Coarse-grained Protein Modeling: Dynamics, Folding Pathways and Mechanical Unfolding

Andrzej Kolinski, Maciej Blaszczyk, and Sebastian Kmiecik

Laboratory of Theory of Biopolymers, Faculty of Chemistry, University of Warsaw
Pasteura 1, 02-093 Warsaw, Poland
E-mail: kolinski@chem.uw.edu.pl

Coarse-grained protein modeling tool, CABS, is used in multiscale modeling of protein dynamics. It is demonstrated that the stochastic (Monte Carlo) dynamics, combined with all-atom refinement of the coarse-grained structures follows observed in experiments folding pathways of small proteins. The model is also used in model studies of chaperonin-assisted protein folding. It is shown that Iterative Annealing Mechanism of chaperonin action, where periodic distortion of the polypeptide chains by non-specific hydrophobic interactions can promote rapid folding and leads to a decrease in folding temperature. It is also demonstrated how chaperonin action prevents kinetically trapped conformations and modulates the observed folding mechanisms from nucleation-condensation to a more framework-like. Finally, the CABS model is used in simulations of mechanical unfolding, providing new ways for interpretation of Atomic Force Microscopy experiments.

1 Introduction

Genomic projects provide huge number of protein sequences. Experimental determination of protein structures, while very successful, stays well behind the sequencing – at the moment the number of known structures is about 75000. Knowledge of three dimensional structures of protein and protein-protein and protein-nucleic acids assemblies is crucial for understanding the molecular mechanisms of life, and consequently for the rational drug design, modern biotechnologies, etc. Fortunately, the much cheaper and faster *in silico* structure determination is now possible for a large fraction of known sequences. Comparative modeling is the most accurate and the most advanced method. Obviously, with the increasing number of the high accuracy experimentally determined structures the range of application of comparative modeling will increase. A more straightforward, template-free, structure calculation of new folds is now possible only for relatively small and topologically simple proteins, although there is a steady progress in this area.

By many years theoretical studies of biomacromolecules have been guided by a well-defined paradigm: **from sequence – to structure – to function**. Recently, an extension of this paradigm becomes evident. More and more researchers try to predict and describe the dynamics of biomacromolecules, especially proteins. Professor Harold Scheraga and his coworkers are among the leaders of this trend. It becomes widely recognized that most of protein functions (enzymatic action, recognition, signaling, transporting, etc.) could not be fully understood without a clear reference to their dynamic context. Thus, the current paradigm of structural biology could be rephrased as: **from sequence** – (via dynamics of folding or biomacromolecular assembly pathways) **to structure – through dynamics – to function**.

2 Motivation

Due to the level of complexity, and the timescales involved, computational study of protein systems by means of classical (atomistic level) molecular dynamics (or stochastic dynamics) are now limited to small systems or/and to localized conformational changes. The problem is summarized in Fig. 1, where the timescales of various dynamic processes are compared with the timeframes accessible to various experimental techniques and a typical timeframe of large-scale molecular dynamics simulations. To make the problem of large scale protein dynamics computationally tractable it is necessary to reduce the level of complexity[1–4] and/or to apply more efficient simulation techniques[5]. Coarse-grained models of protein dynamics span a wide range of designs – from mesoscopic continuous models (like Gaussian Network Models – GNM[6]) through united-residue models (like UNRES[7] or CABS[2]) to hybrid models, where some atoms are treated explicitly, while the remaining atoms are combined into united atoms[8,9]. The first class of models (GNM) are at present applicable to moderate changes of molecular geometry, although the size of the tractable systems is essentially unlimited[6]. The GNM-type models are intrinsically structure-based. Conformational fluctuation can be calculated around a starting structure, usually the native-like state. United residue models enable studies of large conformational transitions, while the system's size limits depend on particular designs of the force fields and sampling schemes. The hybrid models bridge the gap between the united-residue and all-atom simulations. Sometimes the structure-based features are introduced into coarse-grained (and all-atom) models via a specific simplification of molecular interactions. For instance, in so called Go-models[10] only the residue-residue interactions seen in the native state define the protein energy landscape. Such approximation seem to work quite well when applied in simulations of mechanical unfolding[11]. On the other hand, the results of numerous applications of Go-type models to protein folding pathways prediction should be treated with extreme cautions – many proteins do fold through non-native intermediates.

In this contribution we describe a couple of new applications (published and unpublished) of CABS model to prediction of protein dynamics. Previously, CABS has been successfully used in test predictions of equilibrium folding pathways of several small proteins[13–15], including the atomic level reconstruction of folding nuclei. Here, we study effects of external forces/restraints on folding/unfolding dynamics and thermodynamics.

3 Coarse-grained CABS Model

CABS is a coarse-grained protein modeling tool applicable to large scale simulations of protein structure (comparative and *de novo* modeling) protein folding dynamics, mechanisms of multimeric assembly and mechanisms of molecular machinery[2]. Acronym CABS stands for the names of united atoms (residues) employed in the reduced representation of protein conformations (see Fig. 2a). The main chain of a polypeptide is represented by the alpha carbon (Ca) trace, which is restricted to an underlying fine-mesh (spacing: 0.61 Å) simple cubic lattice. Allowing some fluctuation of the canonical Ca-Ca distances the number of allowed orientation of Ca-Ca pseudobonds is equal to 800. This safely eliminates any effects of the lattice anisotropy. Additionally, the model backbone contains pseudoatoms located at centers of Ca-Ca pseudobonds. These support a definition of the main-chain hydrogen bonds. The side chains are represented by up to two united atoms,

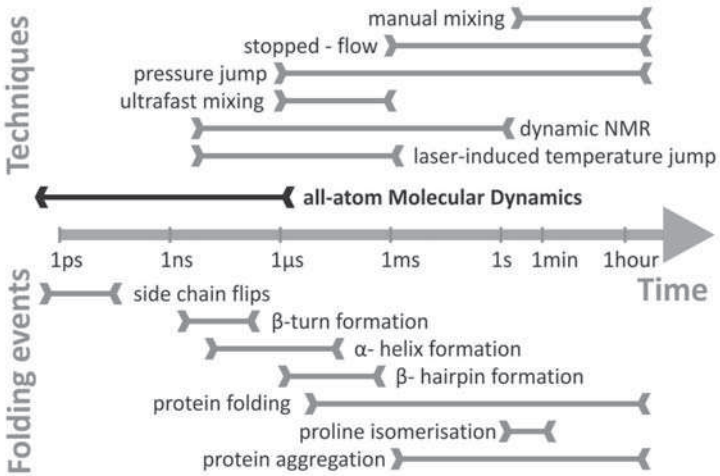

Figure 1. Protein folding dynamics timescale resolutions shown for experimental techniques and all-atom MD (above the axis) and timescales of protein folding events (below the axis)[12].

corresponding to the beta carbons and to geometric centers of the side chains. The side chains are located off-lattice and their positions are defined in the coordinate reference system defined by the Ca-trace (see Fig. 2b), allowing for some rotamer variability.

Motion of CABS polypeptide chains is simulated by means of stochastic dynamics (Monte Carlo Dynamics, MCD) resulting from long series of randomly selected local conformational transitions. These include single Ca moves (see Fig. 2b) accompanied by proper changes of the involved side chains (for clarity the changes of the flanking side chains are not shown in Fig. 2b) and a larger scale transitions, schematically depicted in Fig. 2c. The larger scale moves are attempted less frequently. Lattice representation facilitates very fast computations of conformational transitions leading to speed-up of the model dynamics by about two orders of magnitude in respect to otherwise similar continuous space models. The sampling scheme of CABS depends on specific applications – for structure prediction a multicopy method is recommended (for instance a variant of Replica Exchange Monte Carlo), while for simulations of dynamics a single copy algorithms are easier to interpret (isothermal MCD, or simulated annealing).

Force field of CABS is knowledge-based and consists of statistical potentials derived from structural regularities seen in the known protein structures. The interaction scheme includes: generic short-range conformational propensities mimicking the stiffness of polypeptide chains, sequence-specific short-range (up to four residues along the chain) conformational propensities, a model of highly directional main chain hydrogen bond and the long-range (unbonded) pairwise interactions of the side chains. The side chain interactions are context-specific (depend on mutual orientation of the interacting residues) and thereby account for complex many-body effects (and an averaged implicit solvent effects). Details of the CABS force field design could be found in earlier publications[2] and the numerical data for the potentials could be seen and downloaded from our website (`www.biocomp.chem.uw.edu.pl`).

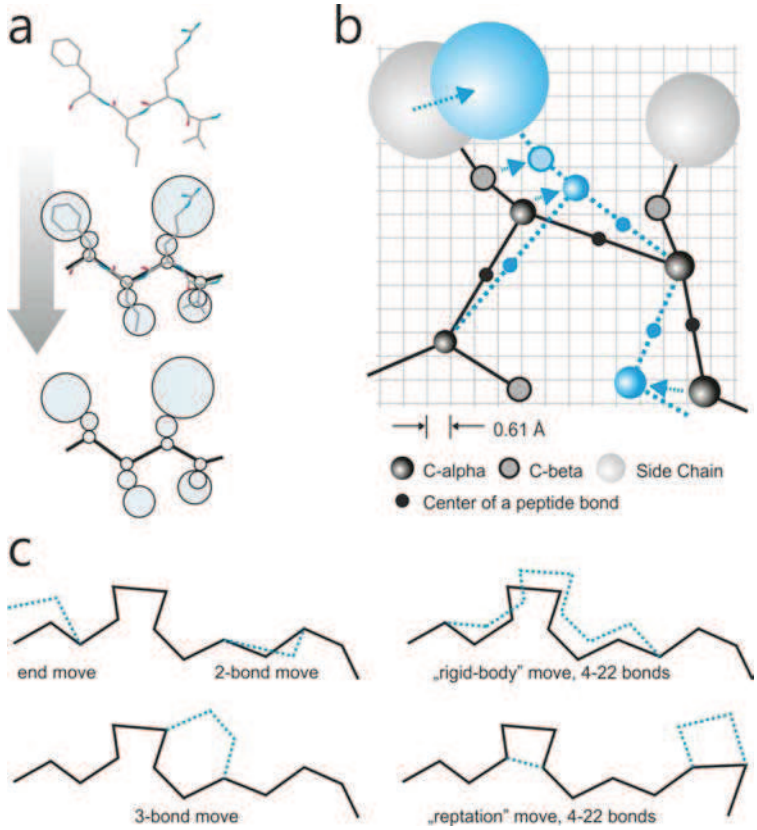

Figure 2. CABS model overview: (a) reduced representation (b) single Ca kink move (c) schematic illustration of larger scale moves of the MCD scheme.

The lattice design of CABS is quite unique. How does it compare with more popular continuous space models? The representation of UNRES[7] is quite similar to that of CABS, except the side chains which in UNRES are single entities in a form of ellipsoids of revolution (two spherical united atoms in CABS). Continuous representation of UNRES enables classical molecular dynamics sampling, although other (sometimes quite sophisticated) sampling schemes were used by the Authors. The main difference is in the designs of the force fields of both models. UNRES force field is physics-based, where parameters of potentials are derived from corresponding atom-level interactions. On contrary, the force field of CABS is fully knowledge-based, attempting to reproduce the very complex patterns of effective interactions in proteins without referring to specific atom-level origins. Both approaches could be comprehensively argued for, depending on a context.

4 Modeling of Chaperonins' Mechanism of Action

Molecular chaperonins are hollow protein chaperones. A single type of a chaperonin (for instance GroEL/GroES complex) can assists in folding of many various proteins. The exact

mechanism of chaperonin action remains largely unknown. That is because difficulties in experimental probing of protein folding mechanisms in such a confined environment. Nevertheless, it is known that for many substrate proteins the interactions with the chaperonin are non-specific and that the periodic changes of hydrophobicity of the chaperonin interior are crucial for its action. Basing on these observations a very simple theoretical model of the chaperonins' mode of action has been proposed[16]. Iterative annealing model (IAM) assumes that the periodic changes of hydrophobicity of the inner walls of a chaperonin complex lead to unfolding of misfolded proteins, consequently enabling them to reach the proper fold. Using the fundamental theoretical concept of IAM we designed a simulation scheme based on CABS. The idea is explained in Fig. 3. The chaperonin is modeled as a spherical object (cylindrical chaperonins were also simulated, without qualitative differences in the results) with thick walls (or a vicinity of the walls) of variable hydrophobicity. In the basic state 9/10 of the simulation time the walls are inert. Periodically (see the right side of Fig. 3) the walls became hydrophobic, attracting the encapsulated protein chain with the strength typical for the hydrophobic interactions within folded proteins (according to the CABS force field).

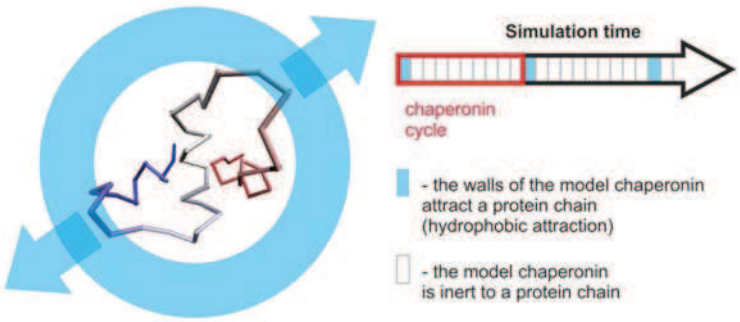

Figure 3. Chaperonin model (see the text).

Chaperonin-assisted folding simulations were performed for two examples: 56 residue B1 domain of protein G and for 46 residue B domain of protein A[16]. The first is an alpha/beta protein and the second a helical protein. These proteins perhaps are not typical chaperonin substrates, but this seems to be irrelevant for the general question about the molecular implications of the IAM model. On the other hand these proteins are paradigms of experimental and theoretical studies of protein folding mechanisms. Since their chaperone-free CABS simulations agree very well with experimental facts, we expect that the observed changes in folding mechanisms within the chaperonin are meaningful. During the simulations we monitored the changes of protein coil/globule volumes, the numbers of native contacts, the secondary structure contents and other structural, dynamic and thermodynamic characteristics. In comparison to the bulk (chaperonin-free) folding simulations the results of simulations of the chaperonin-assisted folding of both proteins could be characterized as follows:

1. Chaperonin significantly (up to two-times) increases the frequency of the folding-unfolding transitions between the near-native structures and the highly denatured structures.

2. Chaperonin decreases the folding temperature (defined as the temperature at which the number of native-denatured transitions is the largest).

3. Chaperonin shifts the folding mechanisms from a sequential, nucleation-and-growth, towards a more prefabricated, framework type assembly. This manifests itself in weakening of typical nucleation sites and a larger volume of intermediates with a significant, native-like, secondary structure content (see Fig. 4).

4. Contrary to the original interpretation of the theoretical IAM model, the simulations show that the chaperonin prevents formation of misfolded conformations, rather than unfolding the already existing misfolds (see Fig. 4).

While very plausible, the specific predictions emerging from these simulations await experimental verification.

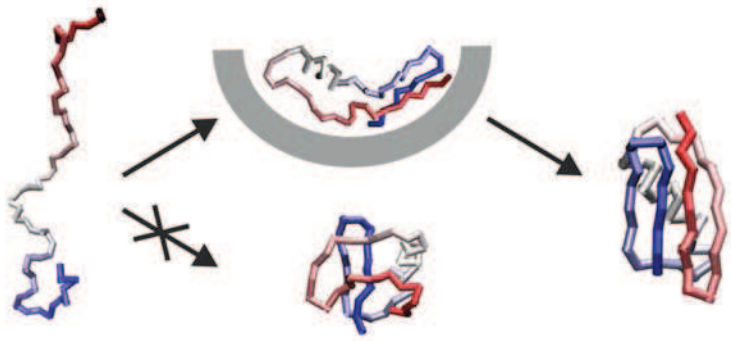

Figure 4. Illustration of the misfolding prevention mechanism within a chaperonin. Pictures from B1 domain of protein G folding simulations[16].

5 *In silico* Atomic Force Microscopy Experiment

The Atomic Force Microscopy is a powerful method for single molecule biophysical experimentations. However, interpretation of the AFM curves requires a model and simulations. Probably, at present the most popular are simulations with Go-type reduced models. As mentioned before, the Go-type approximation, where only native interactions are taken into account, could be non-realistic. Here, we give an example of *in silico* AFM simulation for the CABS model, where the force field do not include the *a priori* knowledge of a specific structure. The simulations correspond to isothermal conditions, well below unfolding temperature. The computational experiment is illustrated in Fig. 5. The molecule (apo-plastocyanin) is attached at one end to the AFM table and at the opposite end to

the molecular cantilever. The cantilever moves apart the table with constant speed and the force exerted on the spring is measured. The resulting force/time (or force/extension) curve exhibits characteristic rapid drops of the force upon unfolding of particular fragments of the structure. Usually, the C-terminal hairpin of the plastocyanin unfolds as the first, immediately followed by unfolding of the N-terminal hairpin (both events are close in time and sometimes the order of the two events changes). The most resistant is the central foldon (a six-member beta barrel) that always unfolds as the last.

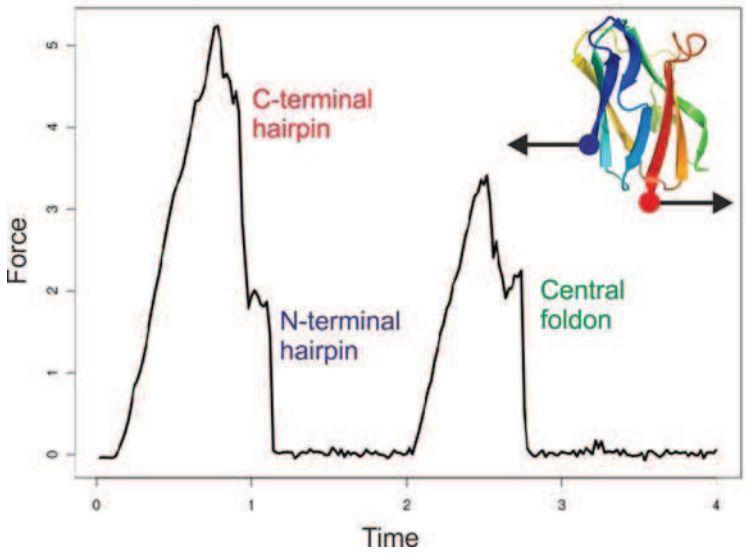

Figure 5. *In silico* AFM experiment with CABS (see the text).

While promising, the CABS assisted interpretation of the AFM experiments requires a precise scaling between the time of the physical AFM experiments and the time of the coarse-grained AFM simulations (which determines a meaningful elongation speed in the simulated experiments). The work along this direction is in progress.

Acknowledgments

This work has been supported by the Polish National Science Center (NN301071140) and the Polish Ministry of Science and Higher Education (NN507326536). Computational part of this work was done using the computer cluster at the Computing Center of Faculty of Chemistry, University of Warsaw.

References

1. H. A. Scheraga, M. Khalili, and A. Liwo, *Protein-folding dynamics: overview of molecular simulation techniques*, Annu Rev Phys Chem, **58**, 57–83, 2007.

2. A. Kolinski, *Protein modeling and structure prediction with a reduced representation*, Acta Biochim Pol, **51**, no. 2, 349–71, 2004.
3. V. Carnevale, C. Micheletti, F. Pontiggia, and R. Potestio, *Bridging the Atomic and Coarse-Grained Descriptions of Collective Motions in Proteins* in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 159–178, Springer, New York, 2011.
4. M. Feig, S. Gopal, K. Vadivel, and A. Stumpff-Kane, *Conformational sampling in structure prediction and refinement with atomistic and coarse-grained models* in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 85–109, Springer, New York, 2011.
5. U. H. E. Hansmann, *Sampling Protein Energy Landscapes The Quest for Efficient Algorithms* in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 209–230, Springer, New York, 2011.
6. T. Haliloglu, I. Bahar, and B. Erman, *Gaussian dynamics of folded proteins*, Physical Review Letters, **79**, no. 16, 3090–3093, 1997.
7. A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Oldziej, K. Wachucik, and H. A. Scheraga, *Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins*, J Phys Chem B, **111**, no. 1, 260–85, 2007.
8. S. M. Gopal, S. Mukherjee, Y. M. Cheng, and M. Feig, *PRIMO/PRIMONA: a coarse-grained model for proteins and nucleic acids that preserves near-atomistic accuracy*, Proteins, **78**, no. 5, 1266–81, 2010.
9. A. Irbäck and S. Mohanty, *Effective All-Atom Potentials for Proteins*, pp. 111–126, Springer, New York, 2011.
10. H. Taketomi, Y. Ueda, and N. Go, *Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions*, Int J Pept Protein Res, **7**, no. 6, 445–59, 1975.
11. J. I. Sulkowska and M. Cieplak, *Selection of optimal variants of Go-like models of proteins through studies of stretching*, Biophys J, **95**, no. 7, 3174–91, 2008.
12. S. Kmiecik, M. Jamroz, and A. Kolinski, *Multiscale approach to protein folding dynamics*, in Multiscale approaches to protein modeling, A. Kolinski (Ed.), pp. 281–294, Springer, New York, 2011.
13. S. Kmiecik and A. Kolinski, *Characterization of protein-folding pathways by reduced-space modeling*, Proc Natl Acad Sci USA, **104**, no. 30, 12330–5, 2007.
14. S. Kmiecik and A. Kolinski, *Folding pathway of the B1 domain of protein G explored by multiscale modeling*, Biophys J, **94**, no. 3, 726–36, 2008.
15. S. Kmiecik, M. Kurcinski, A. Rutkowska, D. Gront, and A. Kolinski, *Denatured proteins and early folding intermediates simulated in a reduced conformational space*, Acta Biochim Pol, **53**, no. 1, 131–44, 2006.
16. S. Kmiecik and A. Kolinski, *Simulation of chaperonin effect on protein folding: a shift from nucleation-condensation to framework mechanism*, J Am Chem Soc, **133**, no. 26, 10283–9, 2011.