ELSEVIER

Studies of folding and misfolding using simplified models

Nikolay V Dokholyan

Computer simulations are as vital to our studies of biological systems as experiments. They bridge and rationalize experimental observations, extend the experimental 'field of view', which is often limited to a specific time or length scale, and, most importantly, provide novel insights into biological systems, offering hypotheses about yet-to-be uncovered phenomena. These hypotheses spur further experimental discoveries. Simplified molecular models have a special place in the field of computational biology. Branded as less accurate than all-atom protein models, they have offered what all-atom molecular dynamics simulations could not — the resolution of the length and time scales of biological phenomena. Not only have simplified models proven to be accurate in explaining or reproducing several biological phenomena, they have also offered a novel multiscale computational strategy for accessing a broad range of time and length scales upon integration with traditional all-atom simulations. Recent computer simulations of simplified models have shaken or advanced the established understanding of biological phenomena. It was demonstrated that simplified models can be as accurate as traditional molecular dynamics approaches in identifying native conformations of proteins. Their application to protein structure prediction yielded phenomenal accuracy in recapitulating native protein conformations. New studies that utilize the synergy of simplified protein models with all-atom models and experiments yielded novel insights into complex biological processes, such as protein folding, aggregation and the formation of large protein complexes.

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

Corresponding author: Dokholyan, Nikolay V (dokh@med.unc.edu)

Current Opinion in Structural Biology 2006, **16**:79–85

This review comes from a themed issue on
Folding and binding
Edited by Mikael Olivberg and Eugene I Shakhnovich

Available online 18th January 2006

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2006.01.001

Introduction

Perturbations of protein structure and dynamics often alter protein function and, in some cases, lead to human disease. As our understanding of protein structure has become increasingly sophisticated, the number of human diseases associated with abnormal protein folding or dynamics has risen significantly. Viewed from the perspective of thermodynamics, mutations or changes in the cellular environment often lead to the modification of the free energy landscape of proteins and the equilibrium distribution over different conformations. Protein misfolding (including formation of partially unfolded protein conformations) can be an important consequence of changes that result in protein destabilization. Normally, cells are equipped with molecular chaperones that reverse the process of protein misfolding, guiding misfolded proteins to their native state. However, large populations of misfolded proteins may impose an insurmountable stress on this protective machinery by depleting chaperones for other misfolded proteins. Misfolded proteins tend to aggregate [1], often forming hyperstable assemblies. Some of these assemblies, amyloid fibrils, display distinctive needle-like fibrillar shapes.

Characterization of the structural and dynamic properties of proteins, especially large multidomain proteins, protein complexes and protein aggregates, is a major challenge of structural biology. Dynamic and structural features of large biomolecules are often 'invisible' to current experimental techniques, because of their inherent resolution limitations in length and time scales. Although there have been important advances in the structural characterization of molecular complexes, such as amyloid fibrils, and in elucidating the mechanisms of their formation, many aspects of this process remain unclear.

This review contrasts traditional and new computational approaches. Also described are recent studies that utilize the synergy of simplified protein models with all-atom models and experiments to yield novel insights into complex biological processes, such as protein folding, aggregation and the formation of large protein complexes.

Approaches to understanding protein folding, misfolding and association

Computational approaches offer a unique opportunity to uncover the atomic structure and biological properties of experimentally challenging molecules and molecular complexes. Such efforts must take into account time scales of protein association and aggregation subsequent to misfolding, which vary from seconds to hours or even months, making them unreachable by traditional molecular dynamics (MD) simulations [2,3]. Novel simulation protocols have been proposed that improve conformational sampling efficiency. Generalized ensemble sampling techniques involving parallel simulations of molecular systems coupled by a Monte Carlo (MC)

protocol [4] have been successfully applied to study protein folding [5]. Although a one to two orders of magnitude increase in the accessible time or length scales has been achieved, such enhancements still permit studies of only small, fast-folding proteins [6].

An alternative strategy employs a simplified protein model to access the large-scale dynamics of protein folding and aggregation. In simplified (coarse-grained) protein models, amino acids are represented by effective particles (beads) that correspond to atoms or groups of atoms [7–12]. The interaction potential between these beads is empirical and can be derived from protein structure [13], experimental strategies [14] or physics [15,16]. A drawback of this approach is that the model of amino acid interaction is not rigorously derived from first physical principles. However, the ability to sample large-scale protein dynamics makes this approach, coupled with all-atom MD simulations, a promising avenue for developing experimentally testable hypotheses.

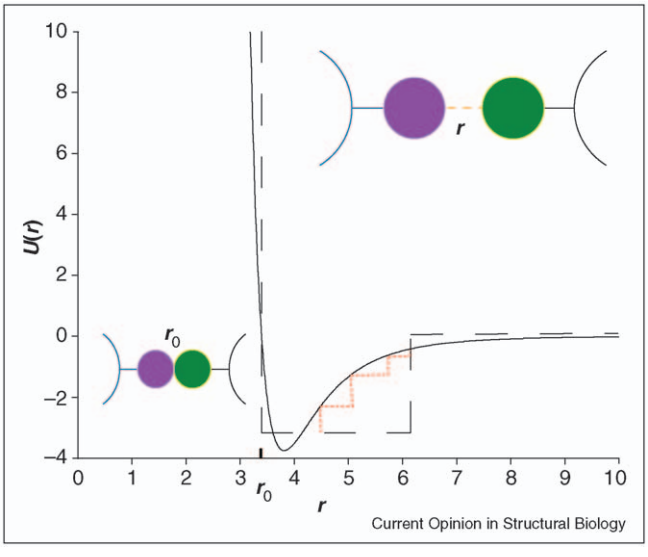
Perhaps one of the most striking recent applications of multiscale modeling methodology is its use in high-resolution *de novo* structure prediction for small proteins by Bradley *et al.* [17••]. By using methods that allow high- and low-resolution sampling, structure prediction <1.5 Å root mean square distance from the native state was achieved for proteins with lengths up to 85 amino acids.

An existing multiscale modeling software package, MMTSB [18], integrates a simplified protein model with the MC simulation engine, MONSSTER [8], and the all-atom MD packages AMBER [3] or CHARMM [2]. In MONSSTER, each amino acid is modeled by one bead corresponding to the sidechain center and the effective beads are positioned on a cubic grid. The lattice model imposes strong constraints on the angles between the covalent bonds and, thus, greatly restricts the conformational space of the protein model. An additional drawback of this restriction is the inability of these models to accurately discern the topological properties of the proteins. Because of several remaining delicate issues, such as the dependence of the dynamics on the MC move set, an independent test of the dynamics using the MD approach is needed.

A suitable simulation approach for simplified protein models is discrete molecular dynamics (DMD) [10,11]. DMD extends the accessible simulation time by using long integration time steps, made possible by the use of a different approximation to the energy function than that used in longer-established MD simulations [2,3]. This approach permits rapid and accurate sampling of the conformational space of proteins and protein complexes [10]. Within the framework of multiscale molecular modeling, the combination of DMD and all-atom MD overcomes the intrinsic limitation of the lattice MC protein models. The DMD algorithm satisfies fundamental physical principles: the energy and momentum conservation. A unique feature of the DMD algorithm is that it is based on the solution of the ballistic equations of motion, rather than Newtonian ones, which reduces the simulation algorithm to an iterative search of the immediate collision events in the system. In DMD, as in any other simulation approach, we approximate interatomic interactions. The simplification that is necessary for the collision-based algorithm and that makes this algorithm up to 10^8–10^9 times faster than traditional MD is the discretization of the interaction potential: interatomic interactions are piece-wise curves, each of the pieces of which is a constant (Figure 1). Although such an approximation can seem limiting for studies of many natural systems, it nevertheless proves to be sufficient in many cases in which DMD has been used. It is important that the discretization of the actual potentials can be made more realistic by having more 'square-wells' steps that mimic an 'actual' potential (Figure 1). However, at least in protein folding studies, just a single step is often sufficient to reproduce the phenomena under investigation.

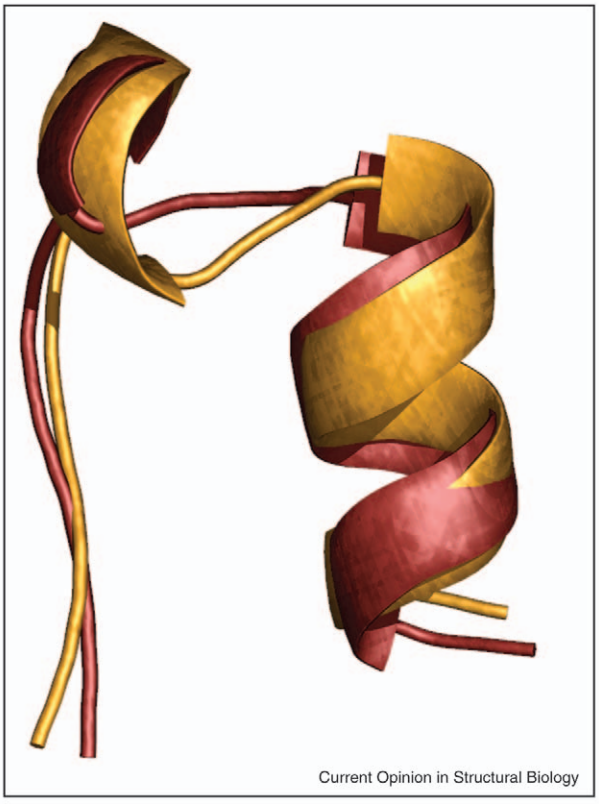
As a validation of the reliability and accuracy of DMD, simulations were used to fold a small 20-residue Trp-cage miniprotein from a fully extended conformation using a simplified interaction model between amino acids [19••]. In these simulations, the Trp-cage model consistently reached conformations within a 2 Å backbone root mean square distance from the corresponding NMR structures (Figure 2), on a par with recent MD simulations of this

Figure 1



The square-well approximation. The Lennard–Jones interaction potential (solid line) is approximated by discrete square-well potentials (dashed lines; more detailed potentials are shown as red dotted lines). The r^{-12} potential is approximated by a reflective wall at a distance corresponding to the sum of the hard-core radii of the interacting atoms.

Figure 2



Superposition of the experimentally (bronze) and computationally (gold) determined three-dimensional structures of the Trp-cage native conformation. The backbone root mean square distance between these two structures was reported by Ding *et al*. [19••] to be less than 2 Å.

protein [20–25]. Despite the simplicity of the interaction potential employed in the folding of the Trp-cage miniprotein, it is potentially transferable to larger proteins. Folding of three designed ccβ peptides (51 residues) [26] in DMD simulations, using the force-field developed for Trp-cage folding, resulted in lowest energy structures within 1.26 Å of the experimentally determined native state (F Ding *et al*., unpublished; see also Update).

The power of multiscale DMD-based modeling approaches has been clearly demonstrated by recent folding simulations of β-amyloid [27], polyglutamine peptides [28•] and the SH3 domain [29]. In these studies, the important computational strategy was the separation of time scales. Protein conformations are adequately sampled at the longer time scales using DMD simulations. The thermodynamic viability of the plausible states determined by the DMD simulations is then tested at the shorter time scales using all-atom MD simulations. In principle, this strategy can be extended to include even longer time scales when considering mesoscopic biological systems, such as chromosomes, or biological pathways in which multiple scales are involved. The accuracy and the predictive power of such approaches can be improved if simulations can be 'navigated' by experiments [14,30], as described next.

Efficient synergy between simplified models and experiment

Perhaps one of the most important uses of simplified models is to extend the 'depth of the field of view' of experiments. Experiments have time and length scale visibility ranges at which the structure and dynamics of molecules may be seen. For example, the length scales of microscopic experiments are limited by the wavelength of light. One approach to extending the experimental visibility range is to build simplified models based on the experimental data obtained within accessible time and length scales, and then perform simulations of the biological systems using these simplified models in the time and length scales that are reachable by experiments.

Recently, Vendruscolo *et al*. [30] and Dixon *et al*. [14] related NMR equilibrium hydrogen exchange protection factors to the interaction parameters of simplified protein models. The experimental data served as constraints in MD simulations. The ensembles of conformations obtained in simulations were consistent with experiment and provided insights into the biology of studied proteins. For example, Vendruscolo *et al*. [30] have found rare fluctuations around the native state of α-lactalbumin. Dixon *et al*. [14] found a transiently populated phosphorylation-competent intermediate of the targeting domain of focal adhesion kinase.

Combining structural data from cryo-electron microscopy (cryo-EM) with molecular modeling enabled Falke *et al*. [31••] to determine the 13 Å resolution structure of the 798 kDa GroEL tetradecamer in both the free state and bound to a single large monomer of glutamine synthetase (GSm) (51 kDa). The authors used the known X-ray crystallographic structure and normal mode flexible fitting analysis based on simplified Cα protein models [32] to describe the changes that are induced in the GroEL structure upon GSm binding.

The interpretation of experimental data often relies on theoretical models. Thus, another important use of MD simulations of simplified protein models is the ability to interpret often seemingly contradictory experimental observations. In this regard, simplified models have been recently used to describe the structure of unfolded protein states. As a result of agreement with the experimentally determined scaling of protein sizes, an ensemble of random-coil conformations with no persistent local and global structure has been used as a theoretical model of denatured proteins. In contrast to this random-coil model of denatured proteins, recent NMR spectroscopy studies of proteins at high chemical denaturant concentrations suggested the presence of significant amounts of

native-like structure. By modeling denatured proteins in DMD simulations, Ding *et al.* [33•] reconciled these seemingly controversial observations. For all studied proteins, the authors found that denatured states indeed have strong local conformational bias toward native states while a random-coil power law scaling of protein sizes is preserved. In addition, Ding *et al.* [33•] explained why the experimentally determined size of the protein creatine kinase does not follow general scaling. It was observed in simulations that this protein exhibits a stable intermediate state, the size of which is consistent with the reported experimental observation.

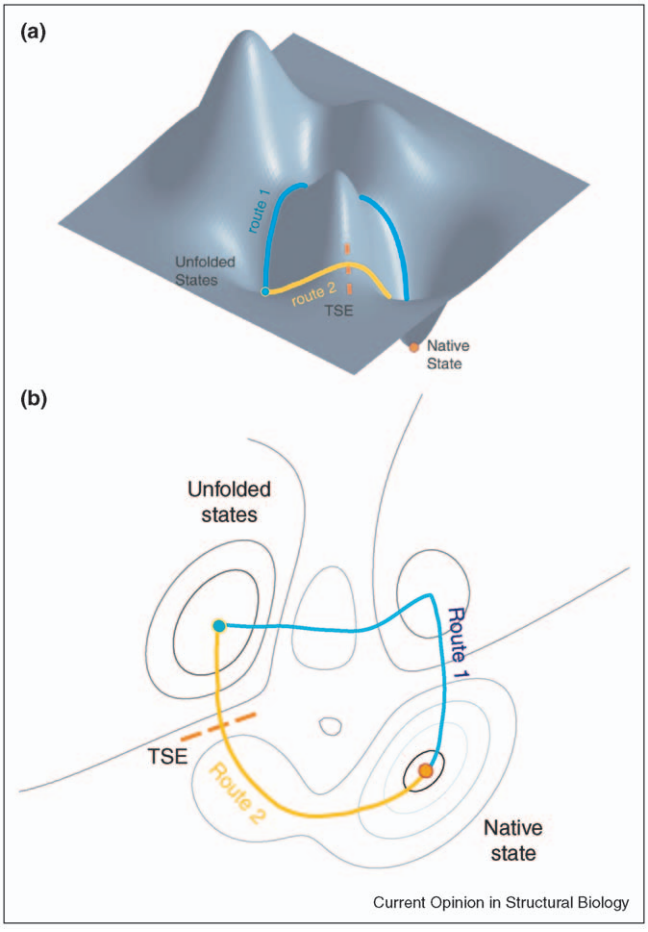
Protein folding under the 'computational' microscope

Developments in molecular modeling approaches have led to significant recent breakthroughs in our understanding of protein folding. Schematic mechanisms of protein folding are now being filled with details. With this information, we can observe new steps in folding pathways and intricate details of the transition states ensemble (TSE) — conformations that appear at the top of the free energy barrier separating the folded and unfolded states of two-state proteins.

Specifically, we learned how important the details of the free energy landscape are. These details include multiple folding pathways and multiple barriers separating the folded and unfolded states (Figure 3). For example, using DMD simulations, Borreguero *et al.* [34] have identified multiple distinct folding pathways of the c-Crk SH3 domain. By studying the folding transitions under various conditions, which are controlled by simulation temperature, the authors determined the kinetic partition temperature — the temperature at which the c-Crk SH3 domain undergoes a rapid folding transition via the path with the lowest free energy barrier. Below this temperature, the model protein folding transition occurred via multiple folding pathways. This finding suggests the possibility that proteins may feature multiple folding pathways.

Using all-atom MC simulations, Shimada and Shakhnovich [35] found that protein G undergoes three distinct folding pathways. Prior experimental studies, based on the monitoring of the burial of the lone tryptophan in protein G, yielded single exponential kinetics, suggesting the two-state nature of protein G folding [36]. Shimada and Shakhnovich demonstrated that the asymmetric location of the tryptophan in the protein structure does not adequately capture folded protein states, thereby 'disguising' the presence of multiple intermediates. Using Brownian dynamics simulations with a protein model made of coarse-grained physical energy functions responsible for sequence-specific interactions and weak Gō-like energies [13], Lee *et al.* [37] corroborated the findings of Shimada and Shakhnovich [35]. A further surprising aspect of protein G folding was reported by Hubner *et al.* [38••], who found that all folding pathways cross the same most dominant free energy barrier. The TSE obtained in their all-atom MC simulations corresponded to a single nucleation event, which results from the 'convergence' of distinct folding pathways.

Figure 3



Three-dimensional **(a)** and contour **(b)** plots of the free energy landscape of a hypothetical protein. Protein folding most likely proceeds via two routes (1 and 2). Although route 2, which connects the unfolded and native states via a smaller free energy barrier (marked by the dashed line), may be the preferable protein folding pathway, under certain conditions, such as salt concentration and pH, route 1 is preferred. In addition, under specific experimental conditions, a local minimum appearing on route 1 of the free energy landscape (b) can be promoted, thereby favoring the formation of the thermodynamically stable folding intermediate.

The protein G story taught us an important lesson: experiments often provide an average view of the folding transition. The choice of monitored parameters may skew the interpretation of experimental data. For example, an innovative methodology for assessing the roles of individual amino acids, known as Φ-value analysis, was proposed by Fersht and colleagues [39,40]. A residue's Φ-value quantifies the impact of this residue's substitution

on the TSE and, therefore, has been suggested to measure the amount of native structure formed by the neighboring residues in the TSE [41]. Although such an interpretation of Φ-values is reasonable and holds for some proteins, such as barnase [42], for other proteins, Φ-values may represent the average participation of residues in the TSE (which incidentally may have its own 'rippled' structure [43]) that appears either on malleable [44••] or on 'convergent' [38••] folding routes. For example, Hubner *et al.* [45•], in their recent study of the SH3 domain, found that the nucleus of this protein — a minimal set of residues that interact at the TSE — is highly polarized, with several residues exposed to the solvent. It is interesting that, despite the differences between the folding TSEs of barnase and the SH3 domain, both studies converge on the nucleation protein folding scenario [46,47], whereby the formation of the nucleus determines the TSE.

The nature of the TSE is also important for understanding protein association, which may compete with folding of the constituent subunits. A recent study by Levy *et al.* [48] demonstrated that the binding mechanism and the nature of the binding TSE can be deciphered from the interactions that stabilize the native fold. It is remarkable that a simplified protein model based on a Cα representation and Gō interaction potential [13] resulted in a comprehensive protein folding and binding landscape.

The future of molecular simulations

Cellular life is organized hierarchically: whereas proteins have specific functions (e.g. specific enzymatic reactions), protein complexes represent higher level functional modules responsible for large-scale events in cellular life (e.g. DNA transcription). Molecular complexes are often dynamic — they appear in various forms at various stages of their functional life. Thus, understanding the detailed structures of these large molecular assemblies and their dynamics is necessary to developing a higher level understanding of cellular biology. Uncovering the structure and dynamics of these complexes will require breakthroughs in molecular modeling that span scales from atoms to protein complexes.

It is clear that simplified protein models will have a prominent role in these approaches because of their ability to reach biologically relevant time and length scales in MD simulations. The intrinsic accuracy limitations of these models can be extended by slower but more accurate all-atom physical models and by experiments. Hence, the development of a united framework for molecular modeling that will integrate simplified and detailed protein models is a promising avenue for future molecular simulations. These simulations would be driven by phenomenological force-fields or experimental constraints at the coarse-grained modeling level, and by a physical force-field at the all-atom level.

Conclusions

The use of simplified protein models has allowed us to glance into the large-scale dynamics of biological molecules. The atomic accuracy of molecular modeling can be recovered by integrating the simulations of simplified models and all-atom simulations. Simplified models can also be 'guided' by experiments in the time scale regime accessible to given experimental approaches. Because of the ability of simplified models to access time and length scales of biological importance, we can ask questions directly pertinent to complex biological phenomena. Hence, in the era of systems biology, the development of simplified models is pivotal to the future of molecular modeling.

Update

Recent work has demonstrated that the force-field developed for the folding of the Trp-cage miniprotein, described in [19••], is transferable to other systems, such as the 17-residue ccβ peptide. Using DMD simulations, Ding *et al.* [49•] showed that three ccβ peptides form a three-helix bundle, which is within 1.26 Å root mean square distance from the crystal structure [26]. Although validation and further development of this force-field is required for application to larger proteins, the success with ccβ peptide folding and assembly is an important indication that force-field transferability may no longer be a limitation of simplified protein models.

The work referred to in the text as (F Ding *et al.*, unpublished) has now been published [49•].

Acknowledgements

I would like to thank F Ding and EI Shakhnovich for helpful discussions. This work is supported in part by Muscular Dystrophy Association grant MDA3720 and research grant number 5-FY03-155 from the March of Dimes Birth Defect Foundation.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Dobson CM: **In the footsteps of alchemists**. *Science* 2004, **304**:1259-1262.
2. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M: **CHARMM: a program for macromolecular energy, minimization, and dynamics calculations**. *J Comput Chem* 1983, **4**:187-217.
3. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt SE, Ferguson DM, Seibel GL, Kollman PA: **AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules**. *Comput Phys Commun* 1995, **91**:1-41.
4. Hansmann UHE: **Protein-folding simulations in generalized ensembles**. *Int J Quantum Chem* 2002, **90**:1515-1521.
5. Snow CD, Sorin EJ, Rhee YM, Pande VS: **How well can simulation predict protein folding kinetics and thermodynamics?** *Annu Rev Biophys Biomol Struct* 2005, **34**:43-69.

6. Snow CD, Nguyen N, Pande VS, Gruebele M: **Absolute comparison of simulated and experimental protein-folding dynamics**. *Nature* 2002, **420**:102-106.

7. Abkevich VI, Gutin AM, Shakhnovich EI: **Impact of local and nonlocal interactions on thermodynamics and kinetics of protein folding**. *J Mol Biol* 1995, **252**:460-471.

8. Skolnick J, Kolinski A, Ortiz AR: **MONSSTER: a method for folding globular proteins with a small number of distance restraints**. *J Mol Biol* 1997, **265**:217-241.

9. Zhou Y, Karplus M: **Folding thermodynamics of a model three-helix-bundle protein**. *Proc Natl Acad Sci USA* 1997, **94**:14429-14432.

10. Ding F, Dokholyan NV: **Simple but predictive protein models**. *Trends Biotechnol* 2005, **23**:450-455.

11. Dokholyan NV, Borreguero JM, Buldyrev SV, Ding F, Stanley HE, Shakhnovich EI: **Identifying importance of amino acids for protein folding from crystal structures**. *Methods Enzymol* 2003, **374**:616-638.

12. Smith SW, Hall CK, Freeman BD: **Molecular dynamics for polymeric fluids using discontinuous potentials**. *J Comput Phys* 1997, **134**:16-30.

13. Go N, Abe H: **Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation**. *Biopolymers* 1981, **20**:991-1011.

14. Dixon RDS, Chen Y, Ding F, Khare SD, Campbell SL, Dokholyan NV: **New insights into FAK signaling and localization based on detection of a FAT domain folding intermediate**. *Structure* 2004, **12**:2161-2171.

15. Miller R, Danko CA, Fasolka MJ, Balazs AC, Chan HS, Dill KA: **Folding kinetics of proteins and copolymers**. *J Chem Phys* 1992, **96**:768-780.

16. Srinivasan R, Rose GD: **A physical basis for protein secondary structure**. *Proc Natl Acad Sci USA* 1999, **96**:14258-14263.

17. •• Bradley P, Misura KMS, Baker D: **Toward high-resolution *de novo* structure prediction for small proteins**. *Science* 2005, **309**:1868-1871.

The authors reported high-resolution structure prediction (<1.5 Å root mean square distance) for small proteins (up to 85 amino acids) using a multiscale approach that sampled low- and high-resolution protein conformations.

18. Feig M, Karanicolas J, Brooks CL: **MMTSB tool set: enhanced sampling and multiscale modeling methods for applications in structural biology**. *J Mol Graph Model* 2004, **22**:377-395.

19. •• Ding F, Buldyrev SV, Dokholyan NV: **Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model**. *Biophys J* 2005, **88**:147-155.

The authors developed a coarse-grained protein model that was used in DMD folding simulations of the Trp-cage from a fully extended conformation. The model consistently reached conformations within a 2 Å backbone root mean square distance from the corresponding NMR structures. These findings suggested that coarse-grained simulations with a proper force-field can be as accurate as detailed all-atom protein models with an MD force-field.

20. Simmerling C, Strockbine B, Roitberg AE: **All-atom structure prediction and folding simulations of a stable protein**. *J Am Chem Soc* 2002, **124**:11258-11259.

21. Snow CD, Zagrovic B, Pande VS: **The Trp cage: folding kinetics and unfolded state topology via molecular dynamics simulations**. *J Am Chem Soc* 2002, **124**:14548-14549.

22. Zagrovic B, Pande V: **Solvent viscosity dependence of the folding rate of a small protein: distributed computing study**. *J Comput Chem* 2003, **24**:1432-1436.

23. Zhou R: **Trp-cage: folding free energy landscape in explicit water**. *Proc Natl Acad Sci USA* 2003, **100**:13280-13285.

24. Pitera JW, Swope W: **Understanding folding and design: replica-exchange simulations of "Trp-cage" fly miniproteins**. *Proc Natl Acad Sci USA* 2003, **100**:7587-7592.

25. Chowdhury S, Lee MC, Xiong GM, Duan Y: ***Ab initio* folding simulation of the Trp-cage mini-protein approaches NMR resolution**. *J Mol Biol* 2003, **327**:711-717.

26. Kammerer RA, Kostrewa D, Zurdo J, Detken A, Garcia-Echeverria C, Green JD, Muller SA, Meier BH, Winkler FK, Dobson CM, Steinmetz MO: **Exploring amyloid formation by a *de novo* design**. *Proc Natl Acad Sci USA* 2004, **101**:4435-4440.

27. Urbanc B, Cruz L, Ding F, Sammond D, Khare S, Buldyrev SV, Stanley HE, Dokholyan NV: **Molecular dynamics simulation of amyloid beta dimer formation**. *Biophys J* 2004, **87**:2310-2321.

28. • Khare SD, Ding F, Gwanmesia KN, Dokholyan NV: **Molecular origin of polyglutamine-mediated protein aggregation in neurodegenerative diseases**. *PLoS Comput Biol* 2005, **1**:230-235.

Using a simplified model of polyglutamine peptides, the authors showed that they form a specific secondary structure, the β-helix. The probability of β-helix formation increases with increasing peptide length, consistent with the etiologies of nine neurodegenerative diseases, such as Huntington's disease. This work provided a biophysical mechanism for the aggregation of polyglutamine-containing proteins in nine neurodegenerative diseases.

29. Ding F, Guo W, Dokholyan NV, Shakhnovich EI, Shea JE: **Reconstruction of the src-SH3 protein domain transition state ensemble using multiscale molecular dynamics simulations**. *J Mol Biol* 2005, **350**:1035-1050.

30. Vendruscolo M, Paci E, Dobson CM, Karplus M: **Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange**. *J Am Chem Soc* 2003, **125**:15686-15687.

31. •• Falke S, Tama F, Brooks CL III, Gogol EP, Fisher MT: **The 13 angstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy**. *J Mol Biol* 2005, **348**:219-230.

Using cryo-EM and simplified protein models, the authors reconstructed, at 13 Å resolution, the structure of the ~800 kDa GroEL chaperonin complex in the unliganded state and bound to a single monomer of the protein substrate GS_m. With the aid of normal mode flexible fitting analysis, the authors uncovered the structural rearrangements that occur in GroEL upon substrate binding and demonstrated that these rearrangements have a significant impact on GroEL as they propagate throughout the structure of the complex.

32. Tama F, Miyashita O, Brooks CL III: **Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM**. *J Struct Biol* 2004, **147**:315-326.

33. • Ding F, Jha RK, Dokholyan NV: **Scaling behavior and structure of denatured proteins**. *Structure* 2005, **13**:1047-1054.

Recent NMR spectroscopy studies of proteins at high chemical denaturant concentrations suggested the presence of significant amounts of native-like structure, in contrast to the classical random-coil model of proteins, which has been supported by several experiments. To reconcile these seemingly controversial observations, the authors employed DMD simulations of simple protein models to examine thermally denatured states of experimentally characterized proteins. It was found that, whereas the sizes of the denatured proteins were fully recapitulated in these simulations, the proteins featured 'flickering' local structural organization. This study directly demonstrated that the presence of local structure in denatured proteins does not contradict the random-coil view of proteins.

34. Borreguero JM, Ding F, Buldyrev SV, Stanley HE, Dokholyan NV: **Multiple folding pathways of the SH3 domain**. *Biophys J* 2004, **87**:521-533.

35. Shimada J, Shakhnovich EI: **The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation**. *Proc Natl Acad Sci USA* 2002, **99**:11175-11180.

36. Alexander P, Orban J, Bryan P: **Kinetic-analysis of folding and unfolding the 56-amino acid IgG-binding domain of streptococcal protein-G**. *Biochemistry* 1992, **31**:7243-7248.

37. Lee SY, Fujitsuka Y, Kim DH, Takada S: **Roles of physical interactions in determining protein-folding mechanisms: molecular simulation of protein G and alpha spectrin SH3**. *Proteins* 2004, **55**:128-138.

38. Hubner IA, Shimada J, Shakhnovich EI: **Commitment and**
•• **nucleation in the protein G transition state**. *J Mol Biol* 2004, **336**:745-761.
Using all-atom MC simulations of protein G, the authors tested the validity of interpreting experimental Φ-values as the fraction of native contacts made by a residue in the TSE. They demonstrated that this definition was unable to uniquely specify a TSE and that a high experimental Φ-value does not guarantee a residue's importance in the TSE. The authors found that protein G undergoes the folding transition via three distinct pathways, all of which converge on a specific nucleus formed by six residues. The predicted nucleus was corroborated by evolutionary analysis and experiments. The authors constructed a unified theoretical model of protein G folding.

39. Matouschek A, Kellis JT, Serrano L, Bycroft M, Fersht AR: **Transient folding intermediates characterized by protein engineering**. *Nature* 1990, **346**:440-445.

40. Matouschek A, Kellis JT, Serrano L, Fersht AR: **Mapping the transition-state and pathway of protein folding by protein engineering**. *Nature* 1989, **340**:122-126.

41. Vendruscolo M, Paci E, Dobson CM, Karplus M: **Three key residues form a critical contact network in a protein folding transition state**. *Nature* 2001, **409**:641-645.

42. Salvatella X, Dobson CM, Fersht AR, Vendruscolo M: **Determination of the folding transition states of barnase by using Φ_I-value-restrained simulations validated by double mutant Φ_{IJ}-values**. *Proc Natl Acad Sci USA* 2005, **102**:12389-12394.

43. Oliveberg M, Tan YJ, Silow M, Fersht AR: **The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding**. *J Mol Biol* 1998, **277**:933-943.

44. Shen T, Hofmann CP, Oliveberg M, Wolynes PG: **Scanning**
•• **malleable transition state ensembles: comparing theory and experiment for folding protein U1A**. *Biochemistry* 2005, **44**:6433-6439.
Using a variational free energy functional, the authors calculated the characteristics (chevron plots) of the TSE for the folding of wild-type and mutant U1A proteins. Remarkably, this simplified theoretical approach predicted the detailed variations of the TSE and changes in the chevron plots agreed well with experiments.

45. Hubner IA, Edmonds KA, Shakhnovich EI: **Nucleation and**
• **the transition state of the SH3 domain**. *J Mol Biol* 2005, **349**:424-434.
Using all-atom MC protein folding simulations and experimental Φ-values as constraints, the authors identified the TSE of the SH3 domain. They demonstrated that averaging over all conformations may consider non-TSE conformations as transition states. Computing transmission coefficients ('probability to fold') for putative conformations allowed the authors to make rigorous conclusions regarding the structure of the folding nucleus. Through analysis of the TSE, the authors discovered that the nucleus has many residues that are solvent exposed.

46. Abkevich VI, Gutin AM, Shakhnovich EI: **Specific nucleus as the transition-state for protein-folding - evidence from the lattice model**. *Biochemistry* 1994, **33**:10026-10036.

47. Fersht AR: **Nucleation mechanisms in protein folding**. *Curr Opin Struct Biol* 1997, **7**:3-9.

48. Levy Y, Cho SS, Onuchic JN, Wolynes PG: **A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes**. *J Mol Biol* 2005, **346**:1121-1145.

49. Ding F, LaRocque JJ, Dokholyan NV: **Direct observation**
• **of protein folding, aggregation, and a prion-like conformational conversion**. *J Biol Chem* 2005, **280**:40235-40240.
The authors demonstrated the transferability of a force-field developed for a simplified protein model. In addition, they reported the direct observation of the prion-like conformational interconversion of peptides into amyloid-like structures, whereby two peptides that have adopted β-hairpin conformations promote such interconversion of the third peptide. This study suggested a possible generic molecular mechanism for the template-mediated aggregation process, originally proposed by Prusiner to account for prion infectivity.