

Protein Folding, Binding and Energy Landscape: A Synthesis

Shu-Qun Liu^{1,2} Xing-Lai Ji^{1,2}, Yan Tao¹, De-Yong Tan³,
Ke-Qin Zhang¹ and Yun-Xin Fu^{1,4}

¹Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for
Southwest Biodiversity, Yunnan University, Kunming,

²Sino-Dutch Biomedical and Information Engineering School,
Northeastern University, Shenyang,

³School of Life Sciences, Yunnan University, Kunming,

⁴Human Genetics Center, School of Public Health,
The University of Texas Health Science Center, Houston, Texas,
^{1,2,3}P. R. China
⁴USA

1. Introduction

Protein folding and molecular recognition and binding provide the basis for life on earth. The native 3D structure of a protein is necessary for its biological function; and the protein-associated molecular recognition and binding is the fundamental principle of all biological processes, thereby unraveling the mechanisms of protein folding and binding is fundamental to describing life at molecular level (Perozzo et al., 2004).

Of particular interest is that protein folding and binding are similar processes with the only difference between them being the presence and absence of the chain connectivity (Ma et al., 1999). Among many models such as diffusion-collision (Karplus & Weaver, 1994), nucleation-condensation (Itzhaki et al., 1995), jigsaw puzzle (Harrison & Durbin, 1985), hydrophobic collapse (Agashe et al., 1995) and stoichiometry models (Mittal et al., 2010) proposed to describe the mechanism of protein folding or, to some extent, protein binding, the "folding funnel" model (Leopold et al., 1992) based on the free energy landscape theory (Bryngelson et al., 1995; Wales, 2003) has now been most widely accepted. In this model, protein folding can be regarded as going down the funnel-like free energy landscape through multiple parallel pathways towards the bottom of the funnel (Noe et al., 2009; Yon, 2002); and protein binding occurring along the rugged free energy surface around the funnel bottom can be viewed as microfunnel fusion (Ma et al., 1999; Tsai et al., 2001) or downward extension of free energy well, which provides the thermodynamic interpretation for the binding models such as lock-and-key (Fischer, 1894), induced fit (Koshland, 1958) and conformational selection (Foote & Milstein, 1994; Monod et al., 1965). Thus, the energy landscape provides a basis to synthesize and explain the mechanisms of protein folding, binding and dynamics. Under the energy landscape theory, the protein folding and binding

are essentially thermodynamically controlled processes involving various types of driving forces, including the enthalpic contribution of noncovalent bond formations, entropic effects such as uptake and release of solvent molecules, burial of apolar protein surface area (hydrophobic effect), restrictions of degrees of freedom of protein and ligand, and loss of the rotational and translational freedom of the interacting partners (Chaires, 2008; Perozzo et al., 2004). Briefly, these two processes, which are driven by a decrease in total Gibbs free energy, are dictated by the mechanism of a delicate balance of the opposing effects involving enthalpic and entropic contributions. It should be noted that it is the thermodynamically driven subtle enthalpy-entropy compensation that leads to the global free energy minimum of the protein-solvent system, but that the specific inter-atomic interactions observed in the folded 3D protein structure are to a large extent the consequence of thermodynamic equilibrium and can not fully define the driving forces for the folding process (Ji & Liu, 2011a). Such a subtle enthalpy-entropy compensation also underlies the mechanisms of both the induced fit and conformational selection models of protein-ligand binding since most of the binding processes occur through both mechanisms: the conformational selection, which is under entropy control, generally plays a role in the initial contact between the protein and ligand and determines the binding specificity; the induced fit, which is mainly driven by the negative enthalpy of bond formation, plays a role in the subsequent conformational adjustment and determines the binding affinity.

Interestingly, we speculate that many other processes can be explained by thermodynamic enthalpy-entropy compensation (Ji & Liu, 2011b), i.e., the Yin and Yang balance in traditional Chinese medicine theory could correspond to the enthalpy and entropy compensation of the second law of thermodynamics; global warming can be considered as the consequence of excessive production of positive entropy (carbon dioxide) from chemically ordered fossil fuel (enthalpy storage), urging people to slow resource consumption to delay the inevitable death by entropy; the cancer occurrence can be considered as the consequence of accumulation of nucleotide mutations (which could be seen as increase in entropy) caused by gradual loss of DNA repair ability, leading to disordered gene expressions and uncontrolled cell proliferation.

In this chapter the problems of protein folding are first introduced and discussed. In order to address these problems, several protein folding models or hypotheses are discussed and compared. Subsequently, the free energy landscape theory, including the concept of the energy landscape, the funnel-like folding landscape and the detailed folding process occurring in such a folding funnel, the roughness/ruggedness of the energy landscape, the amplitude and timescale of the protein dynamics and their associated functions, and the dynamic feature of the energy landscape, are introduced and discussed. The protein-ligand binding models such as the lock-and-key, induced fit and conformational selection are introduced and the driving forces and thermodynamic mechanism underlying these models are discussed in depth based on the free energy landscape theory. Finally, the mechanisms of protein folding and binding are synthesized to elicit a general funnel model of the energy landscape, under which the folding and binding are commonly driven by the decrease in total Gibbs free energy through the entropy-enthalpy compensation, with different steps being dominated by either the entropic or enthalpic contribution. We highlight that it is the free energy landscape theory that synthesizes the protein folding and binding mechanisms together to bring about the common funnel model. In addition, the energy landscape-

associated concepts such as the thermodynamics, kinetics, free energy, enthalpy, entropy, and frustration/ruggedness of the landscape are also introduced. This chapter will facilitate not only the protein engineering studies aimed at modifying protein structure and function but also the rational drug design and understanding of life in the post-genomic era.

2. The problems of protein folding

Protein folding is the process of transformation of one-dimensional linear information encoded in the amino acid sequence into a functional 3D structure. This is essentially a physical chemistry process for which the mechanism remains elusive due to the following three incompletely resolved problems (Dill et al., 2007): i) the thermodynamic question of how a native structure results from the inter-atomic forces acting on an amino acid sequence — the folding code; ii) the kinetic problem of how a native structure can fold so fast — the folding rate; iii) the computational problem of how to predict the native structure of a protein from its amino acid sequence — the protein structure prediction.

2.1 The folding code and thermodynamics

The apparent complexity of folded structures and the abundant diversity of conformational states of the denatured proteins make it difficult to understand and differentiate various types of forces that drive protein folding. A popular opinion, which was first proposed by (Anfinsen & Scheraga, 1975), considers that the protein folding code is the sum of many different small interactions such as electrostatic, van der Waals and hydrogen bonding interactions, which are embodied in the static secondary structures and are mainly distributed locally in the sequence. However, from a statistical mechanics point of view, the folding code can be viewed as the thermodynamic profile of interaction process that represents various types of forces that drive folding, including the changes in folding free energy (ΔG), enthalpy (ΔH) and entropy (ΔS). The predominant force that drives folding is considered the hydrophobic interaction or desolvation, which is essentially the entropic effect and is distributed both locally and non-locally in the sequence. The native secondary structures and the finally observed interactions (such as electrostatic, van der Waals and hydrogen bonding interactions) are the consequence rather than the cause of folding forces (Dill, 1999; Dill et al., 2007), which drive the search for the global free energy minimum of the protein-solvent system through entropy-enthalpy compensation (Ji & Liu, 2011a).

2.2 The folding rate and kinetics

In 1961, Anfinsen's classic experiments on renaturation of denatured ribonuclease showed that small globular proteins can fold spontaneously into their native 3D structures in the absence of any catalytic biomolecules, thus leading to an assumption that the folded proteins exist in the global-minimum free energy state (Anfinsen et al., 1961). Soon afterward, Levinthal (Levinthal, 1968, 1969) recognized that proteins have far too vast conformational spaces to permit a thorough search for the folded native structure in a biologically relevant time. To resolve this "paradox", Levinthal proposed the notion of protein folding pathway through which proteins can search and converge quickly to the native states. Using various experimental techniques including fast laser temperature-jump methods (Yang & Gruebele, 2003), protein engineering technique such as site-directed

mutagenesis that identify those amino acids that control fold speed (Matouschek et al., 1989; Sosnick et al., 2004), fluorescence resonance energy transfer (FRET) methods that can watch the formation of particular contacts (Magg et al., 2006; Schuler et al., 2002), and hydrogen exchange methods that can see the structural folding events (Maity et al., 2005), the timescales of folding events occurring on several model proteins (such as cytochrome *c*, chymotrypsin inhibitor 2, barnase, src, fyn SH3 domains, Trp-cage, etc.) were extensively studied. The results show that the protein folding rates correlate mainly with the topology of the native proteins, exhibiting variable timescales with a range of more than eight orders of magnitude (Dill et al., 2007; Plaxco et al., 1998). The proteins that fold relatively fast usually have more local structural elements such as helices and tight turns, whereas the proteins that fold relatively slowly have more non-local structural elements such as β sheets. However, is there a single pathway or multiple parallel routes by which a protein folds up? How does one characterize the protein folding pathways connecting the denatured and native states? These are two aspects of the kinetic problem of protein folding. Full answers to these questions require the introduction of concept of the "protein folding funnel" (Dill, 1985; Dill & Chan, 1997; Dobson, 2000; Leopold et al., 1992) based on the free energy landscape theory (Bryngelson et al., 1995; Wales, 2003). Briefly, protein folding can be viewed as going down the funnel-like energy landscape via multiple parallel pathways from the vast majority of individual non-native conformations to the native states around the bottom of the funnel. At any stage the protein exists as an ensemble of conformations and can be trapped transiently in many local energy minimum wells.

It must keep in mind that the folding time and pathways are concepts related to the kinetics, which defines the energy barriers between different conformational states; while that the free energy corresponds to the conformational states in relation to thermodynamics, which defines the relative probabilities/populations/lifetimes of the different conformational substates (Henzler-Wildman & Kern, 2007). Both of these concepts belong to the field of statistical mechanics and are of crucial importance in understanding the protein folding mechanism.

2.3 The protein structure prediction

The major motivations for protein structure prediction arise from i) the fundamental desire of human beings to explore natural laws and the needs to test or verify the putative protein folding mechanism; ii) the requirements to determine the functions of encoded protein sequences in the post-genomic era and to make drug discovery faster and more efficient through substitution fast cheap computer simulation for slow and expensive structural biology experiments (Dill et al., 2007). The protein structure prediction methods can be classified as physics-based and bioinformatics-based ones (Dill et al., 2007). The basis for developing the purely physics-based approaches is the thorough recognition and understanding of the protein folding mechanism. An ideal aim of such methods is that they are able to direct an amino acid sequence of a protein to fold into its native 3D structure through mimicking various possible driving forces without knowledge derived from databases. On the contrary, the bioinformatics-based methods rely mainly on various priori information derived from databases such as statistical energy functions, secondary structure and structure classification information of proteins, in addition to the information of the primary amino acid sequence.

2.3.1 Bioinformatics-based methods and CASP

Critical Assessment of Techniques for Protein Structure Prediction (CASP) is a community-wide, worldwide experiment for computer-based protein structure prediction taking place every two years since 1994 (Moult et al., 1995). The primary goal of CASP is to facilitate the development of the approaches for predicting protein 3D structure from its amino acid sequence. To date, there have been nine previous CASP experiments: from CASP1 in 1994 to CASP9 in 2010. Description of these experiments and the full data (such as targets, predictions, interactive tables with numerical evaluation results, dynamic graphs and prediction visualization tools) can be accessed by following the link at: <http://predictioncenter.org/index.cgi>.

Currently, the ranking results show that the most efficient and accurate bioinformatics-based method is the homology modeling (Sali & Blundell, 1993), which has not only the speed to compute approximate folds for large fractions of whole genomes (Pieper et al., 2006), but also sufficient accuracy for the predicted structures generally with errors better than 3 Å if the sequence identity between target and template is higher than 30% (Baker & Sali, 2001). For single-domain globular proteins having $< \sim 90$ amino acids in length, several web server tools can predict the structures often having C_α RMSD (root mean square deviation) values of about 2–6 Å with respect to their experimental structures (Baker, 2006; Bradley et al., 2005; Dill et al., 2007; Zhang et al., 2005). Very recently, a newly developed homology-based structure prediction method achieved 1–2 Å (mean 1.6 Å) C_α RMSD to the reference crystal structures from the full spectrum of test domains in recent CASP experiments (Soundararajan et al., 2010). This approach is mainly based on defining and extracting the fold-conserved PCAIN (protein core atomic interaction network) from distinct protein domain families, which is found to be significantly distinguished among different domain families and as such, can be considered as the “signature” of a domain’s native fold. It should be noted that the bioinformatics methods are mainly limited to structural prediction of the single domains with relatively small size. Predicting the structures of large multidomain or domain-swapped proteins (Dill et al., 2007) and membrane proteins (Bowie, 2005) is the major challenge in the near future.

2.3.2 Physics-based methods

Due to the objective existence of some limitations such as the incomplete understanding of the folding mechanism, inaccurate description of the force-fields and insufficient computational power, there have been only a few successes of physics-based methods in computing structures with very small size in the past decade. For example, in 1988, Duan and Kollman performed nearly one microsecond simulation on the unfolded 36-residue villin headpiece in explicit solvent, yielding a collapsed state with a 4.5 Å C_α RMSD relative to its NMR structure (Duan & Kollman, 1998). In 2003, the IBM blue gene team achieved the nearly native folded structure of the 20-residue Trp-cage peptide with C_α RMSD value of < 1 Å relative to its experimental structure using 92 ns of replica-exchange molecular dynamics in implicit solvent (Pitera & Swope, 2003). In 2010, Folding@home researchers Voelz et al. reported a computer simulation of *ab initio* protein folding on the millisecond timescale for the 39-residue protein NTL9(1–39) using distributed molecular dynamics simulations in

implicit solvent on GPU processors, and observed a small number of productive folding events at a temperature lower than the melting point of the force field (Voelz et al., 2010).

Despite these limited successes in only small simple proteins, we believe that, as was stated by (Dill et al., 2007), *“once physics-based approaches succeed, the advantages would be the ability for us to predict conformational changes, such as induced fit, a common and important unsolved problem in computational drug discovery; the ability to understand protein mechanisms, motions, folding processes, conformational transitions and other situations in which protein behavior requires more than just knowledge of the static native structure; the ability to design synthetic proteins for new applications or to design foldable polymers from nonbiological backbones; and the ability to systematically improve protein modeling based on the laws of physics”*. We also believe that the combination of physics-based with bioinformatics-based structural prediction methods will incubate more powerful approaches that can not only predict accurately and quickly the protein structures but also facilitate the solution of the two key problems relevant to protein folding: the folding code and the folding rate.

3. Protein folding models and hypotheses

Due to the complexity of the protein folding problems, several simplified models have been proposed to explore the conformational and sequence spaces, describe the sequence of events in the folding process, and further, to probe, examine and uncover the protein folding mechanism. These include diffusion collision model, framework model, nucleation-condensation model, Zipping-and-assembly (ZA), hydrophobic collapse model, folding funnel model as well as a stoichiometry-driven protein folding hypothesis.

3.1 Diffusion collision model

In order to discuss the dynamics of protein folding and introduce a possible mechanism by which a protein can make use of a more sophisticated procedure than a simple random search of all conformational possibilities (i.e., the Levinthal paradox), (Karplus & Weaver, 1976, 1994) proposed the diffusion collision model (Figure 1A), which is based on dividing the folding of a protein molecule into parts such that the information stored in the sequence of each part can be used independently. These divided parts, called microdomains, can be portions of incipient secondary structures (such as α -helices, β -strands and connecting turns) or hydrophobic clusters (i.e., hydrophobic amino acids close to each other in the amino acid sequence). In this model, the folding process could start with a polypeptide chain in an extended random coiled state: the microdomains are formed first; then they move diffusively with collisions taking place between them, leading to the coalescence into the multi-microdomain intermediates or higher aggregates; and finally, these multi-microdomain intermediates form the exact tertiary structure through mutual rearrangements. The formation of the microdomains or secondary structures (caused by local interactions) and their interactions rather than interactions between individual amino acids would greatly reduce folding times to reasonable values. However, such a simplified model can neither characterize detailed motions involved in the folding process at the atomic level nor provide complete information on energy barriers and populations of individual conformational states. Therefore, the diffusion collision model is only useful for a

semiquantitative analysis of protein folding dynamics through concentrating on the microdomains.

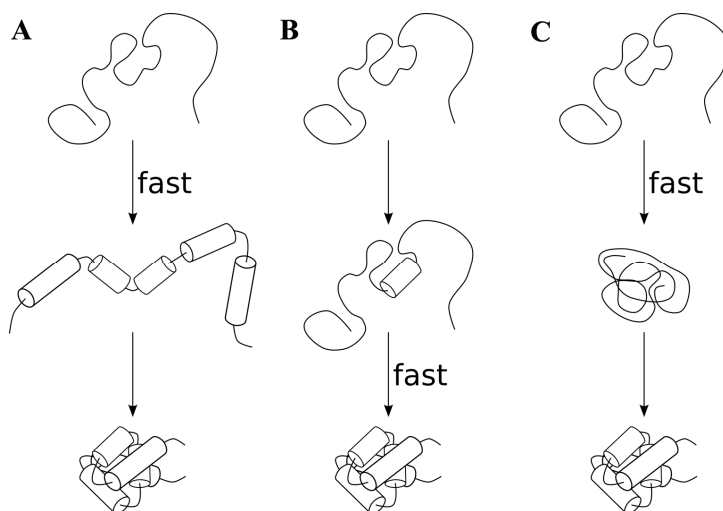


Fig. 1. Schematic representation of several protein folding models. (A) Diffusion collision model. (B) Nucleation condensation model. (C) Hydrophobic collapse model. The fast step of folding is indicated. For details, see the text.

In 1982, Kim and Baldwin (Kim & Baldwin, 1982) proposed the “framework” model for protein folding in which the native secondary structures are formed before the tertiary structures. The framework model is essentially similar to the diffusion collision model since both consider that folding is a hierarchical process in which simple local structures are formed first and this is followed by the formation of more complex structures through interactions or collisions among these simple structures.

3.2 Nucleation-condensation model

The nucleation-condensation mechanism (Figure 1B) suggests that the reaction is initiated by the formation of a marginally stable nucleus that acts as a template or seed from which the native structure would propagate through rapid condensation of further structure around it (Nolting & Agard, 2008). In this way, the number of conformations required to be sampled in the folding reaction is dramatically reduced. An essential feature of the nucleation-condensation model is that the nucleus occurrence requires the simultaneous formation of both secondary and tertiary interactions. This contrasts the diffusion collision model, which, as described above, involves a hierarchical process of structural assembly where the secondary structure elements, guided by local contacts, are initially formed independently of tertiary structure, followed by the packing of secondary structure elements to coalesce into the native tertiary structure. Therefore, the essential difference between these two folding models is whether the folding reaction can proceed in a concerted manner (such as in the nucleation-condensation model) or in a stepwise manner (such as in the diffusion collision and/or framework models).

3.3 Zipping-and-assembly model

ZA model (also termed hydrophobic zipper model) was proposed by (Dill et al., 1993) to explain how a protein finds its native state without a globally exhaustive search. In this model, protein folding reaction starts with the independent formations of local structural pieces along the chain, then those pieces either grow (zip) or coalesce (assemble) with other structures to form the native fold. At first glance, such a folding process is very similar to that of the collision-diffusion model since both describe the hierarchical process of structural assembly. However, these two models differ from each other in some aspects. In the ZA model there is a fast collapse that is driven by nonlocal hydrophobic interactions leading to the concurrent formations of the hydrophobic clusters, helices, and sheets, and the hydrophobic zipping process further result in a broad ensemble of compact intermediate states that have much secondary but little tertiary structure; the subsequent slow process involves breaking/unzipping incorrect (nonnative) hydrophobic-hydrophobic contacts to proceed towards the native structure. The ZA therefore highlights that the folding is dominated by nonlocal interactions, i.e., with collapse as the driving force, and the secondary structures are a consequence of the collapse, rather than its cause (Dill et al., 1995). On the contrary, the diffusion collision model assumes that the local interactions are important factors in reducing conformational search, implying that the major driving forces for protein folding is the hydrogen bonding, electrostatic and van der Waals interactions, leading to the early formations of the secondary structures and the subsequent collapse and assembly into the tertiary structure. The ZA is therefore similar, to some extent, to the nucleation-condensation since in both models the nonlocal interactions dominate the formation of the collapse or the nucleus. An advantage of the ZA model is that it explicitly points out the cause of the collapse — the hydrophobic force. In this way, the ZA model is essentially similar to the hydrophobic collapse model because both models suggest that it is the solvent-mediated hydrophobic force that drives the first stage — the collapse, in the protein folding process.

The ZA mechanism has been recently applied to the folding of proteins through using the AMBER96 force field with a generalized Born/surface area implicit solvent model and replica exchange molecular dynamics sampling. The results show that such a physics-based model can find approximately correct folds for chain length up to ~ 100 amino acids (Ozkan et al., 2007).

3.4 Jigsaw puzzle model

The jigsaw puzzle model, a qualitative mechanism for explaining the folding of the compact domains of proteins, was proposed by (Harrison & Durbin, 1985). They argued that proteins fold by large numbers of different, parallel pathways rather than by a single definitive sequence of events, thus making folding more robust to mutations that do not affect adversely the native structure. Here the existence of multiple protein folding pathways towards the native structure resembles the multiple routes to reach a unique solution in the assembly of a jigsaw puzzle. According to this model, the identification of intermediates represents a kinetic description rather than a structural one since each intermediate consists of heterogeneous species in rapid equilibrium (Yon, 2002). This model presents some similarities with the diffusion-collision model because its kinetics could be obtained through the diffusions and collisions of the "microdomains" if all the elementary microdomains have

similar properties and multi-microdomain intermediates of the same size have similar folding and unfolding rates. The computer simulations based on lattice models support the jigsaw puzzle model (Sali et al., 1994). The multiple pathways folding highlighted by the jigsaw puzzle model is actually consistent with a “new view” of protein folding funnel concept based on the energy landscape theory: the native state of a protein at its global free energy minimum is located at the bottom of funnel-like energy landscape; and each molecule may follow a different microscopic route from the top to the bottom. The folding funnel model and energy landscape theory will be discussed in more detail later.

3.5 Hydrophobic collapse model and folding funnel hypothesis

The hydrophobic collapse model (Figure 1C), first proposed by (Dill, 1985), is mainly based on the observation that the globular proteins' native states often contain a hydrophobic core of apolar amino acid side chains in the interior of proteins, while leaving most of the polar or charged residues on the solvent-exposed protein surface. Therefore this model considers that the folding of a globular protein starts with a rapid collapse of the chain driven by the hydrophobic force, resulting in the formation of the collapsed intermediates, i.e., the molten globule, within which the secondary and/or tertiary structure is subsequently formed. This model highlights that the “earliest event” of the folding is that the side chains of hydrophobic residues are driven together by the action of water, which is in much the same way as aggregation of oil droplets in water (Kauzmann, 1954, 1959; Lum et al., 1999; Stillinger, 1973; R. Zhou et al., 2004). Upon dripping oil into a water solvent or upon recovering the conditions for protein folding, the nonlocal and nonspecific hydrophobic force mediated by the water molecules forces the polypeptide chain or the oil to cluster together rapidly to minimize the surface area of the aggregates and to maximize the entropy of the solvent, thus lowering the total Gibbs free energy of the systems. In the case of the protein folding, the free energy is further lowered by favorable energetic contacts such as the isolation of electrostatically charged side chains on the protein surface and neutralization of salt bridges within the protein's interior. The subsequent step of the protein folding is a slow annealing to the native structure: mutual rearrangement and adjustment of the already formed building blocks through disruption and reformation of various noncovalent bonds, which further lowers the free energy of the system via enthalpy contribution. As mentioned above, the hydrophobic collapse model is similar to the hydrophobic zipper model, both point out that the folding is dominated by water-mediated hydrophobic interaction.

Compared to the hydrophobic collapse model, folding funnel hypothesis characterizes a more complete scenario of how a population of unfolded peptide chains goes down from the top of funnel-like energy landscape to the bottom of the native states via multiple routes. The hydrophobic collapse is just one of the most important events occurring in early-stage folding. Other events such as the formation of the molten globule, trapping of folding intermediates in the local minima, overcoming the local energy barriers, formation of the transition state ensemble, reconfiguration of the tertiary interactions and aspects of thermodynamics and kinetics of protein structure can also be characterized by the folding funnel (these will be introduced and discussed in the “Free energy landscape” section). In the canonical depiction of the folding funnel, the depth of the well represents the energetic stabilization of the native state versus the denatured state; the width of the well represents

the conformational entropy of the system; and the surface outside the well is shown to be relatively flat to represent the heterogeneity of the random coil state.

3.6 Stoichiometry-driven protein folding hypothesis

The stoichiometry-driven protein folding hypothesis has been proposed very recently by Mittal et al., whose statistical analyses of 3718 folded protein structures reveal a surprisingly simple unifying principle of backbone organization, which is interpreted as Chargaff's Rules related to protein folding, i.e., a stoichiometry-driven protein folding (Mittal & Jayaram, 2011b, 2011a; Mittal et al., 2010). One of the interesting findings is that the total number of possible contacts for the C_{α} atom of a given amino acid correlates excellently with its occurrence percentage in the amino acid sequences, leading to a conclusion that protein folding is a direct consequence of a narrow band of the stoichiometric occurrences of amino acids in the primary sequences, regardless of the size and the fold of a protein. On the other hand, the statistical results reveal that there is no "preferential interactions" between amino acids, thus leading to a conclusion that the "preferential interactions" between amino-acids do not drive protein folding.

However, although interesting, the above two conclusions need to be interpreted carefully and more work need to be done to verify this hypothesis and to elucidate the mechanism underlying these statistical phenomena (Berendsen, 2011; Chan, 2011; Ji & Liu, 2011a). For example, how to elucidate the relationship between the stoichiometric occurrences of amino acids and the driving force of protein folding? Is it possible that the higher the occurrence frequency an amino acid has, the greater tendency for it to occupy the core of a folded protein, and the more it contributes to the hydrophobic interaction? Further work is required to examine the relationship between the burial extent of an amino acid and its occurrence percentage. The statistical result of the lack of preferential interactions between amino acids is based on a large number of already folded protein structures, and therefore it is safe to interpret it as the consequence rather than the cause of protein folding. However, it must be noted that the preferential interactions between amino acids are the basis of the development of knowledge-based potentials, which in turn form the underpinning of protein structure prediction by modeling and simulation (Ji & Liu, 2011a; Rackovsky & Schraga, 2011; Sarma, 2011) that are now routinely performed in many laboratories across the globe (Aman et al., 2010; Sklenovský & Otyepka, 2010; Tao et al., 2010). The lack of preferential interactions between amino acids in a large sample set suggests that the non-preferential or random inter-residue interactions might maintain the structural stability of the already folded proteins. However, using only the C_{α} atom as the statistical object may shield the effect of the preferential interactions occurring to a very large extent between the side chain groups of different amino acids (Galzitskaya et al., 2011; Matthews, 2011). It is reasonable to consider that the preferential interactions do not dominate protein folding when the folding process is thought to be under the thermodynamic control such as in the hydrophobic collapse and/or folding funnel models. The entropy effect is the dominant force that drives the early rapid collapses of the peptide chain, contributing the most to the decrease of the free energy of the protein-solvent system. The later slow rearrangement of the local building block elements is driven by enthalpic contribution, i.e., the loss and the formation of noncovalent bonds. We consider that the later slow process inevitably results

in preferential or special interactions between amino acids and contributes, although to a lesser extent compared to the entropy effect, to the decrease of the free energy of the system. Therefore, the stoichiometry-driven folding hypothesis needs further analyses and verification using both theoretical and experimental methods, as has been suggested and discussed by dozens of protein structural chemists in “a conversation on protein folding”, which was organized by professor Sarma (Mittal & Jayaram, 2011b; Sarma, 2011), the Editor-in-chief of the Journal of Biomolecular Structure and Dynamics.

3.7 Association between different protein folding models

The protein folding models or hypotheses described above are not independent and mutually exclusive but rather they are inextricably intertwined and commonly attempt to grasp different aspects of protein folding and to solve one or more of the three problems concerning protein folding: the folding code, folding rate and protein structure prediction. Experimental results have provided some support to each of the models. For example, the diffusion collision and framework models are essentially similar, commonly emphasizing a hierarchical folding process in which the local interactions drive the initial formations of simple structures (Figure 1A). Such a stepwise process is in line with our intuitive thinking for sequential events and facilitates the development of secondary structure prediction methods. In addition, if we consider the early formation of the secondary structures as individual nucleation processes, the diffusion collision model could be thought of as special case of the nucleation-condensation model. The hydrophobic collapse model highlights that the nonlocal hydrophobic force or the entropy effect mediated by the water is the predominant driving force for protein folding (Figure 1C). This model can be considered as another extreme of nucleation-condensation model because the collapse leads to one or more hydrophobic nucleuses composed of both the transient secondary and tertiary structures. Interestingly, the ZA model may be considered as the combined case of both the diffusion collision and the hydrophobic collapse models because of its hierarchical assembly feature and the dominant nonlocal hydrophobic force in collapse and zipping processes. In addition, the hydrophobic collapse is only a specific event occurring in the folding process described by the folding funnel model. The latter emphasizes the multiple parallel routes towards the global free energy minimum, which is also the essential feature of the jigsaw puzzle model. The advantage of the folding funnel model is the characterization of two essential elements of protein folding: the thermodynamics and kinetics, which play crucial roles not only in the protein folding but also in the binding and interaction between protein and its ligands.

4. Free energy landscape

As described above, protein folding is a complex problem and involves a series of complicated processes. Much of the complexity can be described and understood by resorting to a statistical mechanics approach to the energetics of protein conformation: the energy landscape (Bryngelson et al., 1995; Frauenfelder et al., 1991). It has been considered as the most realistic model of protein (Henzler-Wildman & Kern, 2007; Shea et al., 1999) because it provides a complete quantitative description of protein conformational space, including the folded native state (which is generally considered as the state of a global free energy minimum), ensembles of various conformational substates near or far way from the

native state, various unfolded or denatured states, and a large ensemble of folding intermediates (such as molten globule, transition states, trapped nonnative states, etc.). In the case of protein folding, the energy landscape is embodied by the folding funnel model, which provides not only a simple way of understanding why the Levinthal paradox (Levinthal, 1969) is not a real problem, but also a conceptual framework for understanding the different scenarios of protein folding highlighted by different folding models described above. Therefore, the energy landscape and folding funnel views have progressively replaced the classical model of a hierarchical pathway of folding (Yon, 2002). In the case of protein binding and function, it has now been widely accepted that the dynamics of proteins govern ultimately their function (Henzler-Wildman & Kern, 2007). The protein dynamics are defined as any time-dependent change in atomic coordinates, including both equilibrium fluctuations and non-equilibrium effects (Henzler-Wildman & Kern, 2007). The equilibrium fluctuations result in distinct probability distributions of conformational substates over the rugged energy landscape. These conformational substates reside in different energy wells near or far away from the well of the global energy minimum, in which the dominant population of the native state resides. The dynamics at the equilibrium, therefore, are often thought to govern biological function in processes both near and far from equilibrium (Henzler-Wildman & Kern, 2007) such as the protein-ligand binding through conformational selection or induced fit; and most studies focus on these motions and their functional consequences (Liu et al., 2007c; Liu et al., 2008; Liu et al., 2010, 2011; Tao et al., 2010). The non-equilibrium fluctuations arise from the conformation transition between conformational substates, which needs to overcome the energy barrier between them and involves the transition conformational states with transient lifetime and therefore, is hard to detect experimentally. Taken together, the energy landscape has two key factors that can characterize its shape: i) the thermodynamics that defines the relative probabilities of the conformational states; ii) the kinetics that defines the energy barriers between the conformational states (Henzler-Wildman & Kern, 2007) and the transition between them.

4.1 The origin of energy landscape and its concept

The concept of the energy landscape, although being now most familiar in the field of protein folding, is first proposed to be applied to the folded protein more than 30 years ago by Frauenfelder and colleagues (Austin et al., 1975). Using the low-temperature flash photolysis technique, they investigated the kinetics of carbon monoxide (CO) and oxygen (O₂) rebinding to myoglobin (Mb) as a function of temperature and ligand concentration. Moreover, through computer-solving the differential equation for the motion of a ligand molecule over four barriers, the rates for all important steps were obtained. The temperature dependences of the rates yield enthalpy, entropy, and free-energy changes at all barriers (Austin et al., 1975). In order to explain the observations of non-exponential kinetics and the four successive barriers below 210 K, an energy landscape model was proposed, which was described by Frauenfelder et al.: "*The energy landscape describes the potential energy E_c of the protein as a function of conformational coordinates; it is a hypersurface in the high-dimensional space of the coordinates of all atoms in Mb*" (Frauenfelder et al., 1991). Accordingly, the relation of the energy landscape to the myoglobin function was elucidated; and the features of the landscape, such as the existence of a large number of nearly isoenergetic conformations

(conformational substates) and the heights of the barriers between energy wells were characterized (Frauenfelder et al., 1979; Frauenfelder et al., 1991).

In the field of physical chemistry, the energy landscape of a protein-solvent system is defined as an energy function $F(x) = F(x_1, x_2, \dots, x_n)$, where x_1, x_2, \dots, x_n are variables specifying the protein microscopic states (Dill, 1999), which can be all the dihedral angles of the peptide chain, the eigenvector projections derived from essential dynamics analysis (Amadei et al., 1993; Tao et al., 2010), end-to-end distance of the peptide chain, the number of native contacts, and an order parameter that describes the similarity of the structure to the native one and others (Kapon et al., 2008) or any degree of freedom (Dill, 1999). $F(x)$ is then usually the free energy as a function of the conformation of the protein to describe the protein-solvent system, where the entropic part of the free energy comes from all possible solvent configurations and solute conformational states; and the enthalpic part comes from the noncovalent bond formation within the protein interior and formation and loss of bonds (such as the hydrogen bonds and van der Waals interactions) between the protein and solvent. Thus, the protein folding process can be seen as a search for the global free energy minimum value through solving the free energy function. For a protein with a rough energy surface at the bottom of the energy landscape, the locally stable conformations can be found by determining the set of values x_1, x_2, \dots, x_n that gives the local minimum values of the free energy function.

4.2 The funnel-shaped energy landscapes

Although the energy landscapes are highly multi-dimensional, they are generally pictured as a surface in the three dimensions due to the difficulty to draw a multiple-dimensional space. In the 3D energy landscape, the vertical axis represents the free energy and the horizontal axes represent the conformational degrees of freedom of a polypeptide chain (conformational entropy). Random heteropolymers, i.e., a polypeptide chain containing random amino acid sequences, may have either a very rugged energy landscape with too many local minima (Plotkin et al., 1996) or a very large flat energy landscape (Dill & Chan, 1997). Systems like this can either easily get trapped in one of the local minima or never find the minimum when walking randomly on the very flat surface and as such, usually do not have a unique, well-defined and stable conformational state. On the contrary, the real proteins have evolved to contain optimized sequences so that they can fold rapidly and efficiently into well-defined native conformational states (Onuchic & Wolynes, 2004). In order to fold quickly and efficiently, proteins must have the tilt, funnel-shaped energy landscape. Hydrophobic collapse driven by the entropy gain of the water solvent leads to compact chain conformations of the polypeptide and a reduced ensemble of peptide chain conformations, thus narrowing the energy landscape. There are many non-native states with high free energy but only one native state with a global free energy minimum. The ensemble of conformations is further reduced by enthalpic contribution such as the isolation of electrostatically charged or polar amino acid side chains on the protein surface and the neutralization of ion pairing within the interior of the protein, thus further narrowing the landscape. The final step occurring at a narrower part (near the bottom of the landscape) of the energy landscape is the reconfiguration of a very small set of near-native conformations through cooperative formation and mutual adjustment of tertiary interactions, resulting in the unique native conformation located at the narrowest part of the energy landscape.

Therefore, the progressive reduction in dimensionality of the accessible conformational space (or the number of conformational states) makes the protein folding landscape look like a funnel, meaning that many conformations have high free energy and few have low energy. A large number of the denatured states are located on the surface outside or the upper part of the funnel and have high conformational entropy, whereas few native states are located at the narrowest bottom of the funnel and have low conformational entropy. Intuitively, “funnel” also implies that the landscapes are relatively smooth, meaning that the barriers are small so the folding process happens quickly. However, the “real” protein folding funnel should carry information about kinetics or barrier heights or smoothness or any landscape shape feature (Dill, 1999). Depending on the folding rate, properties, flexibility and the native structure of the proteins, folding energy landscapes are divided into the following types (Dill & Chan, 1997).

- i. The idealized protein folding funnel with smooth surface (Figure 2A). This funnel contains no traps or bumps, showing how the many denatured conformations can roll down along different routes to become the fewer compact conformations, and finally to one native conformation. Such a smooth funnel often results in fast folding and two-state (single-exponential) kinetics (Huang & Oas, 1995; Jackson & Fersht, 1991; Schindler et al., 1995; Sosnick et al., 1996).
- ii. The rugged landscape with hills, traps and energy barriers (Figure 2B). Like the real mountain ranges, this rugged funnel can also have much broader array of shapes involving hills, valleys, ridges, channels, moguls, plains and valleys inside valleys, moats, varying slopes and ups and downs of all kinds. Such a rugged funnel often results in slow folding and the folding kinetics is likely multiple-exponential.

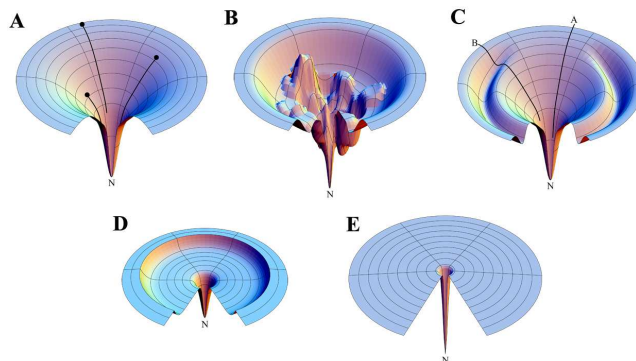


Fig. 2. Different types of free energy landscapes. (A) An idealized funnel-shaped landscape with smooth surface, which could guide the two-state folding kinetics at maximum attainable speed. (B) A rugged energy landscape with hills, traps and energy barriers. Folding kinetics is likely multiple-exponential. (C) The moat landscape illustrates how a protein could involve a fast (via path A) or slow (via path B) folding process. (D) The champagne glass landscape exhibits how conformational entropy can bring about “free-energy barriers” through random wandering on the flat plateau. (E) The Levinthal “golf course” landscape illustrates that a random search could not find the native state of the global energy minimum at the bottom of the “golf hole”. N denotes the native state. Images were cited from (Dill & Chan, 1997).

- iii. The moat landscape on which the “moat” represents the kinetic trap (Figure 2C). As shown in Figure 2C, the A route is a funnel-like “throughway” path, while the folding molecules should have to pass through an obligatory folding intermediate trap when they follow the B route.
- iv. The champagne glass landscape (Figure 2D). This type of landscape is proposed to illustrate that the “bottleneck” or rate limit to folding is due to conformational entropy (Dill & Chan, 1997). The polypeptide chain is delayed en route to the bottom by aimless wandering on the flat plateau to find the remaining routes downhill, indicating how conformational entropy can cause “free-energy barriers” to folding. Here the “free-energy barriers” do not correspond to a process that goes “uphill in free energy” to overcome a free-energy barrier in the conventional sense, but just mean that the folding process is delayed compared to some reference rate.
- v. The Levinthal “golf course” landscape (Figure 2E). This type of landscape is proposed to illustrate Levinthal’s argument (Levinthal, 1968, 1969) that a random search could not find the native state of the global energy minimum located at the bottom of the “golf hole”. It is just like a ball rolling randomly on a very large flat course, taking a long time to find and fall in the hole.

4.3 Protein folding in the funneled energy landscape

Although there exist the smooth funnel-shaped energy landscapes without significant kinetic traps (Figure 2A), the most realistic folding funnel is rugged, guiding multi-exponential slow folding of most proteins (Figure 2B). Under inappropriate folding conditions, the funnel is shallow and therefore the polypeptide chains spend most of their time meandering around the upper part of the funnel, leading to conformational heterogeneity of the unfolded proteins and very slow folding due to the shallow slope (Dill & Chan, 1997). However, once restoration of the appropriate conditions, the funnel becomes stretched down leading to the steep slope, thereby initiating the downhill movements of molecules towards the native state (Figure 3).

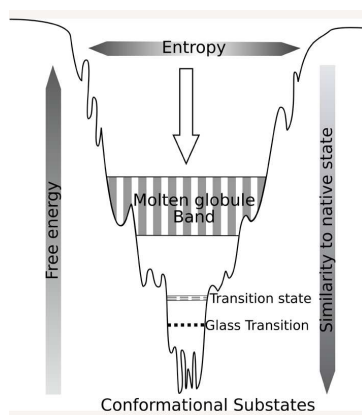


Fig. 3. A rugged funnel-like free energy landscape that describes the detailed processes and various intermediates in the folding reaction. The width of the funnel represents the conformational entropy; the depth of the funnel represents the change in total Gibbs free energy between the denatured and native states. This figure is modified from (Onuchic et al., 1997; Onuchic et al., 1995; Wolynes et al., 1995).

Numerous events occur during the downhill process and lead to the formation of numerous different conformational states, such as the unfolded state, collapse, molten globule, glass transition, transition state, intermediate, native state, and conformational substate. It must be kept in mind that each of these conformations just represents a certain type of thermodynamic state, which exists in an ensemble of conformers and exhibits distinct population time. The conformational states occurring in the folding process should be seen as ensembles but not as specific structures. For example, the intermediate is a very large ensemble of conformational states because it contains all the states with only the exception of two states, the initial unfolded and the final native ones.

A funnel-shaped energy landscape illustrating the detailed processes and various intermediates in protein folding is shown in Figure 3. Just like an ensemble of skiers jumping and skiing down the steep slope (in this case the major driving force is the gravity), the first phase of protein folding is that an ensemble of unfolded polypeptide chains collapses rapidly to a compact ensemble. In this case the driving force is the chemical potential of the reduction in Gibbs free energy, to which the entropy gain via solvent sequestration makes a substantial contribution. Each skier skis along a different route, down the mountainside as shown in Figure 2B and Figure 3 and subsequently enters a rugged region composed of hills, valleys and traps. Correspondingly, the rapid hydrophobic collapse leads to the accumulation of "misfolded" compact intermediates, many of which are trapped in the kinetic traps or local free energy minima. Such an ensemble of compact folding intermediates, which is termed molten globule, contain some transient secondary structures and nonspecific tertiary interactions (Onuchic et al., 1995). The subsequent processes are slow steps, which arise from climbing an uphill slope, then reach a mountain pass and continue to the next downhill search (Dill & Chan, 1997). In these processes, the ensemble of the conformations located at the hilltop can be considered as "transition state", which, in the sense of the energy landscape, is defined as "*the bottleneck processes of flows down different mountainsides*" by (Dill & Chan, 1997) to highlight the aspect of the rates but not the specific structures. In order to overcome the bottleneck, the already collapsed conformations need to be pulled apart. This may be driven by the conformational entropy of the molten globules. In addition, uphill climbing does not necessarily mean a full opening of the compact collapses. This process usually is the breaking of a few non-covalent bonds here or there to resume progress towards the native state. The process of the transient trapping, uphill, downhill, transient trapping can repeat many times, thereby slowing the exploration of the routes towards the native state. Therefore such a repetitive process is a bottleneck for folding and can be viewed as the folding transition state as defined by Dill et al. (Figure 3) (Dill & Chan, 1997). Under thermodynamic conditions near the folding transition midpoint, entropy and energy do not completely compensate for each other; thus, intermediates are not present at equilibrium, i.e., a free thermodynamic energy barrier intercedes (Wolynes et al., 1995). The repetition of the uphills and downhills ultimately leads to the arrival of a so-called glass transition state (Bryngelson et al., 1995; Onuchic et al., 1995). Such a trapping resembles the way a liquid becomes a glass when cooled, remaining fixed in one of many structures and unable to reconfigure to the lowest energy crystal state (Wolynes et al., 1995). In the case of protein folding, therefore, the process of reconfiguring the glass transition to the native state is very slow (Onuchic et al., 1995), requiring a sufficient overall slope of the energy landscape so that the numerous valleys flow in a funnel towards the native structure. In addition, when the folding process enters into the glass transition, the intermediates have

only a few paths to the native structure. The lifetimes of the discrete glass transition states are relatively longer than those of other intermediates but shorter than the average folding time. The formation of native contacts will further reduce both the entropy and enthalpy, with the larger negative enthalpy compensating for the entropy loss, putting the glass transition \rightarrow native state under enthalpy control. For a more detailed and quantitative description of the funnel-like protein energy landscape, please refer to (Onuchic et al., 1997).

4.4 The roughness/ruggedness of energy landscapes

For most of proteins, the funnel-like energy landscapes, either on their tube part or at the bottom, are largely rugged or rough due to the kinetic traps, energy wells and barriers. Rough energy landscapes occur in problems in which there are many competing interactions which can not be satisfied at once (Bryngelson et al., 1995; Kapon et al., 2008). For example, in the entire folding process there are too many "mistaken interactions" that are caused by competition between different chain parts. The initial hydrophobic collapse expels the bound water molecules and leads to a compact ensemble which, although having relatively lower free energy, contains many non-native contacts, energetically costly cavities generated by imperfect fit of side-chains, and locally favorable secondary structures. These inappropriate interactions and orientations have to be disassembled and readjusted in order to form the final tertiary structure. Analogously, in the process of protein-ligand binding, the water network formed around the surface of the interacting partners need to be broken upon interaction and the changes in conformations of the protein or ligand may also be required, and sometimes a large scale conformational changes are required. Such competitive interactions, on the one hand, lead to relatively stable residence of conformational states in wells of the local free energy minima, on the other hand, facilitate conformational motions that can help the molecules to jump out of the wells. In fact, for a bio-macromolecule such as the protein, its folding or binding processes will inevitably lead to a single structure or a complex that can not satisfy completely all of the constraints posed on the system and thus the reactions are said to be frustrated (Kapon et al., 2008; Shea et al., 1999).

Frustration, or ruggedness, a well-known trait of the complex systems such as glasses, solute-solvent systems and even social and economic networks, arises essentially from the fluctuation of free energy caused by non-complementary changes between entropy and enthalpy. For instance for the protein-solvent system, the change in enthalpy is the consequences of the competitive interactions between different parts of the protein and between the protein and solvent; while the change in entropy is caused by the nature of protein and solvent to increase their disorder or randomness. When the enthalpy change can not compensate for the entropy change, the fluctuation of the free energy occurs. In the energy landscapes, the local free energy fluctuations manifest themselves as a series of hills, valleys and traps of various heights and widths. In the case of protein folding, although the local free energy surface is rough or rugged, the entire free energy landscape is funnel-shaped (Figure 2A, 2B and 3), satisfying the need for the global decrease in free energy of the protein-solvent system. In the case of protein function, although most functional studies are concentrated on static X-ray crystallographic structures of proteins, the biologically relevant processes, such as receptor-acceptor binding, enzymatic catalysis, neuron activity and selective passing of ions through ion-channels are essentially dynamic. It is the ruggedness of the energy surface that brings about the dynamic behaviors and molecular motions of proteins. For example, the polar or charged residues residing inside the

hydrophobic core can often bring about local instability and contribute to frustration. Nevertheless, it is such a frustration that serves to exchange ligands or degrade substrates (Tao et al., 2010). Loops are observed to exist universally in protein structures. Long flexible loops often assume a large number of conformations due to their inherent large conformational entropy. However, the large flexibility of loops makes them very useful for contact, recognition, binding and catching of the ligands, in which they are frequently found to play a pivotal role. Interestingly, for HIV gp120 glycoprotein, several long variable loop regions, i.e., V1-V5, located outside the surface of the molecule, shield the antigenic sites and make substantial contribution to the immune escape of HIV (Liu et al., 2007a; Liu et al., 2007b; Liu et al., 2007d, 2008).

4.4.1 Ruggedness, folding rate and protein function

As described in the section “The problems of protein folding”, one of the protein folding problems is the folding rate. The funnel-shaped landscape seems to address this problem. It has been predicted that the speed limit for the folding of small globular proteins in a smooth folding funnel is $\sim 100/n$ μ s, where n denotes the number of amino acid residues in the protein (Kubelka et al., 2004). However, most natural proteins fold at least two or three orders of magnitude slower than the predicted empirical relationship due to the presence of the roughness of the energy landscape. Furthermore, the different degree of the ruggedness raises another question: what is a good speed for a protein to fold (Kapon et al., 2008)? One intuitive answer is that proteins should fold as fast as possible in order to avoid degradation and maintain maximum stability of their folded structure in the cellular environment. However, this raises the question of why roughness, with its detrimental effect on folding rates, has been preserved by evolution. It has been proposed that the answer is the functional requirement of the proteins because God has created proteins to perform a task but not to fold as fast as possible (Gruebele, 2005). It has been shown that some proteins which have smooth energy landscapes and fast folding rates tend to aggregation and proteolysis since they can also unfold very rapidly (Gruebele, 2005; Jacob et al., 1997; Kapon et al., 2008). On the other hand, if the energy landscape is too rough, the protein may also be subjected to similar risks due to the attenuation of diffusion along the energy surface (Kapon et al., 2008). Therefore, every protein must have evolved an optimum energy landscape roughness to satisfy simultaneously the two conflicting requirements: the survival and the function.

The ruggedness in the energy landscapes makes it possible for a protein to assume many states that may be quite distinct from each other in structure but have similar free energy. Furthermore, the transition between states or substates causes protein motions on different time scales and at different amplitude and directionality, which may involve different biological processes. The ability to adopt multiple conformational states also increases the repertoire of molecules with which a ligand can interact and provides robustness against deleterious mutations (Kapon et al., 2008).

4.4.2 Amplitude and timescale of protein motions

The protein dynamics contain two elementary components, i.e., the kinetic component and the structural component, which characterize the timescale and amplitude of the fluctuations, respectively (Ansari et al., 1985; Henzler-Wildman & Kern, 2007). Figure 4 shows a free energy landscape defining these two components. The protein dynamics are

divided into three tiers, i.e., tier-0, tier-1 and tier-2 dynamics that occur on timescales ranging from the “slow” to “fast” timescale.

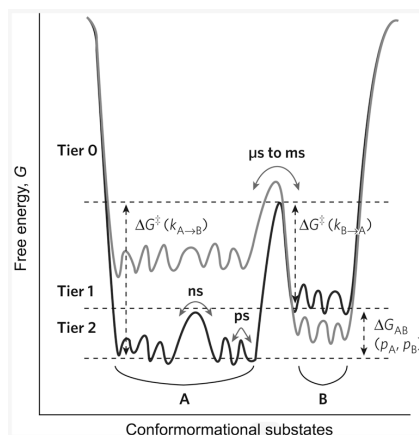


Fig. 4. The energy landscape that defines the amplitude and timescale of protein dynamics. The conformational states are defined as minima in the energy surface and are located in the energy wells. The transition states can be seen as the free energy maxima between the wells and are located at the tops of hills. Tier 0 includes two conformational states, A and B. Conformational transition between them are rare because of the high energy barriers $\Delta G^\ddagger(k_{A \rightarrow B})$ and $\Delta G^\ddagger(k_{B \rightarrow A})$, resulting in a slow fluctuations between these two states with timescale of μs to ms , i.e., tier-0 dynamics. The populations of the tier-0 states, A and B (p_A , p_B), are defined as Boltzmann distributions based on their differences in free energy (ΔG_{AB}). Tier 1 includes several conformational substrates located within tier 0 wells. Tier-1 dynamics are fast fluctuations on timescale of ns to allow fast transition between tier 1 substrates. Similarly, Tier 2 describes faster fluctuations on timescale of ps between a large number of closely related substrates located within tier 1 wells. The difference between the two landscapes, depicted by the dark line and grey line, respectively, indicates that the energy landscape is dynamic, with its shape being affected and changed by physical (such as temperature, pressure, pH, ionic strength, presence of denaturant, etc.) or functional (such as ligand binding and protein mutation) conditions. This figure is modified from (Ansari et al., 1985; Henzler-Wildman & Kern, 2007).

The tier-0 dynamics, which occur on a slow timescale, define fluctuations between kinetically distinct states that are separated by energy barriers of several kT (the product of the Boltzmann constant k and the absolute temperature T). Such slow timescales are evaluated to be microseconds (μs) or slower at physiological temperature; and the corresponding motions are typical larger-amplitude concerted motions that can propagate over the entire molecular structure, leading to a relatively small number of conformational states with relatively large conformational difference between them (Figure 4). The protein molecules within one of these tier-0 states (the largest wells in Figure 4) are also dynamic, fluctuating around the average structure on a faster timescale and exploring a large ensemble of closely related structures. This brings out the tier-1 or even tier-2 states located in smaller wells relative to the tier-0's large wells. An important feature of tier-0 states is that

the conformational transitions between these states are rare because of i) high free energy barrier between tier-0 states; ii) long lifetimes of tier-0 states and stable equilibrium fluctuations around these states; iii) the low probability of the conformation that can initiate the transition. Such large-scale concerted motions occurring on the slow timescale are now receiving more and more attention since they govern many biological processes such as the enzymatic catalysis (i.e., the opening of substrate binding channel), signal transduction and protein-protein interactions (i.e., the conformational selection process). The relatively long lifetimes of individual states within the tier-0 wells make it possible for direct trap and observation of these states, and even for detection of the kinetics of inter-conversion between these states using experimental and computational methods. For the detailed description and discussion of these methods, refer to (Henzler-Wildman & Kern, 2007).

The tier-1 and tier-2 dynamics occur on “fast” timescale and define fluctuations within the wells of a tier-0 state. These many smaller wells within the tier-0 wells are occupied by a large ensemble of structurally similar substates. The energy barriers between them are less than $1 kT$, resulting in more local, small-amplitude picosecond (ps)-to-nanosecond (ns) fluctuations at physiological temperature, with the tier-1 and tier-2 dynamics corresponding to the ns and ps fluctuations, respectively (Figure 4). The dynamics on the fast timescale, which are the consequent of occasional Brownian bombardments, can sample a large number of conformational substates. In other words, the conformational diversity can be considered as the subsequence of the entropy of the system. According to (Henzler-Wildman & Kern, 2007), the tier-1 dynamics can be used to describe ns-timescale fluctuations of a small group of atoms, such as loop motions, while the tier-2 dynamics correspond to ps-timescale fluctuations of local atoms. such as the rotations of the amino acid side chains (Figure 4) . The even higher tiers of dynamics also exist, such as bond vibrations, which occur on the femtosecond (fs) timescale.

4.5 The energy landscapes are dynamic

The concept of a dynamic landscape was first proposed by (Gulukota & Wolynes, 1994) in their statistical model of how chaperones can work to help to form the normal folding funnel of proteins to their native states. The shape of energy landscape can be affected by an individual set of physical environments, such as pH, ionic concentration, presence of denaturant, pressure and temperature, or functional environment such as binding to other molecules (Figure 4). For example, under the denaturing conditions of high temperature or presence of denaturant, the energy landscape of a protein is shallow and its surface is relatively flat, with an ensemble of unfolded polypeptide chains walking randomly around the flat surface. However, upon restoration of the native-like condition, the energy landscape is stretched down and the slope becomes steep. Despite the existence of the hills, valleys and traps, the emergence of the steep wall allows the chains to roll down the funnel. Therefore, switching the conditions between the denaturation and refolding is essentially to change the shape of the energy landscapes of proteins. Furthermore, changes in energy landscape can also be the consequence of the bound state of the protein, that is, whether it is free, or bound to one or more ligands, can change the depth and width of the wells around the bottom of the energy landscape. The works by (Freire, 1999; Todd & Freire, 1999) showed that residues far away from the active site can be stabilized by intermolecular association, which provide an example illustrating that the binding event shifts the energy

landscape of a folding funnel (Tsai et al., 1999b). Changing or manipulating the physical or functional conditions is the most common way to change the relative populations of the conformational states or substates, and the kinetics of the conversion between them (Henzler-Wildman & Kern, 2007). In summary, although the protein function is often inferred from its single stable conformations derived from X-ray crystallographic technique, a deep and complete understanding of the protein function requires analysis of the changes in populations of its conformational substates caused by the dynamic energy landscape upon change in the physical or functional conditions.

5. Protein-ligand interactions and binding mechanisms

A fundamental principle of all biological processes is molecular organization (e.g., protein folding) and recognition, e.g., protein-ligand interactions and binding (Perozzo et al., 2004). Essentially, proteins perform their function through interaction with molecules such as proteins and peptides, nucleic acids, ligands and substrates, and other small molecules such as oxygen or metal ions. Thus, a detailed understanding of biological processes requires to investigate not only the protein folding mechanism but also the mechanism of protein-ligand binding.

Molecular recognition is a process by which biological macromolecules interact with each other or with small molecules to form a specific complex (Demchenko, 2001; Janin, 1995; Otlewski & Apostoluk, 1997). It has two important characteristics: specificity and affinity. The specificity is that biological macromolecules are able to distinguish the highly specific ligand from less specific ones. In order to make an interaction specific, the bonds between correct partners should be strong, while for other partners showing only minor differences in structure they should be weak or even repulsive. Therefore, once a correct ligand binds to its acceptor molecule, the affinity between them should be strong enough to prevent disassociation. The most important aspect of the molecular recognition is, as pointed out by (Demchenko, 2001), that the recognition is usually not a process in itself, but is an element of a more complex, functionally important mechanism such as allosteric regulation of enzyme activity, signal transduction, protein folding or the formation of multi-subunit and supramolecular structures. This requires important and sometimes dramatic changes in properties of the interacting partners. In this section, the protein-ligand binding mechanisms, the relationship between protein folding and binding, and the thermodynamics and kinetics of protein-ligand interaction will be described and discussed.

5.1 Protein-ligand binding mechanisms

The protein-ligand binding mechanisms have evolved from the early “lock-and-key” (Fischer, 1894) to the “induced fit” (Koshland, 1958) and to the now popular “conformational selection” models (Frauenfelder et al., 1991; Ma et al., 1999; Tsai et al., 1999a) (Figure 5). The lock-and-key model (Figure 5A) was first proposed by (Fischer, 1894) to explain binding of a single substrate to the enzyme. In this analogy, the lock is the enzyme and the key is the substrate. The prerequisite of this model is that the enzyme and substrate are both rigid and their surfaces should be complementary, and therefore only the correctly sized key (substrate) fits into the key hole (active site) of the lock (enzyme); the keys with incorrect size or incorrectly positioned teeth (incorrectly shaped or sized substrate

molecules) do not fit into the lock (enzyme) because both the key and the lock can not change their shape and size. However, not all experimental evidence can be adequately explained by the lock-and-key model. Therefore, a modification called the induced fit model was proposed to compensate for the rigid, structurally invariable protein and ligand. The induced fit model (Figure 5B) assumes that the ligand plays a role in determining the final shape of the protein, i.e., the ligand binding induces a conformational change in the protein (Koshland, 1958). However, it seems that the induced fit mechanism is only suitable for the proteins showing minor or moderate conformational change after the ligand bindings. For the proteins undergoing substantially larger conformational changes upon ligand binding, Bosshard (Bosshard, 2001) suggested that *"induced fit is possible only if the match between the interacting sites is strong enough to provide the initial complex enough strength and longevity so that induced fit takes place within a reasonable time"*. Therefore, the induced fit mechanism alone can not explain well the association between ligand and proteins undergoing large conformational changes after binding (Tobi & Bahar, 2005).

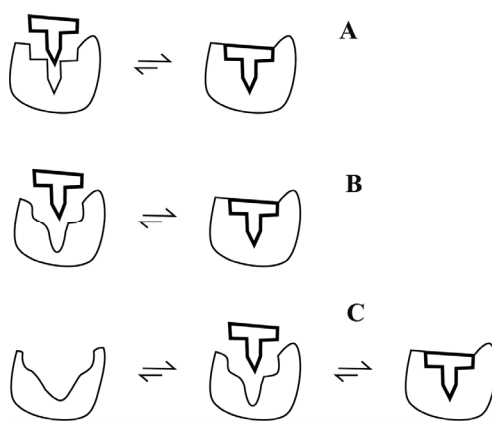


Fig. 5. Mechanisms for protein-ligand binding. (A) lock-and-key model. (B) Induced fit. (C) Conformational selection followed by induced fit. This figure is modified from (Tobi & Bahar, 2005).

In recent years, the conformational selection model (also termed preexisting equilibrium, population selection, fluctuation fit, or selected fit; see Figure 5C) has emerged as an alternative for induced fit and is becoming prevalent. Essentially, such a mechanism is the extension of the folding funnel model, postulating that the native state of a protein is not defined by a single rigid conformation but rather by an ensemble of closely related conformational substates coexisting in equilibrium, and the ligand can select to bind to the most suitable conformer among these conformational substates, shifting the equilibrium toward complex formation. As described above, at the bottom of the folding funnel, the rugged energy landscape near the native state contains several minima corresponding to these conformational substates. The more flexible the protein is, the lower the barrier height between the substates, and the larger the ensemble of acceptable near-native conformations may exist. Generally, the conformational changes in this model are beyond local side chains rearrangements near the binding site, but rather display large concerted motions or entire domain movements.

The early studies by (Berger et al., 1999) and (Foote & Milstein, 1994) demonstrate that conformational selection is responsible for antigen-antibody complex formation. For example, of the two isomeric conformations determined by X-ray crystallography for the SPE7 antibody (L.C. James et al., 2003), only one possesses a promiscuous, low-affinity binding site for haptens. The initial selective binding of haptens to this site is followed by an induced fit, resulting in a high-affinity SPE7-haptens complex. Recently, more and more data from the X-ray and cryo-electron microscope images, kinetics studies, extensive single molecule fluorescence and, in particular, NMR technique show that a unliganded protein can assume a repertoire of conformational states, including conformations corresponding to the bound form (Boehr et al., 2009; Lange et al., 2008; Tobi & Bahar, 2005; Wu et al., 2009).

5.2 The relationship between lock-and-key, induced fit and conformational selection

Interestingly, when considering the lock-and-key and conformational selection models under the background of the funnel-like energy landscape, the former can be viewed as an extremity of the later. The lock-and-key model is useful for explaining binding of a ligand to a very rigid protein, which has a smooth folding funnel with almost no ruggedness around the bottom of the energy landscape. Although a protein with relatively large structural flexibility has the folding funnel with massive ruggedness, the conformational selection model considers that, generally, only one ensemble of conformational substates residing in one energy well is appropriate for selective ligand binding. If we consider such an energy well as a small funnel and neglect the other wells as well as the kinetics of conversion between conformational substates, the binding process occurring in this well can be described by the lock-and-key mechanism.

Furthermore, induced fit and conformational selection are not two independent and exclusive processes but rather they both play a joint role in molecular binding (Perica & Chothia, 2010). What seems to be the general trend is that the conformational selection is more important for the initial recognition/contact between the interacting partners, while the induced fit is more important for the subsequent mutual conformational adjustment. Therefore, we consider that the large-scale concerted motions, i.e., tier-0 dynamics, may govern the conformational selection, while the tier-1 and tier-2 dynamics, which describe the fast-timescale motions (i.e., loop motions and side chain rotation), may play a role in fine-tuning the interactions at the later "induced-fit" stage. The induced fit model can also be viewed as an extreme version of the conformational selection when considering the detailed process of the molecular recognition rather than focusing only on the final conformational difference between the free and bound states. As discussed above, most proteins have a rugged energy surface around the bottom of the landscape, leading to distinct population distributions of different conformational substates. The initial selective collision/contact between the partners may decide whether the binding proceeds forward or the occasional complex disassociates. Only in the case of the approximately "correct" conformational state (which provides the initial complex the interactions with enough strength and longevity so that induced fit takes place within a reasonable time) can "conformational selection" occur, although the already chosen conformational state is significantly distinct from the final bound state. In other words, in the case of molecular recognition between flexible partners, the selectively initial interactions must occur before the event of the conformational adjustment, and therefore the induced fit can be perceived as an extremity of the conformational selection model.

Very recently, Nussinov and her colleagues (Csermely et al., 2010) proposed an extended conformational selection model to integrate the conformational selection and induced fit mechanisms, suggesting that i) the protein binding embraces a repertoire of selection and adjustment processes; and ii) the protein segments whose dynamics are distinct from the rest of the protein can govern conformational transitions and allosteric propagation that accompany the binding processes. A single step of conformational selection generally lowers the entropy barrier and helps to achieve a high degree of specificity (e.g., like the situation under the lock-and-key model). The subsequent mutual conformational rearrangement contributes significantly to the high binding affinity, e.g., the binding interfaces often have a certain degree of flexibility that is advantageous to conformational change and adjustment (Liu et al., 2011; Tao et al., 2010), like the situation in the induced fit mechanism.

The superficial difference between these two mechanisms is whether the binding is “conformational change first” (for the conformational selection) or “binding first” (for the induced fit). However, for a given protein to interact with its ligand, how the binding can be characterized as conformational selection or induced fit? What factors determine which mechanism dominates the binding process? Hammes et al. has proposed a flux criterion by which the sequence of events can be determined quantitatively (Hammes et al., 2009). Through applying the flux calculation to protein-ligand binding, they found that the binding mechanism switches from being dominated by the conformational selection pathway at low ligand concentration to induced fit at high ligand concentration. Interestingly, through establishing a solvable model that describes the conformational transition of the receptor between the inactive (this form prevails in solution in the absence of the ligand) and active (this form is favored while a ligand is loosely bound) forms, Zhou (Zhou, 2010) found that the timescale of the active-inactive conformational transitions has effect on the selection of these two binding mechanisms, i.e., when the active-inactive transition rates increases, the binding mechanism gradually shifts from the conformational selection to induced fit, indicating that the timescale of conformational transitions plays a role in controlling the binding mechanism.

5.3 Examples of protein binding: ubiquitin, aggregation and molecule chaperones

By measuring residual dipolar couplings (RDCs) (Lakomek et al., 2008) resulting from partial alignment in a large number of media, the conformational ensemble of ubiquitin was calculated up to the microsecond timescale at atomic resolution (Lange et al., 2008). The most striking feature of the ensemble is the presence of “bound” conformations in the free form of ubiquitin (Figure 6), which result from the large scale concerted pincer-like motions of the binding interfaces. All the backbone conformations in the available 46 X-ray crystallographic structures of ubiquitin in complex with various binding partners are observed in the solution ensemble, despite the absence of any crystallographic information in ensemble refinement. This provides direct evidence that the conformational ensemble of ubiquitin in its bound states does indeed exist in the solution condition without including protein interaction partners, and that conformational selection is important for protein-protein binding of ubiquitin. However, subsequent conformational changes may be induced fit processes, especially in the side chains, after the initial binding via conformational selection. Accordingly, we could conclude that the side chain flexibility contributes to the

high binding affinity through induced fit, while the conformational selection and highly rigid hot spot residues (Ma et al., 2001) contribute to the high specificity through lowering entropy barrier.

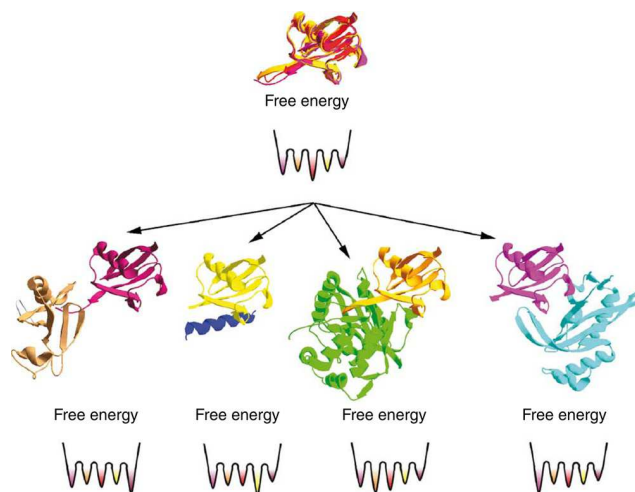


Fig. 6. An example of the conformational selection mechanism responsible for protein-protein binding of ubiquitin. The structures on the top show the NMR-derived conformational ensemble of ubiquitin. Shown in the middle are five ubiquitin crystallographic structures in complex with different ligand proteins (PDB 1F9J, 1S1Q, 1XD3, 2D36 and 2G45). The hypothetical free energy landscapes are shown in the bottom. The color of the free energy well is the same as that of the ubiquitin crystal structure corresponding to the conformational state located within this well. The deeper the well is, the larger population of the corresponding conformational state resides in it. The energy barriers separating the conformations are not known. This figure is modified from (Boehr et al., 2009)

There are several “misfolding diseases” that arise from possible mistakes in protein folding. These diseases, i.e., amyloid (Alzheimer’s disease) and prion diseases (BSE or “mad cow” disease, scrapie, Creutzfeld-jacob disease (Serban et al., 1990)), are characterized by abnormal accumulation of aggregated proteins in the brain and other tissues. For example, the formation of amyloid plaque are thought to be related to the conformational transition of a N-terminal structural region of A β protein, that is, an easy transition from the native α -helical to non-native, prone-to-plaque-formation β -stranded structures (Kosik, 1992). The prion protein also undergoes a conformational transition from a normal cellular isoform containing a preponderance of helical secondary structure to a scrapie isoform containing a greater proportion of β -sheet (Inouye & Kirschner, 1997; Pan et al., 1993). At first glance, the formation of the pathogenic plaque is due to the misfolding of relevant proteins. However, the so-called “misfolding” is in fact the finally observed already folded structure in the plaque but rather than the cause of the protein aggregation. Conformational analysis demonstrates that the A β peptide segment exists as a mixture of rapidly equilibrating extended conformers, including the “native” conformational state and the “natively unfolded” states (Weinreb et al., 1996). Some of the natively unfolded states may initiate

protein-protein interaction, forming a fibril seed facilitating plaque growth. The flexible nature of A β or prion proteins determines that the bottoms of their funnel-like landscapes, despite ruggedness, are relatively shallow, allowing easy transition between conformational substates. It has been proposed that flexible protein interface regions are more prone to aggregation than other surface regions (Pechmann et al., 2009). Therefore, some of the conformers that have proper flexible surfaces, although having a low population time, can interact and aggregate together, biasing the equilibrium towards the plaque conformation and deepening the energy well in which the plaque state is located. Therefore, aggregation is a manifestation of conformational selection of higher energy, less populated monomer states leading to highly polymorphic aggregate species, and is now increasingly considered as the “side-effect” of protein binding (Csermely et al., 2010). As such, the “misfolding decreases” should be called the “diseases of false recognition and binding”.

It seems that protein aggregation is inevitable in the living cell, and therefore there must be some mechanisms capable of repairing or preventing the aggregation. Actually, molecular chaperones can perform this function through temporarily covering aggregation-sensitive surfaces of the relevant proteins. Molecular chaperone is the common name of a diverse family of proteins which help other proteins to fold, refold after transport through biological membrane or maintain the folded state under conditions of stress (Demchenko, 2001). The common feature of molecular chaperones is that they interact only with unfolded or partially folded but not with native proteins, while the formed complex can dissociate into chaperone and a folded protein molecule. The chaperones are not enzymes and do not accelerate the folding reaction, but facilitate the assembly of protein complexes through the temporal occlusion of particular hydrophobic sites in polypeptide sequence, allowing the self-assembly of structure to proceed with the formation of folding determinants at sites that are free from chaperone protection. Molecular chaperones realize their functions through altering the flexibility of the target proteins or through protecting the hydrophobic and flexible interfaces of the potential aggregation-sensitive proteins, causing the distributions and redistributions of different conformational substates along the rugged surface of the dynamic energy landscape.

5.4 Thermodynamics of protein-ligand interaction

The complete understanding of molecular recognition of the proteins and ligands requires characterization of the binding energetics. Like the situation of protein folding, a quantitative description of the forces that drive the protein binding also requires characterization and determination of the complete thermodynamic profile, including the binding free energy (ΔG), the enthalpy (ΔH) and entropy (ΔS) of binding and the heat capacity change (ΔC_p) (Chaires, 2008; Perozzo et al., 2004). These thermodynamic data contain information crucial not only for elucidating the binding mechanism but also for rational drug design through relating the thermodynamic data to the structural data, which alone can not fully describe the driving forces for binding or predict accurately the binding affinity (Weber & Salemme, 2003).

The only direct way to measure the heat change during complex formation at constant temperature is implemented by isothermal titration calorimetry (ITC) (Freire et al., 1990; Wiseman et al., 1989). In this method one binding partner is titrated into a solution

containing the interacting partner, thereby generating or absorbing heat. This heat, reflected by temperature difference between a sample cell and a reference cell, is directly measurable and can be quantified by the calorimeter. Under appropriate conditions data analysis from a single experiment of the ITC allows the simultaneous determination of the equilibrium binding constant (K_b), the standard Gibbs free energy change (ΔG), the changes in enthalpy (ΔH) and entropy (ΔS), as well as the stoichiometry (n) of the association event. Moreover, experiments performed at different temperatures yield the heat capacity change (ΔC_p) of the binding reaction (Chen & Wadso, 1982; Cooper & Johnson, 1994; Perozzo et al., 2004).

The basic thermodynamic relationships are summarized below:

$$\Delta G = -RT \ln K_b \quad (1)$$

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

$$\Delta H_{VH} = -R \left(\frac{\delta \ln K_b}{\delta (1/T)} \right) \quad (3)$$

$$\Delta C_p = \left(\frac{\delta \Delta H}{\delta T} \right)_p \quad (4)$$

$$\Delta H_T = \Delta H_r + \Delta C_p (T - T_r) \quad (5)$$

$$\Delta S_T = \Delta S_r + \Delta C_p \ln \left(\frac{T}{T_r} \right) \quad (6)$$

$$\Delta G = \Delta H_r - T\Delta S_r + \Delta C_p \left[(T - T_r) - T \ln \left(\frac{T}{T_r} \right) \right] \quad (7)$$

where the ΔG , ΔH , ΔS and ΔC_p are the changes in free energy, enthalpy, entropy and heat capacity, respectively; K_b and R are the binding constant and universal gas constant, respectively; T and T_r refer to the temperature and an arbitrary reference temperature, respectively; and the thermodynamic parameters subscripted with "T" or "r" refer to those temperatures. The equilibrium binding constant K_b , which is defined as the ratio of actual product concentration to reactant concentrations, can be used to calculate the Gibbs free energy change ΔG by equation 1. The change in binding enthalpy of protein-ligand interactions can be determined accurately by the ITC method or indirectly from the temperature dependence of equilibrium binding constants and application of the van't Hoff relationship (ΔH_{VH} ; equation 3). Classically, free energy change can be parsed into its enthalpic (ΔH) and entropic components ($-T\Delta S$) as described by equation 2. Equation 4 indicates that the change in enthalpic value is temperature dependent, relating to a nonzero heat capacity change ΔC_p . Equation 7 shows that enthalpy and entropy changes depend on temperature through the nonzero heat capacity change ΔC_p . The relations of temperature dependent ΔC_p to changes in the enthalpy and entropy are described by equation 5 and 6, respectively.

5.4.1 Concepts of the thermodynamics parameters

Gibbs free energy ΔG , a concept that was originally developed in the 1873 by the American mathematician Josiah Willard Gibbs (Gibbs, 1873), can be seen as the approximation of the chemical potential, which is minimized when a system reaches equilibrium at constant pressure and temperature. The Gibbs free energy of binding reaction is the most important thermodynamic description of binding because it determines the stability of any given biological complex. Free energy changes are a function of states, i.e., their values are defined merely by the initial and final thermodynamic states, regardless of the pathway connecting them. Free energy is the key parameter since its value under a particular set of reactant concentrations dictates the direction of the bimolecular equilibria. The negative sign of ΔG means that the binding reaction or conformational transition will proceed spontaneously to an extent governed by the magnitude of ΔG ; the positive sign means that the energy is needed to drive the reaction to form a product, with the magnitude of ΔG specifying the amount of the required energy. The free energy is a balance between enthalpy and entropy, e.g., the observed ΔG can be the same with completely opposing contribution by enthalpy and entropy (see equation 2). An interaction with positive ΔH and ΔS (binding dominated by hydrophobic effect) can produce the same ΔG as an interaction with negative ΔH and ΔS (when specific interactions dominate). Such an enthalpy-entropy compensation makes the binding free energy relatively insensitive to change in molecular details of the interaction process, and thus the consideration of the ΔH and ΔS are crucial for a detailed understanding of the free energy of binding (Eftink et al., 1983; Lumry & Rajender, 1970; Perozzo et al., 2004; Williams et al., 1993).

The enthalpy is a measure of the total energy of a thermodynamic system. It includes the internal energy, which is the energy required to create a system, and the amount of energy required to make room for it by displacing its environment and establishing its volume and pressure. In many chemical, biological, and physical measurements, the enthalpy is the preferred way of expression of system energy changes because it simplifies descriptions of energy transfer. The total enthalpy, H , of a system cannot be measured directly and, as such, the change in enthalpy, ΔH , is a more useful quantity than its absolute value, which is equal to the sum of non-mechanical work done on it and the heat supplied to it. The ΔH is positive or negative in endothermic or exothermic process, respectively. In the protein-solvent system, the enthalpy change reflects the amount of heat energy required to achieve a particular state. For binding reactions, negative enthalpy values are common (but not omnipresent), reflecting a tendency for the system to fall to lower energy levels by bond formation. The binding enthalpy in its strict sense is considered as the noncovalent bond formation at binding interface. However, the heat effect of a binding reaction is a global property of the entire system, including contribution from solvent and protons (Cooper & Johnson, 1994). Therefore, the change in enthalpy of binding must be the result of the formation and breaking of many individual bonds, including the loss of protein-solvent and ligand-solvent hydrogen bonds, electrostatic and van der Waals interactions, the formation of noncovalent bonds between the protein and ligand, and the solvent reorganization near the protein surfaces. These individual components may make either favorable or unfavorable contributions, and the result is a combination of these contributions, with specific interactions dominating the binding enthalpy.

The entropy is a measure of the tendency of a process, such as a chemical reaction, to be entropically favored, or to proceed in a particular direction. It determines that thermal energy always flows spontaneously from regions of higher temperature to regions of lower temperature, in the form of heat. These processes reduce the state of order of the initial systems, and therefore the entropy can be seen as an expression of the disorder or randomness. In the protein-solvent system, entropy measures how easily that energy might be distributed among various molecular energy levels and represents all the other positive and negative driving forces (in addition to enthalpy change) contributing to the free energy. The total entropy change (ΔS_{tot}) of protein-ligand binding can be expressed as the sum of the three entropic terms: the solvent entropy ΔS_{solv} , conformational entropy ΔS_{conf} and the rotation and translation entropy $\Delta S_{r/t}$:

$$\Delta S_{tot} = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t} \quad (8)$$

where the ΔS_{solv} describes the change in entropy resulting from solvent release upon binding; the ΔS_{conf} is a configurational term reflecting the reduction in rotational degrees of freedom around torsion angles of protein and ligand; and the $\Delta S_{r/t}$ represents the loss in translational and rotational degrees of freedom of the acceptor and ligand upon complex formation. The most important contribution to the ΔS_{tot} arises from the solvation term ΔS_{solv} , primarily due to the release of well-bound water molecules on the surfaces of the protein and ligand, and the accompanying burial of the hydrophobic surface area. Since the entropy of hydration of polar and apolar groups is large, the burial of solvent accessible surface area and the solvent release upon binding often make a large, favorable contribution to the entropy increase. The ΔS_{conf} and $\Delta S_{r/t}$ make unfavorable contributions to total entropy, i.e., causing entropy reduction. Although the positive entropy changes resulting from a natural tendency for disruption of order and exclusion of water molecules are common for binding reactions (Chaires, 2008), all binding reactions would have to overcome the inescapable entropic penalties (Amzel, 1997, 2000; Brady & Sharp, 1997) (i.e., the negative $\Delta S_{r/t}$ and ΔS_{conf} upon complex formation) through large solvent entropy gain (positive ΔS_{solv}) or favorable protein-ligand interactions (negative ΔH) if binding is to occur.

The heat capacity (ΔC_p) is the measurable physical quantity that characterizes the amount of heat required to change a substance's temperature by a given amount. The removal of protein surface area from contact with solvent often results in a large negative ΔC_p . Therefore, the heat capacity provides a link between thermodynamic data and structural information of macromolecules through the correlation of ΔC_p and burial of surface area (Perozzo et al., 2004).

5.4.2 Enthalpy-entropy compensation

The phenomenon of enthalpy-entropy compensation has been widely observed in biological systems by using the thermodynamic methods (Eftink et al., 1983; Lumry & Rajender, 1970; Perozzo et al., 2004). Equation 2 shows that the change in free energy is characterized by a linear relationship between the changes in enthalpy and entropy. For example, favorable changes in binding enthalpy are often compensated for by unfavorable changes in entropy and vice versa. In a binding process, the increased bonding, which results in more negative ΔH , will be at the expense of increased order, leading to more negative ΔS . However, if

molecular binding is coupled with local or global protein folding (e.g. in the intrinsically unstructured protein segments or proteins (Tsai et al., 2001; Wright & Dyson, 1999)), the hydrophobic effect could become much more significant than the entropy loss caused by the loss of conformational entropy and rotational and translational entropy of the protein and the ligand. Furthermore, when the binding is stronger and more specific with higher positive interaction enthalpy contribution to the binding free energy, the entropy contribution can become large and positive (Demchenko, 2001). The consequences of the enthalpy-entropy compensation are i) that it does not bring out dramatic change in the binding free energy and as such leads to only small changes in binding affinity over a range of temperatures; ii) that it makes it difficult to distinguish between the entropy-driven or enthalpy-driven binding processes and, therefore, cautions are required to explain the experimental data, design drug, or plan engineering experiments. The factors affecting the enthalpy-entropy compensation include the properties of the solvent (water), the architecture of the ligand-binding site/pocket/cavity, the molecular structure of the ligand, and the changes in intermolecular forces in the binding process (Dunitz, 1995; Gilli et al., 1994). The two thermodynamic parameters, ΔH and ΔS , are correlated with each other through a bridge of the ΔC_p . In order to increase binding affinity of a ligand to a protein of interest, for example, in the rational drug design, the ideal optimization strategy requires refining the enthalpic or entropic contributions to result in a minimal entropic or enthalpic penalty but induce the largest lowering in ΔG , thereby defeating the deleterious effects of enthalpy-entropy compensation at the thermodynamic level.

5.4.3 Enthalpy-driven, entropy-driven or their combination?

Although the entropy-enthalpy compensation weakens the effect on lowering the total free energy of binding, it is important to distinguish whether a binding process is a entropy- or enthalpy-driven one in fields of the medicinal chemistry and drug design because this will facilitate the understanding of the binding mechanisms and help to improve binding affinity or specificity by modifying the acceptor or ligand.

In the case of the simple lock-and-key binding model, we speculate that the entropy change would make a substantial contribution to the binding free energy due to the structural rigidity and the perfect surface complementary between the acceptor and ligand. The large positive ΔS_{solv} is gained when water molecules, which in the unliganded acceptor form a well-defined network in ligand-binding pockets or cavities, are displaced upon ligand binding. Intuitively, such an ideal conformationally selective binding between rigid partners results in only very little loss in ΔS_{conf} and $\Delta S_{r/t}$ since there is no large backbone conformation change and only some interacting side chains are restricted, thereby resulting in a large positive ΔS_{tot} . The subsequent step of binding is the formation of some noncovalent bonds between the acceptor and ligand. Because of the lack of large conformational adjustment, it seems that only a small amount of enthalpy is required to compensate for the minor entropy cost. Nevertheless, the types and strength of the bonds formed between the protein and ligand contribute to the binding affinity. Our speculation is in line with the recent study by Odriozola and coworkers (Odriozola et al., 2008), who showed that the key-lock assembly is solely driven by the entropic contribution even in the absence of attractive forces, pointing out the importance of solvent contribution in the underlying mechanisms of substrate-protein assembly processes.

In the case of the induced fit model, the initial contact between the interacting partners exclude limited amount of water molecules compared to that in lock-and-key model due to the lack of surface complementary between the protein and ligand. Therefore, the ΔS_{solv} is smaller than that in the lock-and-key model. In order to achieve a high binding affinity, a full contact between the interacting partners should be established, and this is accompanied by the conformational adjustment and ordering of the interacting regions and as thus, the large loss in ΔS_{conf} . It is possible that for the induced fit binding, the solvent entropy gain can not overcome the loss in the conformational and rotational and translational entropy terms, thus resulting in a negative ΔS_{tot} (equation 8). Such an entropy cost must be compensated for by the large negative enthalpy arising from the formation of extensive hydrogen bonding, electrostatic and van der Waals interactions. Taken together, we consider that the induced fit binding is an enthalpy-driven process because the dominant driving force is the large negative enthalpy, which compensates for the entropy penalty and renders the binding thermodynamically favorable.

In the case of the conformational selection model, it appears that the hydrophobic interaction makes a large contribution to lowering binding free energy, just like the situation in the lock-and-key model, which, as described above (section 5.2), is the extreme of the conformational selection mechanism. Theoretically, the step of selective binding of ligand to a structurally favorable state of the receptor is an optimal strategy since it maximizes the ΔS_{solv} while minimizing the cost of the inevitable entropy penalty arising from the ΔS_{conf} and $\Delta S_{\text{r/t}}$ to achieve a as large as possible ΔS_{tot} . Therefore, the conformational selection plays a role in lowering the entropic barrier (Lange et al., 2008; Perica & Chothia, 2010) and contributes to the binding specificity. However, because of the flexibility nature of proteins, especially in the ligand-binding regions, the binding will have to overcome the inescapable entropic penalties. The subsequent step of conformational adaption, which is driven by the enthalpic effect of bond formation, contributes not only to further lowering the free energy, but also to binding affinity. As discussed in section 5.2, the conformational selection mechanism is actually a synthesis of the lock-and-key and induced fit models, and therefore the different stages of molecule recognition can be driven either by the entropy or enthalpy, both making variable contributions to lowering the free energy of binding. The essence of the conformational selection is a combinatorial effect of enthalpy and entropy, which not only guarantees simultaneously the binding affinity and specificity of a single or multiple ligands but also accommodates mutations through shift in the dynamic energy landscape and therefore, is evolutionarily advantageous (Ma et al., 2001).

5.4.4 Case studies on entropy-, enthalpy-, and combination-driven bindings

The thermodynamic principle of enthalpy-entropy compensation (Chaires, 2008) has promoted the Freire group to optimize HIV protease inhibitor binding through taking into account the thermodynamics of the binding interactions (Ohtaka & Freire, 2005; Ohtaka et al., 2004; Velazquez-Campoy et al., 2000a; Velazquez-Campoy et al., 2000b). The results show that, although the interaction of the enthalpically or entropically optimized inhibitors with the target protease may have similar free energies, these inhibitors might have different specificities or different pharmacological properties.

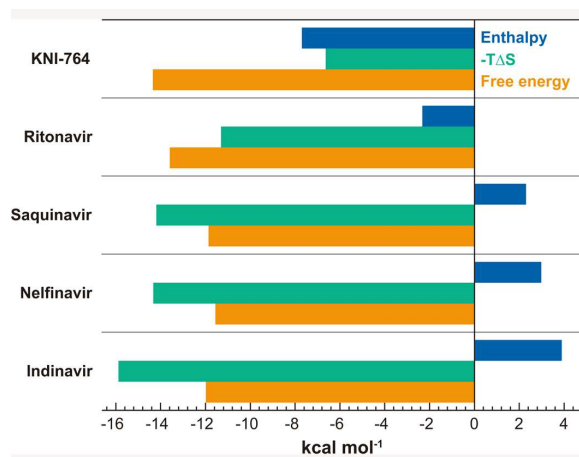


Fig. 7. Thermodynamic profiles of the bindings of the five inhibitors to HIV protease. Free energy, enthalpy and $-T\Delta S$ are shown as orange, blue and green horizontal bars, respectively. This figure is cited from (Chaires, 2008).

Figure 7 shows the thermodynamic parameters of the binding of several inhibitors to HIV protease. For the first-generation inhibitors, i.e., indinavir, nelfinavir, saquinavir, and ritonavir, the binding is entropically driven due to the predominantly positive entropy gain. Since these inhibitors are highly structurally constrained and the protease itself is relatively rigid, the ΔS_{solv} arising from the hydrophobic effect makes a substantial contribution to the binding entropy (ΔS_{tot}). Interestingly, for three (indinavir, nelfinavir and saquinavir in Figure 7) of the four inhibitors, their binding is accompanied by an unfavorable positive enthalpy change. Although such an enthalpy cost can be compensated for by the large entropy gain, the increase in enthalpy is detrimental to binding affinity, as shown by the relatively higher free energy compared to that for ritonavir, whose binding is accompanied by a small favorable negative enthalpy change. In the case of a newer-generation inhibitor, KNI-764, its binding is tighter than bindings of the first-generation inhibitors to the HIV protease by at least 1 kcal mol^{-1} . However, as shown in Figure 7, such a lower free energy is the outcome of the almost equal favorable enthalpic and entropic contributions, suggesting that its binding is driven by both the hydrophobic (entropy effect) and favorable intermolecular forces (enthalpy contribution). Such a binding process may be a conformational selection followed by an induced fit, in line with the extended conformational selection mechanism proposed by (Csermely et al., 2010). The bindings of a series of inhibitors to the HIV protease, which are characterized by being from the pure entropy-driven to entropy-driven plus weak enthalpic contribution to the combinatorial entropy- and enthalpy-driven, suggest that the extremely high binding affinity requires a favorable binding enthalpy, and therefore the enthalpic optimization is one of the most important principles in drug design (Chaires, 2008; Ohtaka et al., 2004). It should be noted that the optimal binding enthalpy does not correlate simply with the number of interactions, but instead reflects mostly the strength and quality of the noncovalent interactions (such as hydrogen bonding, electrostatic and van der Waals interactions) of the inhibitor with the protease.

Temp. (°C)	TK+dT ^a			TK:dT+ATP ^b			TK+dT/ATP ^c		
	$\Delta G^{d,e}$	ΔH^d	$T\Delta S^d$	$\Delta G^{d,e}$	ΔH^d	$T\Delta S^d$	$\Delta G^{d,e}$	ΔH^d	$T\Delta S^d$
10	-7.1	-13.6	-6.5	-8.9	-11.7	-2.8	-18.9	-25.4	-6.5
15	-7.0	-15.8	-8.8	-8.9	-12.1	-3.2	-18.9	-27.9	-9.0
20	-7.1	-17.4	-10.2	-9.1	-13.0	-3.9	-17.2	-30.5	-13.3
25	-7.2	-19.1	-11.9	-9.0	-13.8	-4.8	-16.8	-33.1	-16.3

^a TK was titrated with the dT.

^b TK:dT complex was titrated with the ATP.

^c TK was titrated with a 1:1 mixture of dT and ATP.

^d The unit was in kcal mol⁻¹.

^e ΔG was calculated from equation 1, where the K_b is the binding constant determined by ITC.

Table 1. Thermodynamic parameters of binding of thymidine (dT) and ATP to HSV1 TK at pH 7.5. Data in this table are cited from (Perozzo et al., 2004).

Thermodynamic parameters of the binding of thymidine (dT, which is the substrate) and adenosine triphosphate (ATP, which is a cofactor) to the thymidine kinases (TK) from Herpes simplex virus type-1 (HSV1) were determined using the ITC method (Perozzo et al., 2004; Perozzo et al., 2000; Pilger et al., 1999). In the ternary complex TK:dT:ATP, dT and ATP are located in separate, well defined binding pockets of the enzyme. Table 1 lists the thermodynamic parameters for binding of dT and ATP to TK. The "TK+dT" indicates that the TK is titrated with the dT; the "TK:dT+ATP" indicates that using the ATP to titrate the TK:dT binary complex; and the "TK+dT/ATP" indicates that TK is titrated with a 1:1 mixture of dT and ATP. All these three groups of binding reactions at a temperature range of 10-25 °C are an exothermic reaction driven by a large negative change in enthalpy, which is favorable enough to compensate for the unfavorable decrease in the entropy of the system. As expected, the group "TK+dT/ATP" yields a heat change approximately equals the sum of the heat changes for groups "TK+dT" and "TK:dT+ATP". An interesting phenomenon is that the ΔH and $T\Delta S$ of all these three reactions depend strongly on the temperature, e.g., their values decrease when temperature increases, while ΔG is almost insensitive to temperature due to the entropy-enthalpy compensation. The high flexibility caused by high temperature facilitates mutual conformational adaption between interacting partners, thereby resulting in increased bonds and as such, more negative ΔH , which is counterbalanced by more negative ΔS caused by increased order of binding partner after binding. The ΔS_{tot} of the three groups of reactions at 25 °C are further decomposed into ΔS_{solv} , ΔS_{conf} and $\Delta S_{r/t}$ and are listed in Table 2. It is evident that the favorable positive ΔS_{solv} is overcompensated by the large unfavorable negative ΔS_{conf} . The contribution by $\Delta S_{r/t}$ although unfavorable, is one or two orders of magnitude smaller than the other entropy terms. Interestingly, the sum of the decomposed entropy terms of "TK+dT" and "TK:dT+ATP" is approximately equal to the corresponding change in entropy terms of "TK+dT/ATP". The large positive solvent entropy ΔS_{solv} originates from the water release upon binding. The larger the interacting interfaces or binding cavities are, the more ordered water molecules will be displaced, and the larger the positive ΔS_{solv} will be gained. For example, binding of the larger dT to TK yields a larger positive ΔS_{solv} (92 cal K⁻¹ mol⁻¹) than binding of relatively smaller ATP to TK (36 cal K⁻¹ mol⁻¹). Similarly, the magnitude of the ΔS_{conf} is related to the area of the rigidified interfaces and the number of the fixed side chains/main chains upon binding. The large negative ΔS_{tot} needs a larger negative ΔH to

overcompensate for to lower the binding free energy, suggesting an induced fit binding driven mainly by intermolecular forces. This is in line with the observation that substrate binding to HSV1 TK leads to a conformational closing of the binding sites, bringing thymidine and ATP into an orientation appropriate for catalysis (Perozzo et al., 2004).

Reaction	ΔS_{tot} (cal K ⁻¹ mol ⁻¹)	ΔS_{solv} (cal K ⁻¹ mol ⁻¹)	ΔS_{conf} (cal K ⁻¹ mol ⁻¹)	$\Delta S_{\text{r/t}}$ (cal K ⁻¹ mol ⁻¹)
TK+dT	-39.9	92.0	-124.0	-8.0
TK:dT+ATP	-16.1	36.0	-44.0	-8.0
TK+dT/ATP	-54.7	133.0	-172.0	-16.0

Table 2. Decomposition of total entropy changes ΔS_{tot} into ΔS_{solv} , ΔS_{conf} and $\Delta S_{\text{r/t}}$ for substrate binding to HSV1 TK at 25 °C. Data in this table are cited from (Perozzo et al., 2004).

The difference in the driving force between induced fit/binding-induced folding and conformational selection/constitutive binding can be illustrated by studies on the binding of the kinase-inducible activation (KID) domain of the cAMP response element-binding protein (CREB) to the KIX domain (residues 586-672) of CREB binding protein (CBP), and binding of the proto-oncogene c-Myb to the KIX domain of CBP (Wright & Dyson, 1999). The KID domain binds with high affinity to the KIX domain only when the Ser133 is phosphorylated to form the pKID. The pKID domain is observed to be disordered in its free state but folds into a pair of orthogonal helices upon binding to the KIX domain. Such a binding is enthalpy-driven and entropically disfavored, with the entropic and enthalpic contributions being -6.0 cal K⁻¹ mol⁻¹ and -10.6 kcal mol⁻¹, respectively. The entropic penalty of the folding transition (from coil to helix) is compensated for by the large negative enthalpy of binding, attributed partly to the complementary intermolecular hydrogen bonds formed between the phosphor-Ser133 in KID and the Tyr and Lys side chains in the target protein KIX (Radhakrishnan et al., 1997). In the unphosphorylated state, binding of KID to KIX is very weak (Wright & Dyson, 1999) since the smaller enthalpy of binding (lacking the large number of hydrogen bonds formed by phosphor-Ser133) can not compensate for the entropic cost of the folding transition. Here the binding-induced folding is an extreme case of induced fit. In order to guarantee the binding affinity, there must be strong enough intermolecular interactions to stabilize the folded structure in the complex. In contrast to pKID, the activation domain of the proto-oncogene c-Myb binds constitutively, i.e. without requiring activation by phosphorylation, to the same site on the KIX domain as pKID does. Unlike pKID, the c-Myb activation domain can spontaneously adopt the helical conformation, even as an isolated peptide (Parker et al., 1999). Binding to KIX is now entropically favored ($\Delta S = 7.5$ cal mol⁻¹ K⁻¹), but the enthalpy change ($\Delta H = -4.1$ kcal mol⁻¹) is smaller than for pKID, presumably because the hydrogen bonding interactions made by the phosphoryl moiety of pKID are absent for c-Myb (Parker et al., 1999). Therefore, such a binding can be explained by the conformational selection mechanism driven mainly by the hydrophobic force. Furthermore, the negative binding enthalpy, although smaller than that of pKID binding, is able to consolidate binding affinity. Taken together, we conclude that the binding of different activation domains to the same site of the target protein can either be enthalpy- or entropy-driven, depending on the initial structure and the flexibility of the activation domains. Moreover, in the case of binding-induced folding, certain structural

factors capable of stabilizing the folded structure in the complex are required, i.e., the enthalpic contribution must be favorable enough to compensate for the large entropic loss.

6. Energy landscape contains both the folding and binding funnels

Protein folding and binding are essentially similar processes because of their common essence: the recognition and organization/reorganization of amino acids, either within a single protein or between different proteins. The sole difference between folding and binding is the presence and absence of the chain connectivity between their components, leading to two different terms, i.e., “the intramolecular” and “the intermolecular” recognition (Ma et al., 1999; Tsai et al., 1998). Most studies generally focus on either one process or the other: studies of protein folding focus merely on a single polypeptide; studies of protein binding involve at least two chains. In this section we will show how protein folding and binding are synthesized together by the energy landscape theory to form a common funnel model.

Under the background of statistical mechanical theory, both the protein folding and binding are driven by a decrease in total Gibbs free energy, which is dictated by a mechanism of the delicate balance of opposing effect, i.e., the entropy-enthalpy compensation. The requirement for lowering free energy while reducing conformational space determines that the energy landscape of a protein folding must be funnel-shaped. Analogously, the spontaneous protein-ligand association also lowers the free energy of the system composed of protein, ligand and solvent, while reducing the conformational entropy and rotational and translational entropy of the protein and ligand. Thus the energy landscape of the protein-ligand binding should also be funnel-like, which we term protein binding funnel.

Two hypotheses can be used to illustrate the similarity of the protein folding and binding: i) the hierarchical building block model (Ma et al., 1999; Tsai et al., 1999a) and ii) the folding funnel model. In the building block model, protein folding is viewed as the consequence of a combinatorial assembly of a set of transient “building blocks”. The formation of any building block in a given sequence can be described and guided by a microfunnel-like energy landscape. The mutual recognition between building blocks resembles a fusion of two microfunnel-like landscapes. A compact, stable hydrophobic folding unit is located at the bottom of a subfunnel-like landscape. Such a hydrophobic folding unit, in turn, serves as the basic unit in building a functional single domain or multi-domain protein, an oligomer, or a functional complex. The entire process may be seen as a sequential fusion and modification of these individual microfunnels. In the case of protein binding, the folded monomers for protein-protein association correspond to the building blocks in protein folding. In both cases, structural entities associate with each other and this process is driven by fusing two funnels into a higher dimensional funnel, regardless of the chain connectivity. Therefore, the essence of the hierarchical building block model is that a series of microfunnel fusion events lower the Gibbs free energy of the system and finally achieve a global free energy minimal state or protein-ligand complex.

In the folding funnel model, the first step is the fast hydrophobic collapse driven by the entropy. Concretely, the water molecules squeeze the peptide chain to collapse into molten globule intermediates and exclude themselves from the interior of the collapsed protein, resulting in the maximum of the solvent entropy. At the same time, the heat energy is required to strip the water molecules from the polypeptide surface, leading to an

endothermic reaction and therefore, an enthalpy increase. However, in order to lower the free energy of the system, the increase in entropy must be large enough to overcompensate for the increase in enthalpy (see equation 2), putting this step under entropy control (Griko, 2000). In the molten globule intermediates, some of the native secondary structures may have been formed, while many native contacts, or close residue-residue interactions present in the native state, have yet to form. Therefore, the subsequent step of protein folding is to establish these native inter-atomic contacts through conformational adjustment and adaptation, which may be a slow process because of the many times trial attempts of bond breaking and formation. Finally, the formed tertiary native interactions reduce the conformational flexibility, leading to the loss of the conformational entropy of the protein. Although the native state has lower entropy than the intermediate molten globule, this step also leads to the formation of a large number of favorable specific interactions and therefore, the large negative enthalpy, putting it under enthalpy control. Interestingly, the formation of the protein-protein or protein-ligand complex is similar to the protein folding process. The initial contact of the protein and its ligand will inevitably displace water molecules around the interfaces of both acceptor and ligand. Therefore, this step is driven by the increase of the solvent entropy in a way similar to the hydrophobic collapse of protein folding. An ideal "hydrophobic collapse" may occur for the lock-and-key and to a lesser extent, the conformational selection processes because the interacting interfaces are generally hydrophobic and match well between the protein and ligand. This process can be considered to occur at the bottom of the relatively smooth folding funnel (in the case of lock-and-key model for rigid proteins) or the bottom of the tier-0 wells/microfunnels of the rugged energy landscape (in the case of conformational selection model; see Figure 4). Although the initial trial collisions and contacts may be time-consuming, the correct orientation can lead to a rapid change of the shape of funnel or microfunnel, i.e., the downward extension of the bottom of the funnel/microfunnel driven by the entropy gain of the system. Similar to what occurs to the protein folding process, the next step is the conformational adjustment and mutual adaptation between interacting partners driven by the negative enthalpy, resulting in the final formation of favorable specific bonds. This step further downwards extends and narrows the binding funnel, shifting the equilibrium towards the bound state. The induced fit may occur along the very rugged floor of the folding funnel that includes many free energy minimum wells (i.e., the tier-1 or tier-2 wells as described in section 4.4.2) separated by small energy barriers, thereby allowing easy conformational transition between different substates. This is in agreement with the hypothesis that the fast conformational transition rate is beneficial to induced fit binding (Zhou, 2010). The fast and easy conformational transition may increase the probability of the conformational substates appropriate for ligand binding. As argued by (Tsai et al., 2001), these conformational substates, although having a low population time, still have higher concentrations than the other substates, thus facilitating the selective binding of the ligand. This also suggests that the conformational selection process may play a relatively important role in the induced fit binding. Although the initial solvent entropy gain is small and contributes little to lowering the free energy, such an entropy change triggers the deepening of the small binding funnel, allowing the dominant driving force, the negative enthalpy, to further lower the binding free energy. In this process, the entropy penalty arising from the ordering of interacting partners is counterbalanced by the large negative enthalpy arising

from bond formations, analogous to the enthalpy-driven conformational adjustment step occurring in the folding funnel. Taken together, although different proteins may have different landscapes around the bottom of their folding funnels (e.g., smooth, rugged or very rugged floors), the binding event occurring in one of the energy wells can further extend and deepen the single well into a binding funnel. Binding is also driven by the decrease in total Gibbs free energy through the entropy-enthalpy compensation, with different steps being dominated by either the entropy or enthalpy. If the ligand is a protein, the binding funnel can be viewed as the consequence of fusion of two individual microfunnels or wells; if the ligand is a small compound, the binding funnel can be viewed as extension and deepening of a specific energy well located at the bottom of the protein folding funnel.

7. Conclusion

We conclude that the protein folding and binding are similar processes that are commonly driven by the decrease in total Gibbs free energy of the system, which is dictated by a mechanism of the delicate balance of opposing effects involving entropy and enthalpy contributions, i.e., the effect of entropy-enthalpy compensation. Such a mechanism is described vividly by the free energy landscape theory, which characterizes the protein folding as going down the minimally frustrated, funnel-like energy landscape via multiple pathways towards native states located at the bottom of the funnel, and the protein binding as fusion of the microfunnels/energy wells or as downwards extending the free energy well into a binding funnel at the bottom of the folding funnel. Furthermore, the extent of the ruggedness of the energy landscape, which governs the dynamic behavior of the protein, is also dictated by the entropy-enthalpy compensation. Therefore, the protein folding, binding and its dynamic behavior are all essentially thermodynamically controlled processes that are governed by entropy-enthalpy compensation. The tendency to maximize the entropy of the protein-solvent system is the original driving force for protein folding, binding and dynamics, while the enthalpy change, an opposing force that tends to drive the system to become ordered, can compensate for the entropy change to ultimately allow the system to reach equilibrium at the free energy minimization, either global or local. Therefore, the description of protein folding, binding and dynamics by the leveling free-energy landscape is consistent with the second law of thermodynamics (Sharma et al., 2009). This law, generally understood as the entropy maximum principle, dictates and underlies not only the protein folding and binding, but also other life phenomena (Erwin, 1944) such as the origin of life (Doolittle, 1984; L. James, 1979), biological evolution (John, 2003), cancer occurrence (Ji & Liu, 2011b), among others. Interestingly, entropy maximum principle or the thermodynamics second law can also be used to explain and describe many of the economic and social phenomena. Summarily, a deep understanding of the thermodynamic entropy-enthalpy compensation based on the dynamic energy landscape view can not only help to understand the nature of the forces that drive protein folding and complex formation, but also facilitate the function, protein engineering and drug design studies in the post-genomic era.

8. Acknowledgement

We thank Ms. Sara A. Barton for her careful reading and useful comments. Research described was supported by grants from National Natural Science Foundation of China (No.

31160181 and 30860011) and Yunnan province (2007PY-22 and 2011CI123), and by the foundation for Key Teacher of Yunnan University.

9. References

- Agashe, V.R.; Shastry, M.C.R. & Udgaonkar, J.B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature*, 377, 754-757.
- Amadei, A.; Linssen, A.B.M. & Berendsen, H.J.C. (1993). Essential dynamics of proteins. *Proteins: Struct Funct Genet*, 17, 412-425.
- Aman, M.J.; Karauzum, H.; Bowden, M.G. & Nguyen, T.L. (2010). Structural model of the pre-pore ring-like structure of panton-valentine leukocidin: providing dimensionality to biophysical and mutational data. *J Biomol Struct Dyn*, 28, 1-12.
- Amzel, L.M. (1997). Loss of translational entropy in binding, folding, and catalysis. *Proteins: Struct Funct Genet* 28, 144-149.
- Amzel, L.M. (2000). Calculation of entropy changes in biological processes: folding, binding, and oligomerization. *Methods Enzymol*, 323, 167-177.
- Anfinsen, C.B.; Haber, E.; Sela, M. & White, F.H.J. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA*, 47, 1309-1314.
- Anfinsen, C.B. & Scheraga, H.A. (1975). Experimental and theoretical aspects of protein folding. *Adv Protein Chem*, 29, 205-300.
- Ansari, A.; Berendzen, J.; Bowne, S.F.; Frauenfelder, H.; Iben, I.E.T.; Sauke, T.B.; Shyamsunder, E. & Young, R.D. (1985). Protein states and protein quakes. *Proc Natl Acad Sci USA*, 82, 5000-5004.
- Austin, R.H.; Beeson, K.W.; Eisenstein, L.; Frauenfelder, H. & Gunsalus, I.C. (1975). Dynamics of ligand binding to myoglobin. *Biochemistry*, 14, 5355-5373.
- Baker, D. & Sali, A. (2001). Protein structure prediction and structural genomics. *Science*, 294, 93-96.
- Berendsen, H.J.C. (2011). The relevance of distance statistics for protein folding. *J Biomol Struct Dyn*, 28, 599-602.
- Berger, C.; Weber-Bornhauser, S.; Eggenberger, J.; Hanes, J.; Pluckthun, A. & Bosshard, H.R. (1999). Antigen recognition by conformational selection. *FEBS Lett*, 450, 149-153.
- Boehr, D.D.; Nussinov, R. & Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol*, 5, 789-796.
- Bosshard, H.R. (2001). Molecular recognition by induced fit: How fit is the concept? *News Physiol Sci*, 16, 171-173.
- Bowie, J.U. (2005). Solving the membrane protein folding problem. *Nature*, 438, 581-589.
- Brady, G.P. & Sharp, K.A. (1997). Entropy in protein folding and in protein-protein interactions. *Curr Opin Struct Biol*, 7, 215-221.
- Bryngelson, J.D.; Onuchic, J.N.; Socci, N.D. & Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein Folding: A synthesis. *Proteins: Struct Funct Genet*, 21, 167-195.
- Chaires, J.B. (2008). Calorimetry and thermodynamics in drug design. *Annu Rev Biophys*, 37, 135-151.

- Chan, H.S. (2011). Short-range contact preferences and long-range indifference: Is protein folding stoichiometry driven? *J Biomol Struct Dyn*, 28, 603-606.
- Chen, A. & Wadso, I. (1982). Simultaneous determination of delta G, delta H and delta S by an automatic microcalorimetric titration technique: application to protein ligand binding. *J Biochem Biophys Meth*, 6, 307-316.
- Cooper, A. & Johnson, C.M. (1994). Introduction to microcalorimetry and biomolecular energetics. *Methods Mol Biol*, 22, 109-124.
- Csermely, P.; Palotai, R. & Nussinov, R. (2010). Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem Sci* 35, 539-546.
- Demchenko, A.P. (2001). Recognition between flexible protein molecules: induced and assisted folding. *J Mol Recognit*, 14, 42-61.
- Dill, K.A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry*, 24, 1501-1509.
- Dill, K.A.; Fiebig, K.M. & Chan, H.S. (1993). Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci USA*, 90, 1942-1946.
- Dill, K.A.; Bromberg, S.; Yue, K.; Fiebig, K.M.; Yee, D.P.; Thomas, P.D. & Chan, H.S. (1995). Principles of protein folding - A perspective from simple exact models. *Protein Sci*, 4, 561-602.
- Dill, K.A. & Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* 4, 10-19.
- Dill, K.A. (1999). Polymer principles and protein folding. *Protein Sci*, 8, 1166-1180.
- Dill, K.A.; Ozkan, S.B.; Weikl, T.R.; Chodera, J.D. & Voelz, V.A. (2007). The protein folding problem: when will it be solved? *Curr Opin Struct Biol*, 17, 342-346.
- Dobson, C.M. (2000). The nature and significance of protein folding. In: *Mechanisms of Protein Folding (2nd ed.)*, R.H. Pain, (Ed.), Oxford University Press, Oxford, UK.
- Doolittle, R. (1984). The probability and origin of life. In: *Scientists Confront Creationism*, L.R. Godfrey, (Ed.), pp. 85.
- Duan, Y. & Kollman, P.A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282, 740-744.
- Dunitz, J.D. (1995). Win, some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem Biol*, 2, 709-712.
- Eftink, M.R.; Anusiem, A.C. & Biltonen, R.L. (1983). Enthalpy-entropy compensation and heat capacity changes for protein-ligand interactions: general thermodynamic models and data for the binding of nucleotides to ribonuclease A. *Biochemistry*, 22, 3884-3896.
- Erwin, S. (1944). *What is life - the physical aspect of the living cell*, Cambridge University Press, Cambridge.
- Fischer, E. (1894). Einfluss der configuration auf die wirkung der enzyme. *Ber Dtsch Chem Ges*, 27, , 2984-2993.
- Foote, J. & Milstein, C. (1994). Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci USA*, 91, 10370-10374.
- Frauenfelder, H.; Petsko, G.A. & Tsernoglou, D. (1979). Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature*, 280, 558-563.

- Frauenfelder, H.; Sligar, S.G. & Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science*, 254, 1598-1603.
- Freire, E.; Mayorga, O.L. & Straume, M. (1990). Isothermal titration calorimetry. *Anal Chem*, 62, 950A-959A.
- Freire, E. (1999). The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc Natl Acad Sci USA*, 96, 10118-10122.
- Galzitskaya, O.V.; Lobanov, M.Y. & Finkelstein, A.V. (2011). Cunning simplicity of a stoichiometry driven protein folding thesis. *J Biomol Struct Dyn*, 28, 595-598.
- Gibbs, J.W. (1873). A method of geometrical representation of the thermodynamic properties of substances by means of surfaces. *Trans Conn Acad Arts Sci*, 2, 382-404.
- Gilli, P.; Ferretti, V.; Gilli, G. & Borea, P.A. (1994). Enthalpy-entropy compensation in drug receptor binding. *J Phys Chem B*, 98, 1515-1518.
- Griko, Y.V. (2000). Energetic basis of structural stability in the molten globule state: a-lactalbumin. *J Mol Biol*, 297, 1259-1268.
- Gruebele, M. (2005). Downhill protein folding: evolution meets physics. *C R Biol*, 328, 701-712.
- Gulukota, K. & Wolynes, P. (1994). Statistical mechanics of kinetic proofreading in protein folding *in vivo*. *Proc Natl Acad Sci USA*, 91, 9292-9296.
- Hammes, G.G.; Chang, Y.C. & Oas, T.G. (2009). Conformational selection or induced fit: A flux description of reaction mechanism. *Proc Natl Acad Sci USA*, 106, 13737-13741.
- Harrison, S.C. & Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA*, 82, 4028-4030.
- Henzler-Wildman, K.A. & Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450, 964-972.
- Huang, G.S. & Oas, T.G. (1995). Structure and stability of monomeric λ repressor: NMR evidence for two-state folding. *Biochemistry* 34, 3884-3892.
- Inouye, H. & Kirschner, D.A. (1997). X-ray diffraction analysis of scrapie prion: intermediate and folded structures in a peptide containing two putative α -helices. *J Mol Biol*, 268, 375-389.
- Itzhaki, L.S.; Otzen, D.E. & Fersht, A.R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol*, 254, 260-288.
- Jackson, S.E. & Fersht, A.R. (1991). Folding of chymotrypsin inhibitor 2.1. Evidence for a two-state transition. *Biochemistry*, 30, 10428-10435.
- Jacob, J.; Schindler, T.; Balbach, J. & Schmid, F.X. (1997). Diffusion control in an elementary protein folding reaction. *Proc Natl Acad Sci USA*, 94, 5622-5627.
- James, L. (1979). *GAIA - A new look at life on earth*, Oxford University Press, Oxford.
- James, L.C.; Roversi, P. & Tawfik, D.S. (2003). Antibody multispecificity mediated by conformational diversity. *Science*, 299, 1362-1367.
- Janin, J. (1995). Protein-protein recognition. *Prog Biophys Mol Biol*, 64, 145-166.
- Ji, X.L. & Liu, S.Q. (2011a). Is stoichiometry-driven protein folding getting out of thermodynamic control? *J Biomol Struct Dyn*, 28, 621-623.

- Ji, X.L. & Liu, S.Q. (2011b). Thinking into mechanism of protein folding and molecular binding. *J Biomol Struct Dyn*, 28, 995-996.
- John, A. (2003). *Information theory and evolution*, World Scientific, New Jersey.
- Kapon, R.; Nevo, R. & Reich, Z. (2008). Protein energy landscape roughness. *Biochem Soc Trans*, 36, 1404-1408.
- Karplus, M. & Weaver, D.L. (1976). Protein-folding dynamics. *Nature*, 260, 404-406.
- Karplus, M. & Weaver, D.L. (1994). Protein folding dynamics: the diffusion - collision model and experimental data. *Protein Sci*, 3, 650-668.
- Kauzmann, W. (1954). Denaturation of proteins and enzymes. In: *The mechanism of enzyme reaction*, W.D. McElroy, & B. Glass, (Eds.), pp. 70-120, Johns Hopkins Press, Baltimore.
- Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv Protein Chem*, 14, 1-63.
- Kim, P.S. & Baldwin, R.L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu Rev Biochem*, 51, 459-489.
- Koshland, D.E.J. (1958). Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA*, 44, 98-104.
- Kosik, K.S. (1992). Alzheimer's disease: a cell biological perspective. *Science*, 256, 780-783.
- Kubelka, J.; Hofrichter, J. & Eaton, W.A. (2004). The protein folding "speed limit". *Curr Opin Struct Biol*, 14, 76-88.
- Lakomek, N.A.; Lange, O.F.; Walter, K.F.; Farès, C.; Egger, D.; Lunkenheimer, P.; Meiler, J.; Grubmüller, H.; Becker, S.; de Groot, B.L. & Griesinger, C. (2008). Residual dipolar couplings as a tool to study molecular recognition of ubiquitin. *Biochem Soc Trans*, 36, 1433-1437.
- Lange, O.F.; Lakomek, N.A.; Fares, C.; Schroder, G.F.; Walter, K.F.A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C. & de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*, 320, 1471-1475.
- Leopold, P.E.; Montal, M. & Onuchic, J.N. (1992). Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci USA*, 89, 8721-8725.
- Levinthal, C. (1968). Are there pathways for protein folding? *J Chim Phys*, 65, 44-45.
- Levinthal, C. (1969). How to fold graciously. In: *Mossbauer spectroscopy in biological systems*, P. Debrunner, J. Tsibris, & E. Munck, (Eds.), pp. 22-24, University of Illinois Press, Urbana.
- Liu, S.Q.; Fu, Y.X. & Liu, C.Q. (2007a). Molecular motions and conformational transition between different conformational states of HIV-1 gp120 envelope glycoprotein. *Chin Sci Bull*, 52, 3074-3088.
- Liu, S.Q.; Liu, C.Q. & Fu, Y.X. (2007b). Molecular motions in HIV-1 gp120 mutants reveal their preferences for different conformations. *J Mol Graphics Model*, 26, 306-318.
- Liu, S.Q.; Liu, C.Q. & Fu, Y.X. (2007c). Molecular motions in HIV-1 gp120 mutants reveal their preferences for different conformations. *J Mol Graphics Modell*, 26, 306-318.

- Liu, S.Q.; Liu, S.X. & Fu, Y.X. (2007d). Dynamic domains and geometrical properties of HIV-1 gp120 during conformational changes induced by CD4-binding. *J Mol Model*, 13, 411-424.
- Liu, S.Q.; Liu, S.X. & Fu, Y.X. (2008). Molecular motions of human HIV-1 gp120 envelope glycoproteins. *J Mol Model*, 14, 857-870.
- Liu, S.Q.; Meng, Z.H.; Fu, Y.X. & Zhang, K.Q. (2010). Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. *J Mol Model*, 16, 17-28.
- Liu, S.Q.; Meng, Z.H.; Fu, Y.X. & Zhang, K.Q. (2011). The effect of calciums on the molecular motions of proteinase K. *J Mol Model*, 17, 289-300.
- Lum, K.; Chandler, D. & Weeks, J.D. (1999). Hydrophobicity at small and large length scales. *J Phys Chem B*, 103, 4570-4577.
- Lumry, R. & Rajender, S. (1970). Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers*, 9, 1125-1227.
- Ma, B.; Kumar, S.; Tsai, C.J. & Nussinov, R. (1999). Folding funnels and binding mechanisms. *Protein Eng*, 12, 713-720.
- Ma, B.; Wolfson, H.J. & Nussinov, R. (2001). Protein functional epitopes: hot spots, dynamics and combinatorial libraries. *Curr Opin Struct Biol*, 11, 364-369.
- Magg, C.; Kubelka, J.; Holtermann, G.; Haas, E. & Schmid, F.X. (2006). Specificity of the initial collapse in the folding of the cold shock protein. *J Mol Biol*, 360, 1067-1080.
- Maity, H.; Maity, M.; Krishna, M.M.; Mayne, L. & Englander, S.W. (2005). Protein folding: the stepwise assembly of foldon units. *Proc Natl Acad Sci USA*, 102, 4741-4746.
- Matouschek, A.; Kellis, J.T.J.; Serrano, L. & Fersht, A.R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, 340, 122-126.
- Matthews, B.W. (2011). Stoichiometry versus hydrophobicity in protein folding. *J Biomol Struct Dyn*, 28, 589-591.
- Mittal, A.; Jayaram, B.; Shenoy, S.R. & Bawa, T.S. (2010). A Stoichiometry driven universal spatial organization of backbones of folded proteins: Are there Chargaff's rules for protein folding? *J Biomol Struct Dyn*, 28, 133-142.
- Mittal, A. & Jayaram, B. (2011a). Backbones of folded proteins reveal novel invariant amino-acid neighborhoods. *J Biomol Struct Dyn*, 28, 443-454.
- Mittal, A. & Jayaram, B. (2011b). The newest view on protein folding: stoichiometric and spatial unity in structural and functional diversity. *J Biomol Struct Dyn*, 28, 669-674.
- Monod, J.; Wyman, J. & Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. *J Mol Biol*, 12, 88-118.
- Moult, J.; Pedersen, J.T.; Judson, R. & Fidelis, K. (1995). A large-scale experiment to assess proteinstructure prediction methods. *Proteins: Struct Funct Genet*, 23, ii-iv.
- Noe, F.; Schutte, C.; Vanden-Eijnden, E.; Reich, L. & Weikl, T.R. (2009). Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc Natl Acad Sci USA*, 106, 19011-19016.
- Nolting, B. & Agard, D.A. (2008). How general is the nucleation-condensation mechanism? *Proteins*, 73, 754-764.

- Odrizola, G.; Jiménez-Ángeles, F. & Lozada-Cassou, M. (2008). Entropy driven key-lock assembly. *J Chem Phys*, 129, 111101-111104.
- Ohtaka, H.; Muzammil, S.; Schon, A.; Velazquez-Campoy, A.; Vega, S. & Freire, E. (2004). Thermodynamic rules for the design of high affinity HIV-1 protease inhibitors with adaptability to mutations and high selectivity towards unwanted targets. *Int J Biochem Cell Biol*, 36, 1787-1799.
- Ohtaka, H. & Freire, E. (2005). Adaptive inhibitors of the HIV-1 protease. *Prog Biophys Mol Biol*, 88, 193-208.
- Onuchic, J.N.; Wolynes, P.G.; Luthey-Schulten, Z. & Socci, N.D. (1995). Towards an outline of the topography of a realistic protein folding funnel. *Proc Natl Acad Sci USA*, 92, 3626-3630.
- Onuchic, J.N.; Luthey-Schulten, Z. & Wolynes, P.G. (1997). Theory of protein folding: The energy landscape perspective. *Ann Rev Phys Chem*, 48, 545-600.
- Onuchic, J.N. & Wolynes, P.G. (2004). Theory of protein folding. *Curr Opin Struct Biol*, 14, 70-75.
- Otlewski, J. & Apostoluk, W. (1997). Structural and energetic aspects of protein-protein recognition. *Acta Biochim Pol*, 44, 367-387.
- Ozkan, S.B.; Wu, G.A.; Chodera, J.D. & Dill, K.A. (2007). Protein folding by zipping and assembly. *Proc Natl Acad Sci USA*, 104, 11987-11992.
- Pan, K.M.; Baldwin, M.; Nguyen, J.; Gasset, M.; Serban, A.; Groth, D.; Mehlhorn, I.; Huang, Z.; Fletterick, R.J.; Cohen, F.E. & Prusiner, S.B. (1993). Conversion of α -helices into β -sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA*, 90(10926-10966).
- Parker, D.; Rivera, M.; Zor, T.; Henrion-Caude, A.; Radhakrishnan, I.; Kumar, A.; Shapiro, L.H.; Wright, P.E.; Montminy, M. & Brindle, P.K. (1999). Role of secondary structure in discrimination between constitutive and inducible activators. *Mol Cell Biol*, 19, 5601-5607.
- Pechmann, S.; Levy, E.D.; Tartaglia, G.G. & Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc Natl Acad Sci USA*, 106, 10159-10164.
- Perica, T. & Chothia, C. (2010). Ubiquitin - molecular mechanisms for recognition of different structures. *Curr Opin Struct Biol*, 20, 367-376.
- Perozzo, R.; Jelesarov, I.; Bosshard, H.R.; Folkers, G. & Scapozza, L. (2000). Compulsory order of substrate binding to Herpes simplex virus type 1 thymidine kinase. A calorimetric study. *J Biol Chem*, 275, 16139-16145.
- Perozzo, R.; Folkers, G. & Scapozza, L. (2004). Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *J Recept Signal Transduct Res*, 24, 1-52.
- Pieper, U.; Eswar, N.; Davis, F.P.; Braberg, H.; Madhusudhan, M.S.; Rossi, A.; Marti-Renom, M.; Karchin, R.; Webb, B.M.; Eramian, D.; Shen, M.Y.; Kelly, L.; Melo, F. & Sali, A. (2006). MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res*, 34, D291-295.

- Pilger, B.D.; Perozzo, R.; Alber, F.; Wurth, C.; Folkers, G. & Scapozza, L. (1999). Substrate diversity of Herpes simplex virus thymidine kinase-impact of the kinematics of the enzyme. *J Biol Chem*, 274, 31967-31973.
- Pitera, J.W. & Swope, W. (2003). Understanding folding and design: replica-exchange simulations of "Trp-cage" miniproteins. *Proc Natl Acad Sci USA*, 100, 7587-7592.
- Plaxco, K.W.; Simons, K.T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol*, 277, 985-994.
- Plotkin, S.; Wang, J. & Wolynes, P.G. (1996). Correlated energy landscape model for finite, random heteropolymers. *Phys Rev E Stat Nonlin Soft Matter Phys*, 53, 6271-6296.
- Rackovsky, S. & Schraga, H.A. (2011). On the information content of protein sequences. *J Biomol Struct Dyn*, 28, 593-594.
- Radhakrishnan, I.; Pe rez-Alvarado, G.C.; Parker, D.; Dyson, H.J.; Montminy, M.R. & Wright, P.E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: A model for activator:coactivator interactions. *Cell*, 91, 741-752.
- Sali, A. & Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234, 779-815.
- Sali, A.; Shakhnovich, E. & Karplus, M. (1994). Kinetics of protein folding. A lattice model study of the requirements for folding to the native state. *J Mol Biol*, 235, 1614-1636.
- Sarma, R.H. (2011). A conversation on protein folding. *J Biomol Struct Dyn*, 28, 587-588.
- Schindler, T.; Herrler, M.; Marahiel, M.A. & Schmid, F.X. (1995). Extremely rapid protein folding in the absence of intermediates. *Nat Struct Biol*, 2, 663-673.
- Schuler, B.; Lipman, E.A. & Eaton, W.A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, 419, 743-747.
- Serban, D.; Taraboulos, A.; DeArmond, S.J. & B., P.S. (1990). Rapid detection of Creutzfeldt-Jakob disease and scrapie prion proteins. *Neurology*, 40, 110-117.
- Sharma, V.; Kaila, V.R.I. & Annala, A. (2009). Protein folding as an evolutionary process. *Physica A*, 388, 851-862.
- Shea, J.E.; Onuchic, J.N. & Brooks, C.L.I. (1999). Exploring the origins of topological frustration: Design of a minimally frustrated model of fragment B of protein A. *Proc Natl Acad Sci USA*, 96, 12512-12517.
- Sklenovsk y, P. & Otyepka, M. (2010). *In silico* structural and functional analysis of fragments of the ankyrin repeat protein p18^{INK4c}. *J Biomol Struct Dyn*, 27, 521-539.
- Sosnick, T.R.; Mayne, L. & Englander, S.W. (1996). Molecular collapse: The rate-limiting step in two-state cytochrome *c* folding. *Proteins Struct Funct Genet*, 24, 413-426.
- Sosnick, T.R.; Dothager, R.S. & Krantz, B.A. (2004). Differences in the folding transition state of ubiquitin indicated by f and c analyses. *Proc Natl Acad Sci USA*, 101, 17377-17382.
- Soundararajan, V.; Raman, R.; Raguram, S.; Sasisekharan, V. & Sasisekharan, R. (2010). Atomic interaction networks in the core of protein domains and their native folds. *Plos One*, 5, e9391.
- Stillinger, F.H. (1973). Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *J Solution Chem*, 2, 141-158.

- Tao, Y.; Rao, Z.H. & Liu, S.Q. (2010). Insight derived from molecular dynamics simulation into substrate-induced changes in protein motions of proteinase K. *J Biomol Struct Dyn*, 28, 143-157.
- Tobi, D. & Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proc Natl Acad Sci USA*, 102, 18908-18913.
- Todd, M.J. & Freire, E. (1999). The effects of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins: Struct Funct Genet*, 36, 147-156.
- Tsai, C.J.; Xu, D. & Nussinov, R. (1998). Protein folding via binding and vice versa. *Fold Des*, 3, R71-R80.
- Tsai, C.J.; Kumar, S.; Ma, B. & Nussinov, R. (1999a). Folding funnels, binding funnels, and protein function. *Protein Sci*, 8, 1181-1190.
- Tsai, C.J.; Ma, B. & Nussinov, R. (1999b). Folding and binding cascades: Shifts in energy landscapes. *Proc Natl Acad Sci USA*, 96, 9970-9972.
- Tsai, C.J.; Ma, B.; Sham, Y.Y.; Kumar, S. & Nussinov, R. (2001). Structured disorder and conformational selection. *Proteins: Struct Funct Genet*, 44, 418-427.
- Velazquez-Campoy, A.; Luque, I.; Todd, M.J.; Milutinovich, M.; Kiso, Y. & Freire, E. (2000a). Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor. *Protein Sci*, 9, 1801-1809.
- Velazquez-Campoy, A.; Todd, M.J. & Freire, E. (2000b). HIV-1 protease inhibitors: enthalpic versus entropic optimization of the binding affinity. *Biochemistry*, 39, 2201-2207.
- Voelz, V.A.; Bowman, G.R.; Beauchamp, K. & Pande, V.S. (2010). Molecular simulation of ab Initio protein folding for a millisecond folder NTL9(1-39). *J Am Chem Soc*, 132, 1526-1528.
- Wales, D.J. (2003). *Energy Landscapes*, Cambridge University Press, Cambridge.
- Weber, P.C. & Salemme, F.R. (2003). Applications of calorimetric methods to drug discovery and the study of protein interactions. *Curr Opin Struct Biol*, 13, 115-121.
- Weinreb, P.H.; Zhen, W.; Poon, A.W.; Conway, K.A. & Lansbury, P.T.J. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry*, 35, 13709-13715.
- Williams, D.H.; Searle, M.S.; Mackay, J.P.; Gerhard, U. & Maplestone, R.A. (1993). Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *Proc Natl Acad Sci USA*, 90, 1172-1178.
- Wiseman, T.; Williston, S.; Brandts, J.F. & Lin, L.N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, 179, 131-137.
- Wolynes, P.G.; Onuchic, J.N. & Thirumalai, D. (1995). Navigating the folding routes. *Science*, 267, 1619-1620.
- Wright, P.E. & Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*, 293, 321-331.
- Wu, Z.; Elgart, V.; Qian, H. & Xing, J. (2009). Amplification and detection of single-molecule conformational fluctuation through a protein interaction network with bimodal distributions. *J Phys Chem B*, 113, 12375-12381.
- Yang, W.Y. & Gruebele, M. (2003). Folding at the speed limit. *Nature*, 423, 193-197.

- Yon, J.M. (2002). Protein folding in the post-genomic era. *J Cell Mol Med*, 6, 307-327.
- Zhou, H.X. (2010). From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys J*, 98, L15-L17.
- Zhou, R.; Huang, X.; Margulis, C.J. & Berne, B.J. (2004). Hydrophobic collapse in multidomain protein folding. *Science*, 305, 1605-1609.



Protein Engineering

Edited by Prof. Pravin Kaumaya

ISBN 978-953-51-0037-9

Hard cover, 344 pages

Publisher InTech

Published online 24, February, 2012

Published in print edition February, 2012

A broad range of topics are covered by providing a solid foundation in protein engineering and supplies readers with knowledge essential to the design and production of proteins. This volume presents in-depth discussions of various methods for protein engineering featuring contributions from leading experts from different countries. A broad series of articles covering significant aspects of methods and applications in the design of novel proteins with different functions are presented. These include the use of non-natural amino acids, bioinformatics, molecular evolution, protein folding and structure-functional insight to develop useful proteins with enhanced properties.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Shu-Qun Liu, Xing-Lai Ji, Yan Tao, De-Yong Tan, Ke-Qin Zhang and Yun-Xin Fu (2012). Protein Folding, Binding and Energy Landscape: A Synthesis, Protein Engineering, Prof. Pravin Kaumaya (Ed.), ISBN: 978-953-51-0037-9, InTech, Available from: <http://www.intechopen.com/books/protein-engineering/protein-folding-binding-and-energy-landscape-a-synthesis>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.