

Molecular Simulation of Protein Aggregation

Dusan Bratko,^{1,2} Troy Cellmer,³ John M. Prausnitz,^{2,4} Harvey W. Blanch²

¹Department of Chemistry, Virginia Commonwealth University, Richmond, Virginia 23284

²Department of Chemical Engineering, University of California, Berkeley, California 94720; telephone: (510) 642-1387; fax: (510) 642-1228; e-mail: blanch@berkeley.edu

³Laboratory of Chemical Physics, National Institute of Diabetes, Digestive, and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

⁴Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

Received 7 September 2006; accepted 7 September 2006

Published online 29 November 2006 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21232

ABSTRACT: Computer simulation offers unique possibilities for investigating molecular-level phenomena difficult to probe experimentally. Drawing from a wealth of studies concerning protein folding, computational studies of protein aggregation are emerging. These studies have been successful in capturing aspects of aggregation known from experiment and are being used to refine experimental methods aimed at abating aggregation. Here we review molecular-simulation studies of protein aggregation conducted in our laboratory. Specific attention is devoted to issues with implications for biotechnology.

Biotechnol. Bioeng. 2007;96: 1–8.

© 2006 Wiley Periodicals, Inc.

KEYWORDS: proteins; aggregation prevention; computer modeling

The function of a protein is intimately related to its native state, which is determined by interactions among amino-acid residues (Leach, 1996). Varying degrees of residue hydrophobicity, hydrogen-bonding groups, and the presence of partial charges result in specific potentials among the different residue pairs. Competition between these interactions, subject to topological constraints imposed by chain connectivity, determines the native state, which is generally believed to correspond to the global free-energy minimum of a single chain (Chan and Dill, 1993; Sali et al., 1994). However, in the presence of additional proteins, attractive forces among residues from other proteins can lead to the formation of inter-molecular clusters or aggregates (Asherie et al., 1998; Fink, 1998; Harrison et al., 1999). In this process, termed abnormal aggregation

(Fink, 1998), proteins acquire a tertiary structure incompatible with the biological function of the protein. At sufficient concentration, aggregation results in protein precipitation (Asherie et al., 1998).

Protein aggregation leading to reduced biological activity and insoluble or poorly soluble forms interferes with the manufacture of protein pharmaceuticals. Expression of therapeutic proteins in bacteria often results in intracellular deposits (inclusion bodies); considerable costs are associated with purification, solubilization, and renaturing the protein product (Clark, 2001). Aggregation can also take place during shipping and storage of protein drugs (Chi et al., 2003). Further, protein aggregates are suspected to be the causative agent in over 20 neurodegenerative diseases (Canet et al., 1999; Dobson, 1999; Koo et al., 1999; Tan and Pepys, 1994). Well-known examples include the Alzheimer's and Parkinson's diseases, Bovine spongiform encephalopathy (BSE) or mad-cow disease, and amyotrophic lateral sclerosis (ALS). The enormous impact of aggregation diseases and the growing number of protein drugs have given rise to considerable research of both equilibrium and dynamic aspects of protein aggregation, with the ultimate objective of learning to manipulate the protein chemistry and system conditions in ways that will prevent or slow down the aggregation process.

Molecular simulation offers a unique opportunity to gain novel insight into the process of protein aggregation. Biotechnology stands to benefit from these studies, as computer-assisted screening of mutations and environmental conditions can be used to reduce the experimental burden. In the present perspective, we provide a survey of our recent computational studies and discuss future possibilities for this nascent, but exciting field.

Correspondence to: H.W. Blanch

Contract grant sponsor: National Science Foundation

Contract grant number: BES-0432625

Contract grant sponsor: Virginia Center of Aging ARDRAF Fund

Contract grant sponsor: Office for Basic Sciences of the U.S. Department of Energy



“Simple” Models of Proteins

Computer modeling has played an essential role in developing our conceptual understanding of protein folding. This has been achieved by analyzing computed folding trajectories of model proteins to obtain microscopic insights and thermodynamic information unavailable through experiment. Because all-atom simulations cannot access time scales relevant to most folding events, the majority of simulation studies involve “coarse-grained” protein models. A polypeptide chain is typically treated as a necklace of beads (Chan and Dill, 1993; Sali et al., 1994). In view of the significant increase in computational demands associated with *multi-protein* simulations, this class of models appears ideally suited for qualitative studies of protein aggregation (Bratko and Blanch, 2001). Figure 1a and b shows snapshots of a lattice-model protein we have studied extensively (Bratko et al., 2006; Cellmer et al.,

2005b,c). By confining the amino acids to the lattice, the required computational power is reduced significantly. For studies with this model, we employed a potential function that was characterized by interaction strengths compatible with amino-acid contact probabilities from experiment (Miyazawa and Jernigan, 1985). As with all simulations using coarse-grained models, the solvent is implicit in the potential function. Therefore, the “solvent” can be altered by changes to the potential function. For example, denaturing conditions can be simulated (to a first approximation) by reducing the strength of attractive interactions.

Coarse-grained models capture many features of folding observed experimentally. Figure 1C and D shows plots of heat capacity and number of native contacts versus temperature for the lattice-model 64-mer shown in Figure 1a. A complete description of the model, and the simulation protocol is available in Cellmer et al. (2005b,c). Briefly, the model consists of 64 on-lattice beads, each representing an

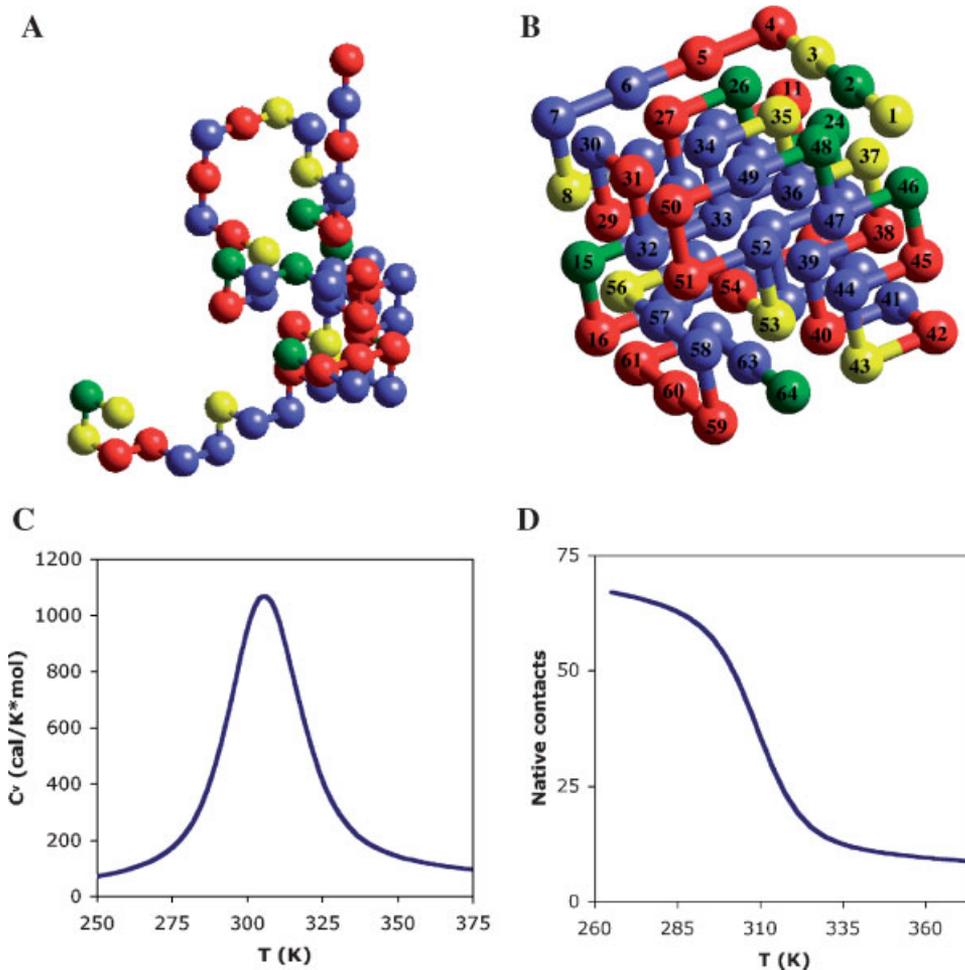


Figure 1. A: A snapshot of an unfolded conformation and lowest potential-energy structure (B) for the lattice-model 64-mer studied in Cellmer et al. (2005b,c). C: Heat capacity as a function of temperature for the lattice-model 64-mer. D: The number of native contacts as a function of temperature for the same model. The 64-mer sequence is KEKSTAGRVASGLDVSACVGLGDIDTLQGSPIAKLKTFFYGNKFNDFEASQAHMIRWPNYLPE. The 64-mer sequence is KEKSTAGRVASGLDVSACVGLGDIDTLQGSPIAKLKTFFYGNKFNDFEASQAHMIRWPNYLPE, blue beads represent non-polar residues, red beads represent polar residues, yellow beads are positively charged, green beads are negatively charged.

individual amino-acid residue. A modified 20-letter Miyazawa–Jernigan potential is used to describe the interactions between the amino acids (beads), as it more correctly reflects the multitude of strong and weak forces that govern protein interactions than other simple potentials commonly used for lattice models. The heat capacity is measured experimentally through calorimetry. With an appropriately chosen energy scale, the numerical values of the computed heat capacity (Fig. 1B) are within the range observed experimentally for proteins. The number of native contacts is experimentally inaccessible, but the data are very similar to that observed in protein denaturation experiments using intrinsic tryptophan fluorescence or circular dichroism.

Many small proteins exhibit a two-state folding character, making a first order transition from unfolded to folded forms. We have attempted to test whether this two-state character is apparent in the lattice-model 64-mer by applying the calorimetric criterion (Cellmer et al., 2005b). Due to difficulties in subtracting the unfolded and folded baselines, the result was inconclusive. However, when the data pertaining to the number of native contacts are fit to a two-state model, with linear fits to the baselines of the folded and unfolded states, very good agreement with two-state behavior is obtained. With the chosen energy scale, we obtain a melting temperature of 310 K and an energy of unfolding of 28 kcal/mol; both numbers are well-within those found by experiment. These data show that despite the simplicity of the models, they can exhibit very protein-like behavior.

Simple models have contributed significantly toward explaining protein folding. For example, they have provided a conceptual solution to the Levinthal paradox, which ponders the ability of proteins to search through the enormous number of possible conformations to reach the native state on time scales that range from hundreds of nanoseconds to minutes. Folding funnels (see Fig. 2), constructed from simulations of coarse-grained models, provide a visual solution to this paradox (Dinner et al., 2000). As a protein makes its way down the funnel, it accumulates native interactions. Because these interactions are, on average, more stable than non-native ones, they persist and systematically reduce the number of possible states. Thus, the protein need not sample all possible configurations on its path to the native state.

Folding Funnels in Multi-Protein Systems

To provide a better understanding of how the presence of additional protein molecules affects folding, we have constructed folding funnels in multi-protein systems (Fig. 2B and C) (Cellmer et al., 2005c). The funnels show how neighboring chains deform the folding funnel by eliminating the favorable free-energy bias towards the native state. While both independent variables (potential energies V_{native} and

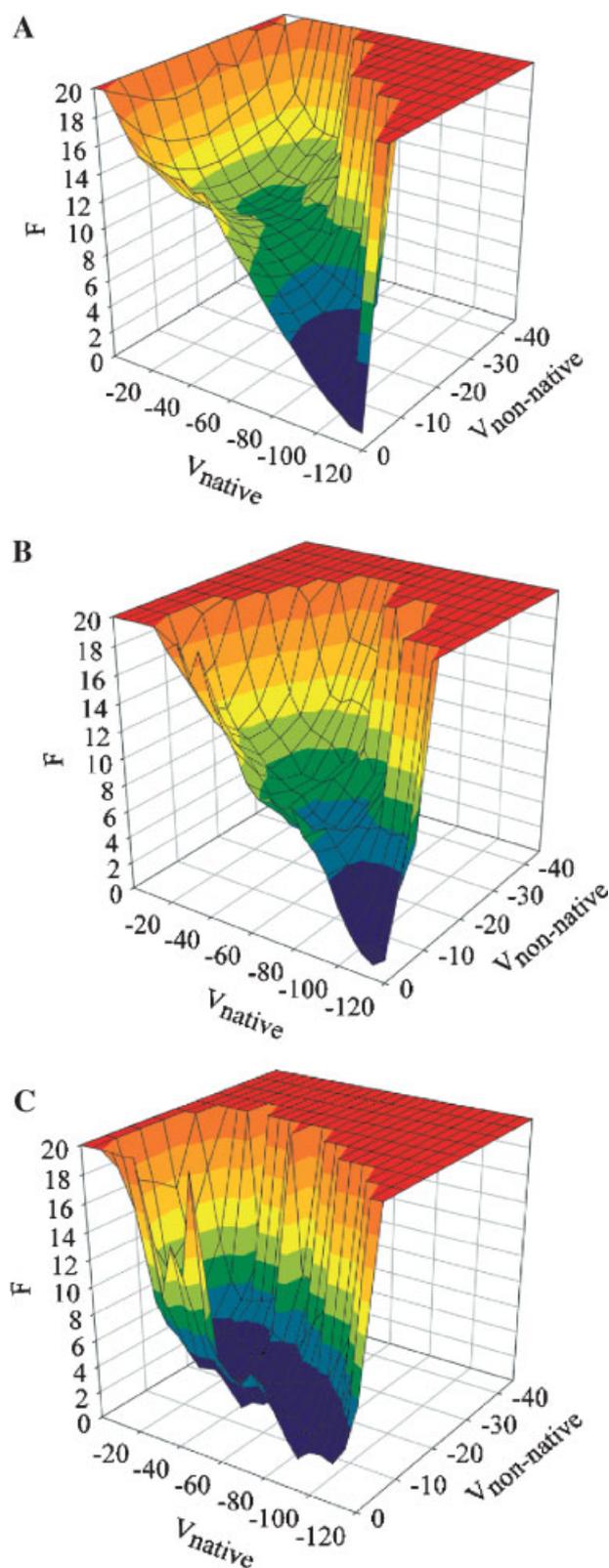


Figure 2. Folding funnels for the lattice-model 64-mer in isolation (A), in a two-chain system (B), and in a four-chain system (C). V_{native} and $V_{\text{non-native}}$ are potential energies arising from native and non-native interactions, respectively. Free energies F , as well as energies of native and non-native interactions, and given in units $k_B T$.

$V_{\text{non-native}}$ in Fig. 2) pertain to intra-molecular data, the inter-molecular effects are reflected in an increasing energetic advantage for misfolded states (reduced V_{native} and increased negative $V_{\text{non-native}}$).

“Hot” Sites for Aggregation

The molecular-level perspective offered by simulations facilitates the identification of the amino acids crucial to folding and aggregation. This information can be exploited to perform mutations that alter the protein’s stability, folding speed, or aggregation propensity. For the previously described lattice-model 64-mer, multi-protein simulations showed that a select group of 12 amino acids controls the aggregation process (Fig. 3A) (Cellmer et al., 2005c). Several of these amino acids are hydrophobic and buried in the native state. Inter-protein association increases as the proteins begin to unfold (Fig. 3B); this behavior is a direct consequence of interactions involving amino acids that are buried in the native state. Similar effects are commonly observed in experiments (Fink, 1998). Proteins exposed to environments where partially folded or unfolded states are heavily populated, become more likely to associate. Other amino acids that control the aggregation process arise from pairs of “charged” residues on the protein surface that form complementary interactions with the same pair of amino acids on another protein. Consequently, these interactions are often present when the protein’s native state is heavily favored (Fig. 3B).

Destabilizing the Native State Alters Protein-Aggregation Propensity

To probe the effects of point mutations on aggregation propensity, we created two mutant proteins, *L36T* and *M54T* from our original lattice-model 64-mer (Bratko et al., 2006). Residue 36 is located in the middle of the 11-residue string (31–41) bound anti-parallel to the string of residues 20–30 (comprising the longest anti-parallel segments) and is coordinated only by relatively close contour neighbors. Residue 54 (Methionine, M) is coordinated by residues distant on the chain contour. Because both *M* and *L* are strongly interacting, substitutions by polar residues prove disruptive to the protein native structure. Substitutions with a residue of intermediate hydrophobicity like threonine, *T*, however, preserve the native structure of the wild-type sequence. Substitutions *L36T* and *M54T* were therefore selected for comparison of model-protein aggregation behavior with that of the unperturbed sequence.

In isolation, all three sequences spontaneously fold into the same native structure and exhibit similar fluctuations around it. Small stability differences, however, translate to quite different behaviors under the influence of additional destabilizing factors such as elevated temperature or increased protein concentration. Under these conditions, the less stable variants become more likely to associate. Similar comments have been made in the context of single-chain simulations of the aggregation-prone *E22Q* mutant of the 10–35 segment of the Alzheimer’s A- β peptide, whose structural fluctuations are noticeably stronger than those for the WT sequence (Massi and Straub, 2001). At a protein

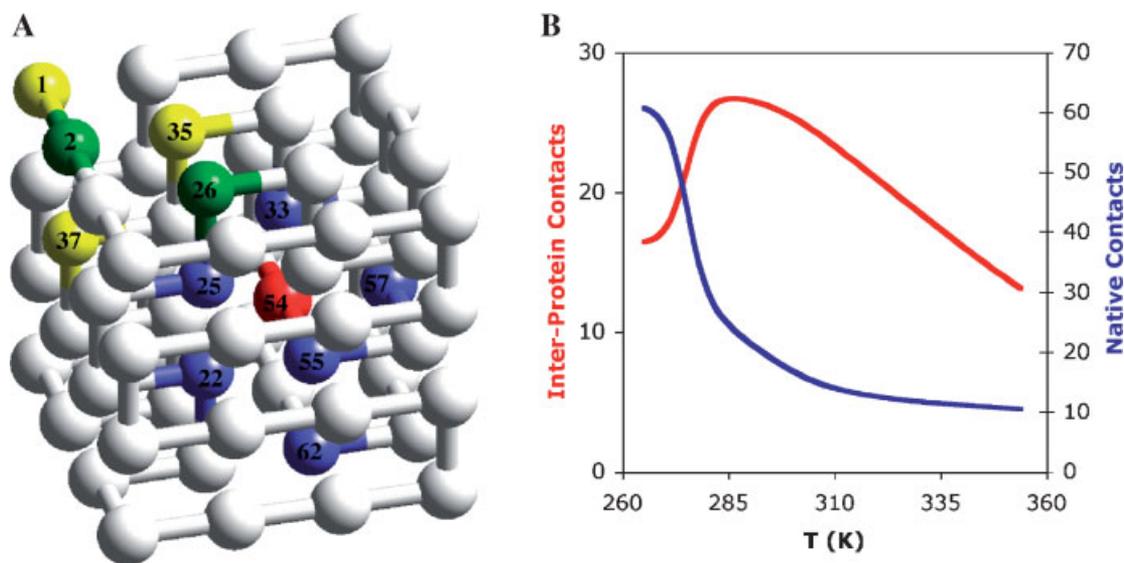


Figure 3. A: Highlighted beads, on average, are involved in interactions that comprise 50% of the total-interaction potential. These are referred to in the text as “hot sites” for aggregation. B: The average number of inter-protein contacts and native contacts (per chain) as a function of temperature for individual chains in *four-chain* systems. Note the decrease in native contacts, and concomitant increase in inter-protein contacts.

volume fraction of $\sim 6\%$, the three variants behave quite differently; the original sequence refolds regardless of initial configuration, while *L36T* and *M54T* unfold and form a cluster of misfolded chains. Figure 4 illustrates the effect of mutations by comparing the free-energy landscapes of two-chain model systems for the three variants. Sequences with nearly identical folding behaviors in isolation exhibit very different affinities for association as a result of moderate single-point substitutions. The landscape of the WT protein *in the two-chain system* retains the funnel-like shape conducive to folding whereas the two mutants develop auxiliary free-energy minima corresponding to misfolded conformations. The shift of these minima toward higher number of inter-protein contacts, N_{inter} , confirms that the misfolded conformations are stabilized by multiple inter-chain interactions (Cellmer et al., 2005b), an effect that increases with protein concentration.

The overall effect of moderate residue substitutions is similar to the influence of weakly destabilizing system conditions such as increases in temperature, concentration (Bratko et al., 2006), or denaturant. Because of strong correlations between the binding states of the specified monomer and the proximity to the folded state (Bratko and Blanch, 2003), strengthening or weakening the bonds of a selected residue can have an effect similar to modulating the overall strength of intra-protein interactions for the whole chain. Subtle mutations can therefore be used in sequence engineering to shift the coexistence lines in the phase diagram of a protein solution without altering the structure of the native state.

The Protein “Refolding” Problem

We have also utilized simulations to address directly a problem of biotechnological interest (Cellmer et al., 2005a). When *E. coli* is chosen as the host for protein production, over-expressed protein often forms intra-cellular aggregates known as inclusion bodies. These aggregates must be dissolved, and the protein refolded in order for it to become biologically active. During the refolding step, aggregation competes with correct folding, and often limits the yield of folded protein. While this problem can be avoided by refolding in a large volume (low protein concentration), such a strategy is often not economically viable. Because little is understood about the underlying mechanisms of the aggregation process it is difficult to prevent its occurrence by rational design.

The “protein-refolding” problem is well suited for computational investigation, as it simply involves a scale-up in the number of chains simulated. For our multi-protein simulations we employed an off-lattice model that folds into a beta-barrel structure (Fig. 5) (Honeycutt and Thirumalai, 1992). Only three different amino acids are used for these simulations. There are “hydrophobic” beads that interact through the Lennard–Jones potential, “hydrophilic” beads that have a purely repulsive potential, and “neutral”, or

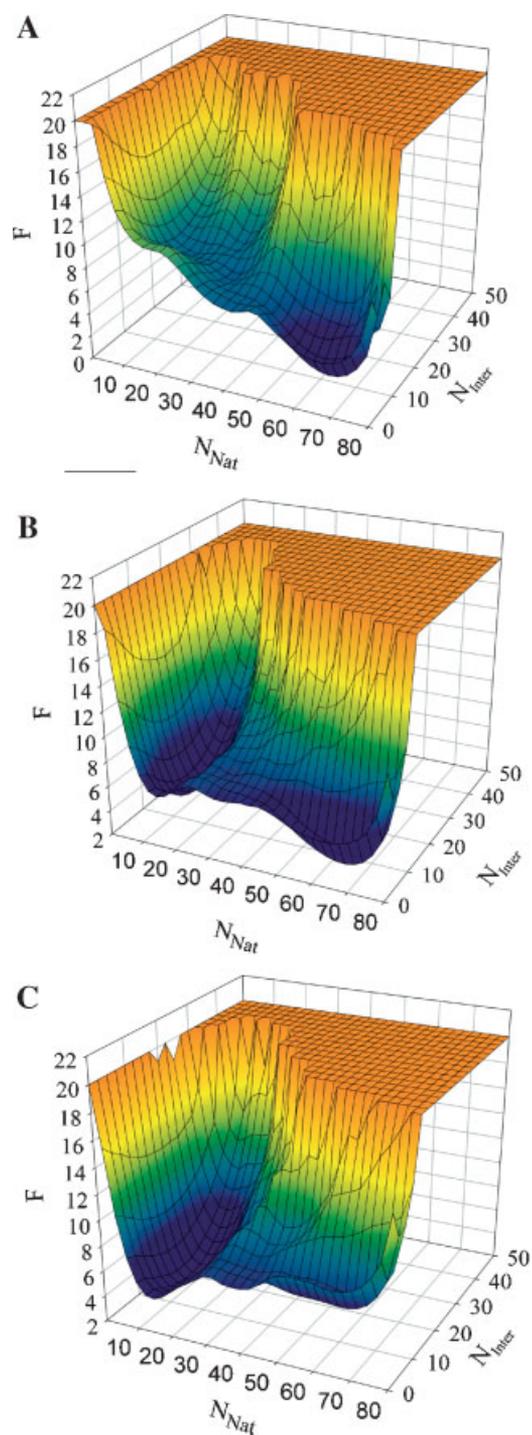


Figure 4. Free-energy landscapes in two-chain systems for (A) the wild-type lattice-model 64-mer, the mutant L36T (B), and the mutant M54T (C). N_{Nat} is the number of native contacts, N_{inter} is the number of inter-protein contacts, and F is the free energy in units $k_B T$.

“glycine” beads that have a purely repulsive two-body potential and a relaxed dihedral potential that allows them to form the turns required for the β -barrel.

Each simulation contained three chains that were equilibrated under the conditions of interest without

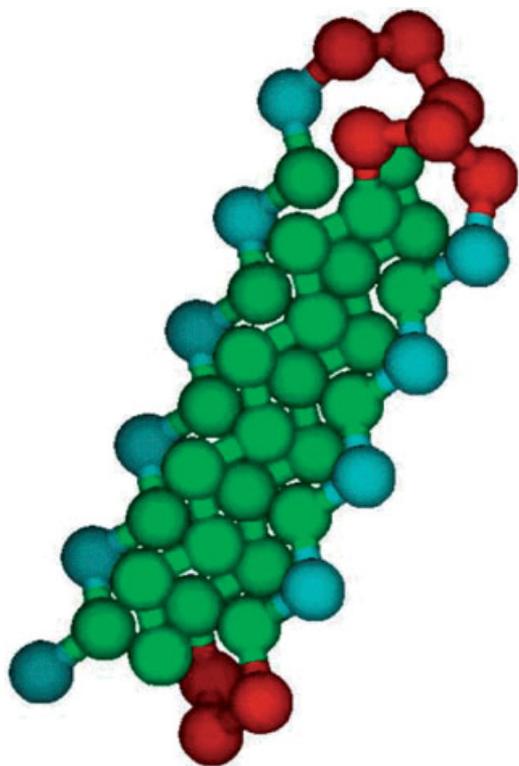


Figure 5. The native structure of the model protein used in these simulations. “Hydrophobic” beads are green, “hydrophilic” beads are blue, and neutral or “glycine” beads are red.

attractive interactions. At time zero, the attractive interactions were turned on, and the simulations were monitored until all the model proteins had folded or aggregated. The refolding yield was calculated by dividing the number of chains that fold by the number simulated. As expected, this number increases as the volume available to the proteins increases. The refolding yield more than triples (15–51%) as the concentration decreases from ~ 3 to ~ 0.3 mM.

As shown in Figure 6B, strands one and three play a dominant role in aggregation. Both strands have nine “hydrophobic” beads, making them the “stickiest” of the four strands (Fig. 6A). Despite having the same level of “stickiness”, the strands have different propensities to associate with other chains. This is likely due to higher-order structural factors, namely that strand three forms intra-protein interactions *earlier* in the folding process than strand one, therefore, limiting its ability to form inter-protein interactions. This result suggests that, for aggregation between proteins in states where some three-dimensional structure is present, the most aggregation-prone segments cannot be identified by amino-acid sequence alone.

To probe the heterogeneity of the aggregation process, we registered each aggregation event, and catalogued events when one pair of strands contributed greater than 50% of the overall interaction potential. Strand one-strand one

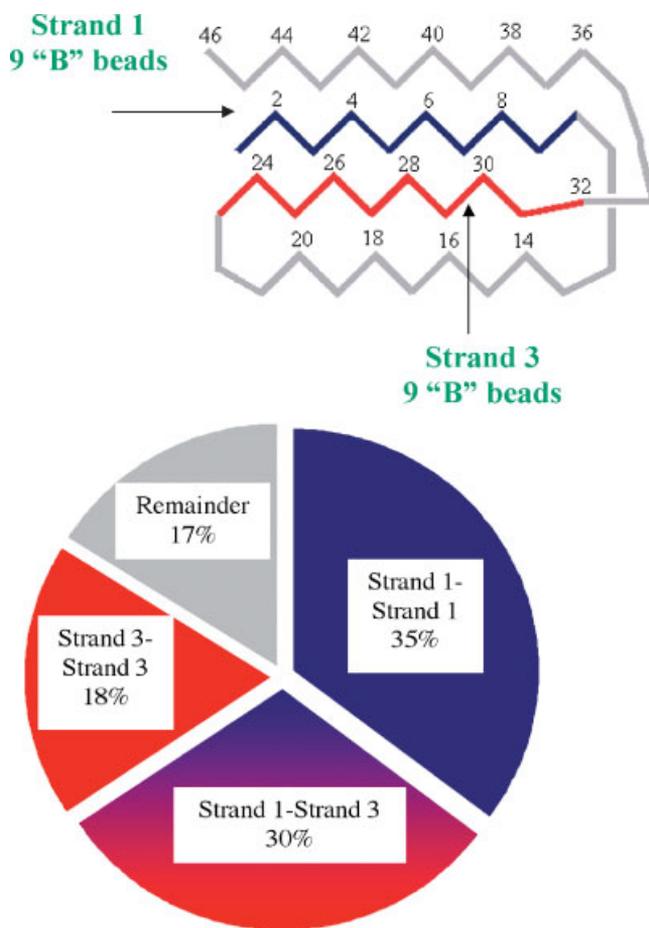


Figure 6. **A:** Schematic of the model protein used in the protein-refolding simulations. **B:** Represents the average contribution of inter-protein interactions involving different strands. For example, when two proteins form an aggregate, on average 35% of the interaction energy comes from interactions between strand one on one protein, and strand one on the other.

interactions dominate 33% of time, strand one-strand three interactions 16% of the time, and strand three-strand three interactions 18% of the time. This result clearly shows that there are different mechanisms for aggregation, although the route involving strand one-strand one interactions is statistically preferred.

Using this information, we designed several mutants where two hydrophobic beads were changed to hydrophilic beads (Fig. 7). We focused on mutations that did not radically alter the stability or native structure of the model protein. Not surprisingly, mutations to strand three improved the refolding yield but not to the same degree as mutations to strand one. This result was expected, given that strand one is more often involved in inter-protein interactions.

This exercise has several implications for experiments. First, as with the lattice-models, aggregation is controlled by a relatively small number of amino acids. Similar behavior has been recently observed in amyloid-fibril formation by

Mutation 1	Mutation2	Refolding Yield
Wild-type	Wild-type	0.15
2 (strand 1)	4 (strand 1)	0.33
2 (strand 1)	6 (strand 1)	0.36
25 (strand 3)	29 (strand 3)	0.19
25 (strand 3)	31 (strand 3)	0.27

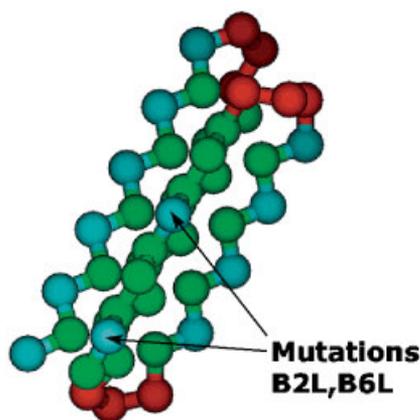


Figure 7. A: Refolding yields for various double mutants. All mutations are of the form “hydrophobic” to “hydrophilic.” B: Native structure of the mutant with the greatest refolding yield.

unfolded proteins and peptides (Chiti et al., 2002), and has facilitated the development of a software package that identifies nucleating regions (Fernandez-Escamilla et al., 2004) of unfolded peptides. While our model possesses two segments with the same degree of “stickiness”, one of these tends to favor aggregation more strongly. This shows that a priori design of mutations from primary-sequence data is complicated in instances where the formation of higher-order structure inhibits certain segments of the protein from inter-protein interactions. We also showed that multiple aggregation pathways can coexist. Thus, some proteins may require mutations to multiple regions of their amino-acid sequence to obtain substantial reductions in aggregation propensity.

Concluding Remarks

For several decades, computer simulations have contributed to the study of protein folding. With advances in computer power, it has become possible to “scale-up” these

simulations to address the problem of protein aggregation. The successes of simple models in capturing the essential physics of folding are encouraging for the study of aggregation, because the two processes share many features.

We present three important conclusions. First, we showed that coarse-grained models of proteins, despite their simplicity, nevertheless are very protein-like. Second, we showed that several features of the aggregation process known from experiment are reproduced by the simple models. Third, we showed that the simulations can be used to “design out” aggregation-prone behavior, as well as provide direction to experimentalists with similar goals.

In addition to our laboratory, several other groups have obtained important results from molecular simulations of protein aggregation (Giugliarelli et al., 2000; Istrail et al., 1999; Ma and Nussinov, 2002a,b; Toma and Toma, 2000). It is not possible to describe the results of all of these studies, but we highlight a few examples here. The Hall group (Nguyen and Hall, 2004a,b, 2005, 2006) has pioneered the use of discontinuous molecular dynamics to study large systems (~100) of polyalanine-like peptides that undergo a spontaneous transition to form amyloid fibrils. The Head-Gordon group has exploited its minimalist model of the small protein, protein L, for multi-protein simulations (Clark, 2005; Fawzi et al., 2005; Sorenson and Head-Gordon, 2002). This work is significant because it offers the possibility for direct comparison to experiment. As protein L forms a stable three-dimensional structure, it should also allow for the direct observation of the conformational changes required prior to non-native aggregation. Such events are very difficult to observe experimentally.

The molecular-level perspective offered by computer simulations will continue to improve our fundamental understanding of aggregation phenomena. Further, it has begun to serve as a complement to experiment in rationally designing sequence alterations that can change a given protein’s propensity to aggregate. As yet, it is not possible to run simulations at an atomic-level of detail and identify with certainty the amino acids that drive the aggregation process. However, studies like those illustrated here can be useful in directing the search for the appropriate mutations before the first experiment is performed. The degree of refinement will continue to improve as the field matures.

For financial support, the authors are grateful to the National Science Foundation under award BES-0432625, Virginia Center of Aging ARDRAF Fund, and to the Office for Basic Sciences of the U.S. Department of Energy.

References

- Asherie N, Pande J, Lomakin A, Ogun O, Hanson SRA, Smith JB, Benedek GB. 1998. Oligomerization and phase separation in globular protein solutions. *Biophys Chem* 75(3):213–227.
- Bratko D, Blanch HW. 2001. Competition between protein folding and aggregation. A three-dimensional lattice model simulation. *J Chem Phys* 114:561–569.

- Bratko D, Blanch HW. 2003. Effect of secondary structure on protein aggregation: A replica exchange simulation study. *J Chem Phys* 118(11):5185–5194.
- Bratko D, Cellmer T, Prausnitz JM, Blanch HW. 2006. Effect of single-point sequence alterations on the aggregation propensity of a model protein. *J Am Chem Soc* 128(5):1683–1691.
- Canet D, Sunde M, Last AM, Miranker A, Spencer A, Robinson CV, Dobson CM. 1999. Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease-related variants. *Biochemistry* 38(20):6419–6427.
- Cellmer T, Bratko D, Prausnitz JM, Blanch H. 2005a. The competition between protein folding and aggregation: Off-lattice minimalist model studies. *Biotechnol Bioeng* 89(1):78–87.
- Cellmer T, Bratko D, Prausnitz JM, Blanch H. 2005b. Thermodynamics of folding and association of lattice-model proteins. *J Chem Phys* 122(17):174908/1–174908/11.
- Cellmer T, Bratko D, Prausnitz JM, Blanch HW. 2005c. Protein-folding landscapes in multichain systems. *Proc Natl Acad Sci USA* 102:11692–11697.
- Chan HS, Dill KA. 1993. The protein folding problem. *Phys Today* 46(2):24–32.
- Chi EY, Krishnan S, Randolph TW, Carpenter JF. 2003. Physical stability of proteins in aqueous solution: Mechanism and driving forces in non-native protein aggregation. *Pharm Res* 20(9):1325–1336.
- Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, Dobson C. 2002. Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol* 9(2):137–143.
- Clark ED. 2001. Protein refolding for industrial processes. *Curr Opin Biotechnol* 12(2):202–207.
- Clark LA. 2005. Protein aggregation determinants from a simplified model: Cooperative folders resist aggregation. *Protein Sci* 14(3):653–662.
- Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 25(7):331–339.
- Dobson CM. 1999. Protein misfolding, evolution and disease. *Trends Biochem Sci* 24:329–332.
- Fawzi NL, Chubukov V, Clark LA, Brown S, Head-Gordon T. 2005. Influence of denatured and intermediate states of folding on protein aggregation. *Protein Sci* 14(4):993–1003.
- Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. 2004. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* 22(10):1302–1306.
- Fink AL. 1998. Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold Des* 3(1):R9–R23.
- Giugliarelli G, Micheletti C, Banavar JR, Maritan A. 2000. Compactness, aggregation, and prionlike behavior of protein: A lattice model study. *J Chem Phys* 113(12):5072–5077.
- Harrison PM, Chan HS, Prusiner SB, Cohen FE. 1999. Thermodynamics of model prions and its implications for the problem of prion protein folding. *J Mol Biol* 286(2):593–606.
- Honeycutt JD, Thirumalai D. 1992. The nature of folded states of globular-proteins. *Biopolymers* 32(6):695–709.
- Istrail S, Schwartz R, King J. 1999. Lattice simulations of aggregation funnels for protein folding. *J Comput Biol* 6(2):143–162.
- Koo EH, Lansbury PT, Kelly JW. 1999. Amyloid diseases: Abnormal protein aggregation in neurodegeneration. *Proc Natl Acad Sci USA* 96:9989–9990.
- Leach YAR. 1996. *Molecular modeling*. London: Longman.
- Ma BY, Nussinov R. 2002a. Molecular dynamics simulations of alanine rich beta-sheet oligomers: Insight into amyloid formation. *Protein Sci* 11(10):2335–2350.
- Ma BY, Nussinov R. 2002b. Stabilities and conformations of Alzheimer's beta-amyloid peptide oligomers (A beta(16–22'), A beta(16–35') and A beta(10–35)): Sequence effects. *Proc Natl Acad Sci USA* 99(22):14126–14131.
- Massi F, Straub JE. 2001. Probing the origins of increased activity of the E22Q “Dutch” mutant Alzheimer's beta-amyloid peptide. *Biophys J* 81(2):697–709.
- Miyazawa S, Jernigan RL. 1985. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* 18:534–552.
- Nguyen HD, Hall CK. 2004a. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc Natl Acad Sci USA* 101(46):16180–16185.
- Nguyen HD, Hall CK. 2004b. Phase diagrams describing fibrillization by polyaniline peptides. *Biophys J* 87(6):4122–4134.
- Nguyen HD, Hall CK. 2005. Kinetics of fibril formation by polyaniline peptides. *J Biol Chem* 280(10):9074–9082.
- Nguyen HD, Hall CK. 2006. Spontaneous fibril formation by polyanilines; discontinuous molecular dynamics simulations. *J Am Chem Soc* 128(6):1890–1901.
- Sali A, Shakhnovich E, Karplus M. 1994. How does a protein fold. *Nature* 369(6477):248–251.
- Sorenson JM, Head-Gordon T. 2002. Protein engineering study of protein L by simulation. *J Comput Biol* 9(1):35–54.
- Tan SY, Pepys MB. 1994. Amyloidosis. *Histopathology* 25(5):403–414.
- Toma L, Toma S. 2000. A lattice study of multimolecular ensembles of protein models. Effect of sequence on the final state: Globules, aggregates, dimers, fibrillae. *Biomacromolecules* 1(2):232–238.