is characteristic of a non-cooperative folding process with significant population of intermediate states (Figure 2a). The main intermediate population appears within the region defined by Q = 0.45−0.75 and RMSD = 0.5−1.5 nm (Figure 2a). This basin of intermediate states conformations corresponds to ~35% of the frames of the Cα system trajectories and ~9% of the all-atom simulation frames. A contact map analysis revealed that these intermediate states

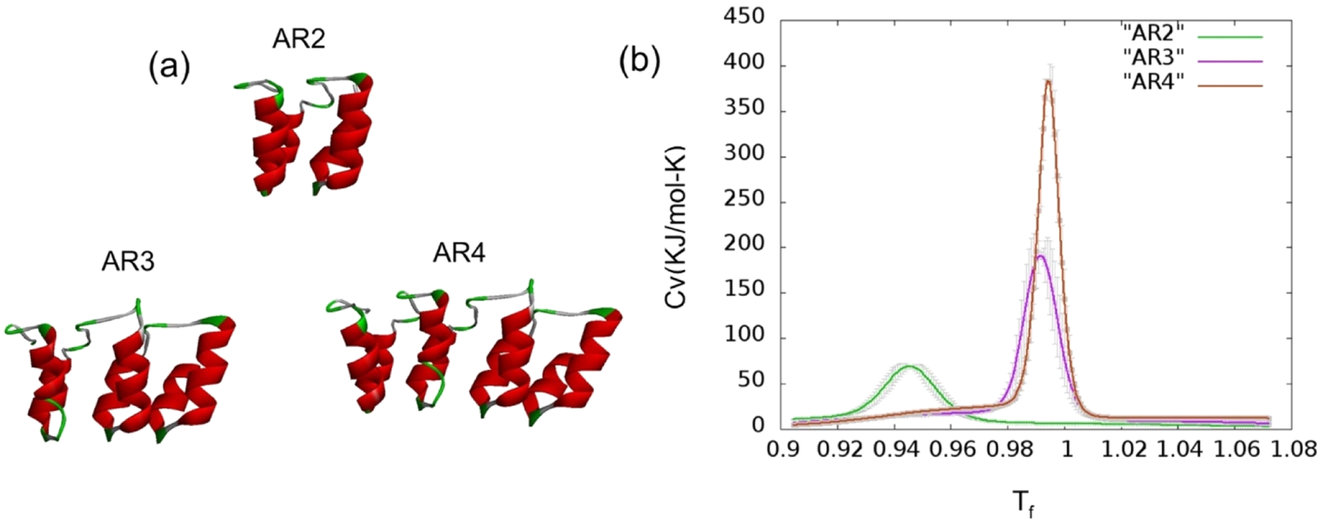


Figure 3. (a) All-atom representations of AR2, AR3, and AR4 with two, three, and four Notch ankyrin repeats, respectively. (b) Specific heat capacity of folding calculated from independent simulations of AR2 (green), AR3 (magenta), and AR4 (brown). The error bars represent variations in C_v values estimated by dividing the trajectories into two equal parts.

are predominately populated by conformations with folded central repeats, in both all-atom and Cα representations (Figure 2c,d).

Thermodynamics Characterization of Notch AR Folding Cooperativity. To explore if nearest-neighbor interactions are responsible for the observed folding cooperativity of Notch AR in all-atom representation, we compared the folding trajectories obtained from Go-model simulations of two (AR2), three (AR3), and four (AR4) ankyrin repeats (Figure 3a). Constant-temperature runs were carried out at over a wide range of temperature for each construct, and the specific heat capacity values were independently calculated using the WHAM algorithm.[39] The change of specific heat over a wide range of temperatures showed that cooperativity and the stability of Notch AR globally increase with the number of repeats, with a drastic difference between AR2 and AR3 but a much subtler change from AR3 to AR4 (Figure 3b). This analysis suggests that two domain repeats are sufficient to obtain a simple two-state folding behavior, but the three central repeats present in AR3 are required to reproduce the folding cooperativity of the longer Notch AR constructs.

Folding Routes of Notch AR and p16^{INK4}. To test the ability of the all-atom model to distinguish the folding routes of repeat proteins with similar topologies, we compared the results obtained for Notch AR to those obtained for p16^{INK4}. Although these two proteins have very similar structures and topologies (Figure 1a,c), they have been shown experimentally to fold through distinct folding pathways.[28] The folding of Notch AR is primarily initiated from the central repeats, with a slight polarization toward the N-terminus,[41] whereas p16^{INK4} is known to fold primarily from the C-term repeats.[44] To gain detailed insights into the folding mechanism of these two model proteins, we analyzed the free-energy landscape in terms of interfacial contacts between the repeat units. In the case of protein with four domain repeats (A, B, C, and D), we defined the three interfaces present between repeats as AB, BC, and CD (Figure 4a).

In line with the contact map analysis presented above (Figure 2c), the interface free-energy surface of Notch AR showed a slight preference for the BC interface compared to the AB interface but a clear difference in stability between the BC and CD interface (Figure 4b,c). This analysis confirms again that the folding nucleus of Notch AR lies in the central repeats with slight polarization toward the N-term unit. In contrast, the folding intermediate state identified for p16^{INK4} shows a much stable BC interface compared to AB and a slight preference for CD over BC interface (Figure 4d,e), suggesting that the folding of p16^{INK4} is initiated from the central and C-term units. Overall, these results demonstrate that the simple all-atom Go-model used here is able to distinguish the folding mechanisms of two proteins with very similar topologies.

Free-Energy Landscape of Fibril Formation. The concept of free-energy landscape established in the field of protein folding can also be applied to understand the mechanisms of amyloid fibrillation.[45] We used our renormalized all-atom Go-model to understand the fibrillation process of the human islet amyloid polypeptide (hIAPP) using the coordinates determined by solid-state NMR as reference structure for the determination of native contacts[35] (Figure 1d). We ran independent simulations with two, three, and four hIAPP peptides over a wide range of temperature and determined T_a as the temperature at which we observed an equivalent number of association and dissociation events. The two-dimensional free-energy landscape analysis performed at $T = T_a$ revealed the presence of intermediate states for all three systems studied, i.e., dimer (Figure 5a), trimer (Figure 5b), and tetramer (Figure 5c). Interestingly, the tetrameric system exhibited two distinct basins of intermediate states in the regions $0.45 < Q < 0.60$ and $1.8 < \text{RMSD} < 3$ (intermediate I) and $0.6 < Q < 0.75$ and $0.75 < \text{RMSD} < 1.8$ (intermediate II), corresponding to ∼6 and ∼5% frames of the trajectories (Figure 5c). A cluster analysis on these basins showed that intermediate I is primarily populated with conformations having fully formed dimer with docked and unfolded peptides, whereas the most dominant conformation in intermediate II corresponds to the trimer of hIAPP with one unfolded peptide (inset of Figure 5c).

Understanding Co-Assembly in Amyloid Formation. In the complex in vivo environment, amyloid peptides and proteins can carry mutations and post-translational modifications that may lead to cross-seeding and co-assembly of fibrils.[46] It has been well documented that mutations in amyloid

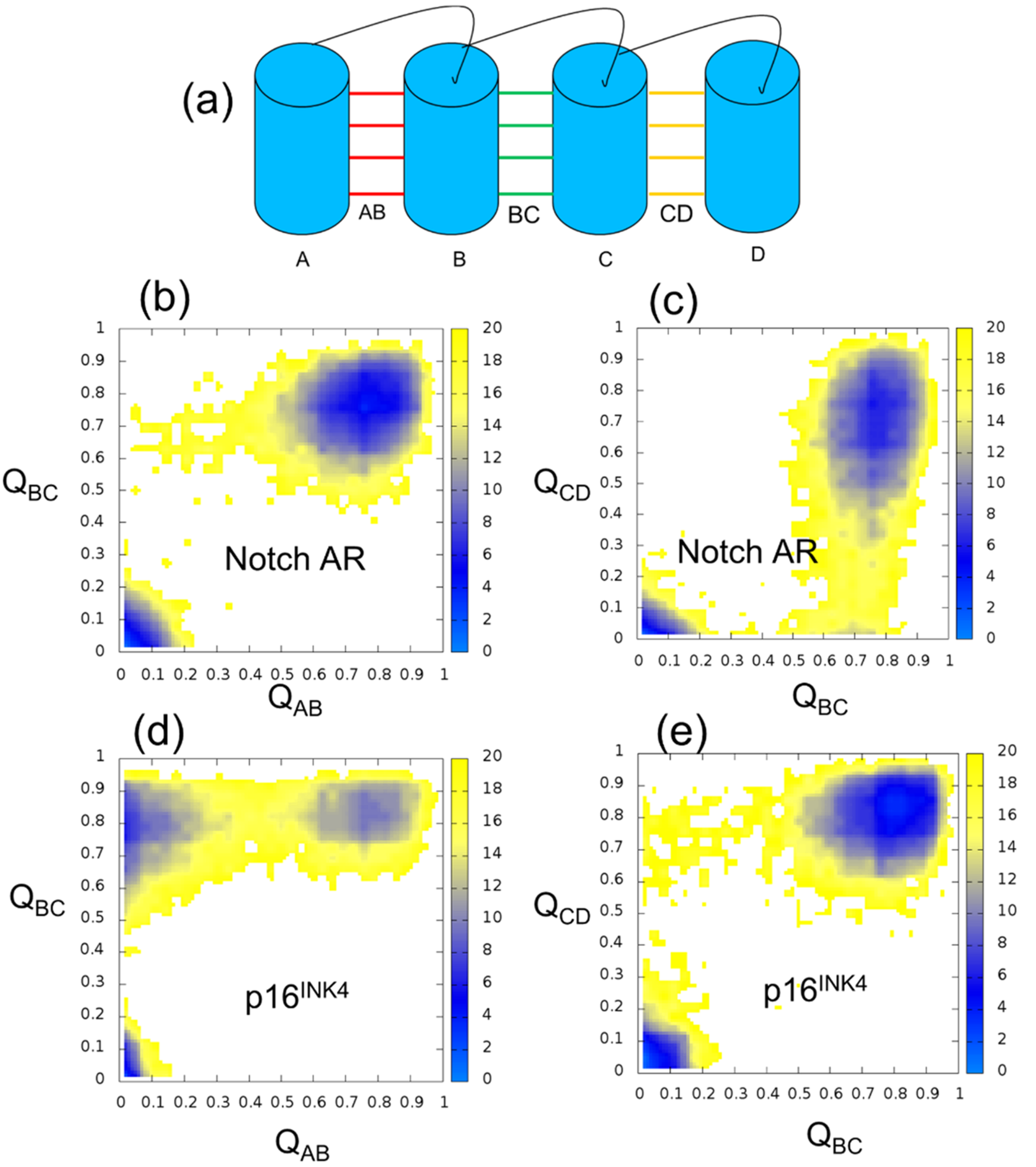


Figure 4. (a) Schematic representation of the four domain repeats representing the interfaces found in both Notch AR and p16^{INK4}. The dashed lines in the figure indicates intermolecular contacts present between repeats. (b) Free-energy surface plots of the formation of AB and BC interface of Notch AR in all-atom representation. (c) BC and CD interface of Notch AR in all-atom representation. (d) p16^{INK4} AB and BC interface in all-atom representation. (e) Free-energy surface of formation of BC and CD interface in p16^{INK4} in all-atom representation. The scale bar in all of the plots represents the free-energy in kJ/mol.

peptides can modulate their fibrillation propensity and their ability to incorporate into pre-existing fibrils.[47] This information is especially important for the design of peptide-based inhibitors of amyloid fibrillation. We examined here the mechanisms of co-assembly by simulating the formation of an hIAPP trimer with two wild-type peptides and one mutant, where phenylalanines at positions 15 and 23 have been replaced by alanines (Figure 6a). The choice of phenylalanine is based on the previous studies highlighting the importance of $\pi-\pi$ interactions in the formation of amyloid fibrils.[48] The

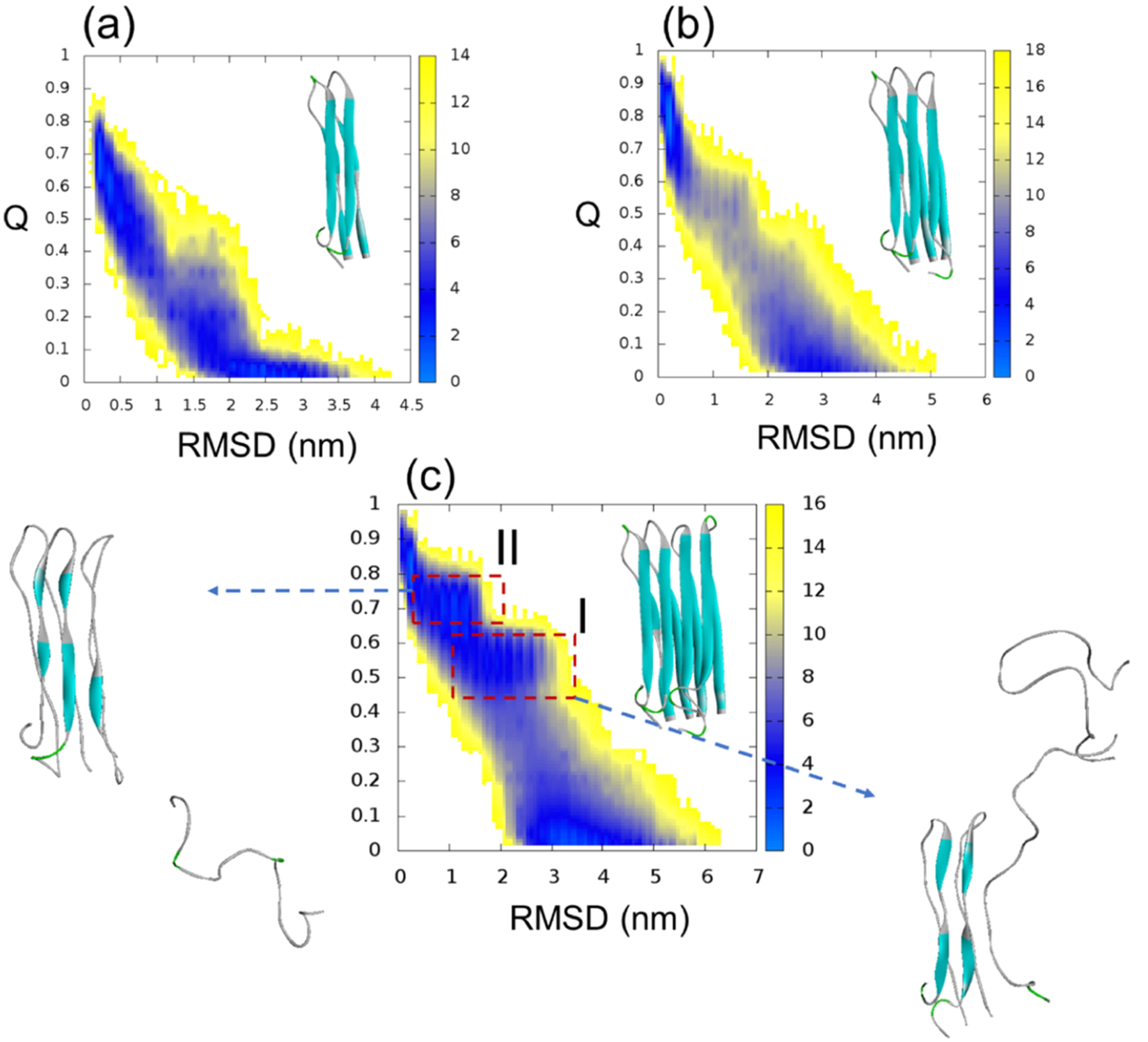


Figure 5. Free-energy landscape of hIAPP fibrillation as a function of fraction of native contacts (Q) and root-mean-square deviation (RMSD). (a) Dimerization of hIAPP; (b) trimerization of hIAPP; and (c) tetramerization of hIAPP. The two distinct intermediate-state basins observed for tetrameric system (intermediate I: $0.45 < Q < 0.60$ and $1.8 < \text{RMSD} < 3$ and intermediate II: $0.6 < Q < 0.75$ and $0.75 < \text{RMSD} < 1.8$) are highlighted with red dashed squares. The main conformations populating these basins are shown as insets (Figures S3 and S4). The scale bar in all of the plots represents the free energy in kJ/mol.

introduction of these two mutations results in a loss of 29 native contacts (about 1.5% of the total number of native contacts present in the trimeric conformation), among which 16 were present at the interface between two peptides. Simulations of this heterogeneous system were carried as previously described, and the temperature of association/dissociation T_a was determined from independent simulations over a broad range of temperatures. Interestingly, the two-dimensional free-energy landscape at $T = T_a$ shows a clear increase in the population of the dimeric intermediate states (labeled "D" in Figure 6b) compared to the homogeneous trimer (Figure 5b). Further analysis on the intermediate-state region $0.5 < Q < 0.7$ and $1.0 < \text{RMSD} < 2.0$ revealed that the conformations populating this basin are composed of ~90% of partially folded homodimers and only ~10% of partially folded heterotrimers.

Overall, the results presented in Figures 5 and 6 confirm that the dimeric state is the smallest stable nucleus on the fibrillation pathway, in good agreement with previous experimental studies, showing that dimerization is indeed a rate-limiting step in the formation of amyloid fibrils.[49] This feature is accentuated by the introduction of destabilizing mutations, such as F15A and F23A, to form a heterotrimeric system, which increases the stability of the dimeric intermediate states relative to the folded trimer (Figure 6).

■ DISCUSSION

Folding Mechanism of Repeat Proteins. Although the fundamental mechanisms underlying the folding of globular

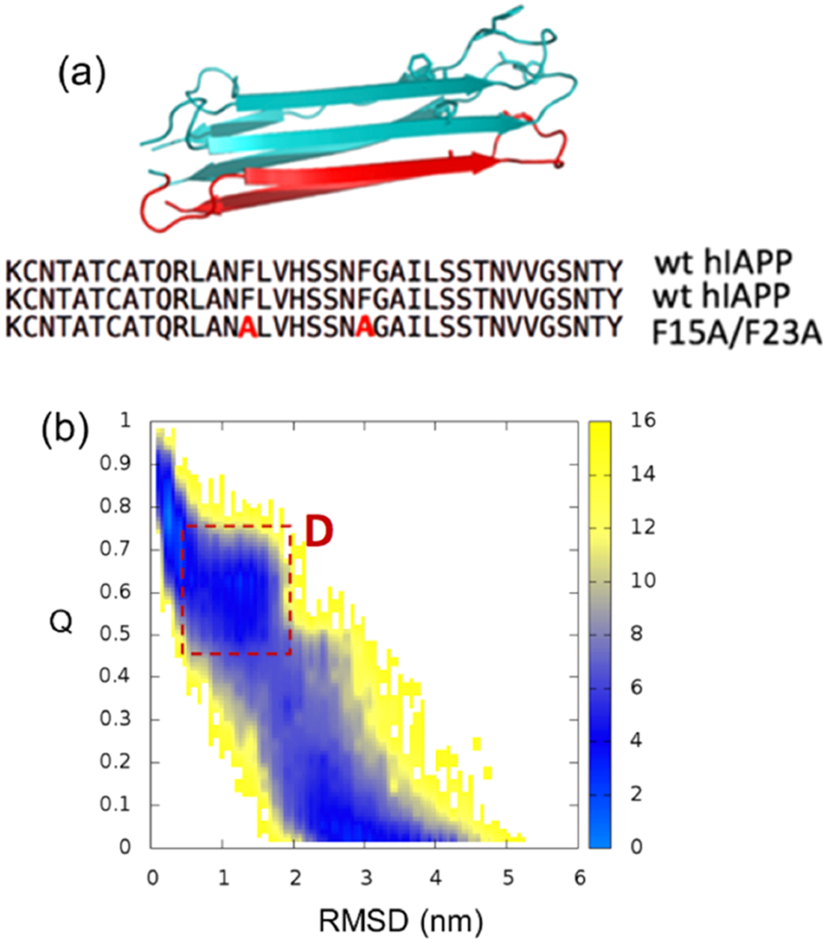


Figure 6. (a) All-atom representation of hIAPP trimer with phenylalanine and mutated residues highlighted with sticks representation; this hIAPP heterotrimer contains two wild-type peptides and one mutated peptide, where phenylalanines 15 and 23 are replaced with alanines. (b) Free-energy landscape as a function of the total number of native contacts Q and RMSD from the reference structure. The intermediate-state region is highlighted with a red dashed square and labeled as D (dimeric intermediate states). The scale bar represents the free energy in kJ/mol.

proteins are well understood from a statistical mechanics point of view,[9−12] the understanding of how pathogenic mutations and post-translational modifications affect the folding of proteins is yet limited. There is therefore a need for computational methods that can accurately describe the mechanisms of protein folding at an atomic resolution, in relevant time scales, without being computationally too intensive. We showed here that the choice of model of representation (Cα vs all-atom) significantly affects the outcomes of folding simulations of the repeat domain protein, Notch AR. The energy landscape and heat capacity analyses demonstrate that all-atom representation increases the overall folding cooperativity of Notch AR (Figures 2 and 3). In addition to the resolution (i.e., Cα vs all-atom), it should be noted that the type of energy function used in Go-model simulations, that is, a three-term LJ potential instead of the classic two-term potential can also strongly influence the folding cooperativity of the simulated system.[17] Moreover, our simulations show that the folding nucleus of Notch AR lies in the central repeats (Figures 2 and 4b,c), in good agreement with the experimental data reported by Mello and Barrick.[41] Our results also demonstrate that a simple all-atom Go-model is able to distinguish the folding pathways of two proteins with very similar topologies, such as Notch AR and p16^{INK4} (Figure 4). The folding mechanism described by our all-atom Go-model for p16^{INK4} was consistent with the previous experimental studies[29,44] and did not reveal any significant population of intermediate states with folded N-terminal repeats, as reported by Ferreiro et al[50] in their Cα Go-model study. We believe that the differences between our results and previous simulations can in part be attributed to the choice of reference structures used for establishing the list of native contacts. By comparing simulations derived from two different p16^{INK4} reference structures, 1BI7[32] (determined by X-ray crystallography) and 2A5E[33] (determined by NMR spectroscopy), we obtained different outputs from the folding simulations ran with Cα models, whereas the all-atom Go-model simulations gave the same results for both reference structures (Figure S5). A close examination of these two structures revealed about 12% more native contacts in the N-terminal repeats (i.e., the AB interface as defined in Figure 4) in the NMR-derived structure 2A5E compared to 1BI7, which explain the N-terminal intermediate states observed from the Cα simulations based on the 2A5E reference structure (Figure S5). These comparisons highlight the extreme sensitivity of Cα models to small differences in the reference structures, whereas in the case of all-atom models, these differences are distributed over a much larger number of contacts and have less impact on the simulations.

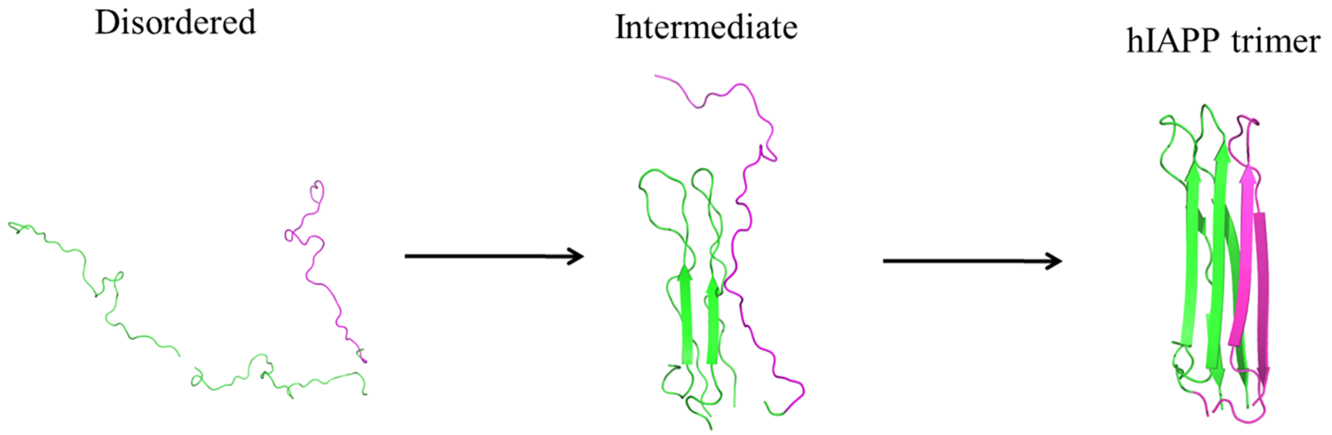


Figure 7. Schematic representation of the formation of an hIAPP trimer from three disordered peptides. The dimeric intermediate states observed in our trajectories represent in this model the docking surface onto which the third peptide can be locked, first in an extended state before a U-shaped conformation leading to the final formation of trimer.

Because all-atom Go-model simulations can successfully reproduce the folding mechanisms of repeat domain proteins, such as Notch AR and p16$^{\mathrm{INK4}}$, we tested if this simple model could also describe the formation of fibrils using hIAPP as a model peptide. Our all-atom Go-model simulations of hIAPP fibril formation suggest that dimers are the smallest intermediate states significantly populated upon fibrillation (Figure 5). The fibril elongation mechanism observed here for hIAPP closely reassembles to lock and dock mechanism described for the amyloid β fibrils.[51] Docking is described as a mechanism by which an incoming monomer loosely associates with a fibril template to form contacts leading to its incorporation in the fibril or to readily dissociate. We believe that the dimeric intermediate states observed here represent the simplest form of template available for the elongation of hIAPP fibrils (Figure 7).

To test the lock and dock hypothesis of hIAPP fibril formation, we designed a heterotrimer formed by two wild-type peptides and one peptide with F15A and F23A mutations. The all-atom Go-model simulations showed a clear increase of the population of dimeric intermediate states relative to the fully formed trimer (Figure 6b compared to Figure 5b) and we noted that 90% of the conformations present in the intermediate states basins were homodimers, and less than 10% heterodimers. This observation suggests that even for a heterogeneous system, such as the one simulated here, these stable homodimers represent the minimal template required for docking of the third peptide and further elongation of the fibril. This simple in silico experiment also highlights the potential of all-atom Go-model simulations for designing peptide-based inhibitors of hIAPP fibrillation that are critically needed for the treatment of type 2 diabetes. Such approach will likely require Go-models with hybrid potentials,[52] including salt bridges, desolvation potential,[53] and hydrogen bonding, for a more realistic depiction of peptide interactions that are currently under development in our group.

CONCLUSIONS

We have presented here a "proof of concept" that all-atom Go-model simulations can accurately reproduce the essential features of protein folding and assembly. We tested the accuracy of our model with two challenging systems, the folding of nonglobular proteins and the formation of an amyloid fibril. The results obtained from the all-atom Go-model simulations were in good agreement with the experimental reports, thereby confirming that structure-based models represent an interesting alternative to computationally intensive molecular dynamics simulations with explicit solvent description. The coarse graining approach offered by structure-based models is especially promising for the study of fibril formation as demonstrated here in the case of hIAPP. Our all-atom model was able to identify the key intermediate states populated on the elongation pathway in the context of homogeneous formation and cross-seeding. We believe that structure-based models have a high potential for the characterization of disease causing mutations and post-translational modifications on protein folding and fibrillation.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b12129.

> Trajectories obtained before and after renormalization of native contacts (Figure S1); central conformations obtained from a cluster analysis of intermediate states ensembles for Notch AR (Figure S2) and hIAPP (Figures S3 and S4); and folding routes of p16$^{\mathrm{INK4}}$ obtained from X-ray-derived and NMR-derived structures (Figure S5) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: jroche@iastate.edu.

ORCID

Julien Roche: 0000-0003-1254-1173

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Iowa State University and the Roy J. Carver Foundation. The authors are grateful to Cathy Royer for her insightful comments.

REFERENCES

(1) Sosnick, T. R.; Barrick, D. The folding of single domain proteins–have we reached a consensus? *Curr. Opin. Struct. Biol.* **2011**, *21*, 12–24.

(2) Englander, S. W.; Mayne, L. The case for defined protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 8253−8258.

(3) Eaton, W. A.; Wolynes, P. G. Theory, simulations, and experiments show that proteins fold by multiple pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E9759−E9760.

(4) Englander, S. W.; Mayne, L. Reply to Eaton and Wolynes: How do proteins fold. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E9761−E9762.

(5) Englander, S. W.; Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 15873−15880.

(6) Sato, S.; Fersht, A. R. Searching for multiple folding pathways of a nearly symmetrical protein: temperature dependent phi-value analysis of the B domain of protein A. *J. Mol. Biol.* **2007**, *372*, 254−267.

(7) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267*, 1619−1620.

(8) Bowman, G. R.; Pande, V. S. Protein folded states are kinetic hubs. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 10890−10895.

(9) Bryngelson, J. D.; Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7524−7528.

(10) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995**, *21*, 167−195.

(11) Wolynes, P. G. Energy landscapes and solved protein-folding problems. *Philos. Trans. R. Soc., A* **2005**, *363*, 453−467.

(12) Onuchic, J. N.; Luthey-Schulten, Z. A.; Wolynes, P. G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545−600.

(13) Go, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183−210.

(14) Clementi, C.; Nymeyer, H.; Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and enroute intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **2000**, *298*, 937−953.

(15) Nymeyer, H.; Garcia, A. E.; Onuchic, J. N. Folding funnels and frustration in off-lattice minimalist protein landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5921−5928.

(16) Gosavi, S.; Leslie, C. L.; Jennings, P. A.; Onuchic, J. N. Topological frustration and the folding of interleukin-1β. *J. Mol. Biol.* **2006**, *357*, 986−996.

(17) Karanicolas, J.; Brooks, C. L., III The origin of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* **2002**, *11*, 2351−2361.

(18) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. U.; Onuchic, J. N. An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **2009**, *75*, 430−441.

(19) Linhananta, A.; Boer, J.; MacKay, I. The Equilibrium properties and folding kinetics of an all-atom Go model of the Trp-Cage. *J. Chem. Phys.* **2005**, *122*, 114901−114916.

(20) Wu, L.; Zhang, J.; Qin, M.; Liu, F.; Wang, W. Folding of proteins with an all-atom Gō-Model. *J. Chem. Phys.* **2008**, *128*, 235103−235111.

(21) Luo, Z.; Ding, J.; Zhou, Y. Temperature-dependent folding pathways of Pin1 WW domain: an all-atom molecular dynamics simulation of a Go model. *Biophys. J.* **2007**, *93*, 2152−2161.

(22) Luo, Z.; Ding, J.; Zhou, Y. Folding mechanisms of individual β-hairpins in a Go model of Pin1 WW domain by all-atom molecular dynamics simulations. *J. Chem. Phys.* **2008**, *128*, No. 225103.

(23) Berhanu, W. M.; Jiang, P.; Hansmann, U. H. Folding and association of a homotetrameric protein complex in an all-atom Go model. *Phys. Rev. E* **2013**, *87*, No. 014701.

(24) Dobson, C. M. Protein Aggregation and its consequences for human disease. *Protein Pept. Lett.* **2006**, *13*, 219−227.

(25) Sciarretta, K. L.; Gordon, D. J.; Meredith, S. C. Peptide-based inhibitors of amyloid assembly. *Methods Enzymol.* **2006**, *413*, 273−312.

(26) Young, L. M.; Ashcroft, A. E.; Radford, S. E. Small Molecule Probes of Protein Aggregation. *Curr. Opin. Chem. Biol.* **2017**, *39*, 90−99.

(27) Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N.-V.; Coté, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; et al. Amyloid β protein and Alzheimer's disease: When computer simulations complement experimental studies. *Chem. Rev.* **2015**, *115*, 3518−3563.

(28) Barrick, D.; Ferreiro, D. U.; Kmives, E. A. Folding landscapes of ankyrin repeat proteins: experiments meet theory. *Curr. Opin. Struct. Biol.* **2008**, *18*, 27−34.

(29) Tang, K. S.; Guralnick, B. J.; Wang, W. K.; Fersht, A. R.; Itzhaki, L. S. Stability and folding of the tumor suppressor protein p16. *J. Mol. Biol.* **1999**, *285*, 1869−1886.

(30) Höppener, J. W. M.; Ahrén, B.; Lips, C. J. M. Islet amyloid and type 2 diabetes mellitus. *N. Engl. J. Med.* **2000**, *343*, 411−419.

(31) Zweifel, M. E.; Leahy, D. J.; Hughson, F. M.; Barrick, D. Structure and stability of the ankyrin domain of the drosophila Notch receptor. *Protein Sci.* **2003**, *12*, 2622−2632.

(32) Russo, A. A.; Tong, L.; Jie-Oh, L.; Jeffrey, P. D.; Pavletich, N. P. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumor suppressor p16INK4a. *Nature* **1998**, *395*, 237−243.

(33) Byeon, I.-J. L.; Li, J.; Ericson, K.; Selby, T. L.; Tevelev, A.; Kim, H. J.; O'Maille, P.; Tsai, M. D. Tumor Suppressor p16INK4A: Determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. *Mol. Cell* **1998**, *1*, 421−431.

(34) DeLano, W. L. *PyMOL*; DeLano Scientific: San Carlos, CA, 2002; p 700.

(35) Luca, S.; Yau, W.-M.; Leapman, R.; Tycko, R. Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. *Biochemistry* **2007**, *46*, 13505−13522.

(36) Noel, J. K.; Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **2010**, *38*, W657−W661.

(37) Noel, J. K.; Levi, M.; Raghunathan, M.; Lammert, H.; Hayes, R. L.; Onuchic, J. N.; Whitford, P. C. SMOG 2: a versatile software package for generating structure-based models. *PLoS Comput. Biol.* **2016**, *12*, No. e1004794.

(38) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A High-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(39) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted histogram analysis method for free-energy calculations on biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(40) Morcos, F.; Jana, B.; Hwa, T.; Onuchic, J. N. Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 20533−20538.

(41) Mello, C. C.; Barrick, D. An Experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14102−14107.

(42) Bradley, C. M.; Barrick, D. The notch ankyrin domain folds via a discrete, centralized pathway. *Structure* **2006**, *14*, 1303−1312.

(43) Rouget, J.-B.; Schroer, M. A.; Jeworrek, C.; Pühse, M.; Saldana, J.-L.; Bessin, Y.; Tolan, M.; Barrick, D.; Winter, R.; Royer, C. A. Unique features of the folding landscape of a repeat protein revealed by pressure perturbation. *Biophys. J.* **2010**, *98*, 2712−2721.

(44) Tang, K. S.; Fersht, A. R.; Itzhaki, L. S. Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure* **2003**, *11*, 67−73.

(45) Zheng, W.; Tsai, M.-Y.; Chen, M.; Wolynes, P. G. Exploring the aggregation free energy landscape of the Amyloid-β Protein (1−40). *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 11835−11840.

(46) Hu, R.; Zhang, M.; Chen, H.; Jiang, B.; Zheng, J. Cross-seeding interaction between β-Amyloid and human islet amyloid polypeptide. *ACS Chem. Neurosci.* **2015**, *6*, 1759−1768.

(47) Young, L. M.; Tu, L.-H.; Raleigh, D. P.; Ashcroft, A. E.; Radford, S. E. Understanding co-polymerization in amyloid formation by direct observation of mixed oligomers. *Chem. Sci.* **2017**, *8*, 5030−5040.

(48) Gazit, E. A possible role for π-stacking in the self-assembly of amyloid fibrils. *FASEB J.* **2002**, *16*, 77−83.

(49) Krishnan, S.; Chi, E. Y.; Wood, S. J.; Kendrick, B. S.; Li, C.; Garzon-Rodriguez, W.; Wypych, J.; Randolph, T. W.; Narhi, L. O.; Biere, A. L.; et al. Oxidative dimer formation is the critical rate-limiting step for parkinson's disease α-synuclein fibrillogenesis. *Biochemistry* **2003**, *42*, 829−837.

(50) Ferreiro, D. U.; Cho, S. S.; Komives, E. A.; Wolynes, P. G. The energy landscape of modular repeat proteins: Topology determines folding mechanism in the ankyrin family. *J. Mol. Biol.* **2005**, *354*, 679−692.

(51) Esler, W. P.; Stimson, E. R.; Jennings, J. M.; Vinters, H. V.; Ghilardi, J. R.; Lee, J. P.; Mantyh, P. W.; Maggio, J. E. Alzheimer's Disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry* **2000**, *39*, 6288−6295.

(52) Sutto, L.; Mereu, I.; Gervasio, F. L. A hybrid all-atom structure-based model for protein folding and large scale conformational transitions. *J. Chem. Theory Comput.* **2011**, *7*, 4208−4217.

(53) Cheung, M. S.; Garcia, A. E.; Onuchic, J. N. Protein folding mediated by solvation: water explusion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 685−690.

DOI: 10.1021/acs.jpcb.7b12129