α -helix to β -hairpin transition of human amylin monomer

Sadanand Singh,¹ Chi-cheng Chiu,² Allam S. Reddy,¹ and Juan J. de Pablo^{1,2,3}

¹Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

²Institute for Molecular Engineering, University of Chicago, Chicago, Illinois 60637, USA

³Argonne National Laboratory, 9700 South Cass Avenue, Building 223, Argonne, Illinois 60439, USA

(Received 29 November 2012; accepted 13 March 2013; published online 18 April 2013)

The human islet amylin polypeptide is produced along with insulin by pancreatic islets. Under some circumstances, amylin can aggregate to form amyloid fibrils, whose presence in pancreatic cells is a common pathological feature of Type II diabetes. A growing body of evidence indicates that small, early stage aggregates of amylin are cytotoxic. A better understanding of the early stages of the amylin aggregation process and, in particular, of the nucleation events leading to fibril growth could help identify therapeutic strategies. Recent studies have shown that, in dilute solution, human amylin can adopt an α -helical conformation, a β -hairpin conformation, or an unstructured coil conformation. While such states have comparable free energies, the β -hairpin state exhibits a large propensity towards aggregation. In this work, we present a detailed computational analysis of the folding pathways that arise between the various conformational states of human amylin in water. A free energy surface for amylin in explicit water is first constructed by resorting to advanced sampling techniques. Extensive transition path sampling simulations are then employed to identify the preferred folding mechanisms between distinct minima on that surface. Our results reveal that the α -helical conformer of amylin undergoes a transformation into the β -hairpin monomer through one of two mechanisms. In the first, misfolding begins through formation of specific contacts near the turn region, and proceeds via a zipping mechanism. In the second, misfolding occurs through an unstructured coil intermediate. The transition states for these processes are identified. Taken together, the findings presented in this work suggest that the inter-conversion of amylin between an α -helix and a β -hairpin is an activated process and could constitute the nucleation event for fibril growth.

© 2013 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4798460>]

I. INTRODUCTION

Human islet amylin polypeptide (human amylin, or hIAPP) is a 37 residue amino-acid polypeptide (with sequence KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY) that is produced in the islet β -cells. Amylin has been implicated in Type II (non-insulin-dependent) diabetes.^{1–3} In healthy individuals, the ratio of human amylin to insulin is maintained at 1:100; diabetic patients, however, exhibit ratios as high as 1:20.⁴ The human amylin peptide shows a propensity to aggregate into amyloid fibrils in *in vitro* experiments. That propensity is concentration dependent, suggesting that higher amylin concentrations in Type II diabetes patients can lead to aggregation of the protein.⁵ Note that amylin aggregates are commonly found as amyloid deposits in pancreatic islets from diabetic patients.⁶ While the causative relationship between aggregation of human amylin peptide and Type II diabetes is not known, it has been established that aggregates of human amylin can induce apoptotic cell-death of the insulin producing β -cells.² Perhaps more importantly, recent experiments have shown that the main toxic species are the pre-fibrillar, early stage oligomers of the peptide.^{5,7} Such observations suggest that a deeper understanding of the aggregation pathways and its early stages could provide valuable insights for development of therapeutic strategies. In this work, we use molecular simulations in explicit water

to examine the conformational dynamics of the hIAPP monomer between distinct metastable free energy minima.

While the results presented in this paper are only concerned with the amylin monomer prior to aggregation, in order to put our findings into perspective it is instructive to briefly discuss recent studies of the hIAPP monomer, its aggregates, and the kinetics of aggregation.^{8–11} For fully formed aggregates, Kajava *et al.*¹² proposed a structure consisting of a three-strand model, with β -strands assigned to residues 12–17, 22–27, and 31–37. Luca *et al.*¹¹ used solid-state NMR to propose a strand-loop-strand morphology for the hIAPP monomers in the fibrils, with residues 8–17 and 28–37 forming the β -sheets. That morphology is also supported by time-resolved two-dimensional infrared spectroscopy (2DIR)¹⁰ and electron paramagnetic resonance (EPR) studies.¹³ Earlier predictions in terms of secondary structure algorithms suggested an α -helical structure for residues located in the N-terminal half of the peptide, and a random-coil structure in the C-terminal half.¹⁴

The aggregation kinetics of human amylin in solution have been studied using thioflavin fluorescence, circular dichroism, light scattering, FTIR and 2DIR techniques.^{15–18} The corresponding aggregation curves exhibit a sigmoidal shape, suggesting the existence of a nucleation event; nucleation is manifest as a time lag, followed by a rapid elongation

phase in which remaining free peptides are incorporated into fibrils.¹⁹ Early work by Westermarck *et al.*²⁰ and Moriarty *et al.*²¹ showed that residues ASP22, GLY24, ILE26, LEU27, and SER28 are important for human amylin aggregation. The rat variant of the amylin protein, which differs from human amylin at six residues, does not form aggregates. Three of the rat mutations are proline substitutions at positions 25, 28, and 29 and are believed to be responsible for the lack of aggregation.²² In contrast, the mutation SER20GLY, found in small numbers of type-II diabetic patients of Chinese and Japanese origin,^{23,24} accelerates formation of human amylin fibrils *in vitro*.²⁵ More recent thioflavin T fluorescence kinetic studies of hIAPP aggregation²⁶ have also highlighted the relevance of residue SER20 on the aggregation process. In particular, it was found that a SER20GLY mutation tends to accelerate aggregation; in contrast, a SER20LYS mutation inhibits aggregation. Andreassen *et al.*²⁷ have also examined the importance of backbone residues in the hIAPP aggregation process. Our own recent studies using 2DIR and simulations⁹ indicate that hIAPP amyloid formation can be inhibited by the presence of rat IAPP. Rat IAPP blocks the N-terminal β -sheet and ultimately forms its own β -sheets on the exterior of the human IAPP fibrils.

In spite of a growing body of work concerning the importance of various residues in the aggregation kinetics of human amylin, not much is known about the folded states of the monomer, and the role that these may play in aggregation. Distinct metastable conformational states in solution exhibit a different propensity towards aggregation. That behavior is evident in simulations using explicit water²⁸ or an implicit solvent.^{29,30} Our own view is that specific residues could play a key role in the pathways for inter-conversion between them, and that such conformational transitions could in fact constitute the nucleation event. Studies of long polyglutamine molecules are consistent with that view,³¹ but comparable studies of conformational transition pathways in amylin have not been carried out before.

Experimental studies of the hIAPP monomer have been limited; the aggregation process is fast, particularly at sufficiently high concentrations to allow for rapid detection, thereby raising considerable challenges for its characterization. Following the thermal denaturing pathway of hIAPP by circular dichroism (CD), Kaye *et al.*³² proposed that the hIAPP monomer adopts two distinct conformers: one non-amyloidogenic and another amyloidogenic, with the latter having higher contributions of both β -sheet and α -helix. That study suggested that, at room temperature, the hIAPP monomer would populate both conformers at equilibrium. Subsequent CD studies demonstrated that the hIAPP monomer adopts primarily a random coil structure.^{16,33–37} Recent NMR studies on human amylin,³⁸ rat amylin and pramlintide peptides, which have sequences similar to that of amylin, suggest that an α -helical structure is adopted near the N-terminus.^{39,40} By interpreting the results of ion mobility mass spectroscopy experiments with a molecular implicit-solvent model, Dupuis *et al.*²⁹ showed that the hIAPP monomer adopts an α -helical conformation between residues 9–17 on the terminus end, and a short β -hairpin between residues 24–28 and 31–35 on the C-terminus. Using a com-

bination of 2DIR experiments and an atomistic explicit-water model, Reddy *et al.*²⁸ reported similar conformations to those reported by Dupuis *et al.*²⁹ Note that previous experiments in the presence of membranes also proposed an α -helical secondary structure for the human amylin protein.^{15,19,41}

The aim of the present study is to use simulations in an explicit water model to arrive at a detailed molecular picture of the inter-conversion pathways of human amylin between the α -helical, β -hairpin, and unstructured coil states. Building on our past work, which was limited to calculation of the free energy difference between those three states,²⁸ here we examine the relative stability of possible monomeric structures by constructing a full-fledged two-dimensional free-energy map of the possible secondary structures, including stable and metastable conformational states. That free-energy map provides an energetic framework on which to rationalize dynamic trajectories for the folding pathways, which we generate by resorting to transition path sampling techniques. As alluded to earlier, the transitions between such conformations could constitute the nucleation event for the onset of aggregation and could be of central importance in the genesis of the disease. By identifying key residues along the relevant pathways, for example, one could conceive of mutations that could potentially inhibit conformational transitions and slow down aggregation.

The β -hairpin conformation, being the simplest anti-parallel β -sheet and one of the principal secondary structural elements of proteins, has been studied extensively in experiments^{42–48} and computer simulations.^{49–52} Previous results have led to two primary hypothesis for the formation of a β -hairpin: one in which the structure is initiated at the turn (zipping mechanism), and another in which a hydrophobic cluster first forms first (hydrophobic collapse mechanism). In the zipping mechanism, turn formation is followed by propagation of structure towards the amino and carboxy termini, with sequential formation of inter-strand hydrogen bonds. In the hydrophobic collapse mechanism, cluster formation precedes the formation of the inter-strand hydrogen bonds. Results from coarse grained simulations suggest that an α -helix to β -hairpin transition occurs through an intermediate random coil conformation.⁵³ In this work, we examine both possible pathways of conversion to a β -hairpin state using atomistic simulations in water. With few exceptions,^{31,52,54} atomistic simulations of long polypeptides in explicit water have focused on the stability of the β -hairpin structure or have resorted to modified potentials whose function is to accelerate the folding dynamics. In this work, large numbers of reactive trajectories along the folding/misfolding pathways of hIAPP were generated for the full potential through a transition-path-sampling (TPS) Monte Carlo approach.⁵⁵ Those trajectories were used to identify the transition state and were examined in the context of the free-energy surface described above.

II. METHODS

The protein was modeled using the GROMOS53a6 force field.⁵⁶ The SPC water model was used for water molecules.⁵⁷ The GROMOS53a6 force field has been extensively used in the literature to study the folding of protein molecules.^{58–60}

FIG. 1. Representative snapshots of human amylin peptide (a) in α -helical conformation, (b) in β -hairpin conformation, and (c) in unstructured coil state.

Our choice is motivated by recent work on hIAPP, where good agreement was found between 2DIR measurements and results from molecular simulations.^{9,10,28} The starting structures for the folded (α -helical) and the misfolded (β -hairpin) configurations (see Figure 1) were obtained from our previous simulation studies.²⁸ Molecular Dynamics simulations in this work were performed using a modified version of the Gromacs molecular simulation package.⁶¹ Long-range electrostatic interactions were treated with a particle-mesh Ewald sum.⁶² All simulations were performed at a temperature of 298 K and at a pressure of 1 bar.

A. Metadynamics and bias-exchange simulations

A free energy map for the monomer as a function of distinct order parameters was generated by means of metadynamics simulations.⁶³ The basic idea behind such an approach is to enhance sampling along order parameters of interest (order parameters that can distinguish all possible configurations, henceforth referred to as “collective variables,” “CVs”); sampling is achieved by modifying the free energy surface through addition of Gaussian potentials that force the system to visit states that are otherwise not easily accessible. The added external force acting on particle i due to such added Gaussian potentials is given by

$$\phi_i \rightarrow \phi_i - \frac{\partial}{\partial s_j} W \sum_{t' < t} \prod_{j=1}^d \exp \left[-\frac{|s_j - s_j'|^2}{2\sigma^2} \right], \quad (1)$$

where ϕ_i denotes the unbiased force on particle i at time t , and W and σ represent the height and the width of each Gaussian; d is the number of CVs used, s_j is the value of the j th CV at time t and s_j' is the value of the j th CV at time t' . As more Gaussians are added at a specific minimum on the free energy surface, the system is forced to escape that minimum and visit other possible states. By construction, in the long-time limit that corresponds to uniform sampling, the negative of the sum of all of the added Gaussians represents the free energy of the system. For a process as complicated as protein folding, the choice of suitable CVs determines the success of the method.⁶⁴

In general, several CVs may be required for correct description of a complex process. Taking a cue from our recent work on the stable structures of amylin in water,²⁸ we posit that the configurations of interest can be expressed in terms of

a combination of variables describing the secondary-structure content of the molecule (α -helix and β -hairpin). Specifically, we use a definition⁶⁵ in which CVs, α_{rmsd} and β_{rmsd} , are given by the total number of beta and alpha blocks whose root-mean-squared deviation (RMSD) from an ideal reference structure is smaller than a specified cut-off value. Ideal structures were extracted from the STRIDE⁶⁶ program database of peptides. The corresponding equations for these CVs can be written as

$$\alpha_{rmsd} = \sum_S n[RMSD(\{\mathbf{R}_i\}_{i \in \Omega_S}, \{\mathbf{R}^0\})], \quad (2)$$

$$n[RMSD] = \frac{1 - [RMSD/0.1]^8}{1 - [RMSD/0.1]^{12}}, \quad (3)$$

where n is a function switching smoothly between 0 and 1, the RMSD is measured in nm, S is the total number of possible blocks, and $\{\mathbf{R}_i\}_{i \in \Omega_S}$ are the atomic coordinates of a set Ω_S of six residues of the protein that define a block, while $\{\mathbf{R}^0\}$ are the corresponding atomic positions of the ideal block. An expression similar to Eq. (2) is used for β_{rmsd} .

To further enhance sampling, the metadynamics formalism is implemented in the context of parallel-tempering,⁶⁷ or replica-exchange simulations.^{64,68} For a system with two CVs, four replicas are simulated simultaneously. The first box is not biased, i.e., no external Gaussians are added. The second and third boxes include added Gaussians to the individual CV dimensions. In the fourth box Gaussians are added in all CVs. Trial exchanges of configurations between all boxes are accepted with probability:

$$p_{ab} = \min \left\{ 1, \exp \left(\beta \left[V_G^a(\mathbf{r}^a, t) + V_G^b(\mathbf{r}^b, t) - V_G^a(\mathbf{r}^b, t) - V_G^b(\mathbf{r}^a, t) \right] \right) \right\}, \quad (4)$$

where p_{ab} is the probability of accepting an exchange of configurations from simulation box “ a ” to simulation box “ b ,” $\mathbf{r}$ is the spatial position of the system, and V_G is the bias potential due to the added Gaussians. The sought after free energy map is obtained from the simulation box that includes biasing in all CVs; the other simulation boxes are used to enhance sampling. A converged value of the free energy map is arrived at through the protocol described in Ref. 69; in that approach, after sufficient sampling of an order parameter (determined through a histogram of visits to distinct order parameter values), the free energy map is averaged over a fixed period of simulation time. In our simulations, the free energy map is obtained from the 120 ns average of a total of 300 ns of simulation time. All simulations were performed with a modified version of the Gromacs⁶¹ simulation package and the PLUMED plugin.⁷⁰

B. Transition path sampling simulations

The aim of our study is to understand the folding pathways of human amylin. To that end, we also rely on transition path sampling (TPS) simulations.⁵⁵ An important attribute of TPS is that it does not require *a priori* knowledge of the reaction coordinate; one can obtain important macroscopic variables that define the reaction coordinate from analysis of the reactive pathways obtained from TPS calculations.^{31,71,72}

TPS simulations require that the stable states between which the reaction occurs be identified in terms of one or more “non-overlapping” order parameters. Here, we used two order parameters referred to as OP1 and OP2. OP1 denotes the fraction of α -helical contacts that are present in a folded configuration, which are defined based on the distance (≤ 6 Å) between the C- α atoms of the i and $i + 4$ residue pair ($7 \leq i \leq 13$). Similarly, OP2 denotes the fraction of β -hairpin contacts that are defined based on the distance (≤ 6 Å) between the C- α atoms of the residues that are involved in backbone hydrogen bonding in the β -hairpin structure. In terms of these two order parameters, the folded or stable states were identified as configurations with (a) OP1 ≥ 0.7 , (b) OP2 ≤ 0.2 and configurations with (a) OP1 ≤ 0.2 , (b) OP2 ≥ 0.7 , respectively. A structure is viewed as a unstructured/random coil when (a) OP1 ≤ 0.2 , (b) OP2 ≤ 0.2 . The boundary conditions for these order parameters were selected on the basis of long (500 ns) equilibrium molecular dynamics simulations started from the three metastable states.

To generate the transition path ensemble (TPE) we used the constant path length TPS algorithm described in Ref. 73; we have previously used this method to study folding of related peptides, such as polyglutamine, and their dimerization.^{31,71} The starting trajectory was obtained from a long MD simulation started in the α -helical conformation; during the course of a simulation, that conformation was able to transform into a β -hairpin structure. The length of the path needed for TPS simulations was set to 2.5 ns for α -helix to β -hairpin and α -helix to random coil transitions and to 10 ns for random coil to β -hairpin transitions. The choice of TPS simulation length was motivated by the simulation length over which a particular transition was observed in the corresponding long MD simulation. This length for the α -helix to β -hairpin transition was also considered appropriate based on Ref. 74, where α -helix to β -hairpin folding/misfolding events were observed to occur within time scales of 1.0–1.5 ns. Finally, the underlying transition state ensemble (TSE) for the folding processes was determined by identifying the frames for which the probability that a random trajectory would end in the α -helical (or β -hairpin or random coil) state was between 0.4 and 0.6.⁵⁴

III. RESULTS AND DISCUSSION

A. Free energy map

The conformational free energy of human amylin is characterized along two order parameters (also referred as CVs): α_{rmsd} and β_{rmsd} (see Sec. II for details). The two-dimensional free energy surface of the monomer in solution is shown in Figure 2. Three stable conformations of the monomer can be identified by the minima in that surface; these conformations are the same as those described in previous works.^{28,29} The relative values of the free energy for the three structures, which were estimated here on the basis of metadynamics simulations, are in quantitative agreement (within 1 kJ/mol) with those calculated in our earlier work on the basis of thermodynamic integration.²⁸ The new crucial piece of information included in Figure 2, which was not available in previous work,

FIG. 2. (A) Free energy map for the different structures of the human amylin monomer along the order parameters α_{rmsd} and β_{rmsd} , as defined in Sec. II. The contour lines have been drawn at an interval of 2.5 kJ/mol, unless specified otherwise. Three low-energy states are labeled as (a) α -helical state, (b) β -hairpin state and (c) unstructured coil state. The black solid curve represents the minimum free energy path joining (a) to (b), calculated by a steepest descent algorithm. The solid and dotted red curves represent the minimum free energy path joining (a) to (c) and the minimum free energy path joining (c) to (b), respectively. The points in the inset of the figure represent transition paths from TPS simulations. Green points correspond to the zipping mechanism pathway between α -helical state and β -hairpin state. White points represent paths for the transition from the α -helical state to the unstructured coil state. Yellow points correspond to the transition pathways for folding from the unstructured coil state to the β -hairpin state. (B) Same as (A) but with results shown in expanded order parameter and energy scales. Roman numbers from I to V represent different states along the minimum free energy transition path for the zipping mechanism. The transition states for the pathways from α to β is denoted by T1, and for α to coil it is denoted by T2.

corresponds to the free energy landscape that surrounds those minima. It is that landscape that dictates the dynamics of folding between distinct states. In Figure 2, the three minima are marked as (a) the α -helical conformation, (b) the β -hairpin conformation, and (c) the unstructured (random) coil structure. It is important to note that the presence of explicit water molecules in the simulation plays a key role in determining the details of that landscape. In particular, when performed in implicit water, simulations of amylin indicate that the β -hairpin is considerably more stable than the α -helix (see supplementary material⁷⁵), thereby leading to drastically different pathways for the folding of amylin.

Using a coarse grained model, Dokholyan *et al.*⁵³ proposed that the transition from the α -helical state to the β -hairpin state occurs through the second pathway, i.e., via the unstructured coil state. Our work, based on fully atomistic explicit water simulations, leads us to propose an additional

possible folding pathway. A minimum free energy path between the α and β conformers can be found through a steepest descent algorithm.⁷⁶ A steepest-descent search starting from the saddle points in the vicinity of three meta-stable states (α -helix, β -hairpin and unstructured coil) result in primarily two pathways of transition from the α -helix to the β -hairpin state. The paths are represented in Figure 2 with solid black, solid red and dotted red lines. In particular, the black line shows a path that goes directly from the α -helical state the β -hairpin state with a gradual loss of α -helicity and a gradual gain of β -hairpin characteristics (referred to as the zipping mechanism⁷⁷). The red lines show a pathway that goes through the unstructured coil structure. The solid red line shows the transition from the α -helix state to the unstructured coil structure, whereas the dotted red lines show the transition from the unstructured coil state to the β -hairpin state. In Secs. III B–III D, we analyze these two pathways in more detail.

B. Pathway 1: Direct conversion from α -helix to β -hairpin

The transition path ensemble (TPE) represents the collection of trajectories that connect the two folded states (α -helical state and the β -hairpin state) of the peptide. A total of 500 reactive trajectories were generated using transition path sampling (see Sec. II for details). The path length for each reactive trajectory was 2.5 ns and frames were saved every 5 ps. The time scale of 2.5 ns was chosen based on observations from long molecular dynamics simulations. Several representative trajectories are shown in the supplementary material,⁷⁵ Figure S2, in terms of their α and β content. The entire range of order parameter (OP1 and OP2, see Methods Section for details) space is explored by the TPE, a feature that is indicative of efficient sampling. The reactive trajectories from the TPE have been represented on the free energy map in the inset of Figure 2(A) (green points). Our atomistic simulations reveal a gradual loss of α -helical contacts, which is accompanied by an increase in β -hairpin content. The folding of the protein from α to β occurs by a zipping mechanism, where contacts between main-chain C- α atoms closer to the turn residues are formed first; contact formation proceeds towards the terminus in a sequential manner. This can be appreciated in the supplementary material⁷⁵ Figure S3, which shows the native β -hairpin contact distance as a function of time for a variety of residues. The process for misfolding of the protein via zipping starts with the aggregation of hydrophobic residues ALA25, ILE26, LEU27 on the N-terminal half of the peptide, leading to a turn-like conformation. The resulting turn consists of residues SER20-ASN21-ASN22-PHE23. The aggregation of the residues is further stabilized by formation of contacts between the main chain atoms. Once these contacts are made, the transformation of the protein proceeds through gradual loss of α -helical contacts and formation of β -hairpin contacts. These results are consistent with site mutagenesis experiments,^{20,21} which showed that residues GLY24, ILE26, and LEU27 are indeed important for aggregation.

FIG. 3. Solvent accessible surface area (SASA) for the hydrophobic residue fragments (residues 16-17 and residues 23-27) as a function of time along the transition path for the zipping mechanism between α -helical and β -hairpin states (dotted black curve). The results have been averaged over 500 trajectories. For reference, a representative trajectory from TPS simulations is also shown in the figure in terms of its α -helical (solid red curve) and β -hairpin content (solid cyan curve). The encircled region represent the transition state ensemble. The configuration shown in the figure represents a typical transition state structure.

The contribution of hydrophobic forces to the stability and folding of the protein can be investigated by considering the solvent accessible surface area (SASA) of the peptide. During the folding process, residues 7-17 change their conformation from an α -helical conformation to a β -hairpin conformation. An important feature of this conformational transition is that at first there is a reduction in the SASA of the hydrophobic residue patches comprising LEU16 and VAL17, and comprising PHE23, GLY24, ALA25, ILE26, and LEU27. Such a collapse leads to a trigger for the formation of the turn in the region of residues 20–23, followed by initiation of the zipping process leading to the β -hairpin formation (see Figures 3 and 4 for details). In Figure 3, we show the SASA for the different sections of the peptide along the trajectories from the TPE. In Figure 4, the hydrophobic collapse is illustrated in terms of representative conformations from an arbitrary TPS trajectory.

The transition paths obtained from TPS simulations can be seen on the free energy map in the inset of Figure 2(A) (green points). The transition paths are distributed along the minimum free energy path obtained via the steepest descent algorithm. The results from different methods are consistent with each other. The transition state ensemble (TSE) is constructed by analyzing all reactive trajectories, and isolating configurations that are equally likely to lead to a folded α -helical state as they are to lead to the folded β -hairpin state (see Sec. II). A representative transition state configuration for this mechanism is provided in Figure 3. The transition state consists of a smaller α -helical segment (at the same location as the α -helical state) and the turn region with the side-chains of the residues collapsed on to each other.

FIG. 4. Representative configurations of amylin showing hydrophobic residues LEU16, VAL17, PHE23, GLY24, ALA25, ILE26, and LEU27. (a) α -helical conformation, (b) intermediate conformation in which all the hydrophobic residues have collapsed, (c) intermediate conformation, when zipping along the turn region begins, and (d) β -hairpin conformation. The arrows in (a) show the direction of hydrophobic collapse that triggers the transition.

C. Pathway 2: Via random coil state

To investigate the transition from the α -helical state to the unstructured coil state we also performed transition path sampling simulations (see Sec. II for details). The initial trajectory was obtained from a long molecular dynamics simulation. A total of 60 reactive trajectories were generated. The path length for each reactive trajectory was 2.5 ns, and frames were saved every 5 ps. As before, the time scale of 2.5 ns was chosen based on observations from long molecular dynamics simulations. The reactive trajectories from the TPE have been represented on the free energy map in the inset of Figure 2(A) (white points). The process occurs by a gradual loss of α -helical characteristics without any gain of β -hairpin character. It starts with the rotation of PHE15 (see Figure 5) towards the inner core of the peptide. The rotation of bulky side chains seems to be the barrier for the process. With such a rotation, the α -helical segment of the peptide becomes unstable and starts to unfold. Note that a hydrophobic collapse is absent here, in contrast to what was observed in the case of the zipping mechanism pathway. Also, such a rotation of PHE15 is absent in the case of the zipping mechanism (see Figure S4 of the supplementary materials⁷⁵). In Figure 5, we show a profile for the number of water molecules within 8 Å of the side chain of PHE15 residue. The number of water molecules around any residue provides a measure of how exposed it is to solvent molecules; smaller values indicate that it is more buried in the interior of the peptide. From our TPS simulations we find that, on average, for all reactive trajectories (see Figure 5) the number of water molecules around residue PHE15 decreases drastically ini-

FIG. 5. Number of water molecules within 8 Å of side chains of PHE15 residue as a function of time along the transition path for the transition from the α -helical state to the unstructured coil state (solid black curve). The results have been averaged over 60 trajectories. For reference, a representative trajectory from TPS simulations is also shown in the Figure in terms of its α -helical content (dotted red curve). The encircled region represents the transition state ensemble and the configuration shown in the figure is representative of a transition state structure.

tially, indicating a sudden rotation towards the interior of the peptide. As the unfolding process continues, the residue again becomes exposed to water molecules. In Figure 6, the rotation of residue PHE15 is illustrated through several representative conformations from an arbitrary TPS trajectory.

The transition paths obtained from TPS simulations can again be superimposed on the free energy map. The corresponding curves are shown in the inset of Figure 2(A) (white points). The minimum free energy path obtained by

FIG. 6. Representative configurations of amylin showing hydrophobic residues PHE15. Panels (b)–(e) represent intermediate configurations for the transition from the α -helical state (a) to the unstructured coil state (f). The arrow in (a) shows the direction of rotation of PHE15 residue that triggers the transition.

a steepest-descent search (see solid red line in Figure 2) was found to be consistent with our TPS-generated trajectories. Once again, the transition state ensemble (TSE) is constructed by analyzing all reactive trajectories, and isolating configurations which are equally likely to lead to a folded α -helical state as they are to lead to the unstructured coil state (see Sec. II). A representative transition state configuration for this mechanism is provided in Figure 6. The transition state consists of a smaller α -helical segment (at the same location as the α -helical state) and the turn region with residue PHE15 collapsed onto the interior on the peptide.

From these TPS results, we conclude that the two reactions described above represent activated processes in the order parameter space considered here. The corresponding energy barriers are analyzed in more detail below. We also examined the conversion from an unstructured coil into a β -hairpin. We found that, in contrast to the transitions from alpha to random coil or from alpha to beta hairpin, this process is diffusive in nature. Both in the case of traditional molecular dynamics and TPS simulations, more than 300 ns are required to go from a random coil to a β -hairpin. The pathway is shown in Figure 2 using yellow points.

We performed a principal component analysis⁷⁸ separately on the set of configurations corresponding to each of the minima identified in our bias-exchange simulations. The results are shown in Figure S5 of the supplementary material.⁷⁵ The two principal components in the case of the α -helical structure show only one large cluster. In the β -hairpin structure and the unstructured coil cases, one can see multiple clusters separated by small energy barriers on the order of $1-2k_B T$, where k_B represents the Boltzmann constant. These results confirm that the random coil and β -hairpin basins are indeed much wider than the α -helical basin.

D. Free energy barriers

Given that the minimum free energy paths on the free energy surface generated by a steepest-descent search (shown in Figure 2) are consistent with the representative TPS trajectories along the α_{rmsd} and β_{rmsd} order parameters, it is of interest to estimate the free energy barrier involved in the corresponding activated processes. Figure 7 shows the free energy profile along the minimum free energy path of the zipping mechanism. By locating the Transition State Ensemble on that path (structure T1 in Figure 7), the free energy barrier for the transition state, i.e., the activation free energy (ΔG^*) for the folding reaction, is found to be 18.5 ± 1 kJ/mol. That barrier corresponds to a rate constant, k_{TST} of 3.55×10^9 s⁻¹. For reference, five conformational states along the transition path, denoted as (I) to (V) in Figures 2, have also been included in the figure. A similar calculation for the unfolding process from the α -helical state to the unstructured coil is shown in Figure 8. The free energy barrier for this reaction is found to be 26.4 ± 1 kJ/mol, which corresponds to a rate constant, k_{TST} of 1.46×10^8 s⁻¹.

It is important to note that these free energy barriers provide only a lower bound for the true energy barriers of the process, as the order parameters chosen here are not true re-

FIG. 7. Free energy (ΔG) along the minimum free energy path connecting the α ($\chi = 0$) and β ($\chi = 1$) states using the zipping mechanism pathway. The horizontal axis (χ) corresponds to the fraction of total distance traveled along the minimum energy path shown by solid black line in Figure 2. Roman numbers correspond to select points on the free energy map shown in the inset of Figure 2. The configuration T1 corresponds to the transition state ensemble as obtained from TPS simulations. The activation energy (ΔG^*), measured as the maximum free energy difference, is found to be 18.5 kJ/mol.

action coordinates. To quantify the proximity of these order parameters to the actual reaction coordinate, we also used the likelihood maximization approach of Peters *et al.*⁷⁹ Briefly, we calculated in an iterative manner the fractional contributions of our order parameters to the actual reaction coordinate by analyzing the statistics of accepted and rejected trajectories of our TPS simulations. Using this approach, we found that our α_{rmsd} and β_{rmsd} parameters, are part of the reaction coordinate for the zipping mechanism transition of α -helix to β -hairpin, α -helix to random-coil transition and random-coil to β -hairpin transition within confidence limits of 70%, 65%, and 20%, respectively. These results indicate that these order parameters provide a reasonable approximation to the reaction coordinate for the first two transitions, but not for the third one (that between a random coil and the β -hairpin state). Arriving at a better description of the hIAPP conversion from a random coil into a β -hairpin remains an unsolved problem.

FIG. 8. Free energy along the minimum free energy path connecting the α ($\chi = 0$) and unstructured coil ($\chi = 1$) states. The horizontal axis (χ) corresponds to the fraction of total distance traveled along the minimum energy path shown by the solid red line in Figure 2. The configuration T2 corresponds to the transition state ensemble as obtained from TPS simulations. The activation energy (ΔG^*), measured as the maximum free energy difference, is found to be 26.4 kJ/mol.

Finally, we note here that any results from molecular simulations are a reflection of the underlying force field used. This is true for equilibrium properties and for dynamic quantities; indeed, recent studies have emphasized the sensitivity of simulated pathways to the underlying force fields.^{80–83} Along with the force field, the choice of the water model and the corresponding water-protein interactions also play an important role^{52,80,84} in folding processes. In this work we used the GROMOS53a6 force field because previous work on rat amylin has shown satisfactory agreement with experimental infrared spectra. New atomistic force fields and explicit water models continue to be proposed and evaluated,⁸⁵ and we are currently pursuing a systematic comparison of pathways and free energy maps for different models in the context of amylin. The results of that effort will be presented in a future publication.

IV. CONCLUSIONS

Conformational transitions from α -helix rich conformations to β -sheet rich fibrils are a common feature of proteins associated with several diseases. A plausible key step in the aggregation pathway is the formation of an intermediate β -hairpin structure. We have presented a free energy surface in terms of the α -helix and β -hairpin content of the human amylin peptide molecule in explicit water; on that surface, one can identify three metastable states, namely

an α -helical, a β -hairpin, and an unstructured coil. Our unbiased atomistic simulations have shown that the human amylin protein monomer can undergo a distinct transition from an α -helical state to a β -hairpin state. In particular, we have shown that such a transition proceeds through either of two mechanisms: (1) a zipping mechanism⁷⁷ or (2) via the unstructured coil state. In the case of the zipping mechanism, the process is initiated by hydrophobic collapse of residues 16–17 and 23–27, leading to formation of a β -turn. That β -turn consists of residues SER20-ASN21-ASN22-PHE23. Our proposed mechanism is consistent with experiments that have emphasized the importance of residues GLY24, ILE26, and LEU27 in the aggregation propensity of hIAPP.^{20,21} It is also consistent with experiments underscoring the critical role of SER20 in amylin fibril formation.^{23–27}

A gradual loss of α -helical contacts occurs in favor of β -hairpin contacts. In the case of the pathway where the loss of α -helical character occurs via the unstructured coil state, residue 15 rotates towards the interior of the peptide, leading to a subsequent loss of α -helicity. The results have been summarized in Figure 9. The two pathways exhibit a similar free energy barrier in the two dimensional order parameter space of α_{rmsd} and β_{rmsd} (see Sec. II for definitions), 18.5 kJ/mol and 26.4 kJ/mol, respectively. The transition from the unstructured coil state to the β -hairpin state is not found to be activated. Our work underscores the importance of explicit water molecules in the kinetics of the folding process.

FIG. 9. Summary of different folding pathways between the α -helical, the unstructured coil, and the β -hairpin states. In the first pathway (shown in blue), the transition occurs through a zipping mechanism, where there is a gradual loss of α -helical character accompanied by a gradual gain of β -hairpin character. The transition state is represented as T1. The reaction involves a hydrophobic collapse of residues 16–17 and 23–27. The typical time scale of the reaction is on the order of 10 ns in molecular dynamics simulations. The transition state ensemble was generated from 2.5 ns TPS simulations. In the second pathway (shown by red arrows), the transition occurs in two steps. First the α -helical state transforms to an unstructured coil state also within a time scale of approximately 10 ns. The transition state ensemble was determined from 2.5 ns TPS simulations. The transition state is represented as T2. The reaction involves the rotation of residue 15 from the solvent-rich region to the interior of the peptide. The second step involves a transition from the unstructured coil state to the β -hairpin state (dotted red); this latter process is diffusion limited and requires on the order of 300 ns in molecular dynamics simulations.

Implicit solvent models lead to different free energy basins, a different surface, and presumably a different transition path between metastable states.

Our work also highlights the importance of the hydrophobic collapse that leads to the transition, a phenomenon that could be missed in implicit solvent simulations. By putting together the free energy surface and the transition state ensemble generated from transition path sampling simulations, we have provided a detailed picture of the folding/misfolding of human amylin monomers. In particular, our results reveal the detailed role that different residues of the peptide play in conformational transitions between metastable states of the amylin monomer. Such detailed information could be of use in identifying important fragments of the peptide that could be altered or targeted to interfere with the molecule's misfolding. In a separate publication, we explore the effect of mutations of the most important residues from this work on the folding/misfolding of the peptide.

ACKNOWLEDGMENTS

The authors are grateful to Jim Skinner and Marty Zanni for helpful discussions. This work was partially supported by the National Institutes of Health under Grant No. 1R01DK088184.

- ¹A. Clark *et al.*, *Lancet* **330**, 231 (1987).
- ²A. Lorenzo, B. Razzaboni, G. C. Weir, and B. A. Yankner, *Nature (London)* **368**, 756 (1994).
- ³P. P. Westermark, A. A. Andersson, and G. T. G. Westermark, *Physiol. Rev.* **91**, 795 (2011).
- ⁴J. D. Knight, J. A. Williamson, and A. D. Miranker, *Protein Sci.* **17**, 1850 (2008).
- ⁵C. S. Goldsbury *et al.*, *J. Struct. Biol.* **119**, 17 (1997).
- ⁶P. Westermark, Z. C. Li, G. T. Westermark, A. Leckstrom, and D. F. Steiner, *FEBS Lett.* **379**, 203 (1996).
- ⁷J. D. Green, C. Goldsbury, J. Kistler, G. J. S. Cooper, and U. Aebi, *J. Biol. Chem.* **279**, 12206 (2004).
- ⁸P. Cao, A. Abedini, and D. P. Raleigh, *Curr. Opin. Struct. Biol.* **23**(1), 82–89 (2013).
- ⁹C. T. Middleton *et al.*, *Nat. Chem.* **4**, 355 (2012).
- ¹⁰L. Wang *et al.*, *J. Am. Chem. Soc.* **133**, 16062 (2011).
- ¹¹S. Luca, W.-M. Yau, R. Leapman, and R. Tycko, *Biochemistry* **46**, 13505 (2007).
- ¹²A. V. Kajava, U. Aebi, and A. C. Steven, *J. Mol. Biol.* **348**, 247 (2005).
- ¹³S. Bedrood *et al.*, *J. Biol. Chem.* **287**, 5235 (2012).
- ¹⁴J. Saldanha and D. Mahadevan, *Protein Eng.* **4**, 539 (1991).
- ¹⁵J. D. Knight, J. A. Hebda, and A. D. Miranker, *Biochemistry* **45**, 9496 (2006).
- ¹⁶C. Goldsbury *et al.*, *J. Struct. Biol.* **130**, 352 (2000).
- ¹⁷J. D. Knight and A. D. Miranker, *J. Mol. Biol.* **341**, 1175 (2004).
- ¹⁸S. H. Shim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6614 (2009).
- ¹⁹S. A. Jayasinghe and R. Langen, *Biochemistry* **44**, 12113 (2005).
- ²⁰P. Westermark, U. Engstrom, K. H. Johnson, G. T. Westermark, and C. Betsholtz, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5036 (1990).
- ²¹D. F. Moriarty and D. P. Raleigh, *Biochemistry* **38**, 1811 (1999).
- ²²J. Green *et al.*, *J. Mol. Biol.* **326**, 1147 (2003).
- ²³S. Sakagashira *et al.*, *Diabetes* **45**, 1279 (1996).
- ²⁴W. J. Lee *et al.*, *Diabetologia* **44**, 2187 (2001).
- ²⁵S. Gebre-Medhin *et al.*, *Biochem. Biophys. Res. Commun.* **250**, 271 (1998).
- ²⁶P. Cao *et al.*, *J. Mol. Biol.* **421**, 282 (2012).
- ²⁷M. Andreassen *et al.*, *Biochim. Biophys. Acta* **1824**, 274 (2012).
- ²⁸A. S. Reddy *et al.*, *Biophys. J.* **99**, 2208 (2010).
- ²⁹N. F. Dupuis, C. Wu, J. E. Shea, and M. T. Bowers, *J. Am. Chem. Soc.* **131**, 18283 (2009).
- ³⁰N. F. Dupuis, C. Wu, J.-E. Shea, and M. T. Bowers, *J. Am. Chem. Soc.* **133**, 7240 (2011).
- ³¹M. Chopra, A. S. Reddy, N. L. Abbott, and J. J. de Pablo, *J. Chem. Phys.* **129**, 135102 (2008).
- ³²R. Kayed *et al.*, *J. Mol. Biol.* **287**, 781 (1999).
- ³³C. E. Higham, E. T. A. S. Jaikaran, P. E. Fraser, M. Gross, and A. Clark, *FEBS Lett.* **470**, 55 (2000).
- ³⁴A. K. Dunker *et al.*, *J. Mol. Graphics Modell.* **19**, 26 (2001).
- ³⁵E. Jaikaran and A. Clark, *Biochim. Biophys. Acta* **1537**, 179 (2001).
- ³⁶S. B. Padrick and A. D. Miranker, *J. Mol. Biol.* **308**, 783 (2001).
- ³⁷S. Jha, D. Sellin, R. Seidel, and R. Winter, *J. Mol. Biol.* **389**, 907 (2009).
- ³⁸I. T. Yonemoto, G. J. A. Kroon, H. J. Dyson, W. E. Balch, and J. W. Kelly, *Biochemistry* **47**, 9900 (2008).
- ³⁹J. A. Williamson and A. D. Miranker, *Protein Sci.* **16**, 110 (2007).
- ⁴⁰J. R. Cort *et al.*, *Protein Eng. Des. Sel.* **22**, 497 (2009).
- ⁴¹S. M. Patil, S. H. Xu, S. R. Sheftic, and A. T. Alexandrescu, *J. Biol. Chem.* **284**, 11982 (2009).
- ⁴²S. R. Griffiths-Jones, A. J. Maynard, and M. S. Searle, *J. Mol. Biol.* **292**, 1051 (1999).
- ⁴³J. F. Espinosa, F. A. Syud, and S. H. Gellman, *Protein Sci.* **11**, 1492 (2002).
- ⁴⁴B. Ciani, M. Jourdan, and M. S. Searle, *J. Am. Chem. Soc.* **125**, 9038 (2003).
- ⁴⁵R. M. Fesinmeyer, F. M. Hudson, and N. H. Andersen, *J. Am. Chem. Soc.* **126**, 7238 (2004).
- ⁴⁶V. Muñoz, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17646 (2004).
- ⁴⁷A. L. Serrano, M. M. Waagele, and F. Gai, *Protein Sci.* **21**, 157 (2012).
- ⁴⁸H. S. Chung, K. McHale, J. M. Louis, and W. A. Eaton, *Science* **335**, 981 (2012).
- ⁴⁹P. Ferrara, J. Apostolakis, and A. Caflisch, *J. Phys. Chem. B* **104**, 5000 (2000).
- ⁵⁰S. Gnanakaran and A. E. Garcia, *Biophys. J.* **84**, 1548 (2003).
- ⁵¹C. M. Santiveri, M. A. Jimenez, M. Rico, W. F. Van Gunsteren, and X. Daura, *J. Pept. Sci.* **10**, 546 (2004).
- ⁵²R. B. Best and J. Mittal, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11087 (2011).
- ⁵³F. Ding, J. M. Borreguero, S. V. Buldyrey, H. E. Stanley, and N. V. Dokholyan, *Proteins: Struct., Funct., Genet.* **53**, 220 (2003).
- ⁵⁴P. G. Bolhuis, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12129 (2003).
- ⁵⁵C. Dellago and P. G. Bolhuis, *Adv. Comput. Simul. Approaches Soft Matter Sci. III* **221**, 167 (2009).
- ⁵⁶C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, *J. Comput. Chem.* **25**, 1656 (2004).
- ⁵⁷P. Mark and L. Nilsson, *J. Phys. Chem. A* **105**, 9954 (2001).
- ⁵⁸C. Camilloni, L. Sutto, D. Provati, G. Tiana, and R. A. Broglia, *Protein Sci.* **17**, 1424 (2008).
- ⁵⁹R. A. Latour, *Biointerphases* **3**, FC2 (2008).
- ⁶⁰J. L. MacCallum, M. S. Moghaddam, H. S. Chan, and D. P. Tieleman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6206 (2007).
- ⁶¹B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4**, 435 (2008).
- ⁶²T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- ⁶³A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
- ⁶⁴S. Singh, M. Chopra, and J. J. de Pablo, *Annu. Rev. Chem. Biol. Eng.* **3**, 369 (2012).
- ⁶⁵F. Pietrucci and A. Laio, *J. Chem. Theory Comput.* **5**, 2197 (2009).
- ⁶⁶M. Heinig and D. Frishman, *Nucleic Acids Res.* **32**, W500 (2004).
- ⁶⁷Q. L. Yan and J. J. de Pablo, *J. Chem. Phys.* **113**, 1276 (2000).
- ⁶⁸S. Piana and A. Laio, *J. Phys. Chem. B* **111**, 4553 (2007).
- ⁶⁹A. Laio and F. L. Gervasio, *Rep. Prog. Phys.* **71**, 126601 (2008).
- ⁷⁰M. Bonomi *et al.*, *Comput. Phys. Commun.* **180**, 1961 (2009).
- ⁷¹A. S. Reddy, M. Chopra, and J. J. de Pablo, *Biophys. J.* **98**, 1038 (2010).
- ⁷²D. Antoniou and S. D. Schwartz, *J. Chem. Phys.* **130**, 151103 (2009).
- ⁷³P. G. Bolhuis, C. Dellago, P. L. Geissler, and D. Chandler, *J. Phys.: Condens. Matter* **12**, A147 (2000).
- ⁷⁴H. W. Wu, S. M. Wang, and B. R. Brooks, *J. Am. Chem. Soc.* **124**, 5282 (2002).
- ⁷⁵See supplementary material at <http://dx.doi.org/10.1063/1.4798460> for additional information and results.
- ⁷⁶C. L. Perrin, *J. Am. Chem. Soc.* **119**, 8748 (1997).
- ⁷⁷S. B. Ozkan, G. A. Wu, J. D. Chodera, and K. A. Dill, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11987 (2007).
- ⁷⁸G. G. Maisuradze, A. Liwo, and H. A. Scheraga, *J. Mol. Biol.* **385**, 312 (2009).

- ⁷⁹B. Peters and B. L. Trout, *J. Chem. Phys.* **125**, 054108 (2006).
- ⁸⁰S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *Biophys. J.* **100**, L47 (2011).
- ⁸¹W. Han and K. Schulten, *J. Chem. Theory Comput.* **8**, 4413 (2012).
- ⁸²R. B. Best and J. Mittal, *Proteins: Struct., Funct., Bioinf.* **79**, 1318 (2011).
- ⁸³T. J. Lane, D. Shukla, K. A. Beauchamp, and V. S. Pande, *Curr. Opin. Struct. Biol.* **23**, 58 (2013).
- ⁸⁴P. S. Nerenberg and T. Head-Gordon, *J. Chem. Theory Comput.* **7**, 1220 (2011).
- ⁸⁵S. Piana, K. Lindorff-Larsen, and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17845 (2012).