Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Pharmacology

From protein folding to protein function and biomolecular binding by energy landscape theory

Alexander Schug[1,2] and José N Onuchic[2]

Protein folding and function are inherently linked sharing a joined funneled energy landscape. In this theoretical framework, the integration of simulations, structural information, and sequence data has led to quantitatively explore, understand, and predict biomolecular binding and recognition, key processes in pharmacology, as a natural extension of the selective self-binding found in protein folding. Computer simulations based on these principles have made valuable contributions to understanding protein and RNA folding, protein–protein interactions, and protein-metabolite/RNA-metabolite interactions.

Addresses
[1] Department of Chemistry, Umeå University, Umeå, Sweden
[2] Center for Theoretical Biological Physics, University of California San Diego, San Diego, USA

Corresponding author: Onuchic, José N (jonuchic@ucsd.edu)

Current Opinion in Pharmacology 2010, **10**:709–714

This review comes from a themed issue on
New technologies – the importance of protein dynamics
Edited by Dennis R Livesay

Available online 14th October 2010

1471-4892/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.coph.2010.09.012

Introduction

During folding a protein performs a biased exploration of its free-energy landscape until it reaches a thermodynamically stable conformation — the native state. Each of the many conformational transitions during this exploration is by itself a complex event, as many intraprotein and protein–solvent interactions need to be formed, broken, and eventually reformed. A biased energy landscape is necessary since an energy landscape of random sequences of amino acids would be too large and rugged to be searched by a simple random walk [1,2,3•].

Nature has found a remarkable simple solution to deal with this challenge. To allow folding in the regime of *ms* to *s*, that is, the times found in biomolecular structure formation, evolution has *funneled the energy landscape* to facilitate efficient folding into the native state [2,3•,4]: unlike in chemical reactions with well-defined discrete intermediate states, proteins fold through an ensemble of converging pathways which taken together define the transition state ensemble. In long evolutionary processes, energetic frustration was removed from a protein's energy landscape (*principle of minimal frustration*) [1,5], smoothening it sufficiently to prevent entrapment in local minima as such local roughness would deter efficient folding and intervene with function.

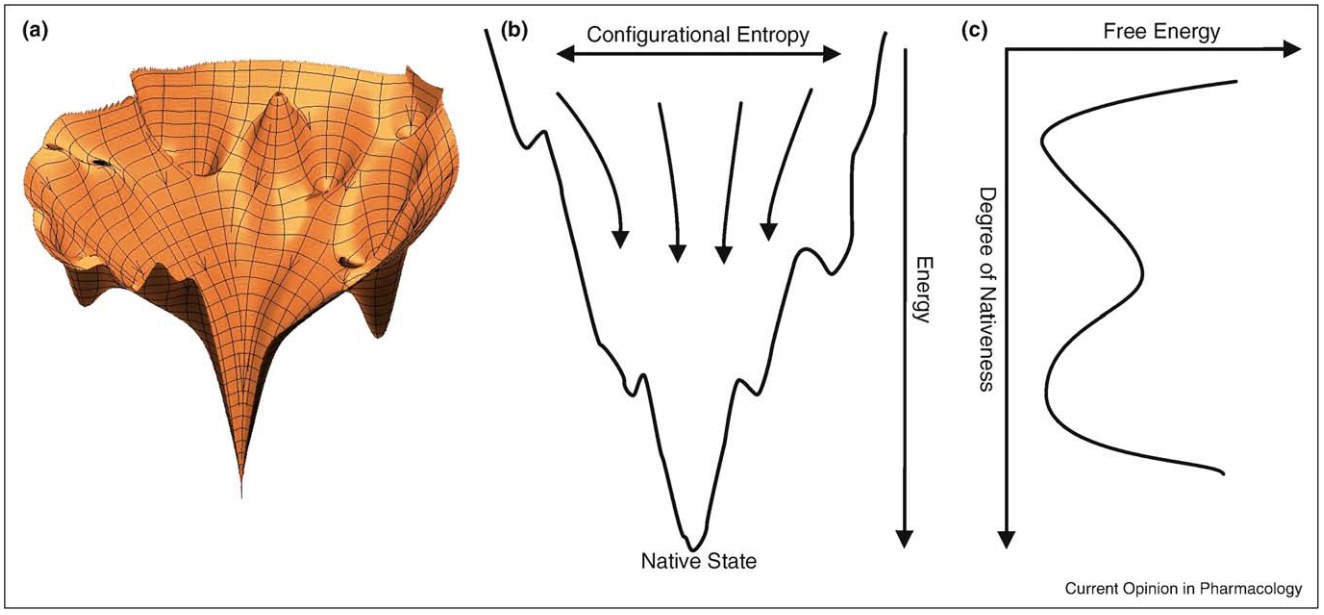
These concepts discovered in protein folding can be directly carried over to binding of two biomolecules, be they proteins, DNA/RNA or small metabolites [6,7••]. Additionally, one finds a competition between affinity and specificity. In the crowded cellular environment such biomolecular interactions need to be stable/affine enough to provide a degree of robustness toward local environmental distortions and changes like those induced by other nearby biomolecules or slight changes in environmental conditions. At the same time, interactions need to make sure that they carry cellular information between the correct partners, that is, they need to be specific. In pharmacology one typically wants to suppress specific interactions by presenting a new binding partner, the drug. The same principles should apply here. If we properly target the right interactions, designed compounds should be at least competitive or even fully block the natural interaction.

Protein/RNA folding and simulations based on energy landscape theory

Evolution smoothened the shape of the energy landscape by ensuring a dominance of interactions present in the native state during the entire folding process. This guiding bias prevents entrapment in local minima representing nonnative folds. It also provides a degree of robustness, permitting protein folding and function despite moderate environmental changes or mutations. Energy landscape theory stipulates that a fully unfrustrated protein, that is, an ideal folder, could be described by only taking interactions present in the native state into account (see Figure 1) [1,3•,4].

It can be shown analytically, that a good folder has a large gap between the two defining temperatures T_F (the folding temperature) and T_G (the glass-transition temperature). T_F is the temperature at which folded and unfolded conformations are equally likely to be adopted and consequently the protein is unfolded for $T > T_F$ and folded for $T < T_F$. Similarly, T_G is the temperature above which a protein efficiently folds into its native state and below which a protein yields to frustrated interactions and becomes entrapped in nonnative minima.

Figure 1



Energy landscape of protein folding. Three-dimensional **(a)** and two-dimensional **(b)** cartoons of a funneled energy landscape for protein folding illustrate the principle of minimal frustration. The ensemble of unfolded configurations occupies the top of the energy funnel. As the protein becomes more native-like and folded its accessible configurational entropy, that is the width of the funnel, decreases which is offset by the gain of energy from increasingly formed native interactions. There is not, however, a unique and single pathway for folding but the whole process functions via an ensemble of converging pathways. Evolution smoothened the surface of the energy landscape sufficiently to prevent entrapment in local minima and ensure the robustness of this molecular self-assembly in a crowded cellular environment. Some local roughness remains due to the limited 20 amino acid code and competing evolutionary constraints introduced by, for example, protein function which often requires binding specificity to other biomolecular partners. **(c)** A typical free-energy landscape for a two-state folding close to folding temperature. Two basins, one for the folded and another for the unfolded state, are separated by a barrier forming the transition state ensemble.

These principles can be molded into native structure-based models,[a] which possess a simplistic Hamiltonian and are based entirely on the structure of the protein. Apart from the typical terms found in molecular dynamics force fields like harmonic terms for bonds, angles, and dihedrals, which have their minima at the value found in the native conformation, an additional attractive contact term is included. This term runs over all pairs of amino acids that interact in the native conformations. This interaction matrix is often called a contact map. In a typical mathematical description each amino acid is represented as a single C_α-bead with van-der-Waals type contact interactions [8,9], although more recent work incorporated Gaussians as contact potentials [10], used $C_\alpha C_\beta$ [11] or all-atoms [12] description. There seems to be good agreement between these coarse-grained and more detailed models [13]. Overall, simulations based on the native structure-based models have shown to be in good agreement with experimental measurements like folding rates [14–16] or phi-values [17] characterizing the transitions state ensemble. Recently, structure-based Hamiltonians have been developed to also simulate RNA folding [18–20] and explore the intricate folding of complex topologies found in knotted proteins [21].

Linking the energy landscape of folding and function

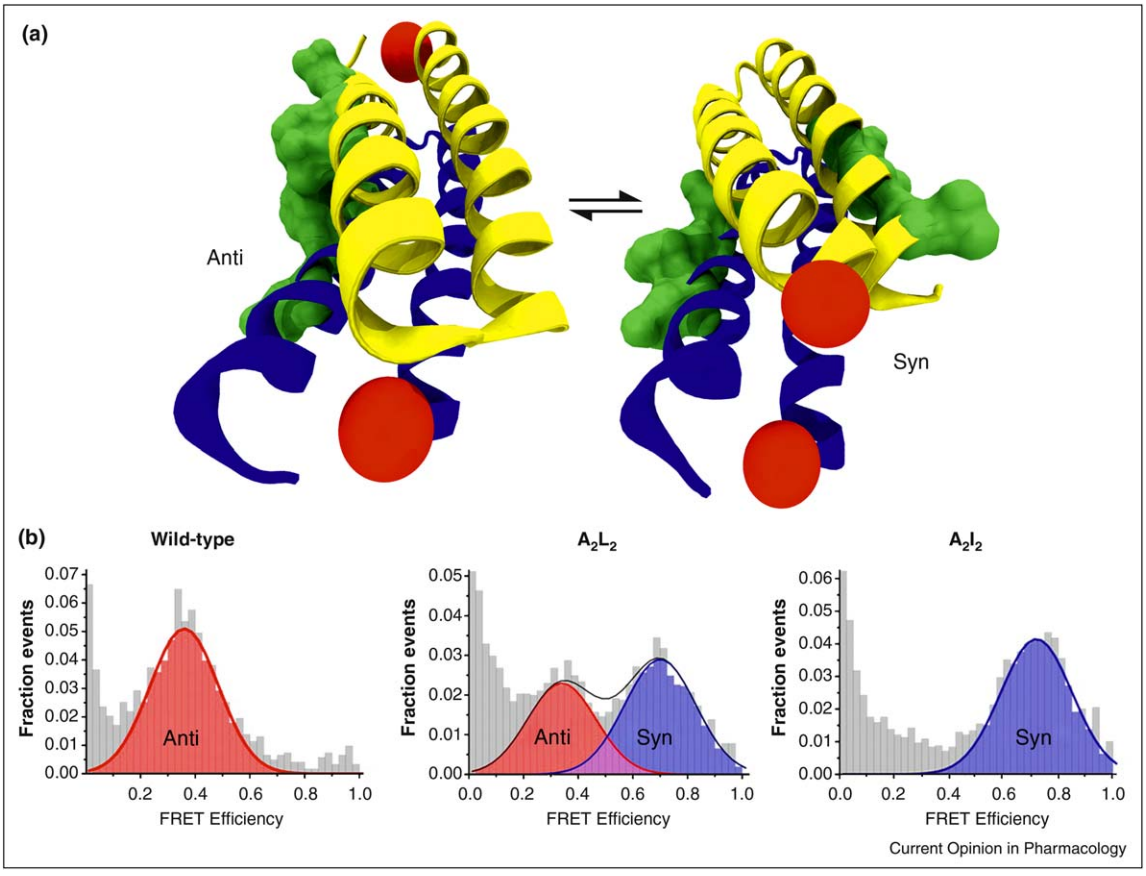
To connect protein folding and function, significant effort has been put into accommodating multiple conformations representing conformational substates on the energy landscape which are associated with biomolecular function [22,23•,24–27,28•]. Striking examples investigated within the framework of energy landscape theory include motor proteins like kinesin and their 8 nm stepping motions — so-called power-strokes — along microtubules [29], the conformational motions of the LID and NMP domains in adenylate kinase [23•], or the competition of two native states for the ROP[b]-dimer (see Figure 2) [22,30].

Technically, the main challenge is the treatment of the dissimilar contact maps found in alternate conformations. Without loss of generality, we will consider only two possible protein conformations for our discussion. In this case, native contacts can be divided into three different

[a] These models are often also referred to as Go-models.

[b] Repressor of primer. ROP is also called ROM (RNA-I modulator).

Figure 2



Configurational competition in protein folding and function. The ROP-homodimer (repressor of primer) possesses an amino acid sequence on the verge of configurational degeneration, making it an ideal test-case to investigate conformational transitions associated with protein function. **(a)** The two wild-type ROP-monomers (left) arrange in an *anti* way forming an RNA-binding interface (green). The sequentially similar mutant A_2I_2-6 (right) does not bind RNA as the RNA-binding interface is disrupted by the *syn*-arrangement of the two monomers. Simulations and considerations based on energy landscape theory predicted that the mutant A_2L_2-6 can, pending on environmental conditions, occupy both anti and syn [22,30]. **(b)** This prediction was validated in single-molecule FRET experiments which are sensitive to the distance of dyes placed on the termini of the ROP-monomers (red) [54•]. Under slight denaturing conditions of 0.6 M GdnHCl the WT and A_2I_2-6 only have single peaks corresponding to spatially far and close dyes, respectively. The mutant A_2L_2-6, however, has a double peak. This mutant can occupy both syn and anti.

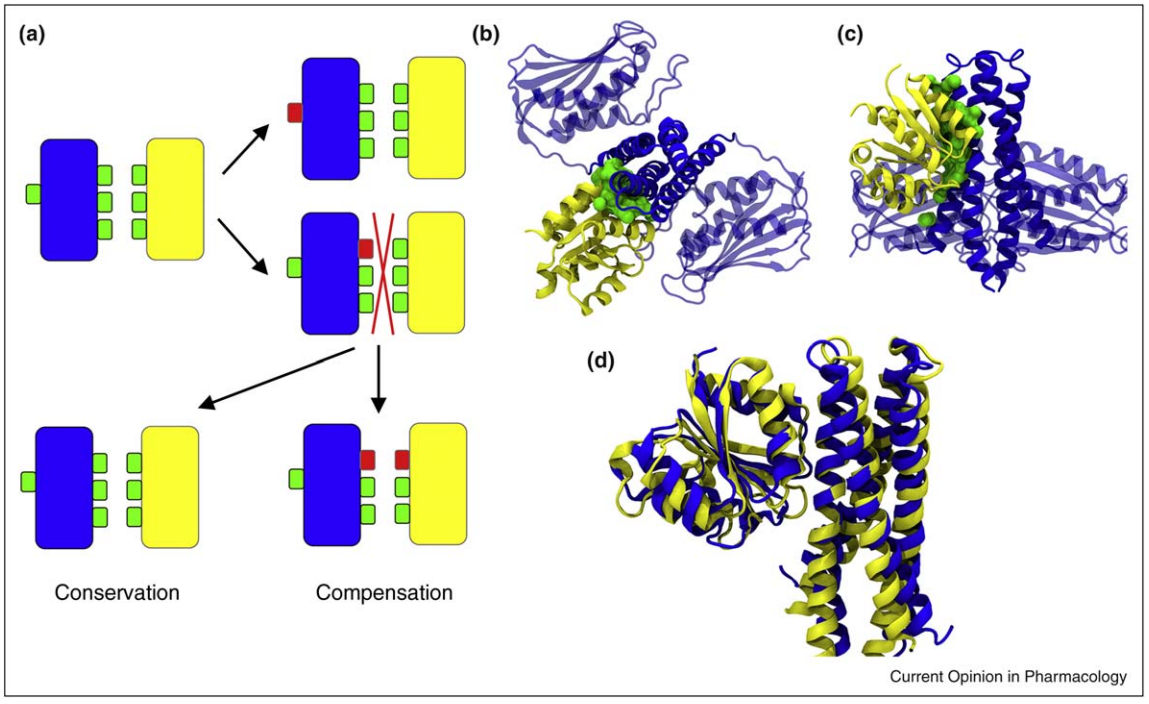
groups [10], pending whether they are realized in only one conformation, shared by two conformations with the same distance, or found in both conformations but with dissimilar distances. Other than the first two groups, the 'degenerated' contacts in the third group cannot be treated by single-welled van-der-Waals contact potentials. Solutions for this challenge include simply disregarding such contacts, thermodynamic weighting [24], or multiwelled Gaussians [10]. A remarkable feature of all these treatments is the relative robustness of the results independent of technical details of implementation.

Biomolecular binding

Biomolecular binding is used to regulate and modify protein function with direct application in ligand screening or drug design. Significant progress is made by ever more powerful experimental tools [31] like high-throughput structure-determination X-ray [32,33] and NMR [34] or indigenous enhanced single-molecule techniques like multicolor smFRET [35]. This increasingly detailed insight from the experimental side into the coupling of protein folding and binding is met by a solid theoretical understanding of the underlying energy landscapes [36,37,38•,39,40].

Two crucial factors dominate the process of binding. The first is the affinity, that is, the thermodynamic stability of molecular association. The second is the specificity of interactions in a crowded cellular environment. In particular high affinity must not satisfy specificity. An *in silico* treatment, for example rational drug design, is an ongoing challenge for more than a decade [41]. Minimal models and energy landscape theory can be used to describe biomolecular binding [6,42] or identify kinetic paths [7••]. In analogy to protein folding theory [3•], one can identify native binding and unbinding phases, and glassy trapping phases [6].

Figure 3



Biomolecular binding by a joined theory approach. **(a)** Two coevolving interacting proteins (blue and yellow) have well-defined interaction residues (green). Mutations (red) of residues far from the surface have little impact on interactions, while a mutation at the surface will negatively impact interactions. Therefore the interaction surface will either be conserved or a second compensating mutation is necessary. Given sufficient sequential information, statistical analysis like direct coupling analysis [51•] can search for such patterns of coevolving amino acid pairs. The two-component system TM0853/TM0468 consists of a histidine kinase (yellow) and its partner response regulator (blue, RR) (**(b)** top and **(c)** side view) is one example for such coevolving proteins. Transient binding (interface in green) facilitates the transfer of a phosphoryl group between the two proteins. Despite many similar copies in the same organism, both proteins achieve high specificity, a requirement in cellular signal transduction. A joined theory approach of direct coupling analysis of ~1000 sequential homologues and simulations based on energy landscape theory could predict the coupled complex [53••] in high agreement (3.5 Å RMSD, **(d)** blue theoretical prediction) with a concurrently published crystal structure (yellow) of the complex [55].

In molecular simulations, biomolecular binding has been shown to modify the folding process. A striking example is the structure formation upon target binding found in natively unstructured transcription factors, as found for the pKID domain of the transcription factor CREB to the KIX domain of CBP. Though the simulations observed both on-pathway and off-pathway intermediates, the binding mechanism was largely dominated by specific native-like interactions even nonspecific interactions modify the rate on binding [43]. Similarly, the SAM-1 riboswitch expresses different folding behavior pending on the presence or absence of its specific binding partner, the SAM[c]-molecule which is binding to an internal binding pocket [18]. As riboswitches are RNA-based genetic control elements [44], insight into the interplay of binding and expression platform promises new antibiotic targets [45].

Inclusion of genomic information

A protein's biological function is often dominated by transient interactions with other proteins with the resulting protein–protein interfaces becoming important targets for drug design [46]. Experimental techniques like NMR or X-ray crystallography are tremendously successful in providing structural information but face problems when resolving transiently bound protein complexes. Structure prediction methods cannot readily close this gap, as database driven methods like homology modeling [47] suffer from the lack of templates while physics-based [48,49] approaches still struggle with the accuracy of their force fields [50] and computationally prohibitive costs.

One can, however, integrate complementary computational/theoretical techniques to simulate molecular docking. The recent growth of genomic data allows meaningful statistical analysis of sequential homologues. As shown for two-component signal transduction systems, the statistical analysis of roughly 1000 sequences of coevolving proteins provides sufficient information to define an intermolecular protein–protein surface [51•] for molecular docking by native structure-based simulations (see Figure 3) [52,53••]. This inclusion of genomic information into molecular simulation might prove useful to deal with insufficient structural information, which is still crucial for meaningful biomolecular simulations.

[c] S-adenosylmethionin.

Summary

Energy landscape theory has been vastly successful in explaining the mechanism and different scenarios governing protein folding as well as how global motions control protein function. The *principle of minimal frustration* combined with the concept of a *funneled energy landscape* has allowed us to only use the information from the native structure to predict the mechanism of protein folding, binding and in many cases of function. These theoretical results have been used to understand and design new experiments and have also made several successful prediction later confirmed in the laboratory. New structural data that include multiple protein structures during the functional activity have even improved the power of these methods. Similar approaches have gone beyond proteins to also include nucleic acids such as RNA. Recent advancements have shown how in cases where structural information is limited, additional information coming from lower resolution structural methods or genomics can be used for similar studies. For example, studies using only structural information from the individual proteins combined with genomic information have successfully predicted transient protein–protein complexes.

Acknowledgements

This work was supported by the Center for Theoretical Biological Physics (CTBP) sponsored by the NSF (Grant PHY-0822283) with additional support from NSF grant MCB-0543906 (J.N.O).

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG: **Funnels, pathways, and the energy landscape of protein-folding — a synthesis**. *Proteins: Struct Funct Genet* 1995, **21**:167-195.

2. Leopld PE, Montal M, Onuchic JN: **Protein folding funnels — a kinetic approach to the sequence structure relationship**. *Proc Natl Acad Sci U S A* 1992, **89**:8721-8725.

3. • Onuchic JN, Wolynes PG: **Theory of protein folding**. *Curr Opin Struct Biol* 2004, **14**:70-75.
A concise introduction into energy landscape theory.

4. Frauenfelder H, Sligar SG, Wolynes PG: **The energy landscapes and motions of proteins**. *Science* 1991, **254**:1598-1603.

5. Bryngelson JD, Wolynes PG: **Spin-glasses and the statistical mechanics of protein folding**. *Proc Natl Acad Sci U S A* 1987, **84**:7254-7528.

6. Wang J, Verkhivker GM: **Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding**. *Phys Rev Lett* 2003, **90**:188101.

7. •• Wang J, Zhang K, Lu H, Wang E: **Dominant kinetic paths on biomolecular binding-folding energy landscape**. *Phys Rev Lett* 2006, **96**:168101.
The dynamics of biomolecular binding are explored in the framework of energy landscape theory.

8. Clementi C, Garcia AE, Onuchic JN: **Interplay among tertiary contacts, secondary structure formation and side-chain packing in the protein folding mechanism: all-atom representation study of protein L**. *J Mol Biol* 2003, **326**:933-954.

9. Clementi C, Jennings PA, Onuchic JN: **Prediction of folding mechanism for circular-permuted proteins**. *J Mol Biol* 2001, **311**:879-890.

10. Lammert H, Schug A, Onuchic JN: **Robustness and generalization of structure-based models for protein folding and function**. *Proteins: Struct Funct Bioinform* 2009, **77**:881-891.

11. Oliveira LC, Schug A, Onuchic JN: **Geometrical features of the protein folding mechanism are a robust property of the energy landscape: a detailed investigation of several reduced models**. *J Phys Chem B* 2008, **112**:6131-6136.

12. Whitford PC, Noel JK, Gosavi S, Schug A, Sanbonmatsu KY, Onuchic JN: **An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields**. *Proteins: Struct Funct Bioinform* 2009, **75**:430-441.

13. Hori N, Chikenji B, Berry NS, Takada S: **Folding energy landscape and network dynamics of small globular proteins**. *Proc Natl Acad Sci U S A* 2009, **106**:73-78.

14. Chavez LL, Onuchic JN, Clementi C: **Quantifying the roughness on the free energy landscape: entropic bottlenecks and protein folding rates**. *J Am Chem Soc* 2004, **126**:8426-8432.

15. Cheung MS, Finke JM, Callahan B, Onuchic JN: **Exploring the interplay between topology and secondary structural formation in the protein folding problem**. *J Phys Chem B* 2003, **107**:11193-11200.

16. Clementi C, Nymeyer H, Onuchic JN: **Topological and energetic factors: what determines the structural details of the transition state ensemble and 'en-route' intermediates for protein folding? An investigation for small globular proteins**. *J Mol Biol* 2000, **298**:937-953.

17. Fersht AR: **Characterizing transition-states in protein-folding — an essential step in the puzzle**. *Curr Opin Struct Biol* 1995, **5**:79-84.

18. Whitford PC, Schug A, Saunders J, Hennelly SP, Onuchic JN, Sanbonmatsu KY: **Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function**. *Biophys J* 2009, **96**:L7-9.

19. Hyeon C, Dima RI, Thirumalai D: **Size, shape, and flexibility of RNA structures**. *J Chem Phys* 2006, **125**:194905.

20. Thirumalai D, Hyeon C: **RNA and protein folding: common themes and variations**. *Biochemistry* 2005, **44**:4957-4970.

21. Bolinger D, Sulkowska JI, Hsu HP, Mirny LA, Kardar M, Onuchic JN, Virnau P: **A Stevedore's protein knot**. *PLOS Comput Biol* 2010, **6**:e1000731.

22. Schug A, Whitford PC, Levy Y, Onuchic JN: **Mutations as trapdoors to two competing native conformations of the Rop-dimer**. *Proc Natl Acad Sci U S A* 2007, **104**:17674-17679.

23. • Whitford PC, Miyashita O, Levy Y, Onuchic JN: **Conformational transitions of adenylate kinase: switching by cracking**. *J Mol Biol* 2007, **366**:1661-1671.
Adenylate kinase's open and closed conformations are taken as basis for structure-based simulations to explore the dynamic interplay of folding and function.

24. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG: **Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations**. *Proc Natl Acad Sci U S A* 2006, **103**:11844-11849.

25. Grant BJ, Gorfe AA, McCammon JA: **Large conformational changes in proteins: signaling and other functions**. *Curr Opin Struct Biol* 2010, **20**:142-147.

26. Grant BJ, Gorfe AA, McCammon JA: **Ras conformational switching: simulating nucleotide-dependant conformational transitions with accelerated molecular dynamics**. *PLOS Comput Biol* 2009, **5**:1000325.

27. Maisuradze GG, Liwo A, Scheraga HA: **How adequate are one- and two-dimensional free energy landscapes for protein folding dynamics?** *Phys Rev Lett* 2009, **103**:238102.

28. • Pisliakov A, Cao J, Kamerlin S, Warshel A: **Enzyme millisecond conformational dynamics do not catalyze the chemical step**. *Proc Natl Acad Sci U S A* 2009, **106**:17359-17364.
Multiscale simulations investigate the coupling between conformational dynamics and the chemical step for adenylate kinase.

29. Hyeon C, Onuchic JN: **Mechanical control of the directional stepping dynamics of the kinesin motor**. *Proc Natl Acad Sci U S A* 2007, **104**:17382-17387.

30. Levy Y, Cho SS, Shen T, Onuchic JN, Wolynes PG: **Symmetry and frustration in protein energy landscapes: a near degeneracy resolves the Rop dimer-folding mystery**. *Proc Natl Acad Sci U S A* 2005, **102**:2373-2378.

31. Renaud J, Delsuc M: **Biophysical techniques for ligand screening and drug design**. *Curr Opin Pharmacol* 2010, **9**:622-628.

32. Brundell T, Patel S: **High-throughput X-ray crystallography for drug discovery**. *Curr Opin Pharmacol* 2004, **4**:490-496.

33. Davis A, St-Gallay S, Kleywegt G: **Limitations and lesons in the use of X-ray structural information in drug design**. *Drug Discov Today* 2008, **13**:831-841.

34. Pellecchia M, Bertini I, Cowburn D, Dalvit C, Giralt E, Jahnke W, James T, Homans S, Kesler H, Luchinat C: **Perspectives in NMR drug discovery: a technique comes out of age**. *Nat Rev Drug Discov* 2008, **7**:738-745.

35. Gambin Y, Deniz AA: **Multicolor single-molecule FRET to explore protein folding and binding**. *Mol Biosyst* 2010, **6**:1540-1547.

36. Kim JN, Keyes T, Straub JE: **Relationship between protein folding thermodynamics and the energy landscape**. *Phys Rev E* 2009, **79**:030902R.

37. Wales DJ: **Energy landscapes: some new horizons**. *Curr Opin Struct Biol* 2010, **20**:3-10.

38. • Zhuravlev P, Papoian G: **Functional versus folding landscapes: the same yet different**. *Curr Opin Struct Biol* 2010, **20**:16-20.
A well-written review exploring the interplay of folding and function energy landscapes.

39. 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.

40. Prada-Garcia D, Gómez-Gardenes J, Echenique P, Fernando F: **Exploring the free energy landscape from dynamics to landscapes and back**. *PLOS Comput Biol* 2010, **18**:154-162.

41. McCammon JA: **Theory of biomolecular recognition**. *Curr Opin Struct Biol* 1998, **8**:245-249.

42. Levy Y, Onuchic JN: **Mechanisms of protein assembly: lessons from minimalist models**. *Acc Chem Res* 2006, **39**:135-142.

43. Turjanski AG, Gutkind JS, Best RB, Hummer G: **Binding-induced folding of a natively unstructured transcription factor**. *PLOS Comput Biol* 2008, **4**:e1000060.

44. Breaker RR: **Complex riboswitches**. *Science* 2008, **319**: 1795-1797.

45. Blount KF, Breaker RR: **Riboswitches as antibacterial drug targets**. *Nat Biotechnol* 2006, **24**:1558-1564.

46. Wells JA, McClendon CL: **Reaching for high-hanging fruit in drug discovery at protein–protein interfaces**. *Nature* 2007, **450**:1001-1009.

47. Eswar N, Eramian D, Webb B, Shen M-Y, Sali A: **Protein structure modeling with MODELLER**. *Methods in Molecular Biology*. Humana Press Inc.; 2008:. pp. 145–159.

48. Schug A, Herges T, Wenzel W: **Reproducible protein folding with the stochastic tunneling method**. *Phys Rev Lett* 2003, **91**:158102.

49. Fujitsuka Y, Takada S, Luthey-Schulten ZA, Wolynes PG: **Optimizing physical energy functions for protein folding**. *Proteins* 2004, **54**:88-103.

50. Best RB, Buchete NV, Hummer G: **Are current molecular dynamics force fields too helical?** *Biophys J* 2008, **95**:07-09.

51. • Weigt M, White RA, Szurmant H, Hoch JA, Hwa T: **Identification of direct residue contacts in protein-protein interaction by message passing**. *Proc Natl Acad Sci U S A* 2009, **106**:67-72.
A statistical analysis of sequential homologues identifies part of a transient binding interface between two proteins.

52. Schug A, Weigt M, Hoch JA, Onuchic JN, Hwa T, Szurmant H: **Computational modeling of phosphotransfer complexes in two-component signaling**. *Methods Enzymol* 2010, **471**:43-58.

53. •• Schug A, Weigt M, Onuchic JN, Hwa T, Szurmant H: **High resolution complexes from integrating genomic information with molecular simulation**. *Proc Natl Acad Sci U S A* 2009, **106**:22124-22129.
A mixed theory approach of molecular simulation and statistical analysis is able to predict a transient two-component signal-transduction protein complex.

54. • Gambin Y, Schug A, Lemke EA, Lavinder JJ, Ferreon AC, Magliery TJ, Onuchic JN, Deniz AA: **Direct single-molecule observation of a protein living in two opposed native structures**. *Proc Natl Acad Sci U S A* 2009, **106**:10153-10158.
Powerful multi-color single molecule FRET measurements validate the prediction from [22] that a specific ROP mutant alternates between two conformations for the same homodimer.

55. Casino P, Rubio V, Marina A: **Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction**. *Cell* 2009, **139**:325-336.