

Molecular Dynamics Simulation of the SH3 Domain Aggregation Suggests a Generic Amyloidogenesis Mechanism

Feng Ding¹, Nikolay V. Dokholyan^{2,3*}, Sergey V. Buldyrev¹
H. Eugene Stanley¹ and Eugene I. Shakhnovich²

¹Center for Polymer Studies
Department of Physics, Boston
University, Boston, MA 02215
USA

²Department of Chemistry and
Chemical Biology, Harvard
University, Cambridge, MA
02138, USA

³Department of Biochemistry
and Biophysics, School of
Medicine, University of North
Carolina at Chapel Hill, Chapel
Hill, NC 27599, USA

We use molecular dynamics simulation to study the aggregation of Src SH3 domain proteins. For the case of two proteins, we observe two possible aggregation conformations: the closed form dimer and the open aggregation state. The closed dimer is formed by “domain swapping”—the two proteins exchange their RT-loops. All the hydrophobic residues are buried inside the dimer so proteins cannot further aggregate into elongated amyloid fibrils. We find that the open structure—stabilized by backbone hydrogen bond interactions—packs the RT-loops together by swapping the two strands of the RT-loop. The packed RT-loops form a β -sheet structure and expose the backbone to promote further aggregation. We also simulate more than two proteins, and find that the aggregate adopts a fibrillar double β -sheet structure, which is formed by packing the RT-loops from different proteins. Our simulations are consistent with a possible generic amyloidogenesis scenario.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: amyloidogenesis; discrete molecular dynamics; dimerization; amyloid fibril; aggregation

*Corresponding author

Introduction

Amyloid fibril is the insoluble aggregate of usually solvable proteins or polypeptides.^{1,2} There are 16 types of human diseases known to be associated with amyloidogenesis.^{1,2} Amyloid fibrils are straight, unbranched, usually 70–120 Å in diameter, and several thousand angstroms in length. Observed types of amyloid fibrils consist of different precursor proteins which share no sequence or structure similarity. However, different types of amyloid fibrils explored by X-ray diffraction^{3–5} show some common core structural features: the presence of a 4.7 Å inter-strand spacing along the fibril axis and a 9–10 Å inter-sheet spacing perpendicular to the axis. Recently, a group of proteins unrelated to any human disease were found to be able to form amyloid fibril structures *in vitro* under denaturing conditions.^{6,7} The ability of proteins with different sequences and native structures to form similar amyloid fibrils suggests that amyloidogenesis is a common feature of proteins in denaturing conditions.⁸

Although many advances have been made in structural characterization of amyloid fibrils and in understanding the mechanisms of their formation, many aspects of this process remain unclear. Due to difficulties in crystallizing amyloid fibrils, the detailed intrinsic structure has yet to be determined from X-ray diffraction. Lack of knowledge of the detailed structure of amyloid fibril makes it difficult to understand aggregation mechanisms. Other experimental techniques have been applied to understand the structure of amyloid fibril core.^{9,10} Designing different fragments of amyloid β -peptide (A β)¹⁰ that can aggregate into fibrils similar to those formed by wild-type peptides shows that residues 14–23 are the basic “bricks” comprising A β amyloid fibrils. However, no direct observations of the amyloid fibril core structures have been reported. The domain swapping mechanism¹¹ posits that two or more protein chains exchange identical domains including a helix, a loop, a single β -strand or an entire domain to form a strongly bound oligomer. In a propagational instead of reciprocal manner, the domain swapping mechanism explains the elongation of amyloid fibrils.^{12–15} However, the domain swapping hypothesis is based on the aggregation of only two proteins into

E-mail address of the corresponding author:
dokh@med.unc.edu

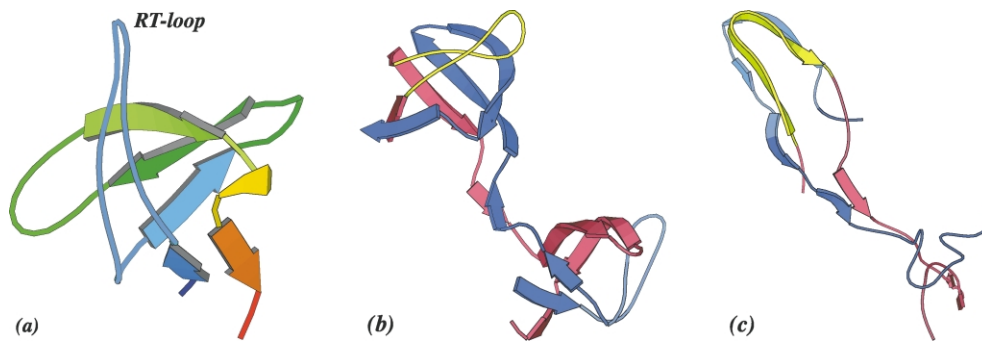


Figure 1. Dimerization of Src SH3 domain. (a) The native state of Src SH3 domain. Molecular dynamics simulations yield two types of aggregates: (b) the closed dimer formed by exchanging RT-loops and (c) an open aggregation state formed by swapping two parts of RT-loop from different proteins. In (b) and (c) the first protein is red and its RT-loop is yellow, while the second protein is blue and its RT-loop is cyan. The pictures are produced with Molscrip.³⁴

dimers, so the mechanism for fibril formation from more than two identical proteins is still unclear.

Due to limitations in computation power to study large protein systems in molecular dynamics simulations, we employ the discrete molecular dynamics algorithm^{16,17}—a computationally fast and dynamically realistic simulation technique for investigations of the protein folding thermodynamics¹⁷ and kinetics.^{18,19} We study the aggregation of Src SH3 domain (PDB entry 1NLO), a globular protein, consisting of 56 amino acids, extensively explored in experiments^{20,21} and computer simulations.²² The longer SH3 fold family homologue PI3-SH3 has been experimentally shown to aggregate into amyloid fibrils under acidic conditions.^{6,23} PI3-SH3, a 84 residue protein with the insertion of a long helical loop between β_2 and β_3 , shares the same fold as Src SH3 domain. The rmsd between the structure of Src SH3 and PI3 SH3 is 1.04 Å with 48 amino acids used for alignment.²⁴ Experimental study of PI3-SH3²⁵ shows a slow refolding with the time constant 2.8 seconds in water, while the folding kinetics still follows a two-state folding scenario. It is a challenge to simulate such a slow folding protein. With a modified $G\ddot{o}$ model, we find the 84 residue SH3 domain follows a two-state folding transition and our simulation of two PI3-SH3 proteins shows the same aggregation scenario as Src SH3 domain (data not shown). However, the requirement to simulate more than two proteins in the aggregation study is extremely time-consuming for PI3-SH3. Thus, we use Src SH3 domain as the model system to study the amyloidogenesis process.

The ability of proteins with different sequences to aggregate into common amyloid fibrils suggests that non-specific hydrogen bonding between the main chain carbonyl oxygen and the amide nitrogen may play an important role in amyloid formation. Here we use the coarse-grained protein model as described by Ding *et al.*,²² with a native-state specific $G\ddot{o}$ interaction potential between C_β atoms representing the side-chains, and non-specific interactions between C_α atoms representing the hydrogen bonding interactions between the backbones of proteins.

The folding kinetics of our Src SH3 domain model, a C_α - C_β model with a $G\ddot{o}$ interaction between side-chains, has been studied by discrete molecular dynamics and has shown agreement with experimental observations.²² We simulate the model proteins near the folding transition temperature, T_f , where the protein has a high probability to be found in the partially folded states. It is most likely to observe amyloid fibril formation under these conditions because complete unfolding was shown to lead mostly to amorphous aggregation²³ while highly stable proteins do not aggregate at all.

The Src SH3 domain under study consists of five β -strands and a long RT-loop (see Figure 1(a)). It was observed that the RT-loop plays a critical role in the folding kinetics of Src SH3 domain despite the low experimental ϕ values.²² Experiments and simulations reveal that the RT-loop is flexible. However, the RT-loop itself is stable and persists in the partially folded states. The unfolding/folding events correspond to the opening/closing of the RT-loop with respect to the rest of the protein.^{22,26} Accordingly, we expect that the RT-loop may play an important role in the amyloid formation of Src SH3 domain.

Dimerization

First, we study the dimerization of two identical Src SH3 domains. The two proteins are confined to a cubic cell with length of 150 Å. Starting from fully unfolded states, we perform simulations to different temperatures. We observe an aggregation temperature threshold, T_a , below which we find aggregation, and $T_a \approx 1.03$ is only slightly higher than folding transition temperature $T_f \approx 0.95$. Ordered aggregations only occur near T_f . At T_f , the time needed for aggregation, τ_a , is of the order of 10^4 time units. This time is significantly smaller than the time needed for a single protein to fold into native state, τ_f , which is in the order of 10^5 time units, indicating that the kinetic barrier for the two-protein system to aggregate is much smaller than the folding barrier of each individual

protein. As temperature decreases, τ_a increases and τ_r decreases. Below a certain temperature threshold $T_c = 0.85$, the two time scales are comparable and we observe the separate folding of the two proteins without dimerization.

For ordered aggregation, we observe a closed form dimer structure by domain swapping (see [Figure 1\(b\)](#)), where the two proteins exchange their RT-loops. We also observe an open aggregation state ([Figure 1\(c\)](#)), which relates to the packing of RT-loops. Unlike the usual domain swapping where the swapped part interacts with the complementary domain from a different protein, in our open aggregation state the first/second strand of the RT-loop from one protein forms contacts with the second/first strand of the RT-loop from another protein. During this process, amino acids in the RT-loops reorient their side-chains. The original interactions that stabilize the RT loop are replaced by similar contacts between the complementary strands of RT-loops from different proteins. Stabilized by hydrogen bonds along the backbone, RT-loops form a β -sheet structure ([Figure 1\(c\)](#)).

The closed dimer has a stable structure with lower potential energy than the more flexible open structure. However, the probability to observe the closed form is lower than the open aggregation state because the entropy of the open structure is higher than that of the closed form. As temperature decreases, the probability to form closed dimers increases. We observe that when the quenching temperature drops below $T_c = 0.85$, the two proteins fold separately, avoiding aggregation. We also find that aggregation process depends on the diffusion coefficient D and the density of proteins ρ : as D and/or ρ increase, T_c decreases, and, therefore the aggregation becomes more likely (data not shown).

The closed dimer has a well-defined 3D structure with the hydrophobic core buried inside, so the closed dimer cannot further aggregate into elongated amyloid fibril. However, according to amyloidogenesis hypothesis,^{12,13} domain swapping may lead to elongated amyloid fibrils if the swapping is not reciprocal but propagational. The open aggregation state is more flexible and has the hydrophobic core exposed. The closely packed RT-loops form a β -sheet structure stabilized by hydrogen bonding interactions and the two exposed ends can accept further condensation to form the fibril structure. In order to test which process leads to the amyloid fibril formation for Src SH3, we study aggregation of more than two proteins in molecular dynamics simulations.

Amyloidogenesis

From the simulation of two proteins, we find that the optimal temperature to observe aggregation is T_r . Next, we perform the molecular dynamic simulations of eight Src SH3 domains in a cubic cell with the length of 300 Å at T_r . All simulations

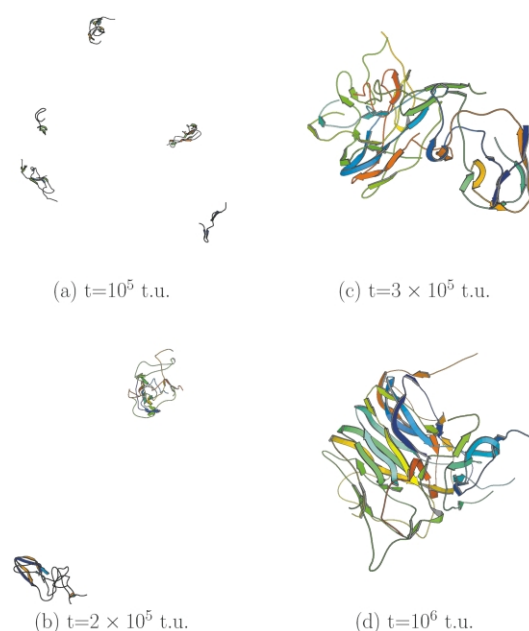


Figure 2. Snapshots during the aggregation of eight SH3 domains at T_r . The simulations start from fully unfolded conformations. The snapshots are taken at different times (a) 10^5 , (b) 2×10^5 , (c) 3×10^5 , and (d) 10^6 time units (t.u.).

from different initial configurations show similar equilibrium structures of SH3 aggregates ([Figure 3](#)). Snapshots during the aggregation ([Figure 2](#)) demonstrate that the initial step of aggregation is the dimerization and the formation of the open states which are the nuclei of aggregation process ([Figure 2\(a\)](#)). Other partially unfolded proteins grow on these open states. Usually, there are more than one nucleus; they merge with each other forming fibrils ([Figure 2\(b\) and \(c\)](#)). Finally, all eight proteins form one aggregate. However, the initial aggregate has only short-range order ([Figure 2\(c\)](#)). As system equilibrates, proteins rearrange themselves so that the equilibrium structure shows distinct long-range order ([Figures 2\(d\) and 3](#)).

At equilibrium, all eight proteins exhibit a tendency to form aggregates with a preferred direction of condensation ([Figure 3\(a\)](#)), which can be identified as the amyloid fibril axis. In the aggregated state, proteins pack their RT-loops on the top of each other by swapping the two parts of the RT-loops. We do not observe the aggregation states formed by propagational domain swapping as described.^{12,13} Packed RT-loops form a double β -sheet structure ([Figure 3\(b\)](#)). As discussed above, the aggregation process involves the reorientation of amino acids along the RT-loop. Thus, the side-chains (usually the hydrophobic residues) from the two parts of one RT-loop directed to each other and the separation of the two β -sheets is around 10 Å ([Figure 3\(b\)](#)). Due to the saturation and angular dependence property of the backbone hydrogen bonds, only the exposed

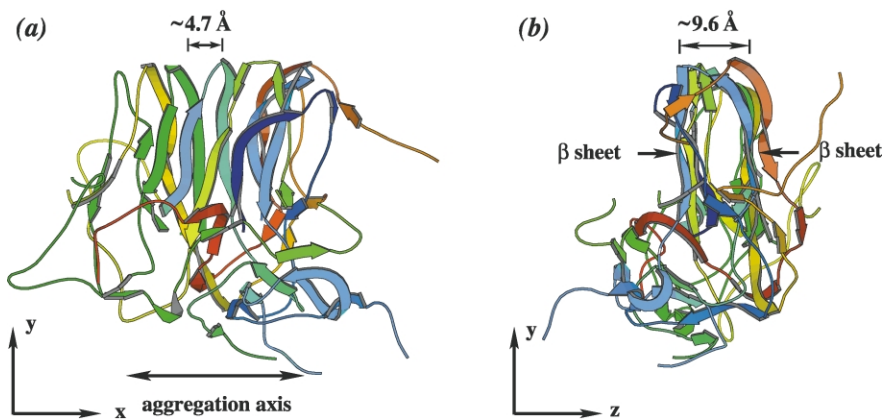


Figure 3. The typical equilibrium aggregation state for eight proteins: (a) xy and (b) zy projections of the aggregate. The preferred aggregation direction is along the x -axis.

proteins on the two ends can allow further aggregation by making backbone hydrogen bonds to form an extended β -sheet structure. Thus, by adding more proteins to the two ends of the aggregate, it continues growth to form an elongated fibril structure—amyloid fibril.

In order to characterize the structure of *in vivo* or *in vitro* amyloid fibrils, experimental X-ray scattering analysis has been widely applied.^{3–5} The X-ray scattering patterns of different amyloid fibrils share the same features: (i) a relatively sharp and intense 4.7 Å meridional reflection, and (ii) a weaker and more diffuse 7–10 Å equatorial reflection. The first peak corresponds to the β -strands spacing along the direction of the fibril axis, and the second, much weaker peak, is understood as the spacing between β -sheets. In order to compare our aggregates to experiments, we compute the X-ray diffraction pattern (Methods) for the aggregation structure obtained from molecular dynamics simulations (Figure 4(a)). We observe a distinguished peak corresponding to 4.7 Å along the meridional direction, which is related to β -strand spacing (Figure 3(b)). We also find a peak along the same direction with the spacing, 9.4 Å,

which is due to the doubling of the β -strand spacing. In the equatorial direction, we observe a weak peak of 9.6 Å which is related to the separation of the two β -sheets (Figure 3(b)). Therefore, the structure of the aggregates derived from the molecular dynamics simulations agrees fully with the experimental observations.

We calculate the correlation function of the structure of our aggregate (Figure 4(b)) and compare it with that of the native structure. Because the native structure of Src SH3 contains five β -strands already, the correlation functions for the aggregation and native state has similar, although weaker, peak at inter β -strand spacing distance, 4.7 Å. However, the aggregation state has a weak long-range peak corresponding to the inter β -sheet spacing, ~ 10 Å, which is not present in the native state.

Discussion

We perform molecular dynamics simulations to study the aggregation of Src SH3 domain. We find that the proteins start to aggregate at the threshold temperature T_a . Simulations at high temperature

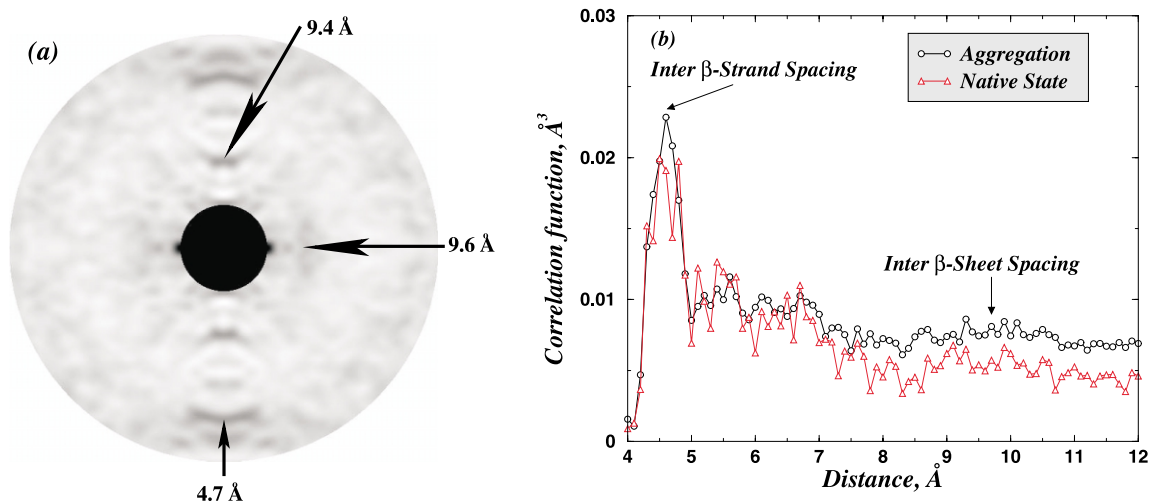


Figure 4. (a) Computed X-ray diffraction pattern of the aggregation structure formed by eight Src SH3 domains from the molecular dynamics simulations. (b) The correlation function of the native state of the Src SH3 domain (Δ) and the aggregate of eight Src SH3 domains (\circ) from Figure 3.

usually produce amorphous aggregates. Near the folding transition temperature $T_f = 0.95$, protein conformations in the unfolded states are partially folded and there are two competing processes: folding and aggregation. When the kinetic barrier for aggregation is smaller than that for folding, we observe ordered aggregation. As temperature decreases, the folding kinetic barrier decreases, so that τ_f decreases. Below some temperature threshold T_c , the partially folded protein states are near transition states at first and then rapidly descend into the native state, by-passing aggregation.

For two proteins, we observe two possible ordered aggregation conformations: a stable closed form dimer and an open aggregation state. The dimer is formed by domain swapping, where the two proteins exchange their RT-loops. We do not find any experimental evidence of the existence of this type of dimer for Src SH3 domain. However, our simulations suggest that it is a possible candidate oligomer state that may be found in future experiments. For the open form structure, the two proteins have their RT-loops packed together by swapping their two stands of the RT-loop. The open dimer structure allows further protein aggregation into fibrils. In our stimulations, we do not observe the aggregation scenario by propagational domain swapping.^{12,13} The closely packed RT-loops are stabilized by hydrogen bonds forming double β -sheets that face each other. From the molecular dynamics simulations we conclude that the core structure of Src SH3 amyloid fibrils is composed of RT-loop. Such a scenario may further be tested by experiments.

Our study reveals a general amyloidogenesis mechanism. Proteins, containing unstable β hairpins or loops, may be vulnerable to aggregation, and these unstable secondary structure elements can serve as the "building block" of amyloid fibrils.^{14,27} In the partially unfolded states, "building blocks" break apart from the rest a single protein. Then, these "building blocks" pack on top of one another by exchanging two complementary strands. During the aggregation process, amino acid residues in building blocks may reorient to make a stable connection, mainly by hydrogen bonding interactions. Due to the saturation and angular dependence of hydrogen bonds, only the exposed two ends can accept further deposition of building blocks from unbounded proteins and form an elongated double β -sheet structure. Thus, the resulting structure has the β -sheets parallel to the fibril axis and the β -strand perpendicular to the direction of fibril.

The presented model is similar in spirit to the β -nucleation model²⁸ to explain prion propagation. An important difference of our model from prion aggregation models is that in the case of prions aggregation, β structure is formed in the fragments of the chain that are α -helical in the monomeric state of the protein, PrP^C , i.e. amyloid formation is accompanied by secondary structure transformations. However, in both cases the phenomenon of

infectivity may be observed whereby existing fibrils may lower the kinetic barriers for monomeric proteins to join them forming larger fibrils.

Finally, we note that our model may shed light on the observation that complex β -topologies are more often found in genomes than more simple meander-like ones.^{29,30} The former are unlikely to have contiguous fragments that are stable by themselves, while the latter, being simple topologies, do have such fragments. They are, therefore, more prone to aggregation/amyloid formation placing organisms that carry such proteins at evolutionary disadvantage.

Methods

Model

We model the Src SH3 domain by beads representing C_α and C_β atoms. In order to mimic the flexibility of real proteins, we apply additional constraints:²² (i) "covalent" bonds between $C_{\alpha i}$ and $C_{\beta i}$, (ii) "peptide" bonds between $C_{\alpha i}$ and $C_{\alpha(i\pm 1)}$, (iii) effective bonds between $C_{\beta i}$ and $C_{\alpha(i\pm 1)}$, and (iv) effective bonds between $C_{\alpha i}$ and $C_{\alpha(i\pm 2)}$, where the subscript i denotes the amino acid residue sequence number.

The principal difficulty to study the protein folding *ab initio* is the lack of knowledge about the energetics between amino acid residues. The native-state specific Gō potential^{16,31,32} has been successfully used to model amino acid residue interactions. It has been shown²² that our coarse-grained model with a Gō interaction potential for the Src SH3 domain can faithfully reproduce the thermodynamic and kinetic properties observed in experiments. Thus, we use the Gō potential to model interactions between C_β atoms for a single protein. In order to reproduce the process of aggregation, we need to simulate more than one protein and to model the interaction between different proteins. For simplicity, we apply the Gō potential for C_β atoms between different proteins by assuming that two amino acid residues that attract to each other in a single protein will also have attraction in different proteins. The cutoff distance to define a contact is set as 7.5 Å.

The ability of proteins with no sequence similarity to aggregate into the same amyloid structure indicates that the non-specific backbone hydrogen bonding interaction may play an important role in the process of amyloidogenesis process. It has been shown³³ that only backbone hydrogen bonds can lead to a cooperative formation of two-dimensional β -sheet. We add to our model a non-specific interaction between any two C_α atoms to model the hydrogen bonding interaction between protein backbones. We set hydrogen bonding interaction range between the C_α atoms to 5.0 Å and the interaction strength as $n\varepsilon$, where ε is the Gō interaction strength and the variable n can be adjusted. In our study, we set $n = 3$ so that the hydrogen bond strength is more stable than the Gō interaction strength. It has been observed in many globular proteins that the number of backbone hydrogen bonds for the each residue does not exceed two (bifurcated hydrogen bonding is very rare and is not considered here). Another important property of the two hydrogen bonds formed by one peptide block is that they are linear. This is a reason for the formation of two-dimensional β -sheet. In the present study, (i) one C_α

atom cannot make more than two effective hydrogen bonds, and (ii) the two hydrogen bonds must be aligned linearly. We perform the simulation of a single protein and the result (data not shown) shows thermodynamic behavior similar to the one described by Ding *et al.*,²² with a cooperative folding transition and a slightly higher folding transition temperature.

We perform discrete molecular dynamics simulation to model the protein system. The number of proteins we studied varies from 2 to 8. The concentration of proteins in our simulation system is usually higher than *in vivo* and *in vitro* conditions, so that the condensation process is much faster and enables us to access the amyloidogenesis process by discrete molecular dynamics. First, we heat the system to high temperatures so that all the proteins are fully unfolded and moving freely inside the system box. Then, we quench the system to the temperature around T_f and wait for the system to equilibrate. The final equilibrium states result in possible amyloid fibril structures of Src SH3 domain.

X-ray

In order to compare with experimental X-ray diffraction patterns, we calculate the intensity of diffraction using the elastic diffraction formula:

$$I(\mathbf{k}_f) = \left| \sum_j \exp(i(\mathbf{k}_f - \mathbf{k}_i) \cdot \mathbf{r}_j) \right|^2 \quad (1)$$

where \mathbf{k}_i is the wave vector of the incoming X-ray, \mathbf{k}_f is the wave vector of the diffracted X-ray, \mathbf{r}_j is the position vector of j th atom, and the summation is over all the atoms in the structure. We align the aggregation structure along the x -axis as in Figure 3 and choose the incoming X-ray with a wavelength of 1 Å along the y -axis. The diffraction intensity is collected in projection of the y - z plane varying the defecting angle, $\theta = \cos^{-1}(\mathbf{k}_f \cdot \mathbf{k}_i / k^2)$, from 0.05 to 0.25 in radian. As in the X-ray diffraction experiments, the amyloid fibril has no preferred orientation in the y - z plane. We rotate the aggregation structure around the x -axis n times by angle $2\pi/n$ and add all the diffraction intensities.

PDB accession numbers

The coordinates have been deposited in the Protein Data Bank with accession code 1NLO.

Acknowledgments

We thank H. Kun, J. M. Borreguero, and C. M. Dobson for helpful discussions. This work was supported by Petroleum Research Fund of the American Chemical Society 37237-AC4, by National Institutes of Health Grant GM52126 (to E.I.S.) and National Institutes of Health National Research Service Award Fellowship GM20251 (to N.V.D.).

References

- Kelly, J. W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step

- assembly pathways. *Curr. Opin. Struct. Biol.* **8**, 101–106.
- Tan, S. Y. & Pepys, M. B. (1994). Amyloidosis. *Histopathology*, **25**, 403–414.
- Bonar, L., Cohen, A. S. & Skinner, M. (1967). Characterization of the amyloid fibril as a cross- β protein. *Proc. Soc. Expt. Biol. Med.* **131**, 1373–1375.
- Eanes, E. D. & Glenner, G. G. (1968). X-ray diffraction studies on amyloid filaments. *J. Histochem. Cytochem.* **16**, 673–677.
- Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B. & Blake, C. C. F. (1997). The common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273**, 729–739.
- Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl Acad. Sci. USA*, **95**, 4224–4228.
- Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999). Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. *Proc. Natl Acad. Sci. USA*, **96**, 3590–3594.
- Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332.
- Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H. & Goto, Y. (2002). Mapping the core of the β_2 -microglobulin amyloid fibril by H/D exchange. *Nature Struct. Biol.* **9**, 332–336.
- Tjernberg, L. O., Tjnnberg, A., Bark, N., Shi, Y., Ruzsicska, B. P., Bu, Z. *et al.* (2002). Assembling amyloid fibrils from designed structures containing a significant amyloid β -peptide fragment. *Biochem. J.* **366**, 343–351.
- Bennett, M. J., Choe, S. & Eisenberg, D. S. (1994). Domain swapping-entangling alliances between proteins. *Proc. Natl Acad. Sci. USA*, **91**, 3127–3131.
- Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. *Nature Struct. Biol.* **8**, 211–214.
- Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abahamson, M. & Jaskolski, M. (2001). Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nature Struct. Biol.* **8**, 316–320.
- Sinha, N., Tsai, C. J. & Nussinov, R. (2001). A proposed structural model for amyloid fibril elongation: domain swapping forms an interdigitating β -structure polymer. *Protein Eng.* **14**, 93–103.
- Staniforth, A. A., Giannini, S., Higgins, L. D., Cornroy, J., Hounslow, A. M., Jerala, R. *et al.* (2001). Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J.* **20**, 4774–4781.
- Zhou, Y. & Karplus, M. (1997). Folding thermodynamics of a three-helix-bundle protein. *Proc. Natl Acad. Sci. USA*, **94**, 14429–14432.
- Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (1998). Molecular dynamics studies of folding of a protein-like model. *Folding Des.* **3**, 577–587.
- Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2000). Identifying the protein folding nucleus using molecular dynamics. *J. Mol. Biol.* **296**, 1183–1188.
- Borreguero, J. M., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002).

- Thermodynamic and folding kinetic analysis of the SH3 domain from discrete molecular dynamics. *J. Mol. Biol.* **318**, 863–876.
20. Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nature Struct. Biol.* **5**, 714–720.
 21. Martinez, J. C., Pissabarro, M. T. & Serrano, L. (1998). Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Struct. Biol.* **5**, 721–729.
 22. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. (2002). Direct molecular dynamics observation of protein folding transition state ensemble. *Biophys. J.* **83**.
 23. Zurdo, J., Guijarro, J. I., Jimenez, J. L., Saibil, H. R. & Dobson, C. M. (2001). Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain. *J. Mol. Biol.* **311**, 325–340.
 24. Singh, A. P., Liu, X. & Brutlag, D. L. (1999). WWW tools for protein structure superposition and classification. *FASEB J.* **13**(7), A1542–A1542. Suppl.
 25. Guijarro, J. I., Marton, C. J., Plaxco, K. W., Campbell, I. D. & Dobson, C. M. (1998). Folding kinetics of the SH3 domain of PI3 kinase by real-time NMR combined with optical spectroscopy. *J. Mol. Biol.* **276**, 657–667.
 26. Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry*, **36**, 15685–15692.
 27. Tsai, C. J., Maizel, J. V. & Nussinov, R. (2000). Anatomy of protein structures: visualizing how a one-dimensional protein chain folds into a three-dimensional shape. *Proc. Natl Acad. Sci. USA*, **97**, 12038–12043.
 28. Morrissey, M. P. & Shakhnovich, E. I. (1999). Evidence for the role of PrPC helix 1 in the hydrophilic seeding of prion aggregates. *Proc. Natl Acad. Sci. USA*, **20**, 11293–11298.
 29. Chothia, C. & Finkelstein, A. V. (1990). The classification and origins of protein folding patterns. *Annu. Rev. Biochem.* **59**, 1007–1039.
 30. Gerstein, M. & Levitt, M. (1997). A structural census of the current population of protein sequences. *Proc. Natl Acad. Sci. USA*, **94**, 11911–11916.
 31. Gō, N. (1983). Theoretical studies of protein folding. *Ann. Rev. Biophys. Bioeng.* **12**, 183.
 32. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **278**, 937.
 33. Guo, C., Cheung, M. S., Levin, H. (2001). Mechanisms of cooperativity underlying sequence-independent β -sheet formation, arXiv:cond-mat/0104065
 34. Kraulis, P. J. (1991). Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Edited by A. R. Fersht

(Received 3 July 2002; received in revised form 4 October 2002; accepted 7 October 2002)