Quarterly Reviews of Biophysics **43**, 3 (2010), pp. 295–332. © Cambridge University Press 2010
doi:10.1017/S0033583510000119 Printed in the United States of America

Protein functional landscapes, dynamics, allostery : a tortuous path towards a universal theoretical framework

Pavel I. Zhuravlev and Garegin A. Papoian*

Department of Chemistry and Biochemistry, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA

Abstract. Energy landscape theories have provided a common ground for understanding the protein folding problem, which once seemed to be overwhelmingly complicated. At the same time, the native state was found to be an ensemble of interconverting states with frustration playing a more important role compared to the folding problem. The landscape of the folded protein – the native landscape – is glassier than the folding landscape; hence, a general description analogous to the folding theories is difficult to achieve. On the other hand, the native basin phase volume is much smaller, allowing a protein to fully sample its native energy landscape on the biological timescales. Current computational resources may also be used to perform this sampling for smaller proteins, to build a 'topographical map' of the native landscape that can be used for subsequent analysis. Several major approaches to representing this topographical map are highlighted in this review, including the construction of kinetic networks, hierarchical trees and free energy surfaces with subsequent structural and kinetic analyses. In this review, we extensively discuss the important question of choosing proper collective coordinates characterizing functional motions. In many cases, the substates on the native energy landscape, which represent different functional states, can be used to obtain variables that are well suited for building free energy surfaces and analyzing the protein's functional dynamics. Normal mode analysis can provide such variables in cases where functional motions are dictated by the molecule's architecture. Principal component analysis is a more expensive way of inferring the essential variables from the protein's motions, one that requires a long molecular dynamics simulation. Finally, the two popular models for the allosteric switching mechanism, 'preexisting equilibrium' and 'induced fit', are interpreted within the energy landscape paradigm as extreme points of a continuum of transition mechanisms. Some experimental evidence illustrating each of these two models, as well as intermediate mechanisms, is presented and discussed.

* Author for Correspondence: G. A. Papoian, Monroe Martin Associate Professor, Department of Chemistry and Biochemistry, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA.
Tel.: 301-405-8667; Fax: 301-314-9121; E-mail: gpapoian@umd.edu

1. Introduction

A few decades ago, the question of how an amino acid sequence can predefine a unique three-dimensional (3D) folded structure for a protein seemed hopelessly intractable. Even then, however, kinetic models were successfully devised to describe the conformational transitions in allosteric proteins such as hemoglobin. Subsequently, the statistical view of the protein folding problem led to the realization that globular proteins are built in a very special way, which tremendously simplifies the folding problem. This, in turn, has allowed nature to evolve complex protein folds.

The energy landscapes of foldable proteins are funneled – the conformations that are structurally similar to the native state are also low in energy, while the native state interactions are minimally frustrated (Bryngelson & Wolynes, 1987; Onuchic *et al.* 1995), diminishing the energetic ruggedness, which could kinetically prevent folding (Bryngelson *et al.* 1995; Chen *et al.* 2008; Dobson *et al.* 1998; Dill *et al.* 2008; Frauenfelder *et al.* 1990, 1991; Hyeon & Thirumalai, 2003; Pande *et al.* 2000; Plotkin & Onuchic, 2002; Scheraga *et al.* 2007; Shakhnovich & Gutin, 1993). At the same time, the native state, which used to be thought of as a single structure (simple and clear), turned out to be a collection of conformations that are explored as the natively folded protein roams the landscape at the bottom of the funnel (Fenimore *et al.* 2004; Frauenfelder & McMahon, 1998; Henzler-Wildman & Kern, 2007; Honeycutt & Thirumalai, 1992; Karplus & McCammon, 1981; McCammon *et al.* 1977). This portion of the landscape (that we will afterwards call a functional landscape) has a much smaller phase volume compared with the whole protein folding phase space, but is much richer in its topography when viewed at high resolution – in other words, it is more rugged and glassy. Ironically, the folding landscape that has an astronomical number of states turned out to be easier to understand conceptually than this small region at the bottom of the funnel. Despite a number of open questions still remaining in protein folding, energy landscape theory and the funnel paradigm ushered in a general understanding and provided a common language, which is used widely by experimentalists and theoreticians (Bae *et al.* 2009; Clore, 2008; Fang *et al.* 2009; Tang *et al.* 2008; Thielges *et al.* 2008). With the functional landscapes, however, it is likely that most of the properties of protein functional dynamics depend on the particular features of the landscape of a specific protein, which might prevent the emergence of a general universal view. However, the energy landscape language can still provide a fruitful common ground for the discussion of protein function and dynamics. These dynamics

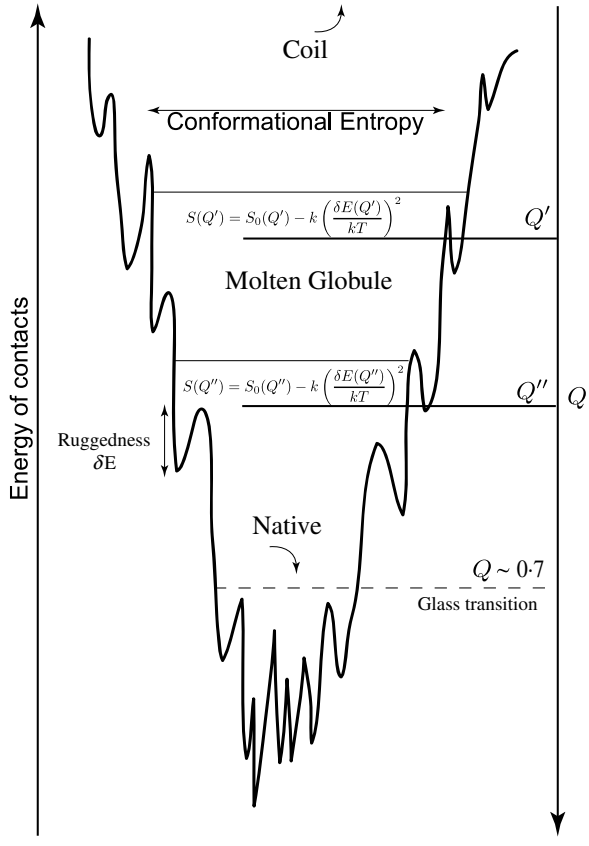

Fig. 1. The protein folding landscape can be schematically drawn as a funnel (Onuchic *et al.* 1995). The vertical axis corresponds to energy of contacts, which is correlated to Q – the similarity to the native state. In the upper region, the protein chain is unfolded with large conformational entropy. In the middle part of the funnel, the molten globule region, the chain is compact, but retains significant entropy. The lower part is a collection of similar low-energy conformations separated by barriers, known as the native state. The funnel can be stratified according to Q. At a fixed Q, configurational entropy is defined by the interplay between total number of states and ruggedness of the landscape for that particular Q stratum. Glass transition temperature T_g corresponds to vanishing of the configurational entropy. In the figure, $T_g(Q'') > T_g(Q')$. The funnel region around $Q \sim 0{\cdot}7$, which corresponds to the onset of native-like conformations, is estimated to have a high glass transition temperature (Onuchic *et al.* 1995).

represent motion in multidimensional manifold riddled with cliffs, ridges, peaks and crevices (also multidimensional). The questions remain the same as ever. How many variables are needed to describe the relevant motions? What are these variables and how to systematically find them?

Because of its minimally frustrated nature, the folding landscape produces kinetics that are often adequately mapped to a single coordinate, the fraction of native contacts, denoted as Q (Plotkin *et al.* 1997). The average energy of contacts decreases as a function of Q, so the energy landscape may be visualized as a funnel (Fig. 1). These statements are not only qualitatively appealing but are also quantitative, based on the generalized random energy model (GREM), which was borrowed from spin glass physics (Derrida, 1985). In the middle part of the funnel, the protein chain is already compact, but very dynamic, still possessing significant entropy (Gast *et al.* 1994; Guo & Thirumalai, 1995; Ptitsyn, 1996; Ptitsyn *et al.* 1990). One consequence of this

is the robust self-averaging property: an ensemble of energies sampled on a small part of the landscape is to a large extent representative of the whole ensemble of energies (Chuang *et al.* 2001; Geissler *et al.* 2004; Onuchic *et al.* 1997). Self-averaging permits the use of the random energy model (REM), which is, in essence, replacing the detailed description of a particular landscape (like a detailed topographical map) with an estimated probability distribution of energy minima and barriers. Configurational entropy in REM is determined by the interplay between total number of possible states and the landscape ruggedness. Therefore, if the funnel is stratified at various Q values, one may compute the configuration entropy as a function of Q, at the given temperature (see Fig. 1). Using reasonable parameters to describe the funnel, it was found that at about $Q \gtrsim 0{\cdot}7$ the configurational entropy is severely diminished, indicating a transition to more glassy dynamics (Onuchic *et al.* 1995). Importantly, the convenient self-averaging character of low Q strata is no longer observed for many dynamical variables during the interconversion of nearly folded states. Therefore, the bottom of the funnel must be described with a detailed map of the functional energy landscape. Q by itself may no longer be the main order parameter, as it is correlated with the folding that has already occurred, while we are interested in functional motions of the folded protein. Therefore, coordinates transverse to Q play an important role. On the positive side, due to small phase volume of the native state ensemble, a brute-force sampling of the functional landscape is feasible, as apparent from numerous experiments and actual protein functioning in biological organisms. One reason for that is that despite the functional landscape being glassy, the protein does not have to actually overcome barriers higher than the one associated with unfolding. From any point of the landscape it can unfold and refold to another point, so the distribution of the barriers (and therefore of timescales) is capped. Indeed, typical excitations of a protein in native state include partial unfolding (Hegler *et al.* 2008). For the same reason, Q is still likely to be a relevant coordinate, but more coordinates are required to describe the motions transverse to Q.

In the most general case, the functional energy landscape has multiple minima, which are hierarchically organized with complicated transition kinetics (Frauenfelder *et al.* 1990; Henzler-Wildman & Kern, 2007). The ultimate goal of computational modeling is to find an approach that is able to predict other native-like states and the kinetics of transitions between all of these states starting from an experimentally determined native structure.

As previously mentioned, the native protein dynamics are immensely multidimensional with several hundred protein residues corresponding to several thousand degrees of freedom, in addition to many more solvent degrees of freedom. However, solvent degrees of freedom are separated by a large timescale gap from the protein conformational motions. For this reason, when discussing the latter, it is often assumed that the solvent degrees of freedom are adiabatically equilibrated around any protein conformation and are present in the description in the implicit way. We follow this convention here: when discussing energy landscapes, by energy, we actually mean free energy of contacts, where solvent degrees of freedom are integrated out. On the other hand, conformational entropy remains explicit and is separate from this 'effective energy'. Even in this case, the energy landscape remains highly multidimensional and native dynamics cannot be interpreted unless projected on a low number of coordinates or represented in a different simplified way.

Proteins possess many structural components: some rigid, some flexible and some liquid. These components move on a multitude of timescales ranging from picoseconds to seconds (Adcock & McCammon, 2006). Likewise, the localization of these motions varies from atomic vibrations to global motions of the whole molecule (García, 1992). When addressing a particular

problem, motions on specific timescales and localizations are usually of interest. Functional allosteric motions, for instance, are typically in the slower range of the possible timescales and are spatially global. This places limitations on the set of motions investigated to describe a particular phenomenon and motivates the projection of the dynamics into a much lower dimensional space. Indeed, techniques like principal component analysis (PCA) or essential dynamics are able to describe much of the non-trivial motion by a mere 3–5% of the total conformational degrees of freedom (Cukier, 2009; Grant *et al.* 2009; Maisuradze *et al.* 2009a; Materese *et al.* 2008; Paschek *et al.* 2008). However, the essential phase subspace may be complicated, even fractal, so it is not clear how to choose these coordinates, starting from a protein crystal structure. Can there be a universal method for characterizing the essential subspace, or dissimilar approaches are better suited for different protein classes and sizes? This is one of the central questions we explore in this review.

One approach is to directly derive coordinates from the spatial motions of the protein of interest, where these motions are obtained either experimentally or from computer simulations (Adcock & McCammon, 2006) or are speculated based on the protein molecular architecture. Often these motions are connected to function, such as binding of oxygen to hemoglobin or ion channel opening and closing. If there is no clear way to correlate the coordinates with functional motions, they may be at least based on the structures of various functional states of the protein. The aforementioned quaternary functional motions, such as respective rearrangements of domains and subunits in the case of multidomain proteins and oligomers, are usually more obvious. However, smaller, single domain globular proteins may perform subtler allosteric switching. In this case, good coordinates may be the ones that correlate with minima or valleys in the energy landscape. At the very least, a good coordinate must evolve slowly on the timescale of interest.

In the first section of the review, we discuss techniques based on the topologies of native conformations or on the difference between two allosteric states, which functional motions have to cover. These are largely variations of normal mode analysis (NMA).

Allosteric switching is, in many cases, a result of ligand binding. In these cases, the ligand degrees of freedom are added to the picture, compromising the goal of decreasing the number for variables used for describing the transition. There exist two simplified representations of allosteric transitions known as 'preexisting equilibrium' (population shift) and 'induced fit' (reaction front) scenarios. They are based on different sets of assumptions. In the second part of the review, we highlight these assumptions and formulate differences and similarities between the two scenarios of allosteric transitions. In particular, it can be shown that the pre-existing equilibrium mechanism is one of the assumptions for the classic Monod–Wyman–Changeux (MWC) model for hemoglobin allostery, and induced fit is the basis for the Koshland–Nemethy–Filmer (KNF) model. Then, we present experimental evidence supporting each of the two scenarios in different cases, discussing in detail a special case of induced fit, which is called the 'communication pathway'.

In the third part, we focus on the techniques that can create a representation of a protein's energy landscape, which allows the subsequent study of specific questions, such as the nature of the dynamics (diffusive or activated) or coupling of chemical and conformational motions in motors. These include PCA, construction of kinetic networks and building free energy surfaces.

Overall, we present here a range of relevant computational techniques that are currently used to study protein functional dynamics. Some of these techniques can only serve as a post-analysis of a full-blown molecular dynamics (MD) simulation, while others have predictive power of their own; others still are in the middle, requiring MD simulations for analysis, but subsequently

providing the means to a deeper investigation of various dynamical problems at a lower cost. In addition, we present evidence for different mechanisms of allosteric switching and interpret them using the energy landscape paradigm. This review, naturally, covers a mere fraction of the relevant literature and presents our subjective view on the problems and methods in protein functional dynamics that we consider important and where a number of outstanding questions remain at the time of this writing.

2. Understanding protein motion using harmonic modes

Native state protein dynamics span a wide range of spatial and temporal scales (Adcock & McCammon, 2006; Henzler-Wildman & Kern, 2007; Karplus & McCammon, 1981, 1983). Fast motions, occurring at short length scales, correspond to atomic vibrations, while the slowest motions at large length scales may be associated with functional or allosteric processes. In large multidomain proteins or protein complexes, visual inspection of many of these motions, such as opening of a channel or moving two domains from a tightly connected state to a loosely connected one, seems to directly provide a mechanistic interpretation of protein function. Hence, this facility of visualization has created a paradigm of thought of a protein as a rigid body on hinges, similar to a mechanical machine. Although such thinking does sometimes provide useful insights in the case of larger protein complexes, possessing a quaternary structure, the smaller, single-domain proteins do not resemble mechanical machines and move more like malleable objects made of clay or even like a liquid.

With the mechanical picture in mind, one realizes that proteins are polymer chains of interacting atoms, and the topology of connections between the atoms is more complicated than in real mechanical machines. In the simplest approximation of these interactions, a Taylor expansion of the potential energy function is truncated at the quadratic term, resulting in a harmonic Hamiltonian. A subsequent analysis of collective modes based on the energy function, called NMA, permits an examination of both the motion frequencies and eigenmodes, which indicate how individual atoms move when a particular mode is activated. In practice, the NMA-related techniques represent atoms (or residues) as point masses connected by harmonic springs, and then find independent (without mutual energy transfer) mechanical harmonic modes.

Even within the harmonic approximation, the protein would not necessarily be moving in these modes: it is frictionally coupled to a solvent bath and therefore evolves according to Langevin, not Hamiltonian dynamics (Zwanzig, 2001). The protein chain dynamics are highly damped by the solvent, and certain chain motions may even be slaved to the solvent dynamics (Fenimore *et al.* 2004; Frauenfelder *et al.* 2007; Henzler-Wildman *et al.* 2007a; Ma, 2005; Samuni *et al.* 2007). Therefore, energy quickly dissipates over the modes and to the solvent on a femtosecond to picosecond timescale, so specific normal modes, including 'plausible' functional modes, quickly loose coherence, with the overall protein configuration motions being highly diffusive.

On the other hand, the molecular architecture may indeed impose significant limitations on collective motions, thus predisposing the molecule toward particular spatial rearrangements. In this way of thinking, normal modes are reflective of the molecular architecture, providing a potentially useful basis set for describing real protein motion, drastically reducing the dimensionality of the original dynamics. In addition, since the normal mode calculations are computationally inexpensive, they may be used to quickly analyze the functional dynamics of a

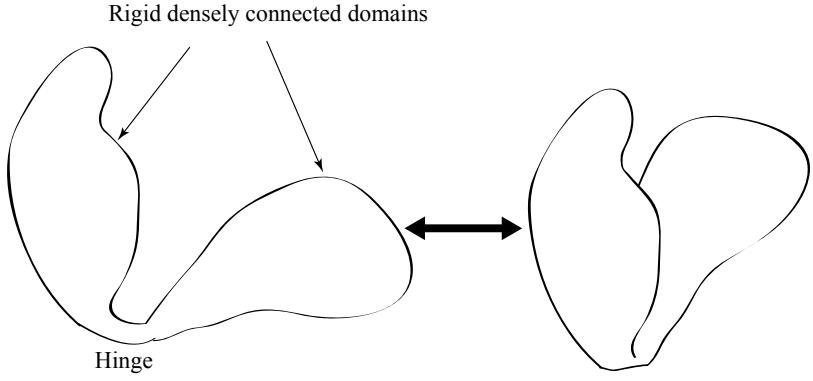

Fig. 2. Some special types of protein molecular architecture allow a mechanistic view on the protein's motions. In this schematic representation, the protein can be thought of as a machine that has two rigid parts connected by a hinge, with motion around that hinge being a single degree of freedom.

large protein complex, which might not be amenable to other more detailed computational techniques. For example, when building a free energy profile between two conformational states of a protein, it is necessary to first devise an appropriate collective coordinate (order parameter). One of the lowest index normal modes, or a linear combination of a few modes, could possibly be used for that purpose if there was a practical criterion that would tell whether a particular order parameter is suitable. For instance, a collective coordinate might be suitable if it changes slowly and in a fashion correlated with the extent of similarity to the starting and final conformations.

In the following subsections, we review a small selection of NMA applications ranging from the cases where the method predictions correlate well with an experimentally known conformational change to cases where NMA is mainly used for guidance. We also discuss approaches that extended NMA beyond its initial domain of applicability.

2.1 Evaluating NMA's predictive power for real proteins

In some cases, the protein architecture – steric interactions arising from spatial distribution of density, coordination number and chain connectivity – is so important that the structure itself may be predictive of other possible functional conformations and even of pathways between them. This occurs more frequently with quaternary motions for large multidomain proteins or protein oligomers. Hemoglobin and kinases like AdK may serve as examples (Arora & Brooks, 2007; Karplus & McCammon, 1983; Whitford *et al.* 2007). Hence, when a large protein can be fruitfully thought of as comprising several rigid bodies that are loosely connected (Karplus & McCammon, 1983; Tama & Brooks, 2006), the large-scale slow motions are likely dominated by relative displacements of these 'rigid bodies' (subunits or domains) as a whole with respect to each other (Fig. 2). Atoms inside the rigid bodies are tightly packed, characterized by high coordination numbers. These cases are identified by a low number of contacts in the interface between the subunits, which are changing in such motions, compared with the nearly unchanged contact maps within the rigid bodies. Consequently, the energy of contacts is not changing significantly in the transition, and therefore the effect of conformational entropy change is expected to be of significant importance.

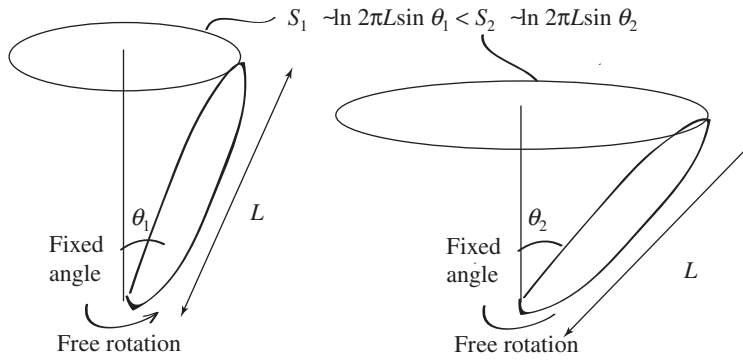

Fig. 3. Entropy contained in a degree of freedom depends on the spatial Jacobean. In this figure, the rotational entropy of the rod becomes larger with larger opening angle, because the size of space available to the rod's end is proportional to the sine of this angle.

To clarify the last statement, let us imagine a rod with one end attached to the origin that can rotate around the z-axis at a fixed angle, θ (Fig. 3). If a slow change of this angle is associated with a conformational motion, then the rotational entropy will be affected too, since it is proportional to $\sin\theta$. Analogously, such 'rigid-bodies-on-hinges' motions in various protein complexes may be partially driven by changes in the protein conformational entropy.

One successful example of the 'rigid-bodies-on-hinges' technique application was a study of hemoglobin, where the first normal mode turns out to be relevant to going from the tense (T) to the relaxed (R2), oxygen-bound state (Xu *et al.* 2003). If one starts from the T state and applies a small perturbation along the slowest mode of motion, as defined by the anisotropic network model (ANM), then the root mean square distance (RMSD) to the R2 state will decrease as the magnitude of this step along the first mode grows, from 3·5 Å (RMSD between T and R2) down to 2·5 Å. In other words, the motion along the first mode from the T state approaches the R2 state. A step in the other direction along the mode takes the protein farther from R2, and there is no way of ascertaining the motion direction, without knowing the R2 structure in advance. But even leaving a choice of two structures is indicative of the method's predictive power. Analyzing the weights with which the coordinates of different residues enter the first mode, authors found that the least mobile atoms are on the interface between α_1 and β_2 subunits, hinting that they mediate a hinge-like motion. The authors suggested that the transition from T to R2 is favored entropically. As the contacts between the subunits (hydrogen bonds and salt-bridges), responsible for maintaining the T state, weaken upon oxygen binding, the hemoglobin transitions to the R2 state primarily because of increased conformational entropy. In the analogy above, this motion is similar to the increase of the angle θ between the rod and the z-axis, along with the corresponding increase in entropy. This view is corroborated by the fact that the $\alpha_2\beta_2$ unit is rotated by more than 15° upon the conformational change from T to R2. Interestingly, when the authors repeated the procedure backwards, starting from R2, then the first mode did not lead to the T state, in either direction. This result is not surprising, as the backwards transition cannot also be driven by an increase in conformational entropy. On the other hand, because of the lack of a detailed mechanism, the picture of entropically driven transition still has to be regarded critically.

In the above example, elastic network modes not only serve as a good basis set for the motion description but are also able to predict the other functional conformation, also suggesting a

specific pathway for the transition. However, it is not clear how often other proteins would produce such favorable outcomes, even if they consist of nearly rigid domains connected by hinges. It is likely that the role played by the architecture is extreme in this example, based on the specific topology of steric interactions and chain connectivity. Furthermore, the hemoglobin transition is well described by the first mode only, which is quite unusual. In other proteins, even if the architecture leads to a transition driven by conformational entropy, a combination of several low-frequency modes typically contributes to the allosteric transition pathway. This diminishes the predictive power of the technique with respect to finding the alternative conformation. It is relatively simple to trace a single mode and produce a series of structures, but it is less practical to cover a space of several dimensions where each point is a possible structure.

Nonetheless, the success of describing allostery with NMA is not unique for hemoglobin. Another similar example is the T→R transition in the chaperonin protein GroEL (Tehver *et al.* 2009). A quantitative measure for the success of a particular mode in the description of the transition (Alexandrov *et al.* 2005; Marques & Sanejouand, 1995; Tehver *et al.* 2009) is the overlap defined as

$$O_M^{\mathrm{T}\to\mathrm{R}} = \frac{|\sum a_{iM}(x_i^{\mathrm{R}} - x_i^{\mathrm{T}})|}{\sqrt{\sum a_{iM}^2 \sum (x_i^{\mathrm{R}} - x_i^{\mathrm{T}})^2}}, \tag{1}$$

where x_i^{T} and x_i^{R} are the atomic coordinates in states R and T, respectively, and a_M is the Mth mode. This quantity equals the absolute value of the cosine of the angle between a_M and $x^{\mathrm{R}} - x^{\mathrm{T}}$, covering a range from 0 to 1. When the cosine is equal to 1, at zero angle, it is indicative of complete coincidence of the mode and the difference between T and R structures. For the T→R transition in a single subunit of GroEL, the overlap for the lowest frequency mode was found to be 0·83 (Tehver *et al.* 2009). Thus, the first mode could have been used to predict the R structure if the latter had not been known.

In addition to detailed studies of allosteric motion for particular proteins, the robustness of the harmonic motion approach has been systematically tested on larger protein datasets (Alexandrov *et al.* 2005; Tama & Sanejouand, 2001). An analysis by Tama *et al.* used a set of 20 proteins for which both 'open' and 'closed' forms are known (Tama & Sanejouand, 2001), calculating overlaps of the normal modes using Eq. (1). The mode that has the maximal overlap is considered 'the most involved' in the conformational change. In 7 out of 20 cases, this was the lowest frequency mode. Thus, in the majority of cases, the use of the first normal mode for the prediction of the other allosteric conformation would not have been successful. Overlap in those seven proteins varies from 0·3 to 0·72 with an average of 0·51. The index of 'the most involved mode' was not very small for all the 20 proteins. For seryl-tRNA synthetase, for instance, the most involved mode had index 18. The overlap of the most involved mode with conformational difference (Eq. 1) was ranging from 0·3 to 0·83, with an average of 0·56. In a different study, Alexandrov *et al.* (2005) filtered all possible pairs of ∼33 000 sequences from Protein Data Bank (PDB) for sufficient sequence identity (>95%), structural dissimilarity (>1·5 Å RMSD) and then, through CATH (Class, Architecture, Topology, Homologous superfamily) database, identified 377 pairs that should have a conformational transition between them. In this study, overlap was defined in a slightly different way: Eq. (1) (calculating a cosine of angle between atom displacement along the normal Mth mode and the difference in position in two real structures) was applied for each atom individually. In addition, as an integral measure, the root mean square of these individual overlaps was computed. When these calculations were

performed without using additional criteria (such as only looking at atoms that move the most), the average angle between the vectors was only marginally different from that between random vectors. However, this study likely included many smaller, single-domain proteins that possess only tertiary motions, which are ill-defined by the protein architecture.

NMA provides a better mode transition overlap when the initial reference conformation is the open form (Tama & Sanejouand, 2001). This is not surprising, as the shape and lack of connections are more pronounced in those forms. Additionally, when a conformational transition is more collective, overlaps also tend to be higher (Tama & Sanejouand, 2001). All these facts indicate that harmonic mode-based techniques are most successful when applied to the problems similar to the hemoglobin or GroEL examples (Tehver *et al.* 2009; Xu *et al.* 2003) – to study quaternary motions of large multidomain proteins or oligomers, preferably with a single hinge region. However, even in these cases, truly successful examples have been rare.

A large number of NMA studies use the reproducibility of the *B*-factors – the experimentally determined values indicating the amplitudes of atomic fluctuations in a crystallized protein – as a way to validate the technique, or for parameterizing the model (Bahar & Rader, 2005; Eyal *et al.* 2007; Hayward & Go, 1995; Kondrashov *et al.* 2006; Riccardi *et al.* 2009; Soheilifard *et al.* 2008; Tehver *et al.* 2009; Zheng, 2008). However, individual atomic fluctuations (characterized by *B*-factors) are mostly defined by the atom's local environment, such as atomic coordination numbers, and in densely packed parts of the protein these local environments will not change significantly regardless of the motions of these parts. Therefore, reproducing *B*-factors is not an impressive test of the validity of applying elastic network models to a particular protein. However, the fluctuations obtained by NMA are sensitive to the cut-off radius of the network model, so *B*-factors can be profitably used for choosing the spring constants or the cut-off radius. Caution is still needed: in protein crystals, the local atomic environment is different than in solutions, because of interactions with neighboring molecules in the lattice; therefore, crystallographic *B*-factors may noticeably differ from the atomic fluctuations in the solution (Kundu *et al.* 2002; Soheilifard *et al.* 2008), while nuclear magnetic resonance (NMR) results show better correspondence with fluctuations in elastic network models (Yang *et al.* 2007). *B*-factors are determined by contributions from motions on all scales, both spatially localized and collective, while the main physical idea of NMA is to study *slow large-scale* motions, which are largely defined by protein architecture. In other words, the springs are representative of the architecture, not the actual energy landscape above the ångström scale, as we discuss further in the section.

2.2 Adiabatic normal modes, double wells and cracking

We now consider a hypothetical protein, having two native conformations A and B with a barrier separating them. Motivated by hemoglobin and other allosteric proteins and enzymes, we are interested in the dynamics of the transition between these states. Problems of this kind are frequently addressed by NMA techniques. Since each of the two states is a free energy minimum, it is natural to depict an approximate one-dimensional free energy surface comprising two quadratic wells with a barrier between them. In reality, the landscape is multidimensional and allows multiple transition paths, but to simplify the problem one may start with two conjugated parabolae, in an approach similar to that used in the electron transfer theory (Marcus & Sutin, 1985, Miyashita *et al.* 2003).

One must always keep in mind that the main reason why the normal modes may serve as suitable reaction coordinates for large-scale protein motions is the fact that these motions are

largely determined by the constraints imposed by the protein architecture. While it is physically reasonable to assume that the native conformation (typically determined by X-ray crystallography) is a minimum of the energy of contacts, and expand these energies to the second order, this harmonic approximation will likely hold in a very narrow vicinity of either structure A or B. Therefore, the NMA springs connecting the atoms are mainly reflecting the architecture of the conformation rather than the actual energy landscape.

If architecture is indeed paramount, NMA may still guide the motion further from the initial state (A) to the final state (B) even when the harmonic approximation fails. It is possible then to adiabatically modulate the normal modes while constructing the transition path, that is, slightly change the structure, as normal modes suggest, and then calculate new normal modes, repeating this process until the final state is reached. This approach has been used with adenylate kinase, where the putative transition between open and closed forms was iteratively followed by NMA steps (Miyashita *et al.* 2003). First, normal modes in state A and their overlaps with the transition displacement vector were calculated. The overlap is a scalar product of a normal mode $3N$-dimensional vector and the difference between target and initial structures ($3N$-dimensional vector whose components are the differences between the atoms' coordinates in states B and A). Then a small step along the highest overlap mode (or a combination of a few highest overlap modes) was taken. Subsequently, the normal modes were recalculated for the new state (conformation), and the new step forward was taken. This procedure generates a sequence of conformations that may reflect the actual transition path, and a reaction coordinate along that path, λ, related to the number of steps along adiabatically changing normal modes. Naturally, the procedure can be also repeated in the opposite direction, starting from state B and going along the modes that overlap with the A–B structural difference (Miyashita *et al.* 2003).

The procedure described above does not guarantee that similar forward and backward trajectories will be generated, nor is it clear how to estimate their correspondence to the actual transition path. Nonetheless, all the conformations from the iterative sequence of steps can be mapped on a traditional reaction coordinate, such as RMSD (root mean square deviation of atomic coordinates) or Q (Plotkin *et al.* 1997) (the fraction of shared contacts with the initial structure), or two coordinates, e.g. Q_A and Q_B (Okazaki *et al.* 2006; Wu *et al.* 2008), or RMSD(A) and RMSD(B), like Miyashita *et al.* did. To eliminate the second coordinate, authors established a correspondence between RMSD(B) and RMSD(A) for each point of the transition path based on the path length (Miyashita *et al.* 2003). It would be interesting to build a 2D free energy surface in terms Q_A and Q_B from all-atom MD (Wu *et al.* 2008), and compute a dominant path on this surface to compare it with the forward and backward paths obtained with the iterative NMA approach.

Another subtle point with this scheme is that in addition to obtaining the transition path itself, one often would like to know the free energy profile along that path to predict the kinetic rates. To address this problem, the authors build on an approach analogous to electron transfer theory and assume two conjugated parabolic potentials. If higher accuracy is required, a free energy profile may also be computed by performing an umbrella sampling along λ using a detailed MD force field. This should be feasible by reversing the correspondence between λ and RMSD. In addition to mapping the adiabatic normal mode pathways in two dimensions, as mentioned above, this approach might lead to a logical extension of the iterative NMA technique, providing higher resolution details.

The next question is how far can normal mode techniques be pushed while still providing meaningful results. Real energy landscapes are almost always poorly known, and parabolic wells are their simplest approximation. Energy landscapes of globular proteins are funnel-like

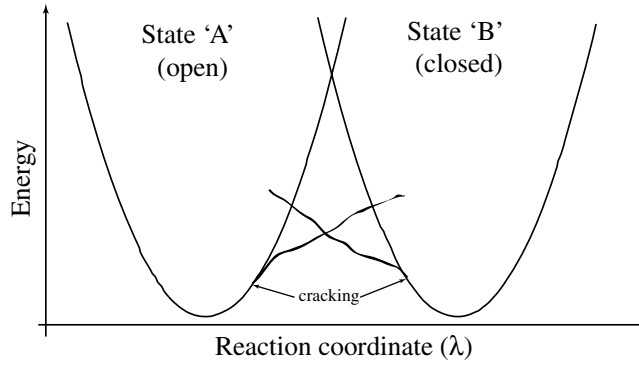

Fig. 4. Cracking is the breaking of some of the native contacts and partial unfolding. If there are two allosteric states, then in harmonic approximation, there are two quadratic surfaces, and the rise up of their slopes may be less favorable than partial unfolding. In this case, the molecule might crack and transition to the other surface at a lower free energy.

(Bryngelson *et al.* 1995; Dill & Chan, 1997; Dobson *et al.* 1998; Frauenfelder *et al.* 1990; Onuchic *et al.* 1995; Papoian *et al.* 2003a; Plotkin & Onuchic, 2002; Yue & Dill, 1995), that is, defined by the native contacts. As these contacts are stretched, the free energy grows, and it is plausible that this growth is quadratic with respect to the atomic coordinate displacements. Furthermore, the spring constants can be tuned to correspond to the actual funnel slope. To save this simple representation beyond the point where the harmonic approximation breaks down, Miyashita *et al.* suggested that the main reason for its violation is that the parabola is too steep, and it becomes more favorable for the chain to locally unfold, or *crack* the contacts, than to continue stretching them (Miyashita *et al.* 2003, 2005). To incorporate this possibility into the model, the authors calculated the strain energy in each conformation along the adiabatic normal modes path as the strain of the springs of the initial structure (A or B). The strain energy turned out to be localized in sequence, thus supporting the cracking hypothesis. The authors assigned an entropy gain to each 'unfolded' residue and compared that to the energy in the case where the residue remained 'folded' (this energy consists of contact energy and strain energy accumulated over adiabatic deformation). Local unfolding changes the free energy surface from quadratic to locally linear, and the same change occurs for the dependence of the barrier height on the free energy difference between the two states: without the cracking it is quadratic, similar to Marcus theory for the electron transfer, while it becomes linear when the cracking is allowed (Miyashita *et al.* 2005) (Fig. 4). This fact may be used to quantitatively test the cracking model.

The idea of conjugating potentials for two different protein states is quite general and may be applied to other models as well. In the case of Gō-like force fields, the conjugation is also related to local unfolding (Okazaki *et al.* 2006). We elaborate on this issue further in this paper.

2.3 Phonons

On very short (picosecond) timescales, proteins do behave like rigid bodies and exhibit phonon dynamics based on atomic vibrations. This range is experimentally accessible through infrared (IR) spectroscopy, and a number of computational techniques permit interpretation of the experimental results. The problem of the extreme locality of the harmonic approximation remains at these timescales. Instantaneous NMA counteracts this problem by generating a number of

solvent cages and performing NMA for each one of them (La Nave *et al.* 2000; Schmitz & Tavan, 2004; Schrader *et al.* 2007; Schulz *et al.* 2009). Energy flow in proteins may be described in a harmonic approximation through quantum heat transfer operators (Leitner, 2008, 2009), or by just observing classical heat flow through a frozen protein globule in MD simulation (Ota & Agard, 2005). These techniques shed light on communication pathways within a protein, which we review in the next chapter.

3. Allostery and communications

Intra-protein communication pathways as a concept are related to the problem of allostery. By allostery we mean the property of a protein to exhibit global conformational change in response to a local perturbation. The most common case of such local perturbation is ligand binding, upon which an allosteric protein changes conformation.

There are two marginal pictures of the allosteric transition mechanisms. They may be traced to the seminal works of Monod *et al.* (1965) and Koshland *et al.* (1966) on the problem of the hemoglobin allosteric transition from the tense (T) to the relaxed (R) state upon oxygen binding. One view states that without the ligand the protein explores both states in thermal equilibrium, with the R state being higher in free energy and hence less populated. When the protein is in the R state, the ligand can bind to it, pulling down the minimum in the energy landscape corresponding to the R state and locking the protein in conformation R. Thus the lock-and-key idea is revisited, although the analogy is not with precise fitting, but with locking of the structure in a new state. This mechanism is currently popular and is known as the 'pre-existing equilibrium mechanism'. Pre-existing equilibrium is assumed in the MWC model for hemoglobin. According to the alternative picture, the protein does not visit the R conformation in the absence of a ligand, but when the ligand binds to an opening on the protein surface, it pushes the protein to the other conformation by propagating a reaction front (the latter language is not yet in common use). When one carefully examines the KNF mechanism of hemoglobin allostery, it becomes apparent that it assumes a reaction-front-like model.

The 'pre-existing equilibrium' concept seems to be very natural – all possible states on the energy landscape must be visited, at least infrequently. So how one could envision a scenario where it fails? If the second state is so high in energy or landscape barriers are so high that it is not visited on a biological (or laboratory) timescales, then the KNF model provides a more plausible physical model of conformational switching. Hence, at a certain threshold value of the free energy difference between two allosteric states, the MWC picture may qualitatively change to the KNF picture. In addition, hemoglobin is an oligomer and an interfacial free energy penalty between subunits in different states is assumed in both models. In MWC, this means that all four subunits change together, while in KNF mixed states do exist despite the penalty, because binding of oxygen to one subunit 'pushes' it to the other state. Sometimes, this is how the difference between the models is formulated in the literature: in the MWC approach, one assumes collective conformational change of all the subunits, while the KNF approach is sequential. From our dynamical perspective, however, emphasizing the difference between 'pre-existing equilibrium' and 'reaction front' scenarios is more physically illuminating. Once again, in MWC model, there is *equilibrium* between T and R conformation in the absence of oxygen (and the interfacial penalty excludes the mixed subunit states), while in KNF, there is no minimum in conformation R without oxygen; this minimum *is created by* binding, hence the name 'reaction front'.

Since both of these models are conceptually appealing, researchers tend to use one or the other when interpreting a particular experiment. Interestingly, in the case of hemoglobin allostery, both models can be used equally well to fit most of the experimental data, although the MWC model is currently slightly favored (Eaton *et al*. 1999). In this section, we briefly review the strength of evidence supporting these limiting case interpretations. On the one hand, it seems unlikely that one of the mechanisms would completely dominate. In particular, for different allosteric proteins different mechanisms should be at work, including mixed scenarios, which are usually not considered. Because of the power of energy landscape paradigm, the research community at this point has largely swung toward the opinion that the MWC picture ('pre-existing equilibrium') is the dominant one, and this model is used most of the time to interpret experiments. In addition, a pure KNF approach indeed seems exotic, as it would require a very intricate design of the protein, which would allow the conformation to be 'pushed' to the other state across the whole molecule by just the binding of a ligand. These questions are further explored next.

3.1 Pre-existing equilibrium

Most of the current literature dedicated to describing the mechanisms of allostery is based on the 'pre-existing equilibrium' approach, related to the MWC model, as discussed earlier. In the last two decades, it has become widely accepted that the native state consists of an ensemble of conformations on an energy landscape rather than a single static state, and the idea of the pre-existing equilibrium among these states is very natural. Likewise, from this perspective, it is more appropriate to talk about the change of the native ensemble, rather than a conformational change (Benkovic *et al*. 2008).

Let us consider the joint energy landscape of a protein and its ligand (Fig. 5, upper panel). With no ligand bound, the protein vigorously explores the left part of this landscape, spending most of the time in conformation A (which is experimentally observed when no ligand is bound), only occasionally visiting conformation B (which is experimentally observed when the ligand is bound), in the state denoted as β on the joint landscape. During one of these visits to the β state, a ligand binds, locking that conformation, so the system moves to state B on the joint landscape. Importantly, the conformational step is much slower than the binding step. Pre-existence of an equilibrium assumes multiple transitions occurring between A and β on microsecond to millisecond timescales. Diffusion of the ligand to the binding site, once the ligand is within an ångström from the binding interface, is on the scale of nanoseconds. For this reason, the MWC mechanism is often called a population shift mechanism, since it can be envisioned as sudden deepening of the minimum at β (becoming B) with simultaneous shallowing of the minimum at A as it goes to α (Arora & Brooks, 2007; Okazaki & Takada, 2008). On the joint landscape, we can think of population shift as the system jumping from the unligated cross-section of the landscape (with A and β) to the ligated cross-section (with B and α). Note that there does not have to be a pronounced minimum at α, since it is not biologically significant.

Figure 5 (lower left) shows another interpretation of an MWC-type allosteric transition that does not consider the joint energy landscape of a protein and its ligand, but only that of the protein. It is analogous to diabatic electron transfer between two surfaces. The surfaces in the case of the MWC switch are the energy landscapes of the protein molecule in unligated and ligated states (they are largely the same as the cross-sections of the joint landscape discussed above). The unligated surface has another minimum at β, but the same conformation has a much

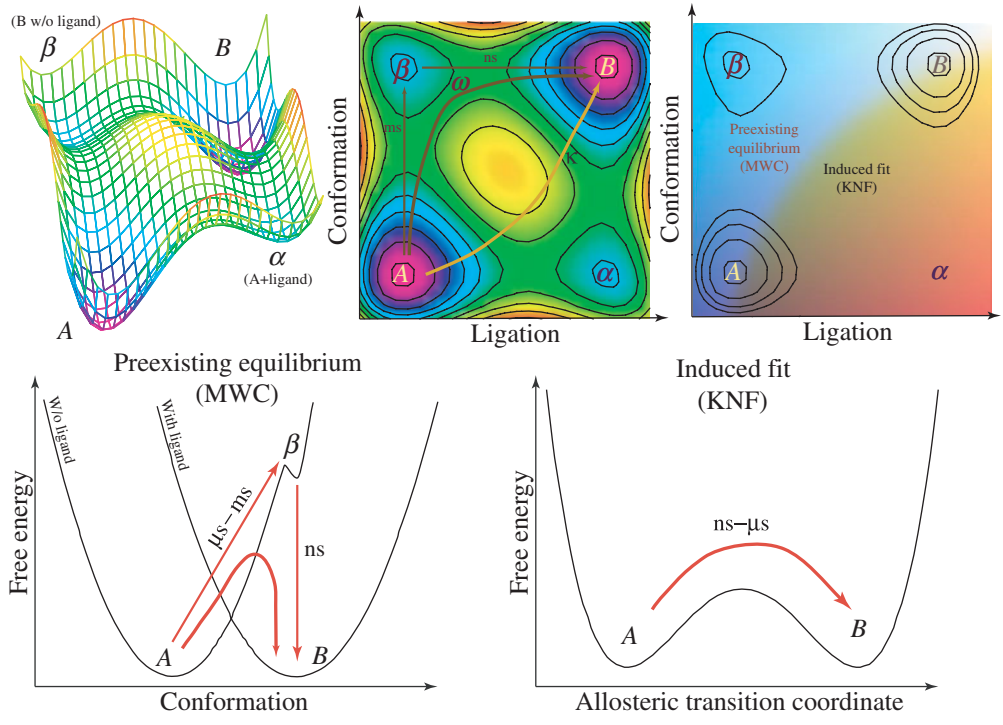

Fig. 5. Upper left: a surface with four minima is the simplest schematic representation of the joint ligand-and-protein energy landscape. Upper middle: in the MWC scheme, the protein is in equilibrium between states A and β in the absence of the ligand and upon ligand binding goes to state B, which is the same conformation of the protein as β but with a ligand. The first step is much slower (μs–ms) than the second (ns). More likely, the MWC transition would follow a less extreme pathway, similar to the one denoted as ω. A minimum α does not have to exist, but it facilitates interpretation of the MWC picture as a 'population shift'. In the KNF scenario, binding starts first, and there is no timescale separation between binding and conformational change. This may happen along a pathway similar to the one denoted as κ. Upper right: the previous statements can be summarized in a phase diagram: the lower right zone of the plot hosts KNF pathways (the start of binding precedes conformational change) and the upper left zone hosts the MWC pathways (conformational change precedes binding). Lower left: to avoid considering the complicated joint energy landscape, the MWC picture may be visualized as a diabatic transfer between two energy surfaces: one with a ligand and one with no ligand, since the switch between the surfaces happens much faster than visitation of the state β on the unligated surface. Such an interpretation also suggests that the transfer will possibly occur earlier than the visit of state β is completed, as it is too high in free energy. Lower right: in the KNF scheme, 1D interpretation is a free energy profile along the transition pathway. If this pathway on a joint landscape is enveloped by a valley, the transition can be thought of as a gradual adiabatic relocation of the minimum in which the protein continues to reside. Note: many ingredients that are used to construct this figure have been employed long ago. However, the authors of this manuscript have introduced some new elements, and the overall synthesis has many new aspects. It has to be seen whether the theoretical description of allostery proposed in this review will be validated by detailed simulations and experimental work.

lower energy when the ligand is bound, on the other surface. Hence, if the ligand becomes available while the protein happens to explore the neighborhood of β, the protein almost instantaneously jumps to the other surface. Such an immediate jump is called a diabatic transfer. Depicting the MWC mechanism in this way suggests that the diabatic transfer may happen prior to the protein chain fully reaching the β state, since two diabatic surfaces are likely to cross at

lower energies before that point. Therefore, if the ligand has approached, the dominant transition pathway may be something like the one denoted as ω or bold arrow (Fig. 5) rather than actually two-step A$\rightarrow\beta$, $\beta\rightarrow$B.

Indeed, it seems to be the limiting case of the MWC picture, when conformation B is visited with extreme precision in the absence of the ligand, having correct conformation of the backbone and local side-chain orientations at the interface primed for binding the ligand. Experimental evidence to date does not possess sufficient structural resolution to ascertain if this is a likely scenario. We consider a number of cases below in detail.

The concept of a pre-existing equilibrium started to develop vigorously after an indication from MD simulations that myoglobin shows the largest conformational fluctuations in the regions that are also the most different in the X-ray structures for allosteric states (Frauenfelder *et al.* 1988). Early experimental evidence has been reviewed by Kern & Zuiderweg (2003). It includes the observation of exchange between active and inactive states in unphosphorylated NtrC by NMR spectroscopy (Volkman *et al.* 2001) (allosteric transition occurs on phosphorylation), a transient loop disengagement in the R-state of fructose-1,6-bisphosphatase (the disengaged loop corresponds to the T-state) revealed by X-ray and fluorescence (Nelson *et al.* 2000), the observation that the conformation of CO-ligated hemoglobin (defined by relative orientation of the α_1/β_2 subunits) shown by NMR is exactly halfway between T and R crystal structures interpreted as dynamic average between them (Lukin *et al.* 2003).

More direct evidence appeared with time. The small angle X-ray scattering curve for wild-type aspartate transcarbamylase changes abruptly at about 60 °C, indicating the transition, while the same curve for a mutant changes gradually and non-monotonically, suggesting that it represents a dynamic average of two states, whose relative populations shift slightly with temperature variation (Fetler *et al.* 2007). The kinetic curves for glucose binding by human glucokinase are best explained by the kinetic scheme that suggests pre-existing equilibrium of states with higher and lower affinity for glucose (Antoine *et al.* 2009; Kim *et al.* 2007). Similar schemes also explain fluorescence anisotropy curves for cyclic AMP (cAMP) receptor protein of *Escherichia coli* (Youn *et al.* 2008).

An MD study of acetylcholinesterase classified 89 PDB structures from five different species and grouped them according to χ_1 and χ_2 angles of Trp279, which is the key residue of the peripheral anionic site that may interact allosterically with the catalytic anionic site and may show conformational change upon binding (Xu *et al.* 2008). The authors find five islands on the (χ_1, χ_2) plot. Some of the groups of PDB structures are seen on the islands, that is, corresponding rotamers are observed in the simulation, so these rotamers are in pre-existing equilibrium. However, the conformational state of the rest of the molecule was not included in the clustering analysis.

Evidence for a pre-existing equilibrium has been also observed *in vivo* (Ayers *et al.* 2007). FABP4 is a protein that performs transport from the cytosol to the cell nucleus. To enter the nucleus it must have a nuclear localization signal (NLS) conspicuously exposed. In FABP4, NLS consists of three residues, and the conformation where NLS is at work, allowing the entrance to the nucleus, is locked by a regulating ligand. However, even without the regulating ligand, some accumulation of FABP4 was observed in the nucleus, suggesting that the corresponding allosteric conformation is visited even in the unligated state (Ayers *et al.* 2007).

It has been shown that a mutation in cyclophilin A can invert relative populations of the pre-existing substates (Fraser *et al.* 2009). This mutation also slows conformational dynamics and decreases catalytic turnover by two orders of magnitude. Further spectroscopic investigation

suggests that enzymatic activity decrease is linked to slowed conformational dynamics. As in the last example, NMR spectroscopy has revealed global conformational rearrangements in G-proteins that are thought to be the reason allowing the protein to interact with diverse ligands (Smrcka *et al.* 2010).

In summary, a small number of works have directly supported the 'pre-existing equilibrium' hypothesis. In many cases, however, only local conformational rearrangements were considered, where it is not possible to positively claim that the other allosteric conformation is visited in its entirety. Nonetheless, the picture of equilibrated ensemble of states is very natural in the energy landscape paradigm. Hence, some studies use this language of ensemble equilibrium and population shift without any pointers as to which allosteric mechanism is relevant to their particular case (Martinez *et al.* 2002; Malmendal *et al.* 1999). This language has even led to a suggestion that allostery is a purely thermodynamic phenomenon (Tsai *et al.* 2009). As previously mentioned, it is true that there is always an ensemble rather than a static conformation, but there is also a biological timescale, which may not accommodate the visitation of the other allosteric state in the absence of the ligand. Therefore, even if these cases are fundamentally the same from the statistical mechanics viewpoint, where finite barrier heights are irrelevant, they may be very different from the biological, or practical, perspective. In other words, the pre-existing equilibrium picture may not hold for all proteins: some would have landscapes with higher barriers, which cannot be crossed on the timescales corresponding to the observed kinetic rates, and therefore binding is needed to physically induce the transition. Visualizing this situation within the energy landscape paradigm, one would observe continuous relocation of the minimum from conformation A to conformation B during the ligand binding process rather than pulling down of a pre-existing local minimum at B. On the other hand, 'relocation of the minimum' is a rather inaccurate term. Because of timescales involved in KNF, the 'relocating minimum' may be very local. On the joint ligand-and-protein landscape, the dominant KNF pathway would probably be similar to the one denoted as κ (Fig. 5). It should be occurring fast, on a nanosecond to microsecond timescale: by definition, the key step in this mechanism is binding (nanosecond timescale), which may be slowed down to some extent by the need for conformational rearrangement. Therefore, if one defines the allosteric transition coordinate along the path κ, the equilibrium transverse to it will only be on the timescales faster than 10^{-8} s. Nonetheless, since there is no timescale separation between binding and conformational change, it is convenient to think of the transition as adiabatic, with a free energy profile along the dominant transition path similar to the one shown in Fig. 5 (lower right).

Thus, the joint ligand-and-protein energy landscape can be roughly partitioned into two areas, corresponding to the two mechanisms, reminiscent of a phase diagram (Fig. 5, upper right). There is no strict boundary between the zones, but largely if the ligation starts earlier than conformational change (the systems move to the right before moving upwards), that corresponds to the KNF mechanism and the opposite corresponds to the MWC scenario.

Structurally, a range of intermediate situations between pure MWC ($A \rightarrow \beta \rightarrow B$) and KNF may also be possible or even likely. During the dynamics without a ligand, conformations other than A may be visited and they may resemble B without being exactly the same (Nelson *et al.* 2000; Xu *et al.* 2008). Resemblance to the B conformation could be complete, including orientations of the side chains in the whole molecule, or only at the level of similarity of the backbones to just a local reorientation at the active site, so that an enzyme could accommodate a ligand – and, subsequently, the induced fit via reaction front would take over from there. On the joint landscape in Fig. 5 (upper right), these scenarios would correspond to the pathways starting in the MWC zone

(upward) and, then turning closer to the KNF zone, maybe even crossing into it. Again, from this perspective the ω pathway seems to be more natural than the extreme A→β→B case.

Returning to the picture of the gradually changing protein energy landscape upon binding in the KNF scheme, the path κ in Fig. 5 corresponds to the case when there is a well-defined reaction coordinate for the transition. In that case, there is a dominant pathway (something like κ), and a free profile along that pathway should be similar to the one given in Fig. 5 (lower right). The landscape has to be intricately built to provide a dominant pathway. Possibly, in some cases, it is partially formed by frustration localization near the binding interface (Ferreiro *et al.* 2007). A single reaction coordinate is not the only option. The total timescale range in the KNF scenario (from the transition time (10^{-7} s–10^{-6} s) to the atomic vibrations (10^{-15} s)) is smaller than the one in the MWC scenario, so there is less space for a timescale gap that could help us to single out slower, important coordinates. It is possible, therefore, that one may need a considerable fraction of the conformational coordinates to properly describe multiple transition pathways.

Finally, intrinsically disordered proteins, which fold upon binding of a ligand, present a special case. Their folding landscape is often weakly funneled (Papoian, 2008), but in the state without a ligand, the backbone conformational entropy wins over the folded active structure, masking it among a myriad of other conformations, despite the fact that this structure can be quite low in energy.

3.2 Communication pathways

The main evidence supporting the reaction front picture is the experimentally found existence of 'communication pathways' across the bodies of allosteric proteins. For some allosteric protein families, which are evolutionarily related, variations of residues in certain positions turned out to be correlated with variations of residues in other positions, sometimes, quite distant in space (in 3D structure). One of the most studied examples is the PDZ domain family. In 1999, Lockless and Ranganathan approached the problem of interaction of distant residues from the evolutionary perspective (Lockless & Ranganathan, 1999). Double-mutant cycle to measure real thermodynamic interactions between the residues had existed for some time, but evolutionary conservation of such interactions in different homologues of the same enzyme would suggest that the interaction is required for function. Based on a large set of sequences a statistical 'free energy' was introduced. Mapping of this statistical mutation free energy (based on 274 PDZ domains) on a 3D structure of PDZ domain resulted in a continuous path consisting of residues with high statistical mutation free energy and going across the whole globule (Lockless & Ranganathan, 1999). The authors also introduced a statistical free energy of interaction between two positions, based on statistics from the same set of 274 sequences. These free energies turned out to be correlated with thermodynamic free energies of residue interactions obtained by mutant cycle analysis for one residue position of one of the PDZ domains. This coincidence points to the assumption that the identified pathway is indeed based on energetic couplings, although without revealing the coupling mechanism, which would be important for enzyme functioning. Subsequent double mutant cycle works on a PDZ domain, however, have challenged the claim of correlation between statistical and thermodynamic free energies (Chi *et al.* 2008).

In 2004, the actual dynamics of the residues along the communication pathway in the PDZ domain were elucidated by NMR (Fuentes *et al.* 2004). In the second PDZ domain from human tyrosine phosphatase 1E (hPTP1E), certain residues (connected by van der Waals interactions, forming thus a contiguous 'path') change their ps–ns side-chain dynamics upon binding of the

ligand. This change propagates up to ∼11 Å away from the ligand. The observed 'path' correlated with the evolutionarily conserved path reported by Lockless & Ranganathan (1999).

The PDZ domain is not the only case where a network of evolutionarily conserved residues is found (Benkovic & Hammes-Schiffer, 2003; Suel *et al.* 2003; Zheng *et al.* 2005). Mutations in residues distant in space from the binding site can change the enzymatic rate quite significantly, by up to a factor of 200 (Benkovic & Hammes-Schiffer, 2003), suggesting some form of communication between these distant residues and the enzyme active site. Whether these communication pathways have been optimized by evolution is an interesting question.

In many works, the communication is envisioned in a mechanistic fashion, using phrases like 'binding energy is not randomly dissipated, but is directed' or 'mechanical deformation' that propagates along the pathway. It is true that sterically connected atoms allow propagation of mechanical or heat waves (Leitner, 2008, 2009; Ota & Agard, 2005), but we recall that, first, the timescale for complete energy dissipation is on the order of fs–ps and, second, binding energy is actually a free energy. Thus, the latter, being directed along a pathway, may, for example, be stored, in side-chain entropy of the residues forming the pathway. The timescale of communication propagation is not known at this time. Mechanistic or quantum heat wave pictures are one possibility, but this mechanism is unlikely if the communication occurs on times longer than several picoseconds. We hypothesize another possible mechanism of front propagation: that the residues along the pathway are frustrated (Bryngelson *et al.* 1995; Ferreiro *et al.* 2007; Onuchic & Wolynes, 2004; Onuchic *et al.* 1995; Whitford *et al.* 2008a) (and thus predisposed to a change) and the binding of a ligand induces a 'flipping' of the pathway resides. An experimentally observed allosteric switching mechanism in caspase-1 may be reinterpreted using this idea (Datta *et al.* 2008).

Caspase-1 is a dimeric enzyme with two cooperatively working active sites on opposite sides of the dimer and an allosteric site at the center. In each subunit, there is a contiguous network of 21 hydrogen bonds formed by nine side chains and a salt bridge in the on state (when the active site is occupied). In the off state, there is a different network, with only 12 hydrogen bonds (only one bond is present in both networks). Thus, the hydrogen-bonding network rearranges upon binding, while maintaining a site-to-site connection (Fig. 6). These residues are well-conserved evolutionarily, and their mutations by alanine scanning show that four specific residues are important for enzymatic activity. Mutation of any of the two residues forming the salt bridge reduces the catalytic activity by 100–200-fold. Two other residues showed minor 4–7-fold activity reduction effect upon mutation (Datta *et al.* 2008).

For the PDZ domain in particular, the results of a very interesting computational work may be viewed as an indirect argument for the fast propagation of the allosteric signal. PSD-95, a member of the PDZ domain protein family, was cooled down to 10 K in an MD simulation without solvent, and then the key active site residue (His76) with ligand and one water molecule bound was heated to 300 K (Ota & Agard, 2005). The absence of explicit solvent is required to weakly hold the other atoms in their places and to introduce a distance-dependent dielectric constant. The authors observed the heat transfer through protein and found it to be highly anisotropic and in agreement with pathways obtained by Lockless & Ranganathan (1999). The timescale of this thermal diffusion was on the order of a few picoseconds. The fact that heat travels this fast means that there is a densely packed 'tube' along the direction of propagation. However, it does not necessarily follow that an allosteric signal propagates as fast, despite following the same pathway. In addition, despite general correspondence, there were differences between pathways obtained in the simulation (Ota & Agard, 2005) and the statistical analysis of

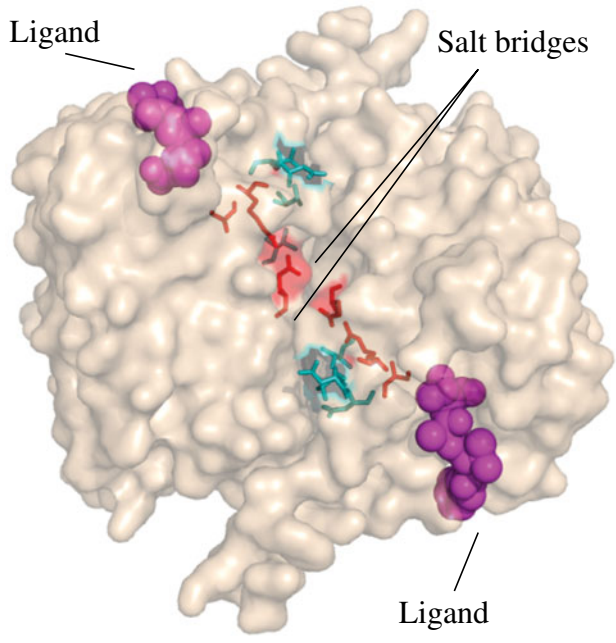

Fig. 6. The dimer caspase-1 has two binding sites connected by a hydrogen bond network that rearrange upon binding. In the figure, the sites are occupied by ligands (shown in purple). The residues participating in the hydrogen bond network are highlighted in red and blue. Red residues form a continuous path, hinting at the possibility of the communication pathway. Of utmost importance are the two residues on the interface between the subunits that form salt bridges. Their mutation to alanine drops the enzymatic rate by a factor of 200 (Datta *et al.* 2008).

sequences (Lockless & Ranganathan, 1999) (although, according to the authors, it may be attributed to the fact that the simulation was done for a particular sequence, PSD-95, while the statistical analysis was done using 274). A similar approach is possible, where some localities in a molecule are 'pumped' by a harmonic perturbation (10–100 ps scale), and other parts are being 'probed' for response (Sharp & Skinner, 2006).

The idea of a 1D communication pathway that is used for allosteric signal propagation is intriguing, but it requires more direct evidence to become an established paradigm. As we argue next, the results described so far largely correlate with the existence of a high-density region(s) within the protein core. The residues in the evolutionarily conserved contiguous path reside in the high-density core and thus have a high burial free energy, which is picked up by double mutant cycle analysis. Therefore, it is also not surprising that they are more conserved, as the change in the core is energetically expensive. The NMR studies do point to a contiguous pathway of residues dynamically responding to the binding, but it does not directly follow that this is a pathway for allosteric conformational change.

Although the pathway concept appeals to our everyday experiences, a front propagating through 3D medium is most naturally expected to be a 2D front. In reality, of course, it can be a manifold of any dimension lower than three (including fractal) down to zero. A zero-dimensional reaction front corresponds to the communication pathway picture and seems quite exotic from this point of view. In fact, many studies, even before 1999, were starting from seeking correlation in structural or dynamical change in distant residues upon binding (e.g. any allostery, not

necessarily a pathway across the protein) (Benkovic & Hammes-Schiffer, 2003; Gianni *et al.* 2006; Hilser *et al.* 1998; Lenaerts *et al.* 2008; Niu *et al.* 2007; Zhuravleva *et al.* 2007). They would often find that many distant residues 'respond' to a residue near the active site, and vice versa. Inside the nascent communication pathway paradigm initiated by Lockless & Ranganathan (1999), such results were formulated as the presence of 'communication networks' inside a protein molecule. For instance, 14 conservative (i.e. deleting methyl groups in the side chain without altering the stereochemistry of the residue) mutations to the second PDZ domain from the tyrosine phosphatase PTP-BL (PDZ2) led to changes in the rate constants, despite the fact that the deleted side chains were not participating in the interactions with the ligand (Gianni *et al.* 2006). Authors state that most of the affected residues form a network of hydrophobic interactions in the core. However, four similar mutations in the third PDZ domain of PSD-95 did not affect the rate constants, so the effect is PDZ2 specific (Gianni *et al.* 2006). Residues showing a change in millisecond dynamics, measured by NMR, in AF-6 PDZ domain upon ligand binding were also named a dynamical network (Niu *et al.* 2007). In barnase, 11 residues that are distant from the binding site show a change in 10–100 ps dynamics upon binding of barstar (Zhuravleva *et al.* 2007).

Thus the paradigm leads to the temptation of representing the complicated distribution of dynamical or structural perturbations inside a protein molecule as a large number of communication pathways. From a MD simulation in the CHARMM force field with some restrictions derived from NMR experiments, Dhulesia *et al.* report the presence of two networks in the second PDZ domain of human tyrosine phosphatase 1E that consist of several hundreds of pathways (Dhulesia *et al.* 2008). With the total number of residues in the PDZ domain being just below 100, it is not clear whether this picture of many intertwined pathways does anything to help envision the reaction front propagation. Assuming that a protein should be envisioned as a spaghetti coil of 'communication cables', perturbing one of them would stir the rest, inducing cross-talk across the whole molecule; but the reported pathways consist only of 4–5 residues and most of the residues are involved in many pathways. In this case, it seems more natural to think of the single higher dimensional fractal manifold formed by the perturbations.

Therefore, caution is needed when exploring communication pathways by either experimental or computational means. However tempting, one should not assume their prevalence in proteins before further direct proof is obtained. For instance, a fast and inexpensive computational method for finding the communication pathways was developed and applied to study the power stroke of myosin (Tang *et al.* 2007). The method is based on seeking the shortest contiguous path of physically interacting residues, which are also evolutionarily conserved. This method is guaranteed to find pathways, but further tests will be needed to examine their relevance to allostery. On the other hand, the hydrogen bond network in caspase-1 is indeed a site-to-site pathway, and that pathway is both conserved evolutionarily and important for activity (Datta *et al.* 2008), but there is no evidence that any reaction front travels along it. In fact, the authors of this alanine scanning study discussed their results in the framework of the pre-existing equilibrium mechanism (Datta *et al.* 2008).

The NMR methods provide the most direct evidence of pathways in a PDZ domain; however, they do not show pathways for all proteins, for example, calmodulin (Igumenova *et al.* 2005) or eglin c (Clarkson *et al.* 2006), where some of the mutations invoke contiguous response, while response to others is non-contiguous. Uncoupling between the dynamics of certain residues upon transition (that had been dynamically coupled before transition) may also play an important role (Clarkson *et al.* 2006; Lenaerts *et al.* 2008). We suggest that signal propagation also may be

engendered by specific distribution of local frustration across the molecule. A recently developed method allows the measurement of frustration localization (Ferreiro *et al.* 2007) and may help to shed light on this issue. Finally, the ps timescale in the case of mechanical signal propagation over a high-density pathway is too removed from the allosteric timescales, which are usually in the ns–ms range.

We conclude this section by suggesting a picture that synthesizes some of the discussed mechanisms: communications may occur via a front of highly frustrated residues, forming a 'frustration tube', with fractal dimensionality likely higher than one, where the tube residue conformations 'switch' upon binding of a ligand. This possibly might be the case for the caspase-1 dimer, discussed above (Datta *et al.* 2008). In general, it may be expected that different mechanisms operate to couple distant residues in various allosteric proteins and enzymes. Even if the original communication pathway paradigm does not prove to be ubiquitous, it has spurred very fruitful discussion among the community, bringing under focus the issue of complicated communication networks in proteins. In particular, these questions are raised directly in the real space, as opposed to the picture based on collective coordinates, which had been more common.

4. Landscapes

Up to this point we have been highlighting various achievements in understanding the protein native state through the lens of the energy landscape paradigm, but the original works themselves did not largely emphasize these ideas. In this section, we review works done with the energy landscape paradigm in mind from the inception. First, we describe the works that try to quantitatively characterize the native state energy landscapes. Since the latter are highly multi-dimensional, the characterization amounts to mapping the landscape at a reduced representation, where a number of different representations and methods of mapping have been employed. Then we present several case studies where landscapes are used to explain protein dynamics and function.

4.1 Revealing the landscape topographies

One straightforward, yet computationally demanding approach to constructing a representation of the protein energy landscape is based on isolating conformational states from the protein dynamics (usually, realized as an MD or Monte Carlo simulation), assigning statistical weights to these states, and computing rates for inter-state transitions. This approach avoids the need for selection of a few collective coordinates, where the latter tends to provide a rigid frame for the whole study. For this reason, the choice of collective coordinates is extremely important and, as discussed earlier, can be highly non-trivial. On the other hand, mapping the conformational space onto a network only requires an order parameter that is good *locally*, providing a criterion for clustering the multitude of recorded structures into network nodes (conformational states). Since these requirements are looser, it is usually easier to find such a criterion. It may be based on analyzing RMSDs (Krivov & Karplus, 2004), contact maps derived from positions of C_α atoms (Hori *et al.* 2009) or mutual overlaps, Q. The edges of the network may be derived by analyzing the actual transitions in a simulation (Caflisch, 2006; Khalili & Wales, 2008; Krivov & Karplus, 2002, 2004; Krivov *et al.* 2008; Rao & Caflisch, 2004) with corresponding rates computed from the number of times a particular transition has occurred in the simulation. Alternatively,

assuming that the structure is correlated with contact energy, one places edges between the nodes whose contact maps are sufficiently close (Hori *et al.* 2009). From the network and estimated rate constants, a master equation can be constructed and studied analytically or solved numerically, using a method such as Gillespie simulations (Berezhkovskii *et al.* 2009; Buchete & Hummer, 2008; Chodera *et al.* 2007; Hinrichs & Pande, 2007; Hori *et al.* 2009).

The master equation approach was developed in the context of protein folding, where the phase space is much larger compared with the size of the functional landscape. Sampling all of it with explicit solvent simulations is beyond computational reach for most of the current supercomputers. Therefore, work using this approach has to date mainly employed force fields with implicit solvent. However, with current state-of-the-art and future computational resources, it should be possible to accurately sample the protein's functional landscapes using explicit solvent.

Experiments by Frauenfelder and Ansari suggested that the basins of the native state ensemble contain sub-basins, which themselves may be subsequently divided further into substates, recursively, thus forming a hierarchical organization of the protein energy landscape (Frauenfelder *et al.* 1990; Henzler-Wildman & Kern, 2007). Based on this idea and somewhat related to the network approach above, the protein's landscape has been projected into a hierarchical tree with the help of a technique based mainly on PCA (Materese *et al.* 2008). This requires a long, straightforward sampling of the landscape in a simulation. PCA is a technique for analyzing trajectories based on coordinate transformation, where the new coordinates are the eigenvectors of the covariance matrix of the old coordinates. Thus, the new collective coordinates (called principal components) behave as independent random variables. They can be ranked by variance, so that the first principal component (PC) describes the largest fluctuations in the trajectory. In this way, most of the protein motions, particularly, large-scale motions (note that functional motions are usually large in scale), are well represented by only the first few PCs. When the simulation trajectory is projected into these PCs, multipeak distributions are observed, with peaks corresponding to different conformational states (García, 1992). On the other hand, the vast majority of PCs are characterized by a narrow unimodal Gaussian-like distribution, corresponding to vibrational fluctuations (García, 1992). A PC analysis of the 70-residue protein eglin c was recently carried out using the CHARMM force field with explicit solvent (Materese *et al.* 2008). The first two PCs, PC1 and PC2, contain the largest conformational fluctuations and turned out to be multipeaked. Several distinct regions of the (PC1, PC2) plot are highly populated, or, represent basins of the free energy surface (Fig. 7).

Furthermore, zooming into a particular basin allows exploration of finer splittings: plotting the population using PCs of higher number as coordinates reveals clearly identifiable sub-basins. Each of them may be scrutinized further in the PCs of even higher number and so on if there is sufficient sampling of the next-level sub-basin from the MD trajectory. Hence, the novelty of this approach was not only to project the protein dynamics into the first two PCs, which is common, but also follow basin splitting in PCs of higher dimensions (Materese *et al.* 2008). For eglin c, five hierarchical tiers were found from a 330 ns simulation. Figure 7 shows a tree-like organization of conformational states, providing microscopic detail to earlier conceptual arguments on the hierarchical organization of the protein native state (Frauenfelder *et al.* 1990; Henzler-Wildman & Kern, 2007).

The tree plays a role similar to that of networks discussed above, that is, shows the connections between the conformational states of the same tier. The tree defines an ultrametric (An ultrametric is a metric that satisfies the following strengthened version of the triangle inequality. Metric is a rule that defines a 'distance' between any two elements of the set. If it is an

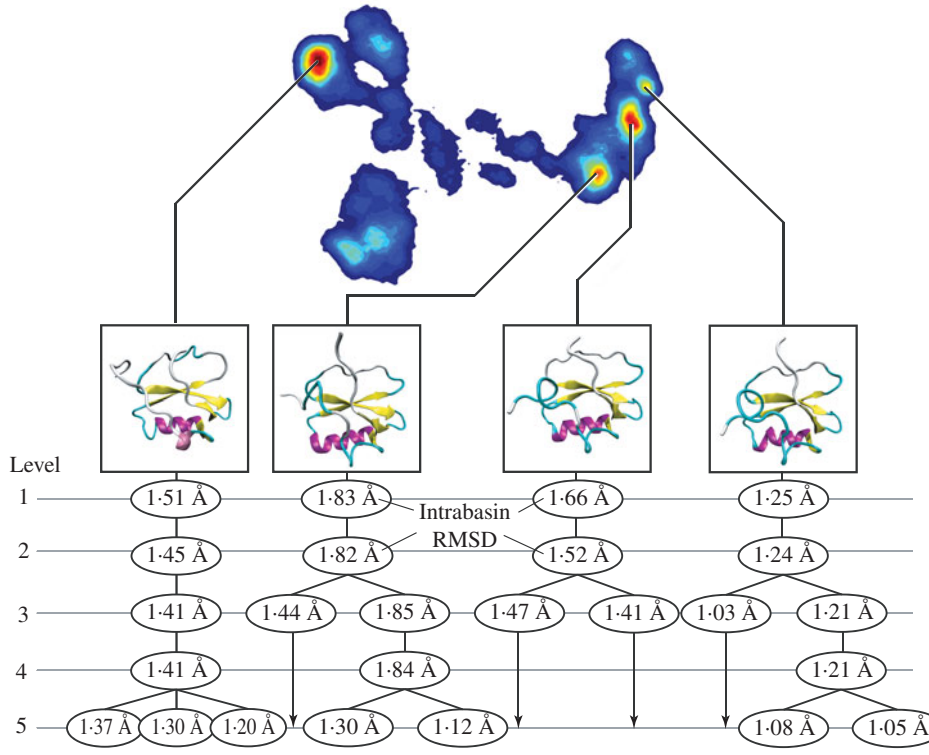

Fig. 7. Tree-like hierarchy of the functional landscape of eglin c is revealed by PCA. In the higher index PCs, the basins show finer splitting, like a basin on a (PC2, PC3) plot that splits into two sub-basins on a (PC3, PC4) plot. Five hierarchical tiers were identified. The number of tiers up to the first common parent basin can serve as ultrametric 'distance' between any two sub-basins on the same tier. This ultrametric can be used as a basis to generalize rate constants in kinetic schemes of landscape mapping from partial sampling of the landscape (Materese *et al.* 2008). We thank Christopher K. Materese for preparing the figure.

ultrametric, then among three pairwise distances between any three elements, two are equal and the third one is smaller. An ultrametric in a tree can be defined as the number of levels upward to the first common ancestral branch.) among its same-level branches, and assuming that the ultrametric reflects actual physical properties of the energy landscape, such as barrier heights, one can use it to assign the transition rates between the conformational states on a particular tier: the closer the states by the tree ultrametric, the higher the transition rate between them. In the next approximation, typical barrier heights or even rates can be calculated for each possible ultrametric distance (there is only as many values as there are levels in the trees) by performing separate simulations, several for each distance. This would be a way for the reconstruction of the energy landscape 'from the bottom'.

This discussion of the populated basins in terms of PCs suggests that PCs may be good coordinates for yet another representation of the protein energy landscape – construction of free energy surfaces (García & Sanbonmatsu, 2001; Maisuradze *et al.* 2009b). They have been used for that purpose in combination with the kinetic network approach, for protein G and SH3 domain (Hori *et al.* 2009). In that study, PCs were based not on Cartesian coordinates of the atoms or dihedral angles, but on the contact order map of the protein, taking advantage of the funneled shape of the energy landscape. Along the same lines, Q – the fraction of contacts shared with the

native state – may also be used. That approach has been used for 20-residue peptide Trp-cage (Wu *et al.* 2008). The authors took two native-like structures from an NMR study and defined *two* Qs, with respect to either of them, using these two coordinates to build a 2D free energy surface, unfurling the native basin at high resolution (Wu *et al.* 2008), which is elaborated below.

4.2 Applications of the landscape paradigm to protein dynamics and function

When a landscape representation is available, detailed investigation may shed light on the nature of protein dynamics and function, such as mechanisms of enzymatic catalysis or molecular motor stepping. A free energy surface in two dimensions is a potent representation of the underlying energy landscape, although building it requires significant computational efforts (Pisliakov *et al.* 2009; Wu *et al.* 2008). As a first step, one has to choose good order parameters as two collective coordinates – that is, the remaining transverse degrees of freedom should be much less important for the questions asked, and, preferably, quickly equilibrate. These are strong requirements, and a choice of coordinates, therefore, is a significant approximation. Would a pair of coordinates be enough for an adequate description? How many coordinates does one need? As an example, Materese *et al.* have shown through sub-microsecond simulations that at least in one case, eglin c, only on the order of 5 out of about 300 conformational coordinates are required for a comprehensive description of the non-trivial part of the native protein dynamics (Materese *et al.* 2008). However, not all features of the protein's conformational dynamics might be of importance for a given problem; hence, lower-dimensional surfaces still play a significant role.

PCA relies on the actual dynamics of a protein system during some period of observation to find the coordinates that optimally represent motions. If there is no initial knowledge about the system dynamics or energy landscape, the choice of collective coordinates must be based on physical plausibility. Q is correlated with folding, but in the bottom of the funnel the motions transverse to Q may also be very important because of the rugged topography of the functional landscape. Interestingly, the choice of two Qs, with respect to two different (but similar) basins in the native state, provides both correlation with the funneled shape of the landscape, as a $Q_1 = Q_2$ diagonal of the ($Q_1 = 0 \ldots 1$, $Q_2 = 0 \ldots 1$) square, and description of the motions transverse to folding, which are transverse to the above-mentioned diagonal (Wu *et al.* 2008). The diagonal direction allows describing native excitations that occur through partial unfolding. The transverse to diagonal motions are resolved non-uniformly throughout the square – they are highly resolved as the native state is approached in the upper-right corner (see Fig. 8) and completely unresolved near the origin (Wu *et al.* 2008). After computing Free energy surface (FES) with the chosen coordinates, Langevin dynamics may be used to run trajectories on this surface, which requires many orders of magnitude less computational effort than a direct simulation of protein dynamics. Therefore, it becomes feasible to run tens of thousands of trajectories, producing various first passage time moments and distributions.

One important question about native basin energy landscapes concerns the nature of conformational dynamics occurring on it. In particular, dynamics are called activated (Debenedetti & Stillinger, 2001) if the molecule spends most of the time in the landscape minima, rarely hopping from a minimum to a minimum, with the average waiting time given by the Arrhenius law. This is a typical situation for a liquid that is slightly supercooled. When the relaxation times become longer than the observation window, the dynamics are considered glassy. On the contrary, if the system spends most of the time cruising from one saddle point to another on the landscape,

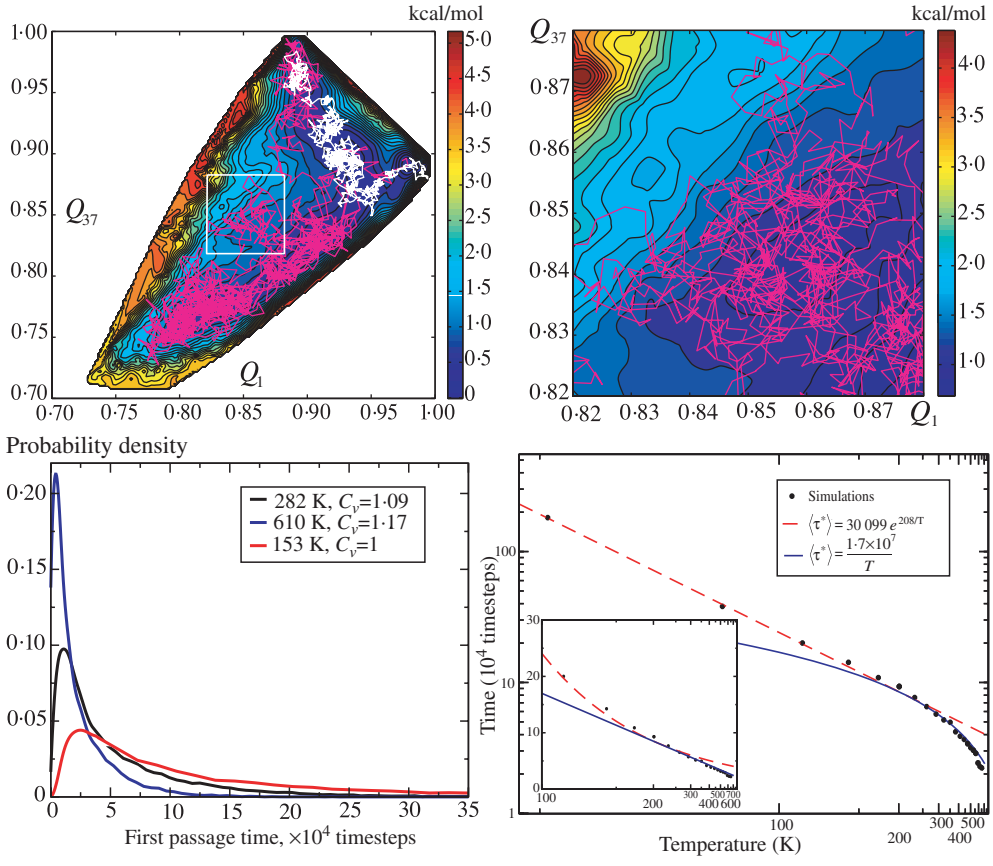

Fig. 8. A free energy surface is constructed for Trp-cage through 2D umbrella sampling. The two coordinates are the fractions of contacts shared with either of two similar structures labeled as '1' and '37'. Magenta and white on the upper panel denote two overdamped Brownian dynamics trajectories. The upper right panel shows a magnified region of the surface, revealing the sensitivity of the trajectory to fine features of the surface, such as shallow traps and low ridges. These features broaden the distribution of first passage times from '37' to '1', but there are no conspicuous deep traps that would engender classification. To get reference for the first passage time distribution at room temperature, two more First-passage time (FPT) distributions were calculated at other temperatures with their coefficients of variation–square-root-of-variance-to-mean ratios (lower left panel). C_v for the room temperature is near 1, which is considered a boundary between 'narrow' and 'broad' distributions. The dependence of the mean first passage time on temperature also changes around room temperature from Arrhenius (corresponding to activated dynamics) to reverse proportionality (corresponding to confined diffusion) (lower right panel). These facts lead to conclusion about the dynamical behavior of Trp-cage at room temperature being borderline between activated and diffusional (Wu *et al.* 2008).

the dynamics are diffusive, similar to those of a normal liquid (La Nave *et al.* 2000; Schulz *et al.* 2009).

To ascertain which of these dynamical regimes is relevant for a small protein Trp-cage, Wu *et al.* computed a 2D free energy surface using two structure-based order parameters, and ran on it hundreds of thousands of Brownian dynamics trajectories (Wu *et al.* 2008). In particular, different sets of simulations were computed at various temperatures. By analyzing the temperature dependence of the mean first passage times to transition from one native-like state to another as well as the shapes of the first passage times distribution, it was concluded that Trp-cage room

temperature dynamics are borderline between diffusive and activated dynamics (Wu *et al.* 2008). To further refine these types of simulations, the dependence of the friction constant on the surface position should be taken into account (Socci *et al.* 1996; Wang *et al.* 2005). The latter inhomogeneity arises from the ruggedness created by the degrees of freedom that are transverse to Q_1 and Q_2. However, these types of simulations are most successful when surface coordinates are appropriately chosen to describe a particular function or process. For instance, Heath *et al.* have elucidated the mechanisms of misfolding in a mutant of a protein S6 with the help of a free energy surface in terms of Q (fraction of native contacts) and the fraction of non-native contacts (Heath *et al.* 2007).

In a different example, where the need for two collective coordinates arose naturally, it was recently suggested from interpretation of experiments that allosteric and catalytic motions are coupled in adenylate kinase (Henzler-Wildman *et al.* 2007a, b; Kern *et al.* 2005). To address this question in detail, Pisliakov *et al.* performed multiscale simulations (Pisliakov *et al.* 2009), building a 2D free energy surface in coordinates relevant to the question: a chemical coordinate (corresponding to the chemical reaction being catalyzed) and a conformational coordinate (corresponding to the conformational motion of the enzyme). Pisliakov *et al.* performed MD simulations on two scales: all-atom and coarse-grained, changing the parameters of the coarse-grained model such that it reproduces the autocorrelation functions computed from atomistic dynamics. Then the analytical 2D surface was built, and the parameters were tuned in a similar fashion to achieve agreement with both of the more detailed models and also two rate constants obtained in FRET and NMR studies (Henzler-Wildman *et al.* 2007b). Studying the dependence of the first passage times from the reactant state to the product state at various barrier heights and friction constants, Pisliakov *et al.* have concluded that their simulations indicate no direct channeling of conformational motion excitations into chemical transformation (Pisliakov *et al.* 2009). This was explained by a timescale separation between picosecond and nanosecond solvent motions and coherent protein modes, on the one hand, and millisecond chemical catalysis step, which is long enough for the coherent protein motions to completely dissipate (Pisliakov *et al.* 2009). These types of questions may be explored in the future with the help of *ab initio* MD simulations, which treat protein motions and chemical rearrangements on an equal footing (Papoian *et al.* 2003b).

When function is strongly coupled to the folding landscape, as may be the case with intrinsically disordered proteins (Papoian, 2008), coordinates related to folding have to be used for building the FES. Using Qs, Turjanski *et al.* (2008) have constructed an FES for a natively unstructured transcription factor, and found that binding of the ligand precedes folding of the transcription factor, indicating a coupling between binding and folding in accordance with NMR experimental data (Sugase *et al.* 2007). In a computational follow-up of an interesting recent finding of catalytic activity in a disordered enzyme (Pervushin *et al.* 2007), radius of gyration and contact order were used for the 2D free energy surface (Roca *et al.* 2008). Subsequent analysis indicated that upon substrate binding the molten globule-like phase partially folds (Roca *et al.* 2008). Coupling between binding and folding is a known and frequent phenomenon (Papoian & Wolynes, 2003). Evolution may have produced natively disordered enzymes to provide finer temporal control over enzymatic activity (as unfolded chains are faster degraded by proteosomes) and acceleration of the binding kinetics through a 'fly-casting' mechanism (Hyeon & Onuchic, 2007; Shoemaker *et al.* 2000) (as an unfolded chain occupies a larger volume) (Papoian, 2008). One of the possible realizations of the coupling between binding and folding is a frustration localization near the binding site. Upon binding, the frustration disappears and

induces a conformational change, resulting in folding (Ferreiro *et al.* 2007). Papoian and Wolynes developed a theory of the coupling between binding and folding energy landscapes (Papoian & Wolynes, 2003), which can be used to study binding and folding thermodynamics and kinetics (Levy *et al.* 2004; Marcovitz & Levy, 2009; Wang *et al.* 2006). Currently, computational power allows construction of the 2D FES even with explicit solvent all-atom force fields, as was applied to study metal-coupled folding of a zinc-finger motif (Li *et al.* 2008).

With the help of free energy surfaces, Hyeon and Onuchic have elucidated the mechanism of chemical energy transduction into mechanical motion in kinesins (Hyeon & Onuchic, 2007). Kinesin is a molecular motor that walks in 8 nm directional steps along microtubules consisting of 13 protofilaments. Kinesin consists of two heads, connected by a neck region, that bind to specific sites on the protofilaments. A conformational switch of the neck, called a power stroke, precedes binding of the tethered (not bound) head to the microtubule, which, in turn, occurs after a non-directional diffusive search for a binding site. In this work, several different free energy surfaces were computed for two different purposes (Hyeon & Onuchic, 2007). First, the authors considered the question of which microtubule-binding site is preferred by a motor head. To address it, Hyeon and Onuchic constructed an FES corresponding to a potential of mean force for the tethered head as a function of its three spatial coordinates. Two such surfaces were computed for both conformational states of the neck region using hybrid Monte Carlo–MD simulations with a coarse-grained force field (Hyeon & Onuchic, 2007). Examination of the surface cross-sections allowed the researchers to identify several basins of attraction corresponding to specific binding sites on the same or different protofilaments in addition to a separate broad entropic basin. The most favorable binding site turned out to be on the same protofilament that hosts the other motor head. Hyeon and Onuchic then combined the two surfaces (for two different conformational states of the neck region) in a single time-dependent surface that would imitate the conformational switching of the neck region during the power stroke. In this combined potential, they ran Brownian dynamics treating the tethered head as a quasi-particle. Depending on the timescale of potential switching (the power stroke), the trajectory may be trapped in the intermediate state, corresponding to the head bound to the nearest binding site (the one on the neighboring protofilament, see Fig. 9). Hyeon and Onuchic also built free energy surface in terms of Q_{p} and Q_{int}, which are the fraction of native contacts in the microtubule-binding motif of kinesin and the fraction of interfacial native contacts between kinesin and microtubule. Examining these surfaces showed that cracking – disruption of some kinesin native contacts – occurs prior to binding. The cracking increases the binding rate because of enhanced flexibility of the partially unfolded structure, through the fly-casting mechanism. Interestingly, the cracking happens right before the moment of binding.

Motivated by the allostery problem, the idea of cracking may also be used to extend the Gō model to a protein that has two well-defined conformational states with a transition between them. This method of constructing protein energy landscapes was used to study glutamine-binding protein (GBP), S100A6 (structural analogue of calmodulin), dihydrofolate reductase and HIV-1 protease (Okazaki *et al.* 2006). Let us assume that there is a Gō-like force field $V_1(R, R_1)$, where R represents the coordinates of an arbitrary conformation and R_1 are the coordinates in the native conformation. If there is another native basin, the corresponding Gō model will be given by a potential $V_2(R, R_2)$. Next, one wants to construct a single potential where both basins are present, though, perhaps with a difference in stability (for example, the second basin is shifted by ΔV). To merge the two surfaces together, techniques from the Marcus theory of electron transfer (Marcus & Sutin, 1985) may be borrowed, and used to 'sew' the two Gō-potentials

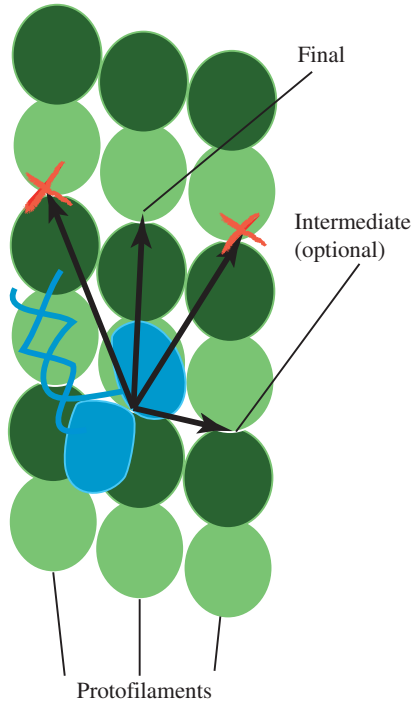

Fig. 9. A microtubule consists of 13 protofilaments made of tubulin heterodimers. Kinesin motors walk on microtubules directionally in 8 nm steps. Hyeon and Onuchic have shown that when one motor head is bound the other one binds to the active site on the same protofilament with a possible intermediate state on the neighboring binding site. (Hyeon & Onuchic, 2007).

together. Introducing a coupling constant Δ, Okazaki *et al.* (2006) obtained a smooth double-basin potential V_{MB} given by

$$\begin{vmatrix} V_1(R,R_1)-V_{MB} & \Delta \\ \Delta & V_2(R,R_2)-\Delta V-V_{MB} \end{vmatrix} = 0,$$

the equation for eigenvalues. The eigenvector identifies the basin in which the protein is residing and therefore may be used to construct a reaction coordinate. Plotting free energy of GBP as a function of this reaction coordinate in addition to its components – the average energy and conformational entropy – led the authors to observe a sudden increase in conformational entropy in the middle region of the transition profile. This is the hallmark of cracking – partial unfolding and breaking some of the native contacts (Hyeon & Onuchic, 2007; Hyeon *et al.* 2009; Miyashita *et al.* 2003, 2005; Okazaki *et al.* 2006; Whitford *et al.* 2007, 2008a, b). Again, it turned out to be useful to plot free energy surfaces in terms of Q_1 and Q_2, that is, the fractions of contacts shared with either one of the native states, respectively. For GBP this surface showed that there is a straight valley between the two minima corresponding to the two states, which means that breaking of contacts specific to the initial basin and formation of contacts specific to the final basin occur simultaneously.

The allosteric transition of S100A6 is more sequential: first, contacts specific to the first basin break and then new contacts are formed, leading to greater unfolding. Different coordinates for the surfaces, Q_{common} and Q_1–Q_2, where Q_{common} is a fraction of the contacts common to both

basins that are present in the conformation, provided another perspective on the transition dynamics. Authors plotted free energy surfaces at the folding temperature and then below the folding temperature and observed the appearance of an intermediate state for S100A6 at the lower temperature (Okazaki *et al.* 2006). Another method of combining several Gō potentials, based on summation of the partition functions, rather than secular equation, has also been reported (Best *et al.* 2005).

A different way to conjugate two structure-based (Gō-like) potentials has been used with adenylate kinase (Whitford *et al.* 2008a). The enzyme consists of three domains: CORE, NMP and LID. The conformational change from open to closed structure consists of motions of the NMP and LID domains. In this study, a potential based on the open structure was modified by the addition of a bias favoring contacts present in the closed structure. A 2D free energy surface was constructed through MD simulations in terms of distances between CORE and NMP centers-of-mass and between CORE and LID centers-of-mass (luckily, adenylate kinase has these natural reaction coordinates). Whitford *et al.* also constructed transition paths for opening and closing using the adiabatic (non-linear) normal mode technique discussed in the first section of the review (Miyashita *et al.* 2003). Interestingly, the pathways largely follow through the valleys of the landscape, especially, the closing transition.

Yet another possible set of coordinates to choose for plotting free energy surfaces may be derived through the Scalable ISOMAP technique (Das *et al.* 2006). From a practical viewpoint, it appears similar to PCA, that is, it reduces the dimensionality of the dynamics, but does so in a non-linear way. Using surfaces with these coordinates, Haspel *et al.* identified contacts that shape the native basins of C3d/Efb-C and two of its mutants (Nurit *et al.* 2010).

4.3 Role of water in shaping the landscape

Many of the works in this section, especially those dealing with large proteins, relied on implicit solvent Hamiltonians (Chen & Brooks, 2007; Pincus *et al.* 2008). However, the role of water in sculpting the functional landscape is extremely important, as elucidated by Materese *et al.* (2008) through examining water-mediated contacts, which may cause splitting of native basins into sub-basins. The presence of water-mediated contacts is frequent in the native state (Zong *et al.* 2006; Papoian *et al.* 2004). The role of water in protein folding has also been appreciated for a long time (Cheung *et al.* 2002; Head-Gordon & Brown, 2003; Kauzmann, 1959; Kaya & Chan, 2003; Lucent *et al.* 2009; van der Vaart *et al.* 2000), where the hydrophobic effect is considered to be the primary driving force for folding (Dyson *et al.* 2006). However, water-mediated hydrophilic interactions were also found to play an important role (Papoian *et al.* 2004). Dynamic coupling between water and proteins usually occurs on fast timescales. When a protein is effectively caged by solvent, the motions of the former become slaved to those of the latter (Austin *et al.* 1975; Frauenfelder *et al.* 2007). Such coupled collective modes can be probed by THz spectroscopy (Ebbinghaus *et al.* 2008; He *et al.* 2008).

On the other hand, functional motions of a protein chain are slow, so there is a timescale gap between the protein and solvent motions. Therefore, in most of the problems discussed in this review, the picture of a chain moving in a thermally averaged field of solvent motions is rather appropriate. In this way, water smoothes the energy landscape (Zhuravlev *et al.* 2009), which, in turn, allows us to talk about effective interaction energies, where solvent degrees of freedom are integrated out. Although implicit solvent models are physically justified by these considerations and are frequently used, it is desirable to account for the role of water as accurately as possible,

for instance, by including water-mediated contacts through a second well in potentials between polar groups, or through the introduction of non-pairwise potentials that are dependent on residues' water exposure (Papoian *et al.* 2004; Zong *et al.* 2006). The resulting representation of the long-range hydrophilic contacts is more rigorous and improves structure prediction (Papoian *et al.* 2004; Zong *et al.* 2006). Alternatively, explicit solvent simulations may be used, which are computationally much more demanding.

5. Summary and outlook

With protein folding generally understood (Bryngelson & Wolynes, 1987; Bryngelson *et al.* 1995; Chen *et al.* 2008; Dill *et al.* 2008; Dobson *et al.* 1998; Frauenfelder *et al.* 1990, 1991; Hyeon & Thirumalai, 2003; Onuchic *et al.* 1995; Pande *et al.* 2000; Plotkin & Onuchic, 2002; Scheraga *et al.* 2007; Shakhnovich & Gutin, 1993), the attention of scientists is turning more and more to the functional transitions and native dynamics of proteins. Many of the ideas from protein folding can be helpful in understanding the native landscape as well; the available computational power is constantly growing (Dror *et al.* 2009); new experimental techniques are appearing – all of this encourages pursuing a detailed study of the native dynamics, which is required because of the glassy rugged nature of this landscape. Luckily, native dynamics take place within a much smaller phase volume.

In this article, we broadly overviewed a range of computational techniques and ideas, as well as experimental results, that caught our attention, (re)interpreting them from the energy landscape paradigm perspective. One of the frequently encountered important questions both in computational and experimental studies is that of the number and nature of variables that can properly describe the protein's dynamics and function. PCA can show how many and which degrees of freedom are needed to describe the relevant part of the dynamics. NMA can provide coordinates correlated with topologies of allosteric states and in rare cases can even predict the other allosteric conformation. Normal modes can be enhanced through a non-linear adiabatic technique to describe putative transition paths between the allosteric states and account for a partial unfolding of chains known as *cracking*.

The energy landscape paradigm serves as a good general basis for uniting the terminology that exists in the area studying the allosteric transitions: MWC, pre-existing equilibrium, population shift, 'binding follows folding' scenarios as opposed to KNF, induced fit, reaction front, 'binding precedes folding' types of scenarios. We highlighted some of the experimental results that contribute to the understanding of different mechanisms from the energy landscape viewpoint. Despite the immense multidimensionality of protein functional motions, it may turn out that the native excitations may be classified into a small number of classes. At this point, we know of cracking excitations that push the molecule up the funnel, propagation of reaction fronts across the molecule or mechanistic hinge opening motions dictated by the protein's architecture. Interestingly, the first two examples are well localized, while the latter example represents global rearrangements. This is one property (local versus global) that can be used in the classification of the excitations. There may be motions of intermediate localization in the molecule, including even fractal. Other classification properties for excitation types may include distinct timescales, change in frustration, or change in entropy. Detailed studies of the landscapes, motions, and distributions of frustration in different proteins followed by analyses of the factors that favor certain classes of rearrangements can bring us closer to the cataloguing of

various types of functional motions. If such a task is feasible, one can envision a table of possible protein motions along with the rules explaining in which cases which types of motions are expected to dominate. Ultimately, perhaps by looking at the protein's architecture and topology and the chart of the spatial distribution of frustrations in the protein, one could predict motions that contribute to function.

To this end, the techniques that have been rapidly developing can help to elucidate the reasons leading to a particular type of excitation. For instance, free energy surfaces may reveal tendencies for partial unfolding, while identification of the inter-connected set of native sub-basins and examination of the contacts shaping them can give insight into the frustration distributions. Several such techniques along with numerous interesting examples of their application were presented in the last section of this review. Although a universal theory of protein functional dynamics does not yet exist, we are hopeful that various recent approaches outlined in this review, along with yet to be discovered techniques, will coalesce into a self-consistent theoretical framework, which, in turn, will be highly predictive and will be used to provide detailed mechanistic insights and interpretations to many experiments.

6. Acknowledgements

We gratefully acknowledge support from the Camille Dreyfus Teacher-Scholar Award, the Arnold and Mabel Beckman Foundation Beckman Young Investigator Award and the National Science Foundation CAREER Award CHE-0846701. We thank Christopher K. Materese for his help with editing the manuscript.

7. References

Adcock, S. A. & McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical Reviews* **106**, 1589–1615.

Alexandrov, V., Lehnert, U., Echols, N., Milburn, D., Engelman, D. & Gerstein, M. (2005). Normal modes for predicting protein motions: a comprehensive database assessment and associated web tool. *Protein Science* **14**, 633–643.

Antoine, M., Boutin, J. A. & Ferry, G. (2009). Binding kinetics of glucose and allosteric activators to human glucokinase reveal multiple conformational states. *Biochemistry* **48**, 5466–5482.

Arora, K. & Brooks, C. L. (2007). Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proceedings of the National Academy of Sciences, USA* **104**, 18496–18501.

Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. (1975). Dynamics of ligand binding to myoglobin. *Biochemistry* **14**, 5355–5373.

Ayers, S. D., Nedrow, K. L., Gillilan, R. E. & Noy, N. (2007). Continuous nucleocytoplasmic shuttling underlies transcriptional activation of PPAR gamma by FABP4. *Biochemistry* **46**, 6744–6752.

Bae, S.-H., Legname, G., Serban, A., Prusiner, S. B., Wright, P. E. & Dyson, H. J. (2009). Prion proteins with pathogenic and protective mutations show similar structure and dynamics. *Biochemistry* **48**, 8120–8128.

Bahar, I. & Rader, A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Current Opinion in Structural Biology* **15**, 586–592.

Benkovic, S. & Hammes-Schiffer, S. (2003). A perspective on enzyme catalysis. *Science* **301**, 1196–1202.

Benkovic, S. J., Hammes, G. G. & Hammes-Schiffer, S. (2008). Free-energy landscape of enzyme catalysis. *Biochemistry* **47**, 3317–3321.

Berezhkovskii, A. M., Hummer, G. & Szabo, A. (2009). Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *Journal of Chemical Physics* **130**, 205102.

Best, R. B., Chen, Y.-G. & Hummer, G. (2005). Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* **13**, 1755–1763.

Bryngelson, J. D. & Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences, USA* **84**, 7524–7528.

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* **21**, 167–195.

Buchete, N.-V. & Hummer, G. (2008). Coarse master equations for peptide folding dynamics. *Journal of Physical Chemistry B* **112**, 6057–6069.

Caflisch, A. (2006). Network and graph analyses of folding free energy surfaces. *Current Opinion in Structural Biology* **16**, 71–78.

Chen, J. & Brooks, C. L. (2007). Can molecular dynamics simulations provide high-resolution refinement of protein structure? *Proteins* **67**, 922–930.

Chen, Y., Ding, F., Nie, H., Serohijos, A. W., Sharma, S., Wilcox, K. C., Yin, S. & Dokholyan, N. V. (2008). Protein folding: then and now. *Archives in Biochemistry and Biophysics* **469**, 4–19.

Cheung, M. S., García, A. E. & Onuchic, J. N. (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences, USA* **99**, 685–690.

Chi, C. N., Elfstrom, L., Shi, Y., Snall, T., Engstrom, A. & Jemth, P. (2008). Reassessing a sparse energetic network within a single protein domain. *Proceedings of the National Academy of Sciences, USA* **105**, 4679–4684.

Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A. & Swope, W. C. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *Journal of Chemical Physics* **126**, 155101.

Chuang, J., Grosberg, A. Y. & Kardar, M. (2001). Free energy self-averaging in protein-sized random heteropolymers. *Physical Review Letters* **87**, 078104.

Clarkson, M. W., Gilmore, S. A., Edgell, M. H. & Lee, A. L. (2006). Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry* **45**, 7693–7699.

Clore, G. M. (2008). Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement. *Molecular Biosystems* **4**, 1058–1069.

Cukier, R. I. (2009). Apo adenylate kinase encodes its holo form: a principal component and varimax analysis. *Journal of Physical Chemistry B* **113**, 1662–1672.

Das, P., Moll, M., Stamati, H., Kavraki, L. E. & Clementi, C. (2006). Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proceedings of the National Academy of Sciences, USA* **103**, 9885–9890.

Datta, D., Scheer, J. M., Romanowski, M. J. & Wells, J. A. (2008). An allosteric circuit in caspase-1. *Journal of Molecular Biology* **381**, 1157–1167.

Debenedetti, P. & Stillinger, F. (2001). Supercooled liquids and the glass transition. *Nature* **410**, 259–267.

Derrida, B. (1985). A generalization of the random energy model which includes correlations between energies. *Journal of Physics Letters – Paris* **46**, 401–407.

Dhulesia, A., Gsponer, J. & Vendruscolo, M. (2008). Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *Journal of the American Chemical Society* **130**, 8931–8939.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural and Molecular Biology* **4**, 10–19.

Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. (2008). The protein folding problem. *Annual Review of Biophysics* **37**, 289–316.

Dobson, C. M., Sali, A. & Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angewandte Chemie International Edition* **37**, 868–893.

Dror, R. O., Arlow, D. H., Borhani, D. W., Jensen, M. Ø., Piana, S. & Shaw, D. E. (2009). Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proceedings of the National Academy of Sciences, USA* **106**, 4689–4694.

Dyson, H. J., Wright, P. E. & Scheraga, H. A. (2006). The role of hydrophobic interactions in initiation and propagation of protein folding. *Proceedings of the National Academy of Sciences, USA* **103**, 13057–13061.

Eaton, W. A., Henry, E. R., Hofrichter, J. & Mozzarelli, A. (1999). Is cooperative oxygen binding by hemoglobin really understood? *Nature Structural and Molecular Biology* **6**, 351–358.

Ebbinghaus, S., Kim, S. J., Heyden, M., Yu, X., Gruebele, M., Leitner, D. M. & Havenith, M. (2008). Protein sequence- and pH-dependent hydration probed by terahertz spectroscopy. *Journal of the American Chemical Society* **130**, 2374–2375.

Eyal, E., Chennubhotla, C., Yang, L.-W. & Bahar, I. (2007). Anisotropic fluctuations of amino acids in protein structures: insights from X-ray crystallography and elastic network models. *Bioinformatics* **23**, i175–i184.

Fang, C., Frontiera, R. R., Tran, R. & Mathies, R. A. (2009). Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy. *Nature* **462**, 200–204.

Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Young, R. D. (2004). Bulk-solvent and hydration-shell fluctuations, similar to alpha- and beta-fluctuations in glasses, control protein motions and functions. *Proceedings of the National Academy of Sciences, USA* **101**, 14408–14413.

Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. (2007). Localizing frustration in native proteins and protein assemblies. *Proceedings of the National Academy of Sciences, USA* **104**, 19819–19824.

Fetler, L., Kantrowitz, E. R. & Vachette, P. (2007). Direct observation in solution of a preexisting structural

equilibrium for a mutant of the allosteric aspartate transcarbamoylase. *Proceedings of the National Academy of Sciences, USA* **104**, 495–500.

Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D. & Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673.

Frauenfelder, H. & McMahon, B. (1998). Dynamics and function of proteins: the search for general concepts. *Proceedings of the National Academy of Sciences, USA* **95**, 4795–4797.

Frauenfelder, H., Parak, F. & Young, R. D. (1988). Conformational substates in proteins. *Annual Review of Biophysics and Biophysical Chemistry* **17**, 451–479.

Frauenfelder, H., Alberding, N. A., Ansari, A., Braunstein, D., Cowen, B. R., Hong, M. K., Iben, I. E. T., Johnson, J. B. & Luck, S. (1990). Proteins and pressure. *Journal of Physical Chemistry* **94**, 1024–1037.

Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science* **254**, 1598–1603.

Frauenfelder, H., Fenimore, P. W. & Young, R. D. (2007). Protein dynamics and function: insights from the energy landscape and solvent slaving. *IUBMB Life* **59**, 506–512.

Fuentes, E., Der, C. & Lee, A. (2004). Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *Journal of Molecular Biology* **335**, 1105–1115.

García, A. (1992). Large-amplitude nonlinear motions in proteins. *Physical Review Letters* **68**, 2696–2699.

García, A. E. & Sanbonmatsu, K. Y. (2001). Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **42**, 345–354.

Gast, K., Damaschun, H., Misselwitz, R., Miller-Frohne, M., Zirwer, D. & Damaschun, G. (1994). Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. *European Biophysics Journal* **23**, 297–305.

Geissler, P. L., Shakhnovich, E. I. & Grosberg, A. Y. (2004). Solvation versus freezing in a heteropolymer globule. *Physical Review E, Statistical, Nonlinear, and Softmatter Physics* **70**, 021802.

Gianni, S., Walma, T., Arcovito, A., Calosci, N., Bellelli, A., Engstrom, A., Travaglini-Allocatelli, C., Brunori, M., Jemth, P. & Vuister, G. W. (2006). Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure* **14**, 1801–1809.

Grant, B. J., Gorfe, A. A. & McCammon, J. A. (2009). RAS conformational switching: simulating nucleotide-dependent conformational transitions with accelerated molecular dynamics. *PLoS Computational Biology* **5**, e1000325.

Guo, Z. & Thirumalai, D. (1995). Kinetics of protein folding: nucleation mechanism, time scales, and pathways. *Biopolymers* **36**, 83–102.

Hayward, S. & Go, N. (1995). Collective variable description of native protein dynamics. *Annual Review of Physical Chemistry* **46**, 223–250.

He, Y., Ku, P. I., Knab, J. R., Chen, J. Y. & Markelz, A. G. (2008). Protein dynamical transition does not require protein structure. *Physical Review Letters* **101**, 178103.

Head-Gordon, T. & Brown, S. (2003). Minimalist models for protein folding and design. *Current Opinion in Structural Biology* **13**, 160–167.

Heath, A. P., Kavraki, L. E. & Clementi, C. (2007). From coarse-grain to all-atom: toward multiscale analysis of protein landscapes. *Proteins* **68**, 646–661.

Hegler, J. A., Weinkam, P. & Wolynes, P. G. (2008). The spectrum of biomolecular states and motions. *HFSP Journal* **2**, 307–313.

Henzler-Wildman, K. & Kern, D. (2007). Dynamic personalities of proteins. *Nature* **450**, 964–972.

Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. (2007a). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916.

Henzler-Wildman, K. A., Thai, V., Lei, M., Ott, M., Wolf-Watz, M., Fenn, T., Pozharski, E., Wilson, M. A., Petsko, G. A., Karplus, M., Hubner, C. G. & Kern, D. (2007b). Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844.

Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. (1998). The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proceedings of the National Academy of Sciences, USA* **95**, 9903–9908.

Hinrichs, N. S. & Pande, V. S. (2007). Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *Journal of Chemical Physics* **126**, 244101.

Honeycutt, J. D. & Thirumalai, D. (1992). The nature of folded states of globular proteins. *Biopolymers* **32**, 695–709.

Hori, N., Chikenji, G., Berry, R. S. & Takada, S. (2009). Folding energy landscape and network dynamics of small globular proteins. *Proceedings of the National Academy of Sciences, USA* **106**, 73–78.

Hyeon, C. & Onuchic, J. N. (2007). Mechanical control of the directional stepping dynamics of the kinesin motor. *Proceedings of the National Academy of Sciences, USA* **104**, 17382–17387.

Hyeon, C. & Thirumalai, D. (2003). Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *Proceedings of the National Academy of Sciences, USA* **100**, 10249–10253.

Hyeon, C., Jennings, P. A., Adams, J. A. & Onuchic, J. N. (2009). Ligand-induced global transitions in the catalytic domain of protein kinase A. *Proceedings of the National Academy of Sciences, USA* **106**, 3023–3028.

Igumenova, T. I., Lee, A. L. & Wand, A. J. (2005). Backbone and side chain dynamics of mutant calmodulin-peptide complexes. *Biochemistry* **44**, 12627–12639.

Karplus, M. & McCammon, J. A. (1981). The internal dynamics of globular proteins. *CRC Critical Reviews in Biochemistry and Molecular Biology* **9**, 293–349.

Karplus, M. & McCammon, J. A. (1983). Dynamics of proteins: elements and function. *Annual Review of Biochemistry* **52**, 263–300.

Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry* **14**, 1–63.

Kaya, H. & Chan, H. S. (2003). Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity: how adequate is native-centric topological modeling? *Journal of Molecular Biology* **326**, 911–931.

Kern, D. & Zuiderweg, E. R. P. (2003). The role of dynamics in allosteric regulation. *Current Opinion in Structural Biology* **13**, 748–757.

Kern, D., Eisenmesser, E. Z. & Wolf-Watz, M. (2005). Enzyme dynamics during catalysis measured by NMR spectroscopy. *Methods in Enzymology* **394**, 507–524.

Khalili, M. & Wales, D. (2008). Pathways for conformational change in nitrogen regulatory protein c from discrete path sampling. *Journal of Physical Chemistry B* **112**, 2456–2465.

Kim, Y. B., Kalinowski, S. S. & Marcinkeviciene, J. (2007). A pre-steady state analysis of ligand binding to human glucokinase: evidence for a preexisting equilibrium. *Biochemistry* **46**, 1423–1431.

Kondrashov, D. A., Cui, Q. & Phillips Jr, G. N. (2006). Optimization and evaluation of a coarse-grained model of protein motion using X-ray crystal data. *Biophysical Journal* **91**, 2760–2767.

Koshland, D. E., Nemethy, G. & Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5**, 365–385.

Krivov, S. & Karplus, M. (2002). Free energy disconnectivity graphs: application to peptide models. *Journal of Chemical Physics* **117**, 10894–10903.

Krivov, S. V. & Karplus, M. (2004). Hidden complexity of free energy surfaces for peptide (protein) folding. *Proceedings of the National Academy of Sciences, USA* **101**, 14766–14770.

Krivov, S. V., Muff, S., Caflisch, A. & Karplus, M. (2008). One-dimensional barrier-preserving free-energy projections of a beta-sheet miniprotein: new insights into the folding process. *Journal of Physical Chemistry B* **112**, 8701–8714.

Kundu, S., Melton, J. S., Sorensen, D. C. & Phillips, G. N. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophysical Journal* **83**, 723–732.

La Nave, E., Scala, A., Starr, F., Sciortino, F. & Stanley, H. (2000). Instantaneous normal mode analysis of supercooled water. *Physical Review Letters* **84**, 4605–4608.

Leitner, D. M. (2008). Energy flow in proteins. *Annual Review of Physical Chemistry* **59**, 233–259.

Leitner, D. M. (2009). Frequency-resolved communication maps for proteins and other nanoscale materials. *Journal of Chemical Physics* **130**, 195101.

Lenaerts, T., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., Schymkowitz, J. W. H. & Rousseau, F. (2008). Quantifying information transfer by protein domains: Analysis of the Fyn SH2 domain structure. *BMC Structural Biology* **8**, 43.

Levy, Y., Papoian, G. A., Onuchic, J. N. & Wolynes, P. G. (2004). Energy landscape analysis of protein dimers. *Israel Journal of Chemistry* **44**, 281–297.

Li, W., Zhang, J., Wang, J. & Wang, W. (2008). Metal-coupled folding of Cys2His2 zinc-finger. *Journal of the American Chemical Society* **130**, 892–900.

Lockless, S. W. & Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299.

Lucent, D., England, J. & Pande, V. (2009). Inside the chaperonin toolbox: theoretical and computational models for chaperonin mechanism. *Physical Biology* **6**, 015003.

Lukin, J. A., Kontaxis, G., Simplaceanu, V., Yuan, Y., Bax, A. & Ho, C. (2003). Quaternary structure of hemoglobin in solution. *Proceedings of the National Academy of Sciences, USA* **100**, 517–520.

Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **13**, 373–380.

Maisuradze, G. G., Liwo, A. & Scheraga, H. A. (2009a). How adequate are one- and two-dimensional free energy landscapes for protein folding dynamics? *Physical Review Letters* **102**, 238102.

Maisuradze, G. G., Liwo, A. & Scheraga, H. A. (2009b). Principal component analysis for protein folding dynamics. *Journal of Molecular Biology* **385**, 312–329.

Malmendal, A., Evenas, J., Forsen, S. & Akke, M. (1999). Structural dynamics in the C- terminal domain of calmodulin at low calcium levels. *Journal of Molecular Biology* **293**, 883–899.

Marcovitz, A. & Levy, Y. (2009). Arc-repressor dimerization on DNA: folding rate enhancement by colocalization. *Biophysical Journal* **96**, 4212–4220.

Marcus, R. & Sutin, N. (1985). Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta* **811**, 265–322.

Marques, O. & Sanejouand, Y. H. (1995). Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins* **23**, 557–560.

Martinez, K. L., Gohon, Y., Corringer, P.-J., Tribet, C., Merola, F., Changeux, J.-P. & Popot, J.-L. (2002). Allosteric transitions of Torpedo acetylcholine receptor in lipids, detergent and amphipols: molecular

interactions vs. physical constraints. *FEBS Letters* **528**, 251–256.

Materese, C. K., Goldmon, C. C. & Papoian, G. A. (2008). Hierarchical organization of eglin c native state dynamics is shaped by competing direct and water-mediated interactions. *Proceedings of the National Academy of Sciences, USA* **105**, 10659–10664.

McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature* **267**, 585–590.

Miyashita, O., Onuchic, J. N. & Wolynes, P. G. (2003). Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proceedings of the National Academy of Sciences, USA* **100**, 12570–12575.

Miyashita, O., Wolynes, P. G. & Onuchic, J. N. (2005). Simple energy landscape model for the kinetics of functional transitions in proteins. *Journal of Physical Chemistry B* **109**, 1959–1969.

Monod, J., Wyman, J. & Changeux, J.-P. (1965). On the nature of allosteric transitions: a plausible model. *Journal of Molecular Biology* **12**, 88–118.

Nelson, S. W., Iancu, C. V., Choe, J. Y., Honzatko, R. B. & Fromm, H. J. (2000). Tryptophan fluorescence reveals the conformational state of a dynamic loop in recombinant porcine fructose-1,6-bisphosphatase. *Biochemistry* **39**, 11100–11106.

Niu, X., Chen, Q., Zhang, J., Shen, W., Shi, Y. & Wu, J. (2007). Interesting structural and dynamical behaviors exhibited by the AF-6 PDZ domain upon Bcr peptide binding. *Biochemistry* **46**, 15042–15053.

Nurit, H., Geisbrecht, B. V., Lambris, J. & Kavraki, L. E. (2010). Multi-scale characterization of the energy landscape of proteins with application to the C3d/Efb-C complex. *Proteins: Structure, Function, and Bioinformatics* **78**, 1004–1014.

Okazaki, K. & Takada, S. (2008). Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proceedings of the National Academy of Sciences, USA* **105**, 11182–11187.

Okazaki, K., Koga, N., Takada, S., Onuchic, J. N. & Wolynes, P. G. (2006). Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. *Proceedings of the National Academy of Sciences, USA* **103**, 11844–11849.

Onuchic, J. N. & Wolynes, P. G. (2004). Theory of protein folding. *Current Opinion in Structural Biology* **14**, 70–75.

Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. (1995). Toward an outline of the topography of a realistic protein-folding funnel. *Proceedings of the National Academy of Sciences, USA* **92**, 3626–3630.

Onuchic, J. N., Luthey-Schulten, Z. & Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry* **48**, 545–600.

Ota, N. & Agard, D. A. (2005). Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. *Journal of Molecular Biology* **351**, 345–354.

Pande, V. S., Grosberg, A. Y. & Tanaka, T. (2000). Heteropolymer freezing and design: towards physical models of protein folding. *Reviews in Modern Physics* **72**, 259–314.

Papoian, G. A. (2008). Proteins with weakly funneled energy landscapes challenge the classical structure-function paradigm. *Proceedings of the National Academy of Sciences, USA* **105**, 14237–14238.

Papoian, G. A. & Wolynes, P. G. (2003). The physics and bioinformatics of binding and folding: an energy landscape perspective. *Biopolymers* **68**, 333–349.

Papoian, G. A., DeGrado, W. F. & Klein, M. L. (2003a). Probing the configurational space of a metalloprotein core: an ab initio molecular dynamics study of Duo Ferro 1 binuclear Zn cofactor. *Journal of American Chemical Society* **125**, 560–569.

Papoian, G. A., Ulander, J. & Wolynes, P. G. (2003b). Role of water mediated interactions in protein–protein recognition landscapes. *Journal of American Chemical Society* **125**, 9170–9178.

Papoian, G. A., Ulander, J., Eastwood, M. P., Luthey-Schulten, Z. & Wolynes, P. G. (2004). Water in protein structure prediction. *Proceedings of the National Academy of Sciences, USA* **101**, 3352–3357.

Paschek, D., Hempel, S. & García, A. E. (2008). Computing the stability diagram of the Trp-cage miniprotein. *Proceedings of the National Academy of Sciences, USA* **105**, 17754–17759.

Pervushin, K., Vamvaca, K., Vogeli, B. & Hilvert, D. (2007). Structure and dynamics of a molten globular enzyme. *Nature Structural and Molecular Biology* **14**, 1202–1206.

Pincus, D. L., Cho, S. S., Hyeon, C. & Thirumalai, D. (2008). Minimal models for proteins and RNA: from folding to function. *Progress in Molecular and Biological Transactions* **84**, 203.

Pisliakov, A., Cao, J., Kamerlin, S. & Warshel, A. (2009). Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proceedings of the National Academy of Sciences, USA* **106**, 17359–17364.

Plotkin, S. S. & Onuchic, J. N. (2002). Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. *Quarterly Review of Biophysics* **35**, 205–286.

Plotkin, S. S., Wang, J. & Wolynes, P. G. (1997). Statistical mechanics of a correlated energy landscape model for protein folding funnels. *Journal of Chemical Physics* **106**, 2932–2948.

Ptitsyn, O. B. (1996). How molten is the molten globule? *Nature Structural and Molecular Biology* **3**, 488–490.

Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990). Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Letters* **262**, 20–24.

Rao, F. & Caflisch, A. (2004). The protein folding network. *Journal of Molecular Biology* **342**, 299–306.

Riccardi, D., Cui, Q. & Phillips Jr, G. N. (2009). Application of elastic network models to proteins in the crystalline state. *Biophysical Journal* **96**, 464–475.

Roca, M., Messer, B., Hilvert, D. & Warshel, A. (2008). On the relationship between folding and chemical landscapes in enzyme catalysis. *Proceedings of the National Academy of Sciences, USA* **105**, 13877–13882.

Samuni, U., Dantsker, D., Roche, C. J. & Friedman, J. M. (2007). Ligand recombination and a hierarchy of solvent slaved dynamics: the origin of kinetic phases in hemeproteins. *Gene* **398**, 234–248.

Scheraga, H. A., Khalili, M. & Liwo, A. (2007). Protein-folding dynamics: overview of molecular simulation techniques. *Annual Reviews in Physical Chemistry* **58**, 57–83.

Schmitz, M. & Tavan, P. (2004). Vibrational spectra from atomic fluctuations in dynamics simulations. I. Theory, limitations, and a sample application. *Journal of Chemical Physics* **121**, 12233–12246.

Schrader, T. E., Schreier, W. J., Cordes, T., Koller, F. O., Babitzki, G., Denschlag, R., Renner, C., Lweneck, M., Dong, S.-L., Moroder, L., Tavan, P. & Zinth, W. (2007). Light-triggered beta-hairpin folding and unfolding. *Proceedings of the National Academy of Sciences, USA* **104**, 15729–15734.

Schulz, R., Krishnan, M., Daidone, I. & Smith, J. C. (2009). Instantaneous normal modes and the protein glass transition. *Biophysical Journal* **96**, 476–484.

Shakhnovich, E. I. & Gutin, A. M. (1993). Engineering of stable and fast-folding sequences of model proteins. *Proceedings of the National Academy of Sciences, USA* **90**, 7195–7199.

Sharp, K. & Skinner, J. J. (2006). Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. *Proteins* **65**, 347–361.

Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. (2000). Speeding molecular recognition by using the folding funnel: the y-casting mechanism. *Proceedings of the National Academy of Sciences, USA* **97**, 8868–8873.

Smrcka, A. V., Kichik, N., Tarrago, T., Burroughs, M., Park, M.-S., Itoga, N. K., Stern, H. A., Willardson, B. M. & Giralt, E. (2010). NMR analysis of G-protein beta gamma subunit complexes reveals a dynamic G(alpha)-G beta gamma subunit interface and multiple protein recognition modes. *Proceedings of the National Academy of Sciences, USA* **107**, 639–644.

Socci, N. D., Onuchic, J. N. & Wolynes, P. G. (1996). Diffusive dynamics of the reaction coordinate for protein folding funnels. *Journal of Chemical Physics* **104**, 5860–5868.

Soheilifard, R., Makarov, D. E. & Rodin, G. J. (2008). Critical evaluation of simple network models of protein dynamics and their comparison with crystallographic B-factors. *Physical Biology* **5**, 26008.

Suel, G., Lockless, S., Wall, M. & Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Structural and Molecular Biology* **10**, 59–69.

Sugase, K., Dyson, H. J. & Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025.

Tama, F. & Brooks, C. L. (2006). Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. *Annual Review of Biophysics and Biomolecular Structure* **35**, 115–133.

Tama, F. & Sanejouand, Y. H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein Engineering* **14**, 1–6.

Tang, C., Louis, J. M., Aniana, A., Suh, J.-Y. & Clore, G. M. (2008). Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease. *Nature* **455**, 693–696.

Tang, S., Liao, J.-C., Dunn, A. R., Altman, R. B., Spudich, J. A. & Schmidt, J. P. (2007). Predicting allosteric communication in myosin via a pathway of conserved residues. *Journal of Molecular Biology* **373**, 1361–1373.

Tehver, R., Chen, J. & Thirumalai, D. (2009). Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle. *Journal of Molecular Biology* **387**, 390–406.

Thielges, M. C., Zimmermann, J., Yu, W., Oda, M. & Romesberg, F. E. (2008). Exploring the energy landscape of antibody-antigen complexes: protein dynamics, flexibility, and molecular recognition. *Biochemistry* **47**, 7237–7247.

Tsai, C.-J., Sol, A. D. & Nussinov, R. (2009). Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Molecular Biosystems* **5**, 207–216.

Turjanski, A. G., Gutkind, J. S., Best, R. B. & Hummer, G. (2008). Binding-induced folding of a natively unstructured transcription factor. *PLoS Computational Biology* **4**, e1000060.

van der Vaart, A., Bursulaya, B. D., Brooks, C. L. & Merz, K. M. (2000). Are many-body effects important in protein folding? *Journal of Physical Chemistry B* **104**, 9554–9563.

Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. (2001). Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429–2433.

Wang, J., Zhang, K., Lu, H. & Wang, E. (2005). Quantifying kinetic paths of protein folding. *Biophysical Journal* **89**, 1612–1620.

Wang, J., Zhang, K., Lu, H. & Wang, E. (2006). Dominant kinetic paths on biomolecular

binding-folding energy landscape. *Physical Review Letters* **96**, 168101.

Whitford, P. C., Miyashita, O., Levy, Y. & Onuchic, J. N. (2007). Conformational transitions of adenylate kinase: switching by cracking. *Journal of Molecular Biology* **366**, 1661–1671.

Whitford, P. C., Gosavi, S. & Onuchic, J. N. (2008a). Conformational transitions in adenylate kinase – allosteric communication reduces misligation. *Journal of Biological Chemistry* **283**, 2042–2048.

Whitford, P., Onuchic, J. N. & Wolynes, P. G. (2008b). Energy landscape along an enzymatic reaction trajectory: hinges or cracks? *HFSP Journal* **2**, 61–64.

Wu, S., Zhuravlev, P. I. & Papoian, G. A. (2008). High resolution approach to the native state ensemble kinetics and thermodynamics. *Biophysical Journal* **95**, 5524–5532.

Xu, C., Tobi, D. & Bahar, I. (2003). Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T↔R2 transition. *Journal of Molecular Biology* **333**, 153–168.

Xu, Y., Colletier, J. P., Jiang, H., Silman, I., Sussman, J. L. & Weik, M. (2008). Induced-fit or preexisting equilibrium dynamics? Lessons from protein crystallography and MD simulations on acetylcholinesterase and implications for structure-based drug design. *Protein Science* **17**, 601–605.

Yang, L.-W., Eyal, E., Chennubhotla, C., Jee, J., Gronenborn, A. M. & Bahar, I. (2007). Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. *Structure* **15**, 741–749.

Youn, H., Koh, J. & Roberts, G. P. (2008). Two-state allosteric modeling suggests protein equilibrium as an integral component for cyclic AMP (cAMP) speci_city in the cAMP receptor protein of *Escherichia coli*. *Journal of Bacteriology* **190**, 4532–4540.

Yue, K. & Dill, K. A. (1995). Forces of tertiary structural organization in globular proteins. *Proceedings of the National Academy of Sciences, USA* **92**, 146–150.

Zheng, W. (2008). A unification of the elastic network model and the Gaussian network model for optimal description of protein conformational motions and fluctuations. *Biophysics Journal* **94**, 3853–3857.

Zheng, W., Brooks, B., Doniach, S. & Thirumalai, D. (2005). Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure* **13**, 565–577.

Zhuravlev, P. I., Materese, C. K. & Papoian, G. A. (2009). Deconstructing the native state: energy landscapes, function, and dynamics of globular proteins. *Journal of Physical Chemistry B* **113**, 8800–8812.

Zhuravleva, A., Korzhnev, D. M., Nolde, S. B., Kay, L. E., Arseniev, A. S., Billeter, M. & Orekhov, V. Y. (2007). Propagation of dynamic changes in barnase upon binding of barstar: an NMR and computational study. *Journal of Molecular Biology* **367**, 1079–1092.

Zong, C., Papoian, G. A., Ulander, J. & Wolynes, P. G. (2006). Role of topology, nonadditivity, and water-mediated interactions in predicting the structures of alpha/beta proteins. *Journal of American Chemical Society* **128**, 5168–5176.

Zwanzig, R. (2001). *Nonequilibrium Statistical Mechanics*. Oxford: Oxford University Press.