The Sequence-Dependent Unfolding Pathway Plays a Critical Role in the Amyloidogenicity of Transthyretin†

Mingfeng Yang,‡ Boyan Yordanov,§ Yaakov Levy,¶ Rafael Brüschweiler,¶ and Shuanghong Huo*†

Gustaf H. Carlson School of Chemistry and Biochemistry, Clark University, Worcester, Massachusetts 01610, Department of Physics, University of California at San Diego, La Jolla, California 92093, and National High Magnetic Field Laboratory, Florida State University, Tallahassee, Florida 32310

Received May 18, 2006; Revised Manuscript Received August 2, 2006

ABSTRACT: Human transthyretin (TTR) is an amyloidogenic protein whose aggregation is associated with several types of amyloid diseases. The following mechanism of TTR amyloid formation has been proposed. TTR tetramer at first dissociates into native monomers, which is the rate-limiting step in fibril formation. The monomeric species then partially unfold to form amyloidogenic intermediates that subsequently undergo a downhill self-assembly process. The amyloid deposit can be facilitated by disease-associated point mutations. However, only subtle structural differences were observed between the crystal structures of the wild type and the disease-associated variants. To investigate how single-point mutations influence the effective energy landscapes of TTR monomers, molecular dynamics (MD) simulations were performed on wild-type TTR and two pathogenic variants. Principal coordinate analysis on MD-generated ensembles has revealed multiple unfolding pathways for each protein. Amyloidogenic intermediates with the dislocated C strand—loop—D strand motif were observed only on the unfolding pathways of V30M and L55P variants and not for wild-type TTR. Our study suggests that the sequence-dependent unfolding pathway plays a crucial role in the amyloidogenicity of TTR. Analyses of side chain concerted motions indicate that pathogenic mutations on “edge strands” disrupt the delicate side chain correlated motions, which in turn may alter the sequence of unfolding events.

The amyloidoses are diseases characterized by fibril deposition in the extracellular space, where a soluble protein/peptide is transformed into an insoluble fibrillar structure. Recent experiments have demonstrated that, in addition to the 20 or so human proteins directly associated with amyloid diseases, many other proteins can also be induced to form amyloid fibrils in vitro, and thus, amyloid formation may be a property common to natural polypeptide chains (1, 2). Despite the fact that the precursor proteins have diversified sequences and native structures, all amyloid fibrils exhibit a similar overall structure, cross-β (3), suggesting that these proteins may share common structural features upon partial denaturation (4, 5). Therefore, understanding the structural details of the assembly competent intermediates on the denaturing pathway will provide insights into the mechanism of amyloidogenesis in general.

In this work, we focus on human transthyretin (TTR).¹ TTR is a plasma protein responsible for carrying thyroxine in plasma and cerebrospinal fluid and indirect transportation of retinal and does so by binding to retinal binding protein (RBP) (6). The amyloidogenicity of TTR putatively causes three types of amyloid diseases: senile systemic amyloidosis (SSA), familial amyloid cardiomyopathy (FAC), and familial amyloid polyneuropathy (FAP) (7, 8). The structure of TTR has been determined by high-resolution X-ray crystallography (9, 10). In its native state, TTR is a homotetramer with a molecular mass of 55 kDa. Each monomer consists of 127 amino acids arranged in a β-sandwich-like structure comprising β-strands DAGH and CBFE (Figure 1). Two monomers form a dimer by direct hydrogen bonding and water bridges between two adjacent H strands and F strands. Two dimers related by a crystallographic 2-fold axis give rise to a tetramer whose dimer–dimer interface is stabilized predominantly by the hydrophobic interactions.

To date, 85 human variants of TTR have been identified, of which 74 are associated with FAP (11). Among the numerous pathogenic variants, Leu55 → Pro (L55P) is the most amyloidogenic and Val30 → Met (V30M) is the most prevalent. Extensive biophysical studies over the years have gradually elucidated the pathway of TTR amyloid formation. At first, the homotetramer dissociates into native monomers, which is shown to be the rate-limiting step in fibril formation.

† This work was supported by NIH (Grant 1R15 AG025023-01 to S.H.), NSF Grant MCB-0211512 to R.B., and a NSF Major Research Instrumentation grant (DBI-0320875) and partially supported by the Petroleum Research Fund (PRF Grant 39205-G4).
‡ To whom correspondence should be addressed. Fax: (508) 793-7533. E-mail: shuo@clarku.edu.
† Clark University.
§ Current address: Biomedical Engineering Graduate Program, Boston University, Boston, MA 02215.
¶ University of California at San Diego.
‡ Florida State University.
stability and/or dissociation barrier (thermodynamic stability and dissociation kinetics of tetramers affect every stage of amyloid formation. Studies of the partially different (23 alone is not sufficient for amyloid formation (24). Further, though an only subtle structural difference can be implemented in AMBER has been proven to be reliable in generalizing Born (GB) solvation model implemented in AMBER7 (34). Following the designation of Case and co-workers, we used the parm99MOD2, GB\(\text{OBC}\) (I), mBondi2 model. Because GB describes the instantaneous solvent dielectric response, it eliminates the time-consuming step of water reorganization, equivalent to smoothing the energy landscape. In addition, due to the lack of viscosity associated with explicit water molecules, the simulation may quickly converge or explore more conformational space. The drawback of this model is that the simulated events take place on a faster time scale than the MD simulation with explicit water. Implicit solvent models can facilitate folding time by a factor of more than 2 (35). While the discrete representation of water is essential in some cases (36), the GB model implemented in AMBER has been proven to be reliable in several studies, e.g., the pK\(_a\) prediction (37), protein–DNA binding (38), and computational protein design (39). The ionic strength was set to zero. Since TTR is not a highly charged system, the influence of ionic strength on the shape of the energy landscape is not significant. The protonation state of ionizable groups was treated at neutral pH. Like many amyloidogenic proteins, TTR, and especially its amyloidogenic mutants, is very sensitive to a change in pH (40). However, it is very challenging to simulate this pH sensitivity even though there has been some success with constant-pH simulation (41). The experimental pH range to which the simulation corresponds when all of the ionizable groups are

Materials and Methods

Conformational Sampling. The molecular dynamics simulations were started from the monomeric form extracted from the X-ray structures [WT-TTR, PDB entry 1DVQ (28); V30M-TTR, PDB entry 1TTC (29); and L55P-TTR, PDB entry 5TTR (30)]. To be consistent with our previous simulations and other ongoing simulations, we chose the X-ray structure of WT-TTR resolved by Klabunde et al. (28). The crystal structures resolved at difference resolution are almost identical with the maximum C\(_\text{α}\) root-mean-square deviation equal to 0.25 Å. The X-ray unresolved residues, residues 1–9 (of the N-terminus) and 125–127 (of the C-terminus) of WT-TTR and L55P-TTR, were rebuilt with the full-length V30M-TTR structure as a template and refined with a simulated annealing technique. All simulations were performed using AMBER (31) and the modified parm99 force field (32, 33). To enhance sampling and accelerate the simulations, the aqueous environment was mimicked by a generalized Born (GB) solvation model implemented in AMBER7 (34). Following the designation of Case and co-workers, we used the parm99MOD2, GB\(\text{OBC}\) (I), mBondi2 model. Because GB describes the instantaneous solvent dielectric response, it eliminates the time-consuming step of water reorganization, equivalent to smoothing the energy landscape. In addition, due to the lack of viscosity associated with explicit water molecules, the simulation may quickly converge or explore more conformational space. The drawback of this model is that the simulated events take place on a faster time scale than the MD simulation with explicit water. Implicit solvent models can facilitate folding time by a factor of more than 2 (35). While the discrete representation of water is essential in some cases (36), the GB model implemented in AMBER has been proven to be reliable in several studies, e.g., the pK\(_a\) prediction (37), protein–DNA binding (38), and computational protein design (39). The ionic strength was set to zero. Since TTR is not a highly charged system, the influence of ionic strength on the shape of the energy landscape is not significant. The protonation state of ionizable groups was treated at neutral pH. Like many amyloidogenic proteins, TTR, and especially its amyloidogenic mutants, is very sensitive to a change in pH (40). However, it is very challenging to simulate this pH sensitivity even though there has been some success with constant-pH simulation (41). The experimental pH range to which the simulation corresponds when all of the ionizable groups are
protonated may highly depend on the force field. Therefore, we concentrate on the neutral-pH study in this work.

Initially, the system was minimized for 1000 steps and then heated slowly from zero to its target temperature at a rate of 100 K/5 ps. The time step was 1 fs. The temperature of the system was regulated using the Berendsen coupling algorithm (42). The SHAKE algorithm was used to constrain the bond length involving H atoms (43). The long-range interaction was truncated at 15 Å. During the heating process, a restraint with a force constant equal to 5.0 kcal mol\(^{-1}\) Å\(^{-2}\) was added to ensure the conformity between the protein conformation and its initial structure. In the following equilibration, this restraint force was gradually reduced to 3.0 kcal mol\(^{-1}\) Å\(^{-2}\), then to 1.0 kcal mol\(^{-1}\) Å\(^{-2}\), and eventually to 0.1 kcal mol\(^{-1}\) Å\(^{-2}\) for each 50 ps. Finally, an unfolding simulation was performed for 2 ns (2000 conformations) without restraint. Another independent trajectory at the same temperature was collected with the same protocol using a different seed of a random number generator.

Construction of a meaningful energy landscape of a protein requires effective sampling of the conformational space ranging from the native state to the unfolded state. Herein, we explore the unfolding process with molecular dynamics (MD) simulations at high temperatures. Two 2 ns independent trajectories were collected at each of the following temperatures (a total of 36 ns for the wild type and its variants): 350, 375, and 400 K. At 350 K, we found only partial unfolding, in line with the far-UV CD data of WT-TTR (variants): 350, 375, and 400 K. At 350 K, we found only partial unfolding, in line with the far-UV CD data of WT-TTR. The center of each cluster (the conformation with the most neighbors) was gradually annealed to 300 K in 5 ps and then heated slowly from zero to its target temperature at a rate of 100 K/5 ps. The time step was 1 fs. The temperature of the system was regulated using the Berendsen coupling algorithm (42). The SHAKE algorithm was used to constrain the bond length involving H atoms (43). The long-range interaction was truncated at 15 Å. During the heating process, a restraint with a force constant equal to 5.0 kcal mol\(^{-1}\) Å\(^{-2}\) was added to ensure the conformity between the protein conformation and its initial structure. In the following equilibration, this restraint force was gradually reduced to 3.0 kcal mol\(^{-1}\) Å\(^{-2}\), then to 1.0 kcal mol\(^{-1}\) Å\(^{-2}\), and eventually to 0.1 kcal mol\(^{-1}\) Å\(^{-2}\) for each 50 ps. Finally, an unfolding simulation was performed for 2 ns (2000 conformations) without restraint. Another independent trajectory at the same temperature was collected with the same protocol using a different seed of a random number generator.

The center of each cluster (the conformation with the most neighbors) was gradually annealed to 300 K in 5 ps and then subjected to 2000 steps of minimization. The rmsd between these annealed and minimized structures and the initial structures representing the clusters ranges from 1.4 to 4.6 Å. The total annealing process was equivalent to a 9.4 ns MD run, and the total minimization steps were 3.7 \(\times\) 10\(^6\) for the wild type and its variants. Finally, the minimized conformations were used to construct the energy landscape. As a result, the redundancy in the MD-generated conformations was effectively removed without affecting the topology of the mapped energy landscape.

Principal Coordinate Analysis. To visualize the energy landscape, principal coordinate analysis (PCoorA) (45) was used to project conformations onto a two-dimensional plane. This method has been shown to effectively preserve the essential conformational subspace (46, 47). First, a distance matrix \(A\) with element \(A_{ij} = -1/d_{ij}\) was set up, where \(d_{ij}\) is the distance between conformations \(i\) and \(j\); namely, \(d_{ij}\) is a similarity measurement, and the distance matrix represents the pairwise similarity between conformations in an ensemble. Then, matrix \(A\) was “centered” by

\[
A_{ij}^* = A_{ij} - \langle A_{ij} \rangle_i - \langle A_{ij} \rangle_j + \langle A_{ij} \rangle_{ij} \tag{1}
\]

where \(\langle \ldots \rangle_i\) is the mean over all specific indices \(k = i, j, \text{ and } ij\). This centering process guarantees a zero root of matrix \(A^*\). In this work, we had a 603 \(\times\) 603 matrix, a 696 \(\times\) 696 matrix, and a 571 \(\times\) 571 matrix for the wild type, L55P-TTR, and V30M-TTR, respectively. Finally, centered matrix \(A^*\) was diagonalized. The resulting eigenvectors formed a new coordinate set on which the corresponding eigenvalues (normalized) represent the percentage of the projection of the original distribution. One natural advantage of PCoorA is that the average projection error \(\delta\) can be quantitatively calculated from the normalized eigenvalues, \(\lambda\) (45). When the conformations are projected onto the first \(s\)-dimensional subspace, the average deviation of the 3N-dimensional distance \(d_{ij}\) from the distance \(d_{ij}^s\) calculated in the \(s\)-dimensional subspace is given by

\[
\delta = 1 - \sum_{i=1}^{s} \lambda_i = \langle d_{ij}^2 \rangle - \langle (d_{ij}^s)^2 \rangle_{ij} \tag{2}
\]

where \(\langle \ldots \rangle_{ij}\) is the average over all pairwise distances in the ensemble.

The choice of the measurement of similarity is crucial in PCoorA, because it has a significant effect on the resulting projection. In the original work of Becker (45), rmsd was used to describe the similarity of pairwise conformations \(d_{ij} = \text{rmsd}(i,j)\). In assessment of the shape of an energy landscape, it is important that \(d_{ij}\) be well-correlated with the effective energy. However, it has been reported that rmsd does not correlate with energy very well (48, 49). In this work, we used “power distance” as the measure of similarity, which is defined as

\[
d_{ij} = \frac{\sum_{a<b} |(r_{ab}^i)^{m} - (r_{ab}^j)^{m}|}{\sum_{a<b} |(r_{ab}^i)^{m} + (r_{ab}^j)^{m}|} \tag{3}
\]

where \(i\) and \(j\) designate two conformations, \(r_{ab}\) is the distance between atoms \(a\) and \(b\), and \(m\) is a parameter to control the relative weight of small distances (49). It has been shown that the power distance provides a better correlation with energy and is therefore more suitable for investigating the shape of the energy landscape (49).

Isotropically Distributed Ensemble Analysis. Isotropically distributed ensemble (IDE) analysis (50) was used to characterize the correlated molecular motions of the wild type and its variants in their native states. Details of the method have previously been described (50). In this work, the motions of backbones and side chains were evaluated separately, and the positions of backbone and side chain units of a residue were represented by their corresponding center of mass. For a protein containing \(n\) amino acids, we first constructed an \(n \times n\) matrix \(P\) with elements \(P_{ij} = \langle f_i|f_j\rangle\), where \(f_i\) is the position vector of the backbone or side chain unit of residue \(i\) originated from the center of mass of the protein. The angular bracket denotes averaging over the trajectory. The cross-correlation coefficient \(r_{ij}\) of residues \(i\) and \(j\) is defined as \(P_{ij}/(P_{ii}P_{jj})^{1/2}\) with the overall rotational modes to be eliminated. If \(r_{ij} = 1\), two residues \(i\) and \(j\) are highly correlated. In contrast, an \(r_{ij} = -1\) indicates strong anticorrelated motion. The IDE analysis was performed on
the monomeric conformations previously simulated at 300 K using the same initial structures and force field as this work (24).

RESULTS

Energy Landscapes of WT-, V30M-, and L55P-TTR. The energy landscape was constructed by depicting the effective energy (potential energy of the protein plus the solvation free energy estimated by GB) as a function of conformational similarity represented by the first two principal coordinates of the system, \( q_1 \) and \( q_2 \). Along \( q_1 \), the variance between all of the pairwise conformations is the largest, and along \( q_2 \), the variance is the second largest. The average accuracy of the projection can be evaluated according to the accumulative sum of the eigenvalues (normalized) (45). Figure 2A shows that only the first 50 dimensions can retain more than 70% of the original distribution; in other words, the error caused by projecting the full dimensional space onto the 50-dimensional subspace is less than 30%, which is consistent with earlier analyses of peptide systems (45–47). However, the accuracy is sensitive to the measure of similarity. For the power distance used in this study, the accuracy decreases as more weights are given to smaller distances (larger \( m \)); when \( m \) is equal to 6, the first two dimensions capture only \( \sim 43\% \) of the actual information, whereas when \( m \) is 3, the accuracy increases to \( \sim 48\% \). The overall topologies of the distributions, however, remain similar. As shown in Figure 2B, the sampled conformations are distributed along two distinct pathways in the first three dimensions regardless of the \( m \) value. Noteworthy is the fact that the two pathways are still separated even if the conformations are projected onto the first two-dimensional plane. The energy landscapes presented below are constructed on the basis of \( m \) equaling 3. Table 1 lists the accuracy of the projection of each protein onto the first two- and three-dimensional subspace. The 2D projections capture \( 47–60\% \) of the normalized sum of all eigenvalues, and the 3D projections capture \( 54–65\% \) of the sum for the three proteins, indicating that the projections are, at the very least, qualitatively representative.

Figure 3 depicts the energy landscapes for the wild type, V30M-TTR, and L55P-TTR. On the three energy landscapes, the partially unfolded conformations are connected along multiple pathways to the native state basin (N in Figure 3) that possesses the lowest effective energy. The native conformations deviate slightly from the corresponding monomeric form extracted from the X-ray structure, with the
Strands, and then advances toward multiple directions; the unfolding of other edge strands, including the H and F of the secondary structure of the D strand followed by the unfolding of all three proteins is initiated by the disruption that found in the native tetramer (23). Moreover, given the fact that extensive intermolecular hydrogen bonding and hydrophobic interactions between monomers in the tetrameric state are absent in the isolated monomer, such a minute structural difference is within expectation.

It should be noted that the presented energy landscapes reflect only the overall shape of the effective energy surface, lacking the barrier information (45). The empty spaces between basins correspond to poorly sampled regions and associate with high energy barriers (45). If one starts from the native conformation, WT-TTR unfolds along three distinct pathways (I, II, and III). Pathway I is far from the other two in the first two principal coordinates, while pathways II and III are adjacent to each other, allowing certain transitions between them (Figure 3A). The two second-lowest energy basins, basin 4 on pathway I and basin 12 on pathway III, are energetically similar to the native state basin. To characterize the structural changes along the pathways, we analyzed the \( \beta \)-sheet content as a function of the first two principal coordinates. As shown in Figure 4A, the number of residues pertaining to \( \beta \)-sheets decreases from ca. 50 to ca. 30 along each of the three pathways and finally to ca. 10. The effective energy correlates with the secondary structure contents very well in that the two second-lowest energy basins in Figure 3A correspond to the two second-highest \( \beta \)-sheet contents in Figure 4A. In contrast to that of the wild type, the effective energy landscapes of the V30M and L55P mutants (panels B and C of Figure 3, respectively) contain only two unfolding pathways for each. For V30M-TTR, pathways I and II totally overlap at the outset and start to diverge at basin 3 (Figure 3B), while the two pathways of L55P-TTR are distinct from the very beginning (Figure 3C). The effective energy increases sharply along the pathways of V30M-TTR. Correspondingly, the \( \beta \)-sheet contents are quickly lost along both pathways of V30M-TTR (Figure 4B). Whereas the decrease in \( \beta \)-sheet content along pathway II of L55P-TTR is significantly smaller than that of pathway I (Figure 4C), which is consistent with the topology of the effective energy landscape in that three second-lowest energy basins, basin 6–8 in Figure 3C, are present along pathway II while the minima along pathway I possess much higher energies than the native state. The difference between the topologies of the effective energy landscapes of the three TTR proteins suggests that single-point mutations indeed influence the mechanism of TTR unfolding, and in different manners.

Unfolding Pathways of TTR Are Sequence-Dependent. The representative snapshots along the unfolding pathways are presented in Figure 5. If one starts from the native structures, the unfolding of all three proteins is initiated by the disruption of the secondary structure of the D strand followed by the unfolding of other edge strands, including the H and F strands, and then advances toward multiple directions; the remaining stable core mostly consists of the N-terminus of the C strand and the BE strands before the protein completely unfolds. However, the detailed unfolding processes of the three proteins are significantly different from each other as reflected by the diverse on-pathway intermediate states.

For WT-TTR (Figure 5A), the DAGH sheet unfolds from the edge strands, the D and H strands, along pathway I quickly and becomes random coils, whereas the CBEF sheet remains almost intact (basin 3 is the most important intermediate conformation along pathway I). The most stable core is the N-terminus of the C strand and the BE strands (Figure 5A, 5). Pathways II and III stay in proximity and are nearly parallel, allowing transitions between one another. Pathway II begins with the disruption of the H strand and the D strand is followed by the elongation of the D strand toward the upstream of the protein main chain. After then, the elongated D strand across the sheet joins the CBEF strands to form the D’CBEF sheet yet still keeps contact with the A strand (Figure 5A, 6). As a result, the AG strands are dragged to rotate clockwise relative to the newly formed D’CBEF sheet. The F strand then unfolds, and the D strand moves away from the A strand completely (Figure 5A, 7). Because of the conformational similarity between conformers 8 and 12, pathways II and III are observed to converge on the energy landscape. The unfolding of the newly formed
Figure 5: Representative conformations sampled on the unfolding pathways: (A) WT-TTR, (B) V30M-TTR, and (C) L55P-TTR. The numbering scheme is the same as that used in Figure 3. The α-helix and eight β-strands are colored as follows: blue for A (residues 12–18), red for B (residues 29–35), green for C (residues 41–48), yellow for D (residues 54 and 55), black for E (residues 67–73), magenta for the helix (residues 74–81), brown for F (residues 91–97), olive for G (residues 104–112), and orange for H (residues 115–123).
D’CBEF sheet continues after the AGH sheet becomes random coil completely (Figure 5A, 9). Unlike pathway II, the initial event on pathway III involves the separation of the D strand from the AGH strands and the unfolding of the F strand, instead of the H strand as on pathway II (Figure 5A, 10). Next, the AGH sheet drifts to the right, farther from the D strand that begins to interact with the C strand (Figure 5A, 11). Meanwhile, the AGH sheet undergoes unfolding and refolding and rotates clockwise relative to the new D’CBEF sheet (Figure 5A, 12). So far, pathway III reaches the intermediate state as seen in pathway II (Figure 5A, 8), where the D’CBEF sheet formed by the elongated D strand and the original CBEF sheet is packed with the AG strands (Figure 5A, 12). After the AGH sheet unfolds mostly (Figure 5A, 13), the remaining stable core consists of the BE strands and the N-terminal part of the C strand (Figure 5A, 5).

The unfolding pathway of V30M-TTR is different from that of the wild type in terms of the intermediate state conformations. After the D strand of V30M-TTR moves away from the A strand, it shifts two residues upstream of the protein and interacts with the C strand (Figure 5B, 2). The shift in β-strand D was also observed in the highly amyloidogenic TTR triple mutant G53S/E54D/L55S-TTR (51). Then the unfolding pathways diverges as soon as the H and C strands start to unfold from their C-termini (Figure 5B, 3). The C strand—loop—D strand motif completely dislocates from the core of the protein along pathway I and becomes a random coil (Figure 5B, 4), followed by the unfolding of the B strand with the AG and EF strands to form two small β-sheets (Figure 5B, 5) before the protein unfolds completely (Figure 5B, 6). Along pathway II, only the C-terminus of the C strand, the CD loop, and the D strand become a random coil (Figure 5B, 7), while the N-terminus of the C strand remains in close contact with the B strand. The unfolding process continues in the AG strands (Figure 5B, 8), resulting in the final conformation (Figure 5B, 9) which is similar to the unfolding product of WT-TTR (Figure 5A, 5).

Unfolding pathway I of L55P-TTR resembles to a great extent the unfolding pathway II of V30M-TTR because both pathways exhibit the shift in the β-strand D and the following D’CBEF sheet (Figure 5C, 2) as well as the partially dislocated C strand—loop—D strand motif (Figure 5C, 3), whereas along pathway II neither the completely nor the partially dislocated C strand—loop—D strand intermediate is present. Instead, the H strand unfolds at a very early stage. The newly formed D’CBEF sheet packs with the AG sheet (Figure 5C, 6), and then the AG sheet unfolds completely (Figure 5C, 7), resembling the conformation in basin 9 on pathway II of WT-TTR.

Val30 → Met and Leu55 → Pro Mutations Disrupt the Side Chain Correlated Motions. To understand how the pathogenic mutations that play a crucial role in the unfolding process affect the concerted motions in the native monomeric state, we have investigated the correlated motions of the monomers in their native states. Previous studies have shown that for simulations of native states, regions of secondary structure were found to move in a correlated matter, whereas the significant anti-correlated motions for residues close in space were predicted to be destabilizing to protein structure (24, 52). As shown in Figure 6, the neighboring strands in the same β-sheet move in a correlated manner, with the exception of no correlated motion between the D and A strands that are thus proven to be weakly coupled. The patterns of the correlated motions between the side chains are similar to that of the backbones, though weaker. In other words, the number of side chains with a correlation coefficient greater than 0.5 is less than that of the backbones. The correlations between the backbones of neighboring strands are very strong for WT-TTR (Figure 6A, upper triangle), and the V30M and L55P mutants exhibit almost identical patterns (data are not shown). Significant differences, however, are observed in the side chain correlated
motions between the C and B strands, between the B and E strands, and between the E and F strands. For WT-TTR, the side chain motions of ten and nine pairs of residues in the CB strands and BE strands show large positive correlations ($r_{ij} > 0.5$), respectively. These correlations are so strong that the C and E strands are also positively correlated. For both V30M- and L55P-TTR, nevertheless, the side chain motions of only four pairs of residues in the CB strands are identified to be correlated, and three pairs between the B and E strands. These data suggest that the Val30 → Met substitution located at the N-terminus of the B strand disrupts the concerted motion of its surrounding side chains, and this disruption propagates through the CBEF sheet with no obvious effects on the main chain correlated motion. Coincidentally, residue 55 is in the proximity of the CB strands. Therefore, this substitution leads to a similar disruption of the side chain correlated motions.

**DISCUSSION**

**Exposure of the AB Strands for Aggregation.** Studies on the TTR fibril structures have revealed that the A, B, and E−H strands retain their nativelike conformation, and the aggregated intermolecular interface is formed between strands A and A′, B and B′, F and F′, and H and H′ with the C strand−loop−D strand region protruding from the fibril core (18−20). The formation of a new interface in the fibril state between the A and A′ strands as well as that between the B and B′ strands requires exposure of the AB strands that are protected by the CD strands in the native monomer. As seen in Table 2, the disruption of the D strand takes place in the first step of unfolding in each protein, where the D strand moves apart from the A strand and adopts a surface loop. This initial step was also reported in our previous simulation results and directly validates early speculation that the separation of the D strand from the core of the protein initiates the conformational changes and facilitates amyloidogenesis (24, 25). Afterward, the D strand can be re-formed and stabilized by moving to the side of the CBEF sheet via the close interaction with the C strand. The following unfolding of the C strand occurs in the fashion of unzipping from the B strand starting at its C-terminus, and the N-terminus unzips in a later stage. This observation is not surprising because the N-terminus of this strand is stabilized by three pairs of strong hydrophobic contacts, Val32 with Phe44, Lys35 with Trp41, and Phe33 with Pro43, whereas the C-terminus of this strand lacks this kind of contact (Figure 1). This unfolding sequence is in line with the hydrogen exchange experiments. The C-terminal residues of the C strand are found to be weakly protected, and the N-terminal W41 and E42 residues are highly protected (53−55).

As pointed out by Richardson et al., Pro43 and the β-bulge at Phe44 and Ala45 located in the C strand had been selected by evolution to protect the native state against edge-to-edge aggregation (56). Even after the AB strands are exposed, the protruding AB loop can still prevent the association of A and A′ strands as well as the interaction of B and B′ strands, which is probably nature’s negative design, too. In contrast to the AA′ and BB′ interface, the negative design at the edges of the FH strands involves only the protective charged residue, Arg103, in the FG loop, which is relatively weak. The HH′ and FF′ interface can be formed with fewer conformational changes (20, 24). It is tempting to assume that the formation of a nativelike interface between the H and H′ strands as well as between the F and F′ strands takes place earlier in the fibril formation than that between the A and A′ strands as well as between the B and B′ strands. In the experiments at high concentrations, the HH′ and FF′ interfaces may form before the monomer continues to unfold, and the newly formed interface can in turn stabilize other strands. Therefore, the subsequent unfolding as summarized in Table 2 that suggests the B and E strands could be more stable than the A, G, and H strands, leading to the N-terminus of the C strand and the BE strands as the final core, may be not relevant to amyloid formation. Another scenario is that the HH′ interface is stabilized by the α-sheet structure as proposed previously (57, 58), but we have never seen an α-sheet with an AMBER/GROMOS model. There is strong evidence that surface charges can exert a profound influence on the stability of the protein (59). As shown in Figure 1, there are four titratable groups under acidic conditions, His31, Glu42, Glu72, and Glu92, on the CBEF sheet, but only two such residues, Asp18 and Glu54, exist on the DAGH sheet. Additionally, if the N- and C-termini are excluded, seven more such residues, Asp38, Asp39, Glu66, Asp74, His88, Glu89, and His90, are located in the loop regions on the CBEF side, while only two more such residues, Glu51 and His56, are in the loop regions on the DAGH side. Therefore, the selective destabilization of one half of the β-sandwich (54), the CBEF sheet, can be simply explained by the fact

<table>
<thead>
<tr>
<th>initial disruption of the D strand</th>
<th>subsequent disruption</th>
<th>intermedidates</th>
<th>final state</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT-I</td>
<td>a-10</td>
<td>a-6/b-2</td>
<td>a-3</td>
</tr>
<tr>
<td>WT-IIb</td>
<td></td>
<td>b-4/b-7</td>
<td>a-7</td>
</tr>
<tr>
<td>WT-IIc</td>
<td></td>
<td>c-4</td>
<td>c-4</td>
</tr>
<tr>
<td>V30M-I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V30M-IIb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L55P-I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L55P-IIb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a-6/b-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b-4/b-7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a-7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c-4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a-5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The typical conformations corresponding to Figure 5 are listed. The strands are named from A to H. Boldface and italic type indicates the complete unfolding of the corresponding strand(s). D′ denotes the re-formed D strand on the CBEF side, D, and D denote the elongated and shifted D strand, respectively. Cs denotes the N-terminus of the C strand. The square root sign denotes the existence of an event. Roman numbers refer to the pathways shown in Figure 5.
that the CBEF sheet contains more titratable groups under acidic conditions than the DAGH sheet, instead of the α-sheet as previously suggested (57, 58).

The Dislocated C Strand–Loop–D Strand Intermediate Is on the Unfolding Pathways of V30M-TTR and L55P-TTR but Not on That of the Wild Type. It is believed that the amyloid fibrils of TTR result from the self-assembly of amyloidogenic intermediates that in turn are formed by partially denatured monomers. The purpose of carrying out PCoorA is to capture the partially denatured monomeric intermediates that may lead to an amyloid deposit. The mapped unfolding pathways are summarized in Table 2, among which the intermediate states with the C strand–loop–D strand partially or completely dislocated are the most interesting. The C and D strands are torn apart from the rest of the protein and turned into random coil, whereas the A, B, and E–G strands retain their secondary structure and tertiary structure (4 and 7 in Figure 5B and 3 in Figure 5C).

We call the structures with these characteristics the dislocated C strand–loop–D strand intermediate. This intermediate is captured on the unfolding pathways of both V30M and L55P mutants, but not on the pathway of WT-TTR. This observation is supported by the fact that the wild-type monomer is nonamyloidogenic unless the pH is reduced (23), whereas the V30M and L55P mutants can form amyloid protofibrils under physiological conditions after a long incubation (40, 60). The wild-type monomeric intermediate which was capable of amyloid deposits at low pH was found to be predominantly a β-sheet structure with conformational rearrangements in the C strand–loop–D strand region without a large exposed hydrophobic surface area (16). The conformations of the intermediate states identified in this work are consistent with all of these structural features characterized experimentally. The exposure of the C strand–loop–D strand region was also proposed to be in the initial phase of amyloid formation on the basis of the epitope mapping of TTR mutants that exhibit amyloidogenic folding (17). Therefore, we conclude that the C strand–loop–D strand dislocation can be facilitated either by the pH environment or by point mutations. Further along the unfolding pathway after the C strand–loop–D strand dislocation, the F strand unfolds, leaving the AG strands and the BE strands to form two two-strand β-sheets (8 in Figure 5B and 4 in Figure 5C). Such an unfolding sequence is consistent with the earlier D–H exchange experiments on V30M and L55P mutants in the native state at neutral pH that revealed a stable core consisting of strands A, B, E, and G and the AB loop (55).

For the unfolding of the wild type, its C strand is stable until most residues of the molecule lose their tertiary structure. Thus, no dislocated C strand–loop–D strand intermediates are produced. It has been shown that the rate-limiting step of TTR amyloid formation is the dissociation of a tetramer into monomers (12, 13). Surprisingly, the V30M-TTR tetramer dissociates slower than the wild type (12), but it is V30M-TTR instead of the wild type that forms amyloid fibril after a long incubation at neutral pH (40, 60). Therefore, the denaturation pathway of monomers plays a crucial role in the amyloidogenic potential besides the tetrameric dissociation thermodynamics and kinetics. Though wild-type transthyretin is nonamyloidogenic under neutral-pH conditions, it has been identified as the components of amyloid fibrils related to SSA (8). Considering SSA appears only in the aged population, we speculate that the physiological changes, e.g., pH, with age may cause the deposition of wild-type TTR, which is supported by the fact that acidic conditions can facilitate WT-TTR fibril formation (61).

The Sequence-Dependent Unfolding Pathways May Originate from the Disruption of Side Chain Correlated Motions. Our analyses of the concerted motions near native states have revealed the substantial difference in the side chain dynamics between the wild type and its pathogenic variants which may account for the variance of their unfolding pathways. The side chain motions of the CB strands in the wild type are strongly correlated, as are those of the BE and EF strands. The greater amount of collective fluctuations that correspond to contact interactions can be considered to reflect a larger stability. With the V30M mutation in the B strand, such correlations are significantly reduced, reflecting that the stability of the CBEF strands is lowered. For the same mutant, we have uncovered the dislocated C strand–loop–D strand intermediates on the unfolding pathway. To our initial surprise, similar reductions are also seen with the Leu55 → Pro mutation that takes place in the D strand. This could be attributed to the fact that side chain 55 is in the proximity of the CB strands. The statistical analysis of the different substitutions of TTR has pointed out that the amyloidogenic variants promote the growth of an intermolecular β-sheet via altering structure or stability of the “edge strands” (62).

Our calculations on the concerted motions have complemented the early analyses and linked the disruptions of the delicate side chain concerted motions in the edge strands to the unfolding pathway that leads to the formation of amyloidogenic intermediates. Lines of evidence support the hypothesis that collective atomic fluctuations are linked to binding affinity and specificity (63). Herein, we further hypothesize that the collective atomic fluctuations in the native state are linked to an unfolding mechanism. Of course, more work needs to be done to test the hypothesis.

In summary, we projected partially unfolded ensembles of the wild type and its two pathogenic variants, V30M- and L55P-TTR, onto the two-dimensional plane and visualized their unfolding pathways. The amyloidogenic intermediates with the dislocated C strand–loop–D strand motif were identified only on the unfolding pathways of the pathogenic variants. This observation suggests that the effect of the pathogenic mutations may be to alter the denaturation pathways, which in turn arise from the disruption of side chain correlated motions.

ACKNOWLEDGMENT

The molecular graphics images were generated using the Chimera package (64) from the Computer Graphics Laboratory, University of California, San Francisco, CA (supported by NIH Grant P41 RR-01081). We thank the National Center of Supercomputing Applications and the Scientific Computing and Visualization group at Boston University (Boston, MA) for providing part of the computational resources.

REFERENCES


BI0609927