Folding for binding or binding for folding?

Amedeo Caflisch

Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich, Switzerland
Phone: (41 1) 635 55 21
FAX: (41 1) 635 68 62
email: caflisch@bioc.unizh.ch

This article was published in *Trends in Biotechnology* **21**, 423-425, 2003.
It is posted with permission from Elsevier.

November 20, 2003

Abstract

A number of proteins are (partially) unstructured at physiological conditions and assume a well-defined three-dimensional structure only upon binding to another protein or macromolecular complex. Numerous examples have been found especially among proteins involved in regulatory functions in eukaryotic cells. Recent computer simulations by Verkhivker et al. have yielded an atomic detail picture of the folding of p27 and its concomitant binding to the phosphorylated cyclin A-cyclin dependent kinase 2 (Cdk2).

1. Introduction

The biological function of a protein is strictly related to its three-dimensional structure in the folded state. This is a dogma in molecular biology which might have to be slightly refined because recent experimental evidence indicates that there exist soluble proteins which lack a well-defined folded structure or contain disordered segments [1, 2]. A computational approach trained on known proteins with intrinsic disorder has predicted unstructured regions of 40 or more consecutive residues in between 36% and 63% of the genome of five eukaryotes [3]. Some of these might be false positives, e.g., long loops distant from the active/binding site in otherwise rather rigid enzymes/receptors. Yet, these high frequencies of occurrence are significant. In their recent simulation study Verkhivker et al. [4] have investigated the coupled folding and binding of p27 to the cyclin A-Cdk2 complex which is involved in cell cycle regulation. Their analysis yields an atomistic picture of the transition state ensemble (TSE) and the sequence of events for folding/binding. Additional computational studies of concomitant folding and binding are expected in the near future.

2. Computer simulations of protein unfolding/unbinding

Verkhivker et al. have used Monte Carlo simulations (Box 1) with a simple atomistic model of the energy. They have simulated unfolding/unbinding of p27 at high temperature starting from the crystal structure of the tertiary complex and have inferred the

sequence of events of folding/binding by assuming the principle of microscopic reversibility [4]. Using the same assumption, several molecular dynamics simulations in the last decade have yielded interesting results on protein unfolding/folding and insights to guide additional experiments [5, 6, 7, 8]. In the Monte Carlo simulation study of Verkhivker et al. the cyclin A-Cdk2 complex was kept rigid which could have a significant effect on the trajectories. Despite the approximations in the energy model, e.g., lack of an explicit electrostatic energy term, and the fact that the p27 conformations sampled at high temperature (and with rigid partners) might be rather different than those accessible at physiological conditions, the authors propose a hierarchy of structural events for p27 with beta-hairpin and beta-strand folding preceding alpha-helix formation. The simulation results are interesting but await experimental validation, e.g., by site-directed mutagenesis and rate measurements as is usually done on small proteins (phi-value analysis [9]).

3. Atomic picture of the transition state ensemble

Following an idea of Li and Daggett [10], Verkhivker et al. define the putative TSE conformations by first clustering the structures saved along the trajectories and then isolating those that undergo large structural modifications in short time intervals [4]. This is justified by the fact that the TSE is unstable and conformations close to it are characterized by rapid structural changes. Importantly, the authors have validated the TSE by running several short trajectories from the putative TSE conformations obtained by cluster analysis and checking that about one half of them fold/bind and the other half unfold/unbind [4]. In fact, being at the top of a barrier a small random perturbation should push the system on either side with 50% probability. This type of TSE validation has been used in the past in Monte Carlo simulations of a very simplified protein model (residues approximated as beads on a cubic lattice) [11], as well as in atomistic molecular dynamics simulations of folding of the src SH3 domain [12]. However, while the folding of several small proteins (less than about 100 residues) has been shown to be two-state with only the unfolded and folded state populated at equilibrium, the binding of two proteins could be a kinetically more complicated process even in the case of well structured partners [13]. For intrinsically flexible proteins, the kinetics might really be different from two-state behavior. The more complex behavior might explain the fraction of short trajectories that oscillate around the TSE without either folding or unfolding which the authors justify as "potential ruggedness on the top of a broad activation barrier" [4].

4. Advantages of disorder

The large majority of proteins have a folded state more favorable by only 5-15 kcal/mol with respect to the unfolded state. Hence, from a physical point of view an unstructured state more favorable than the folded state is not difficult to realize. The advantages of (partial) disorder, especially for eukaryotic cells, could be at least twofold. Most children (and some research scientists as well) do not like to keep order in their room (desks) not only because it is tedious but also because they can more easily visually recognize and reach the toys they want to play with (papers and documents to read). This metaphor illustrates a possible advantage of unstructured proteins in efficiently binding to other macromolecules. Along these lines, Shoemaker et al. [14] have proposed the "fly-casting" mechanism where an intrinsically disordered protein would have a greater "capture radius" than a folded protein with limited conformational flexibility. Another advantage could be a more precise and efficient control of multiprotein complexes and macromolecular recognition. Let us assume that protein C is functional only if bound to B, and B is intrinsically disordered and folds upon binding to A. In this way, the function of C is regulated by the presence of both A and B, where B can be more efficiently

controlled by the cell through proteolysis [1,2]. In other words, B has to bind to A to fold and fulfill its function (which is an example of the second half of the title of this article). The tumor-suppressor p53 is an interesting but also very important example since in about half of human cancer types p53 is inactivated as a consequence of mutations in the p53 gene. The transactivation domain of p53 assumes a helical conformation only upon binding to the cellular oncoprotein MDM2 [15]. It is important to note that p53 is a highly connected node in the network of protein interactions [16] and its coupled folding and binding might be essential for the regulation of transcriptional activation.

5. Future directions

In the near future it might be interesting to investigate the behavior of p27 with (partial) flexibility in the cyclin A-Cdk2 complex and eventually compare with high temperature molecular dynamics simulations. Since the structural details of coupled folding and binding for intrinsically disordered proteins are not clear, we expect additional simulation studies, by molecular dynamics or Monte Carlo on other biomolecular complexes including protein folding coupled to DNA or RNA binding. One important question concerns the possible disadvantages of intrinsically unstructured proteins which could be more prone to misfold and eventually aggregate. Simulations of peptide aggregation have been performed and analyzed in detail [17, 18, 19] but the competition between ordered aggregation and binding to the right partner has not (yet) been investigated by computational approaches. Comparison with experimental data will be of utmost importance to validate the simulation studies. As in the case of protein folding, strong synergies will emerge from a combined experimental and simulation analysis of the coupling of folding and binding.

Box 1. Monte Carlo and molecular dynamics simulations

Monte Carlo simulation is a stochastic approach to sample configuration space [20]. The molecules evolve by random (hence the name Monte Carlo) perturbations, e.g., in the torsional angles of a macromolecule or the position of a water molecule, and new configurations are accepted according to a criterion that keeps the system at equilibrium (at a given temperature in the simulations of Verkhivker et al. [4]). In molecular dynamics simulations the system evolves in time according to the Newton equation of motion. Both simulation approaches require a starting conformation (e.g., the X-ray or NMR structure) and a potential energy function whose parameters are usually derived from experimental data (e.g., vibrational spectra, structures and thermodynamic parameters of small molecules in the condensed phase) and ab initio quantum mechanical calculations. There are three main types of applications of biomolecular simulation approaches [21]. The first uses either Monte Carlo or molecular dynamics (or a combination of both) to simply sample configuration space, e.g., to determine peptide or protein conformations with data obtained from NMR or in X-ray structure refinement. In the second type of applications the system is sampled at equilibrium, i.e., the space is sampled such that each configuration is weighted with the appropriate Boltzmann factor. The Monte Carlo simulations of Verkhivker et al. [4] belong to this category. Finally, in the third type the actual dynamics is investigated and this can be done only by molecular dynamics because not only the system is sampled at equilibrium but the details of the time evolution, i.e., the natural dynamics, have to be accurately simulated. Typical examples of molecular dynamics simulations include peptide folding [22], protein unfolding [9], functional fluctuations in acetylcholinesterase [23], the spontaneous aggregation of phospholipids into a membrane [24], and ion conduction through the K+ channel [25].

Acknowledgments I thank A. Cavalli and R. Curcio for reading and commenting the manuscript. Work in my research group is supported in part by the National Competence Center for Research in Structural Biology.

References

[1] Wright, P. E. and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. 293, 321–331

[2] Dyson, H. J. and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. 12, 54–60

[3] Dunker, A. K. and Obradovic, Z. (2001) The protein trinity - linking function and disorder. Nature Biotechnology 19, 805–806

[4] Verkhivker, G. M. et al. (2003) Simulating disorder-order transitions in molecular recognition of unstructured proteins: Where folding meets binding. Proc. Natl. Acad. Sci. USA. 100, 5148–5153

[5] Caflisch, A. and Karplus, M. (1994) Molecular dynamics simulation of protein denaturation: Solvation of the hydrophobic cores and secondary structure of barnase. Proc. Natl. Acad. Sci. USA. 91, 1746–1750

[6] Daggett, V. et al. (1996) Structure of the transition state for folding of a protein derived from experiment and simulation. J. Mol. Biol. 257, 430–440,

[7] Lazaridis, T. and Karplus, M. (1997) "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 278, 1928–1931

[8] Levy, Y. and Caflisch, A. (2003) Flexibility of monomeric and dimeric HIV-1 protease. J. Phys. Chem. B 107, 3068–3079

[9] Daggett, V. and Fersht, A. R. (2003) Is there a unifying mechanism for protein folding? Trends in Biochemical Sciences 28, 18–25

[10] Li, A. and Daggett, V. (1994) Characterization of the transition state of protein unfolding by use of molecular dynamics: Chymotrypsin inhibitor 2. Proc. Natl. Acad. Sci. USA. 91, 10430–10434

[11] Du, R. et al. (1998) On the transition coordinate for protein folding. J. Chem. Phys. 108, 334–350

[12] Gsponer, J. and Caflisch, A. (2002) Molecular dynamics simulations of protein folding from the transition state. Proc. Natl. Acad. Sci. USA. 99, 6719–6724

[13] Taylor, M. G. et al. (1998) Kinetic epitope mapping of the chicken lysozyme-HyHEL-10 Fab complex: Delineation of docking trajectories. Protein Science 7, 1857–1867

[14] Shoemaker, B. A. et al. (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proc. Natl. Acad. Sci. USA. 97, 8868–8873

[15] Kussie, P. et al. (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 274, 948–953

[16] Vogelstein, B. et al. (2000) Surfing the p53 network. Nature 408, 307–308

[17] Ma, B. and Nussinov, R. (2002) Stabilities and conformations of Alzheimer's β-amyloid peptide oligomers ($A\beta_{16-22}$, $A\beta_{16-35}$, and $A\beta_{10-35}$): Sequence effects. Proc. Natl. Acad. Sci. USA. 99, 14126–14131

[18] Gsponer, J. et al. (2003) The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion sup35. Proc. Natl. Acad. Sci. USA. 100, 5154–5159

[19] Klimov, D. and Thirumalai, D. (2003) Dissecting the assembly of $A\beta_{16-22}$ amyloid peptides into antiparallel β sheets. Structure 11, 295–307

[20] Frenkel, D. and Smit, B. (2002) Understanding Molecular Simulations Academic Press, San Diego

[21] Karplus, M. and McCammon, J. A. (2002) Molecular dynamics simulations of biomolecules. Nature Struct. Biol. 9, 646–652

[22] Ferrara, P. and Caflisch, A. (2000) Folding simulations of a three-stranded antiparallel β-sheet peptide. Proc. Natl. Acad. Sci. USA. 97, 10780–10785

[23] Tai, K. et al. (2001) Analysis of a ten-nanosecond molecular dynamics simulation of mouse acetylcholinesterase. Biophys. J. 81, 715–724

[24] Marrink, S. J. et al. (2001) Simulation of the spontaneous aggregation of phospholipids into bilayers. J. Am. Chem. Soc. 123, 8638–8639

[25] Berneche, S. and Roux, B. (2001) Energetics of ion conduction through the K^+ channel. Nature 414, 73–77