Protein Engineering vol.12 no.9 pp.713–720, 1999

Folding funnels and binding mechanisms

Buyong Ma[1], Sandeep Kumar[2], Chung-Jung Tsai[2] and Ruth Nussinov[2,3,4]

[1]Laboratory of Experimental and Computational Biology and [2]Intramural Research Support Program—SAIC, Laboratory of Experimental and Computational Biology, NCI–FCRDC, Frederick, MD 21702, USA and [3]Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

[4]To whom correspondence should be addressed, at the US address. E-mail: ruthn@ncifcrf.gov

The long-held views on lock-and-key versus induced fit in binding arose from the notion that a protein exists in a single, most stable conformation, dictated by its sequence. However, in solution proteins exist in a range of conformations, which may be described by statistical mechanical laws and their populations follow statistical distributions. Upon binding, the equilibrium will shift in favor of the bound conformation from the ensemble of conformations around the bottom of the folding funnel. Hence here we extend the implications and the usefulness of the folding funnel concept to explain fundamental binding mechanisms.

Keywords: binding/conformational ensembles/energy landscape/folding/funnels/lock-and-key vs induced fit/misfolding

Introduction

The concept of energy landscapes and folding funnels is important not only for understanding the folding of the polypeptide chain. It is equally critical for the understanding of the folded protein and its function, through its intra- and intermolecular recognition and binding. Here we focus on the implications of the presence of a population of conformers at and around the bottom of the folding funnel. The more flexible the molecule, the larger is the ensemble of diverse conformers and the lower are the barriers between them. We argue that pre-existing subpopulations of conformational isomers preferentially bind to their corresponding ligand, whether a subunit, another protein, an antigen or a substrate, without the need for the long-postulated induced fit mechanism. Similarly, this approach rationalizes broad range versus specific binding, crystal effects and intra- and intermolecular domain swapping. Hence, we extend the concept of the folding funnel and show how its consequences may straightforwardly replace long-held notions of the mechanisms in binding. Funnels with rugged bottoms portray and lead to non-specific molecular associations or, in the extreme case, to amyloids. On the other hand, smoother single or a few minima with high barriers imply rigid binding. On the theoretical side, since the complexity of the energy landscape increases rapidly with the size of the system, funnels constructed for binding can be expected to be complicated (Tsai *et al.*, 1999).

Protein folding and binding mechanisms: the usefulness of the funnel model

Over the last few years, growing attention has been directed to the kinetics of protein folding (Baldwin, 1994, 1995; Dill and Chan, 1997). This has been the outcome of both a significant enhancement in the experimental methodologies and a theoretical framework laying out the fundamental conceptualization (Bryngelson and Wolynes, 1989; Shakhnovitch and Gutin, 1993; Karplus and Shakhnovitch, 1992; Boczko and Brooks, 1995; Dill *et al.*, 1995; Karplus *et al.*, 1995; Onuchic *et al.*, 1995; Wolynes *et al.*, 1995; reviewed in Dill and Chan, 1997; Karplus, 1997; Lazaridis and Karplus, 1997; Dill, 1999; Tsai *et al.*, 1999). Folding is currently viewed as a parallel process whereby an ensemble of molecules goes downhill through the energy funnel, with a higher probability of going through some obligatory steps (Lazaridis and Karplus, 1997; Gruebele and Wolynes, 1998; Martinez *et al.*, 1998) representing transition-state ensembles. The energy landscapes in protein folding have been depicted in terms of hills, corresponding to high-energy conformations and valleys, having more favorable conformations than those in their vicinity. Around the bottom of the valley there is a population of conformations. If the landscape is smooth, the native protein may be expected to have small fluctuations, with only small changes in the conformations. However, if the energy landscape is rugged, the ensemble of structures would include conformations which may be completely different, depending on the extent of the ruggedness.

While this view of protein folding is widely accepted, not all of its implications and ramifications with regard to other processes have been explored. Here we question some of the commonly held notions which can be straightforwardly explained in this light. We illustrate that these can be put within this general framework, tying concepts cast in protein folding with those long cast in protein binding. Moreover, since protein molecules always function via binding, the fact that they exist in solution as an ensemble of conformational isomers has numerous consequences.

Protein molecules are inherently flexible entities. While the entire molecule is flexible, inspection of the structures immediately reveals that some structural parts are more rigid than others. The more rigid parts may be more compactly packed, have a stronger hydrophobic effect and have a larger stabilizing electrostatic contribution. Movements of the backbone of such structural domains, subdomains or any structural part will result in larger displacements of these structural units than side-chain motions. On the other hand, side-chain motions can in turn bring about movements of the backbone. If the structural units are unstable, such thermal motions of the backbone would result in an entirely flexible molecule which does not retain any of its native fold. However, if the structural units are stable, a range of conformational isomers may be observed, depending on the extent of its flexibility and the locations of the more flexible joints. Via interactions with other molecules, conformational isomers may form amyloids and/or domain-swapped structures, where domains from the same molecule switch to pair with sister domains from a different molecule (Bennett *et al.*, 1994, 1995). Backbone

flexibility leading to such movements is also likely to result in less specific binding. Further, it may manifest itself in 'crystal packing' effects. It may also be at the origin of 'induced fit' in binding. We shall come back to these in more detail later.

Here we confine ourselves to movements of the backbone. We argue that the ensembles of conformational isomers, populating the rugged bottom of the folding funnel, are the source of the observed movements of structural units with respect to each other on their flexible joints. While local optimization of the complexed, bound structures would follow intermolecular recognition and interaction, they are the outcome rather than the source. The 'lock' is not rigid when the 'key' is inserted into it. Rather, the 'lock' exists in a range of conformations, some of which fit the 'key'. Similarly, the receptor does not undergo an 'induced fit'. It too exists in a structural ensemble, with some conformations fitting the 'key'. The fact that the crystal structure of the isolated unbound receptor differs from that of the receptor when it is bound to its respective ligand only implies that under these crystallization conditions unbound 'open' receptor conformation is the most populated one. However, the ligand may bind to an alternative conformation of the receptor which is in the 'correct' configuration, with the equilibrium adjusting itself in favor of the bound receptor conformation. Viewed in this light, these two frequently used 'lock-and-key' and 'induced fit' terms are simply the outcome of the distributions of interconvertible of conformations. As such, they may be considered as reflecting the same fundamental phenomenon.

The origin of the long-held views on 'lock-and-key' binding versus 'induced fit' and similarly of 'crystal effects' derive from earlier years, when protein folding was considered to follow a single pathway. It was believed that, depending on the temperature, a single, most stable structure would exist. However, many experiments have recently shown that folded proteins may assume a very large number of conformational substrates (Frauenfelder *et al.*, 1991; Frauenfelder and Leeson, 1998; Dill, 1999; Tsai *et al.*, 1999). When considered now in the context of multiple pathways and of ensembles of conformations, these and other such molecular recognition and binding processes can be explained by a straightforward linkage of the current theory of folding and intermolecular binding.

Extension of the funnel model: from small peptides to proteins and on to complexes

The single most important point about the 'new view' of protein folding is the stipulation that protein molecules glide down the energy landscape via multiple pathways during protein folding (Bryngelson and Wolynes, 1989; Shakhnovitch and Gutin, 1993; Karplus and Shakhnovitch, 1992; Boczko and Brooks, 1995; Dill *et al.*, 1995; Karplus *et al.*, 1995; Onuchic *et al.*, 1995; Wolynes *et al.*, 1995; Dill and Chan, 1997; Karplus, 1997; Lazaridis and Karplus, 1997; Frauenfelder and Leeson, 1998). This model can be used consistently for a range of short and long polypeptide chains and for intra- and intermolecular associations.

Qualitatively, there is no difference between the energy landscapes for small peptides, large proteins and protein complexes. Even though the complexity of the landscape increases very rapidly with the size of the system, the funnel model still applies, manifesting the wide distribution of potential conformational states. That would be the case regardless of whether it is the folding of a large, e.g. multi-domain, protein or the binding of several subunits to form a complex. Binding and folding are similar processes, with the sole difference between them being the absence or presence of the chain connectivity between their components. Both are hierarchical processes and can be a cooperative all-or-none two-state process or a three-state process, depending on the nature of their interactions. These processes can be divided into the formation of 'building blocks', domains, folded proteins and formation of protein complexes (Tsai *et al.*, 1998). The entire process may be viewed as sequential fusion and modification of individual funnels.

In solution, peptides frequently co-exist in a wide range of conformations, corresponding to shallow and rugged energy landscapes. The conformational diversity of peptides may resemble early events in protein folding, especially the formation of what we call the 'building blocks' of the structure. Protein folding can then be described as a combinatorial assembly of such a set of transient, highly populated contiguous fragments (Tsai *et al.*, 1998). A 'building block' may be a single secondary structure element or a contiguous segment of interacting elements. Unlike a compact, independently folding hydrophobic folding unit (Tsai and Nussinov, 1997a,b; Tsai *et al.* 1998) whose thermodynamic stability derives from cooperative interactions between the building blocks, the building block itself is likely to be unstable. Mutual stabilization between building blocks is dictated by the 'non-local' coding of the protein sequence. A building block may have a variable size and its stability is derived from the local interactions within the fragment. A building block may alter its conformation, twisting or opening, losing its intrafragment interactions. While a given building block sequence is likely to have several alternative conformations, we see merely the final, static, most populated structure under specified conditions. This definition is consistent with the experimental observation of a stable structure, with native and non-native interactions in the denatured state (Wang and Shortle, 1996). Clearly, not all fragments of the sequence have highly populated conformations and hence not all contiguous segments can be defined as building blocks.

If a protein–protein complex (or dimer) is a result of a three-state binding process between two already folded proteins, the funnel landscape of the complex can be considered as the fusing of two individual two-state folding funnels. The bottom of the newly fused funnel is now occupied by a collection of favorable associations of two conformations, with each located at the bottom of one (or both) of the two individual folding funnels. Figure 1 depicts a schematic drawing of such a fused funnel shape. If the native complexed conformation is not the most stable one and hence is not the one depicted by the bottom of the funnel, it can most likely still be identified. However, this depends on two considerations: first, how close is the native conformation of the bound, complexed protein to the one present at the bottom of its (respective protein) funnel and how easily it may reach it; and second, on the population time of the new individual 'deformed' (bound) conformation relative to its native unbound conformation. In the 'lock-and-key' rigid binding case, the funnel bottom is an association of two proteins in their most populated native conformations. In the 'induced fit' case, the above two conditions are not severe enough to prevent such a 'deforming' binding process. The population time of the bound-shape isomer may be high and, additionally, the two alternative 'open' and 'closed' conformations may be close on the energy surface. On the

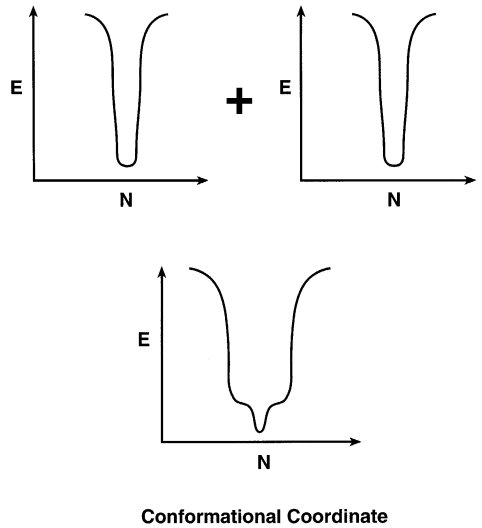

Fig. 1. Schematic illustration of the landscapes of folding and of binding. The energy landscape of the bound proteins case is a fusion between two individual folding funnels. Two conformations, one from each of the individual funnels, associate to form a new minimum at the bottom of the fused funnels. Here we portray a simple folding-binding case. However, in the more rugged case, around the bottom of the new funnel there could exist an ensemble of complexes, the outcome of alternative associations of different, single-molecule conformers. In general, fusing of rugged folding funnels will yield rugged folding–binding funnels. This is expected as folding funnels with rugged bottoms imply flexible molecules existing in an ensemble of conformational isomers. The difference in their energies may be relatively small, yet the structures may be diverse. These may associate in different ways.

other hand, in the domain swapping case, a flip of, say, a single secondary structure element out of its native conformation may result in a conformation with a significantly lower population time. This explains why the conversion of a 'regular' homodimer toward a domain swapped dimer may take hours or days.

As additional components are added to the system under consideration, such as from dimer to trimer, to tetramer and so on, the energy landscape becomes progressively more complicated. However, if the association between any two components has a funnel-like energy landscape without significant trapping bumps, the build-up of an oligomer of any size will still illustrate a funnel-like landscape.

While much attention has been focused on the steepness of the slope of the hill and on the extent of its bumpiness and its local crests and crevices, here we focus on the neighborhood of the bottom of the funnel. The landscape around the bottom can be smooth with a single steep canyon, a few minima or, alternatively, it can be very rugged. The barriers separating the minima around the rugged bottom can also vary in height. Below we analyze the implications of the energy landscape near the bottom of the funnel with regard to a range of binding processes. Figure 2 depicts schematically this correlation between different bottom landscapes and the observed phenomena. Figure 2a shows a single global minimum. This may be the case for an extremely stable protein, such as an extreme hyperthermophile. Figure 2b illustrates a 'lock-and-key' type binding. Here we see specific binding, with relatively little conformational change. Hence the landscape is schematically shown as having only a few minima. Conformations that lie nearby on the energy landscape have similar geometries. Figure 2c depicts a rugged bottom, with rather low barriers separating them. This may correspond to binding of the type commonly dubbed an 'induced fit' or to different crystal forms frequently considered to be crystal effects. Depending on the range of minima and on the barrier heights, we may have a case corresponding to non-specific binding. Non-specific binding, as observed in the non-specific germline antibodies binding a host of invading antigens, may fall into this category. Figure 2d illustrates rough bottoms with higher barriers. Such cases may correspond to domain swapping as in the bovine seminal ribonuclease case or in the more extreme case, where the structures are further away in the funnel, to misfolding, such as in the amyloids. Hence the ruggedness correlates with flexibility. The larger the flexibility, the larger is the population of the conformers observed around the bottom and the lower is the energy required to flip between them. Hence this model of protein folding with its multiple pathways leading to ensembles of conformers is equally useful for binding, eliminating the need to invoke a conformational change induced by the intermolecular interactions.

In the 'building block' folding model described above, a building block is a contiguous fragment in a given sequence, having a relatively high population time in one particular conformation as compared with other conformations. Its formation can be considered to be driven by a funnel-like energy landscape. Hence, if the association between any two building blocks in folding or in binding is also a funnel-like energy landscape resulting from a fusion of two funnel landscapes, then the overall folding (and binding) landscape has a funnel-like shape. Figures 1 and 2 illustrate schematically landscapes of folding and binding.

Hinge-bending 'domain' motions

Most of the motions we have discussed above fall into the so-called 'hinge-bending domain motions' category. Here 'domain' is not meant in the usual sense of an independent structural entity that would keep its conformation even if sliced out of the monomer. Rather, what is meant is a 'structural part' or any subdomain that moves around a swiveling point with respect to another structural entity. The parts themselves are not entirely rigid, although they are relatively so compared with the swiveling region. Swiveling is not a complete rotation in 3D space. Too much freedom is likely to hinder protein function (Sun and Sampson, 1998). Hence, through evolution, particular residues are likely to have been selected to be at and around the hinge and possibly at interdomain interfacial boundaries, limiting both the extent of the motions and at the same time leading to preferred rotational directions.

The existence of an ensemble of hinge-bent conformational isomers around the rugged bottom of the energy funnel, such as in the case of different crystal forms or as in bound and unbound states, suggests low-energy barriers between them. Hence, while the conformations of the structural parts which move as relatively rigid bodies may be expected to be relatively stable, the interactions at the interdomain boundaries may be of a different nature. Some inkling into this problem may be obtained from thermostable proteins. Analysis of salt bridges in an extremely thermostable protein and its comparison with its mesophilic counterpart has recently shown that while there was a large difference in the number of salt bridges between this homologous pair of proteins, this difference appears to be confined to salt bridges within the hydrophobic folding units (Kumar *et al.*, 1999b, submitted). Consistently, it has been suggested that breaking a salt bridge involves overcoming a high conformational energy barrier (Waldburger *et al.*, 1996).

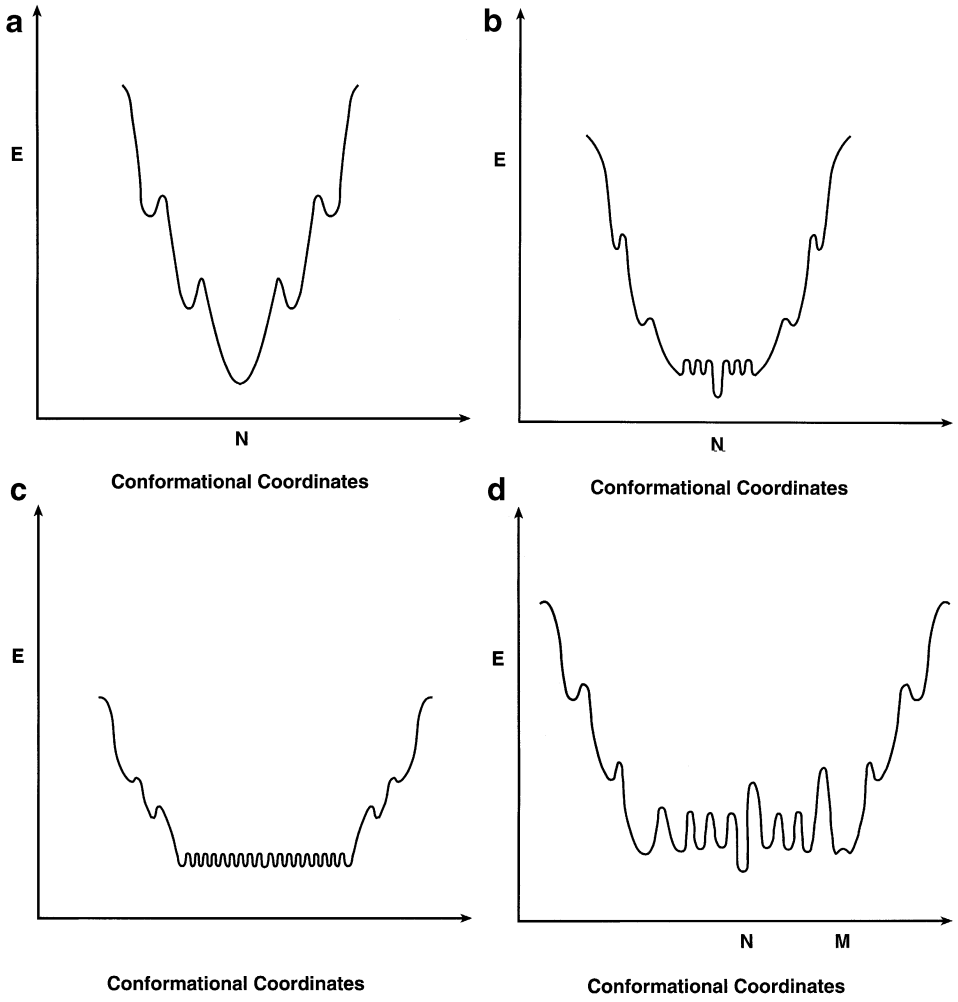

Fig. 2. Schematic depiction of different landscapes around the bottom of the funnels and their correlation with binding mechanisms. In general, the ruggedness correlates with flexibility. (**a**) A smooth funnel, with a single minimum. This may be a simplified case of an extremely stable protein. (**b**) A 'lock-and-key' type binding. Here we may expect relatively little conformational change. The landscape is shown schematically as having few minima, with the conformations nearby on the energy landscape very similar geometrically. (**c**) A rugged bottom, with rather low barriers separating them. This may correspond to binding of the type known as 'induced fit' or to different crystal forms frequently considered to be crystal effects. Depending on the range of minima and on the barrier heights, we may have a case like that corresponding to non-specific binding. (**d**) Rough bottoms with higher barriers. These types of cases may correspond to domain swapping or, in the more extreme case, where the structures are further away in the funnel, to misfolding, such as in amyloids or in aggregation.

Below we review some specific examples of hinge-bending conformational isomers in binding which may be the outcome of the presence of the ensemble of conformers around the rugged bottom of the folding funnel.

Conformational isomers and the lock-and-key versus induced fit binding

Bound and unbound states of proteins depicting 'lock and key' (Fischer, 1894) and 'induced fit' (Koshland, 1958) have been provided by X-ray crystallography. These examples include both allosteric and non-allosteric binding. Allosteric transitions between R (relaxed) and T (taut) states in aspartate transcarbamyltransferase (Stevens and Lipscomb, 1992), fructose-1,6-bisphosphatase (Ke *et al.*, 1991), glycogen phosphorylase (Barford and Johnson, 1989), hemoglobin (Perutz *et al.*, 1998), phosphofructokinase (Schirmer and Evans, 1990) and *lac* repressor (Lewis *et al.*, 1996) can be thought of as examples of conformer selection upon binding of substrate(s). These transitions mostly involve movements of subunits with respect to each other and positional shifts in the cofactor binding site.

Conformational changes in bound and unbound forms of proteins that do not involve any allosteric cofactors have also been studied. Examples include aspartate receptor (Milburn *et al.*, 1991), *Bam*H1 endonuclease (Newman *et al.*, 1995), immunoglobulin VL–VH movements (Herron *et al.*, 1991; Stanfield *et al.*, 1993; Rini *et al.*, 1993) and *Saccharomyces cerevisiae* PPR1 Zn-finger DNA binding protein (Marmorstein and Harrison, 1994). Again, in these cases most of the conformational change involves movements of subunits with respect to one another.

In both allosteric and non-allosteric binding, the deceptively rigid snapshot images of protein motion observed by crystallography can actually be explained by the presence of several low barrier conformational isomers in solution around the bottom of the funnel and shifts in chemical equilibrium in favor of the bound states of the proteins upon substrate–ligand binding.

These examples of protein–substrate binding indicate a predominance of the induced fit mechanism. However, the lock-and-key mechanism has also been observed in the case of Fab D1.3–lysozyme complex (Braden *et al.*, 1996). Unlike

other antibody–antigen complexes, this complex exhibits almost perfect complementarity. The interfaces appear to be preformed and only deform by ~0.5 Å upon complex formation (Amit *et al.*, 1986; Bhat *et al.*, 1990; Braden *et al.*, 1996). The rigidity of the antibody molecule indicates the presence of fewer conformational isomers in this case and hence a smoother energy landscape bottom.

Specificity versus broad range in binding

The immunoglobulins constitute an ideal case for studies of specific versus non-specific binding. Whereas mature immunoglobulins are highly specific, the germline ones bind a broad range of antigens. Hence a difficult although extremely intriguing question is what the origin of this difference is. Is there also a concomitant difference in the extent of the molecular flexibility, with the germline antibody being more flexible than its mature descendent?

In 1994, based on a kinetic analysis, Foote and Milstein (1994) proposed that antibodies do not have a single conformation at their combining site. They suggested that differences observed in crystal structures between bound and unbound forms could arise from the direct interactions of the antibody with its antigen through either induced fit or, alternatively, by preferential ligand binding to a pre-existing subpopulation of antibody isomers. In their insightful paper they note that 'this distinction is of considerable immunological consequence, since two isomers in spontaneous equilibrium would both form part of the humoral repertoire, whereas induced conformation existing only in an immune complex would not contribute to diversity in the same way'. Thus, a few conformations may exist, with the ligands binding preferentially to one form. However, a conformer 'unsuitable' for binding to one antigen may be the preferred one for a second antigen, with a different structure.

Recently, Wedemeyer *et al.* (1997) published an extremely interesting paper, bearing directly on these issues. They solved the X-ray crystal structures of two antibodies, an affinity matured antibody and its apparent corresponding germline antibody. Nine somatic mutations differentiate between these two molecules. Each of these antibodies was crystallized twice, in its free, unbound form and complexed with a hapten antigen. A comparison between the free and complexed forms of the germline showed a significant structural change, suggesting that an 'induced fit'-like association had taken place. On the other hand, the affinity matured antibody showed very little conformational change between the bound and unbound forms, conforming to the classical 'lock-and-key' type of rigid binding. Correspondingly, a comparison between the bound form of the germline and the free form of the mature antibody showed the two to be highly similar. Hence Wedemeyer *et al.* postulated that the somatic mutations have served to 'preconfigure' the mature antibody to complement the structural features of its hapten antigen. On the other hand, as they note, this ability of the germline to bind the hapten, reconfiguring its active site in response to binding, reflects conformational flexibility, expanding substantially the structural diversity of the germline repertoire.

Taken together, the germline, non-specific antibody exists in a range of conformations. The one that binds the invading antigen is the one whose structure is complementary to that of the antigen. If the barriers between the conformers in the ensemble are low, the equilibrium of the population is kept by the low-energy interconversion of the conformations near the one that binds the antigen and thereby drive the reaction in this direction. The more non-specific the antibody, the more flexible it may be expected to be. The more flexible it is, the broader the range of conformations it may adopt. In non-specific, broad range-binding antibodies, the energy surface at the funnel bottom is rugged, with numerous minima reflecting the large number of conformational isomers, with low barriers between the conformers, allowing fluctuations from one to the other. On the other hand, during antibody maturation, mutational events take place, rigidifying some conformers with favorable geometries. These mutations may have the effect of 'smoothing' the rugged bottoms of the funnels and thus reducing the conformational flexibility of the mature antibody.

Another well studied example of a broad range of binding is that of the proteolytic enzymes, such as the aspartic proteinase family. A hinge-based motion has been observed there (James *et al.*, 1982; Sali *et al.*, 1989; Abad-Zapatero *et al.*, 1990; Erickson *et al.*, 1995; Silva *et al.*, 1996). Recently, Lee *et al.* (1998) have shown a pH-dependent modulation of binding of cathepsin D, implicating a pH-dependent switch in the population times.

DNA binding proteins which bind to variable sequences are also likely to display a range of conformations around the bottom of the funnel, while still binding to the DNA with high affinity. For example, both NMR and crystallographic evidence indicate that the arms of Trp repressor which bind DNA are highly flexible (Lawson *et al.*, 1988; Arrowsmith *et al.*, 1989) and do not occupy unique locations on the DNA (Carey, 1989). This flexibility of the DNA binding domain of the *trp* repressor is essential for recognition of different operator sequences (Gryk *et al.*, 1996).

Crystal packing effects

To illustrate our point, we discuss two examples. Recently, the crystal structure of the extracellular portion of the rabbit tissue factor (r-TF) has been solved (Muller *et al.*, 1998). The extracellular portions of both the human tissue factor (h-TF) (Muller *et al.*, 1996) and the rabbit TF consist of two fibronectin type II domains, which are connected by an α-helix. However, the two r-TF molecules in the asymmetric unit have been observed to differ in the orientation of the two domains with respect to each other, illustrating an unexpected hinge of 12.5° around an axis cutting across the molecule at residue 106 in its linker unit. Muller *et al.* (1998) note that while it could be argued that this hinge-bending motion could have been induced by crystal packing forces, they are not in favor of such an explanation. Alternatively, it may be the other way round, namely, that the crystal packing constraints 'stabilize the conformational diversity'. Figure 3a illustrates a superpositioning of the two chains of r-TF (PDB code 1a21) and Figure 3b shows a superpositioning of h-TF (PDB code 2hft) and r-TF. While chain A of r-TF superimposes perfectly on h-TF, chain B of r-TF illustrates the hinge bending. Consistently, Huang *et al.* (1998) have compared the crystal structures of the unbound h-TF (2hft) with the h-TF complexed with a FAB fragment (PDB code 1ahw). They have also observed hinge bending between the two domains, although a smaller angle is observed in this case (7°).

Rose *et al.* (1998) analyzed domain flexibility in retroviral proteases. They compared a crystal structure of a mutant unliganded SIV protease in a new $P3_221$ space group which is in a more 'open' conformation than any other retroviral protease crystallized to date with other unliganded and liganded

HIV proteases. They found that five domains of the SIV dimer move as rigid bodies with respect to each other. However, whereas Wilderspin and Sugrue (1994) suggest that the 'open' unliganded HIV-1 protease structure is an artifact of crystal packing as an explanation of the origin of its difference from the unliganded SIV 'closed' conformation, Rose *et al.* (1998) propose the opposite. Their interpretation is that the unliganded SIV structure is the outcome of the crystal contacts. Here we argue that the origin of both is the population of hinge-bending conformational isoformers. The conformer which is crystallized is the one whose conformation is optimal for the association, with the equilibrium shifting itself in its direction.

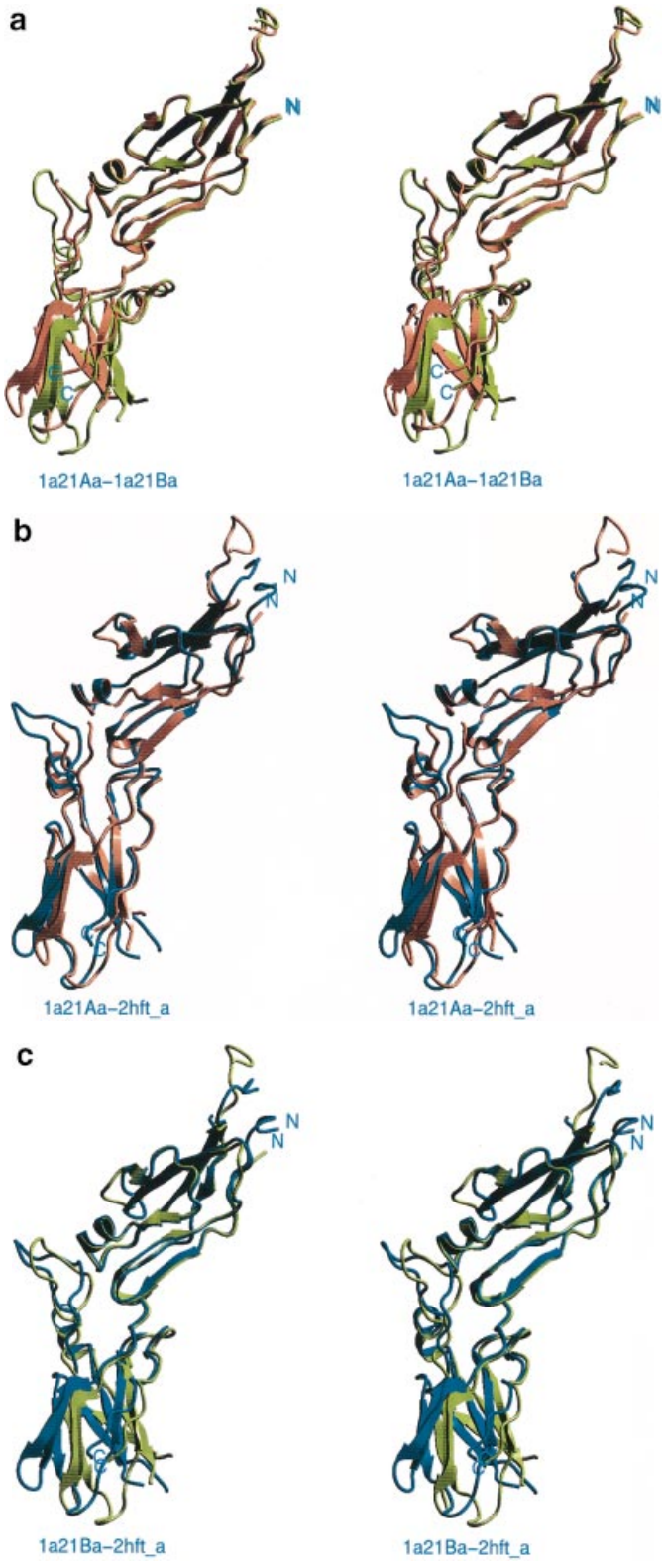

Fig. 3. Crystal packing effects. (**a**) A ribbon stereo superpositioning of the two chains of r-TF (PDB code 1a21A and 1a21B). (**b**), (**c**) Superpositioning of h-TF (PDB code 2hft) and r-TF. While chain A of r-TF superimposes perfectly on h-TF, chain B of r-TF illustrates hinge bending. In order to show the domain movement resulting from crystal packing, the superpositions are calculated based on only the N-terminal domain part. r-TF chain A and chain B are in pink and light green, respectively, and h-TF is in blue. Here we suggest that such 'crystal packing effects' are a simple outcome of conformational isomerism around the bottom of the funnel. The conformer which is crystallized is the one whose conformation is favorable for binding under these conditions, with the equilibrium shifting in its favor.

Domain swapping

Recently, Eisenberg and colleagues presented an inspiring hypothesis for the origin of protein oligomerization (Bennett *et al.*, 1994, 1995). Domain-swapped oligomers have been proposed to arise when a segment of a monomeric protein is exchanged by an analogous segment from a sister monomer. Domain swapping can take place in oligomeric proteins, between their subunits or between domains within the same subunit. They further made an attractive proposition that during evolution domain-swapped dimers have been cleaved to form two separate stable monomers, which subsequently associate to form the oligomer. These may not bear a clear trace of their evolutionary origin. For some of the domain-swapped cases, such as in the case of bovine seminal ribonuclease (BS-RNase), it has been shown that there are two types of dimeric associations: one with swapped N-terminal segment and the other without the swapping. The two conformations co-exist, with the swapping occurring after the non-swapped dimer forms (Piccoli *et al.*, 1992; D'Alessio, 1995). Both conformations may be expected to populate the floor of the funnel. Swapping will be observed, depending on their relative stabilities and the barrier heights. Figure 2d illustrates schematically the shape of the folding funnel for a swapping case.

Misfolding

In real life, the outcome of folding and binding events is determined by thermodynamic and kinetic considerations. As described above, the folding and the binding processes go through an overall funnel-shaped landscape. Either of these can be further decomposed into sub-funnel landscapes of parts of a combinatorial assembly process. Misfolding occurs either because the funnel bottom of the (previous) native conformation is replaced by the misfolded conformation (such as due to a mutation) or since a large bump in the funnel-shaped landscape traps the misfolded conformation. The former reflects the thermodynamic control of misfolding, whereas the latter is controlled by kinetics. If there is an already preformed misfolded nucleus, such as a seed of an amyloid or of an aggregate, the misfolding event will still be controlled either by thermodynamics or by kinetics. As in the protein binding complexed case described above, the funnel will be reshaped. However, in this case here the reshaping is due to the existence of the misfolded protein seed in the overall folding/binding process which is described by the funnel.

The above situation may be illustrated by two kinetic models for the transformation and propagation of the prion protein in bovine, from PrP^{c} to PrP^{Sc} (Prusiner, 1991; Jarrett and Lansbury, 1993). In the first model, the rate-limiting step is the irreverisble autocatalytic conversion of a monomeric PrP^{c} to a monomeric PrP^{Sc}, followed by fast oligomerization of PrP^{Sc} (Prusiner, 1991). This model corresponds to a funnel where native and misfolded conformations are separated by a large barrier. The second model proposes a fast equilibrium between monomeric PrP^{c} and monomeric PrP^{Sc} precursor with the rate-limiting step being the formation of a PrP^{Sc} oligomer. Once formed, the oligomer acts as nucleus for the growth of amyloids (Jarrett and Lansbury, 1993). In this model both the native and the misfolded conformations are populated around the bottom of funnel without being separated by a large barrier.

Chaperones help proteins in climbing up the hill, out of a conformational minimum. In this way they assist correct folding *in vivo*. In particular, this appears to be the case for complex, non-sequentially folding proteins. In the absence of a chaperone, non-native associations may take place, creating a deep crevice and hence trapping misfolded conformations. The chaperone prevents such a kinetic misfolding trap through two different mechanisms: either through its general assistance in the rapid uphill climb or by preventing it from *a priori* getting into the trap. Misfolding may also be triggered by a 'helper' protein. In the case of the prion it is its own self, while in a different conformation. This is a mechanism which is obliquely opposite to that of the chaperone. Amyloid forms when the bottom of funnel is populated by the polymerized state. The funnel landscape would then illustrate a shift from a single soluble molecule into a multiple-component, complicated system.

Domain swapping and misfolding are the outcome of similar phenomena, reflecting metastable states. Around the bottom of the funnel both the swapped and the misfolded conformers can be similarly represented by minima which are well separated by relatively high barriers from the native 'functional' conformers. Domain swapping manifests itself in intra- or intermolecular protein binding while misfolding occurs on the level of protein folding. However, at higher concentrations, misfolding is also expressed in binding, such as in the formation of amyloids or in aggregation. Hence, in the case of misfolding, the native conformation may be more stable than the single misfolded conformation. However, once there exist misfolded conformations which form an amyloid or an aggregate, the bound misfolded conformation is the more stable one. Consistently, the GroEL and HSP104 chaperones promote conversion of PrP^{c} to PrP^{Sc} in the presence of partially denatured PrP^{Sc} (DebBurman *et al.*, 1997).

The energy landscape around the bottom of the funnel and biological activity

The funnel concept is embodied in viewing folding as going downhill via multiple, parallel routes. At any stage the molecule exists in an ensemble of conformations, transiently trapped in local minima. However, since proteins function through interactions with identical or different molecules, it is extremely important to consider not solely the folding process and the numerous downhill pathways down the energy surface of the funnel, but also the vicinity of the bottom of the funnel. Here we argue that the implications of the 'new view' of protein folding cast new light on some long-held concepts in protein (or any macromolecular) binding. Molecular flexibility can be portrayed as a rugged energy surface around the bottom of the funnel. The larger the flexibility, the greater is the population of diverse conformers and the lower are the barriers between them. Conversely, the more rigid the molecule, the fewer are the minima and, in particular, the higher are the barriers between them. Non-specific binding fits into the first scenario and specific, rigid binding fits into the second. Clearly, this does not necessarily imply that specific recognition and binding always require protein rigidity. Extreme rigidity may interfere with biological function and hence be unfavorable to binding (Shoichet *et al.*, 1995; Ogata *et al.*, 1996; Pieper *et al.*, 1997; Wallon *et al.*, 1997; Zavodszky *et al.*, 1998).

There is considerable evidence implying a correlation between flexibility and a range of conformational isomers. First, Ditzel *et al.* (1996) have elegantly shown that an antibody which was polyreactive at 37°C became monoreactive when the temperature was lowered to 4°C. An additional Fab showed a similar, though less marked trend. Consistently, we have recently observed a correlation between the melting temperatures of thermophiles and the difference in the number of salt bridges between thermophiles and their close mesophilic homologs (Kumar *et al.*, 1999a, submitted). Additionally, the number of salt bridges decreases in the psychrophiles (Feller *et al.*, 1994) living in the Antarctic. We have further carried out a detailed analysis of the electrostatic energy contributions of the salt bridges in glutamate dehydrogenases from an extreme thermophile (the hyperthermophilic archaeon *Pyrococcus furiosus*; Yip *et al.*, 1995) and its comparison with its mesophilic (from *Clostridium symbiosum*) homolog (Kumar *et al.*, 1999b). Many of the extra salt bridges in the thermophilic which are absent in the mesophilic enzyme are around the active site of the protein. The electrostatic contribution of salt bridge energies in *P.furiosus* glutamate dehydrogenase stabilizes the protein. In contrast, the salt bridges in the mesophilic *Clostridium symbiosum* homolog contribute only marginally to protein stability. This is largely due to the difference in the protein environment around the salt bridges in the two proteins. The larger number of salt bridges cooperatively enhances their strength. These results indicate that salt bridges and their networks rigidify the protein structure particularly around the active site at high temperatures. In agreement with these results, recently Zavodszky *et al.* (1998) studied the 3-isopropylmalate dehydrogenase (IPMDH) from the thermophile *Thermus thermophilus*. The melting temperature of this enzyme is 17°C higher than that of its mesophilic *Escherichia coli* homolog. However, when its activity was tested at room temperature, it was found to be non-functional owing to its increased flexibility at that temperature. Further corroboration comes from both Ichiyoshi and Casali (1994) and Ditzel *et al.* (1996), who have shown that by grafting complementary determining region (CDR) 3, particularly HCDR3, a broad range of recognition can be conferred on the antibody.

Hence conformational diversity around the bottom of the funnel may provide an elegant and simple solution to a range of binding processes, without necessitating an 'induced fit'-type mechanism. Conformational isomerism is a general phenomenon, borne by the laws of statistical thermodynamics applied to protein folding and binding. Novel experiments may be needed to detect this phenomenon.

Acknowledgements

We thank Dr Jacob Maizel for encouragement and for helpful discussions. The research of R.Nussinov in Israel has been supported in part by grant

number 95-00208 from BSF, Israel, by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by a Magnet grant, by a Ministry of Science grant and by Tel Aviv University Basic Research and Adams Brain Center grants. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-56000. The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

Abad-Zapatero,C., Rydel,T.J. and Erickson,J. (1990) *Proteins*, **8**, 62–81.
Amit,A.G., Mariuzza,R.A., Phillips,S.E.V. and Poljak,R.J. (1986) *Science*, **233**, 747–753.
Arrowsmith,C.H., Carey,J., Treat-Clemons,L. and Jardetzky,O. (1989) *Biochemistry* ,**28**, 3875–3879.
Baldwin,R.L. (1994) *Nature*, **369**, 183–184.
Baldwin,R.L. (1995) *J. Biomol. NMR*, **5**, 103–109.
Barford,D. and Johnson,L.N. (1989) *Nature*, **340**, 609–616.
Bennett,M.J., Choe,S. and Eisenberg,D. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 3127–3131.
Bennett,M.J., Schlunegger,M.P. and Eisenberg,D. (1995) *Protein Sci*., **4**, 2455–2468.
Bhat,T.N., Bentley,G.A., Fischmann,T.O., Boulot,G. and Poljak,R.J. (1990) *Nature*, **347**, 483–485.
Boczko,E.M. and Brooks,C.L. III (1995) *Science*, **269**, 393–396.
Braden B.C., Fields B.A., Ysern X., Goldbaum F.A., Dall'Acqua W., Schwarz F.P., Poljak R.J. and Mariuzza R.A. (1996) *J. Mol. Biol*., **257**, 889–894.
Bryngelson,J.D. and Wolynes,P.G. (1989) *J. Phys. Chem.*, **93**, 6902–6915.
Carey,J. (1989) *J. Biol. Chem.*, **264**, 1941–1945.
D'Alessio,G. (1995) *Nature Struct. Biol*., **2**, 11–13.
DebBurman,S.K., Raymond,G.J., Caughey,B. and Lindquist,S. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 13938–13943.
Dill,K.A. (1999) *Protein Sci.*, **6**, 1166–1180.
Dill,K.A. and Chan,H.S. (1995) *Protein Sci.*, **4**, 561–602.
Dill,K.A. and Chan,H.S. (1997) *Nature Struct. Biol* **4**, 10–19.
Ditzel,H.J., Itoh,K. and Burton,D.R. (1996) *J. Immunol*., **157**, 739–749.
Erickson,J.W., Baldwin,E.T., Bhat,T.N. and Gulik,S. (1995) In Takahashi,K. (ed.), *Aspartic Proteinases: Structure, Function, Biology and Biomedical Implications*. Plenum Press, New York.
Feller,G., Payan,F., Theys,F., Qian,M., Haser,R. and Gerday,C. (1994) *Eur. J. Biochem.*, **222**, 441–447.
Fischer,E. (1894) *Ber. Dtsch. Chem. Ges*., **27**, 2985–2991.
Foote,J. and Milstein,C. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 10370–10374.
Frauenfelder,H. and Leeson,D.T. (1998) *Nature Struct. Biol.*, **5**, 757–759.
Frauenfelder,H., Sligar,S.G. and Wolynes,P.G. (1991) *Science*, **254**, 1598–1603.
Gruebele,M. and Wolynes,P. (1998) *Nature Struct. Biol.*, **5**, 662–665.
Gryk,M.R., Jardetzky,O., Klig,L.S. and Yanofsky,C. (1996) *Protein Sci.*, **5**, 1195–1197.
Herron,J.N., He,X.M., Ballard,D.W., Blier,P.R., Pace,P.E., Bothwell,A.L., Voss,E.,Jr and Edmundson,A.B. (1991) *Proteins*, **11**, 159–175.
Huang,M., Syed,R., Stura,E.A., Stone,M.J., Stefano,R.S., Ruf,W., Edgington, T.S. and Wilson,I. (1998) *J. Mol. Biol.*, **275**, 873–894.
Ichiyoshi,Y. and Casali,P. (1994) *J. Exp. Med.*, **180**, 885–895.
James,M.N.G., Sielecki,A., Salituro,F., Rich,D.H. and Hofmann,T. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 6137–6141.
Jarrett,J.T. and Lansbury,P.T.,Jr (1993) *Cell*, **73**, 1055–1058.
Karplus,M. (1997) *Folding Des.*, **2**, S69–S75.
Karplus,M. and Shakhnovitch,E. (1992) In Creighton,T. (ed.), *Protein Folding*. Freeman, New York, pp. 127–195.
Karplus,M., Sali,A. and Shakhnovitch,E. (1995) *Nature*, **373**, 664–665.
Ke H.M., Liang J.Y., Zhang Y.P. and Lipscomb W.N. (1991) *Biochemistry*, **30**, 4412–4420.
Koshland,D.E.,Jr (1958) *Proc. Natl Acad. Sci. USA*, **44**, 98–123.
Lazaridis,T. and Karplus,M. (1997) *Science* **278**, 1928–1931.
Lawson,C.L., Zhang,R.G., Schevitz,R.W., Otwinowski,Z., Joachimiak,A. and Sigler,P.B. (1988) *Proteins*, **3**, 18–31,.
Lee,A.Y., Gulnik,S.V. and Erickson,J.W. (1998) *Nature Struct. Biol*., **5**, 866–871.
Lewis,M., Chang,G., Horton,N.C., Kercher,M.A., Pace H.C., Schumacher M.A., Brennan,R.G. and Lu P. (1996) *Science*, **271**, 1247–1254.
Marmorstein,R. and Harrison,S.C. (1994) *Genes Dev.*, **8**, 2404–2412.
Martinez,J.C., Pisabarro,M.T. and Serrano,L. (1998) *Nature Struct. Biol.*, **5**, 721–729.
Milburn,M.V., Prive,G.G., Milligan,D.L., Scott,W.G., Yeh,J., Jancarik, J, Koshland,D.E.,Jr and Kim,S.H. (1991) S*cience*, **254**, 1342–1347.
Muller,Y., Ultch,M.H. and De Vos,A.M. (1996) *J. Mol. Biol.*, **256**, 144–159.
Muller,Y., Kelley,R.F. and De Vos,A.M. (1998) *Protein Sci.*, **7**, 1106–1115.
Newman,M., Strzelecka,T., Dorner,L.F., Schildkraut,I. and Aggarwal,A.K. (1995) *Science*, **269**, 656–663.
Ogata,K. *et al.* (1996) *Nature Struct. Biol.*, **3**, 178–187.
Onuchic,J.N., Wolynes,P.G., Luthey-Schulten,Z. and Socci,N.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 3626–3630.
Perutz,M.F., Wilkinson, A,J., Paoli,M. and Dodson,G.G. (1998) *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 1–34.
Piccoli,R., Tamburrini,M., Piccialli,G., Di Donato,A., Parente,A. and D'Alessio,G. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1870–1874.
Pieper,U., Hayakawa,K., Li,Z. and Herzberg,O. (1997) *Biochemistry*, **36**, 8767–8774.
Prusiner,S.B. (1991) *Science*, **252**, 1515–1522.
Rini,J.M., Stanfield,R.L., Stura,E.A., Salinas,P.A., Profy,A.T. and Wilson,I.A. (1993) *Proc Natl Acad Sci USA*, **90**, 6325–6329.
Rose,R.B., Craik,C.S. and Stroud,R.M. (1998) *Biochemistry*, **37**, 2607–2621.
Sali,A., Cooper,J.B., Foundling,S.I., Hoover,D.J. and Blundell,T.L. (1989) *EMBO J*., **8**, 2179–2188.
Schirmer,T. and Evans,P.R. (1990) *Nature*, **343**, 140–145.
Shaknovich,E.T. and Gutin,A.M. (1993) *Protein Engng*, **6**, 793–800.
Shoichet,B.K., Baase,W.A., Kuroki,R. and Matthews,B.W. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 452–456.
Silva,A.M. *et al.* (1996) *Proc. Natl Acad. Sci. USA*, **93**, 10034–10039.
Stanfield,R.L., Takimoto-Kanimura,M., Rini,J.M., Profy,A.T. and Wilson,I.A. (1993) *Structure*, **1**, 83–93.
Stevens,R.C. and Lipscomb,W.N. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5281–5285.
Sun,J. and Sampson,N.S. (1998) *Protein Sci*., **7**, 1495–1505.
Tsai,C.J. and Nussinov,R. (1997a) *Protein Sci.* **6**, 24–42.
Tsai,C.J. and Nussinov,R. (1997b) *Protein Sci*., **6**, 1426–1437.
Tsai,C.J., Xu,D. and Nussinov,R. (1998) *Folding Des*., **3**, R71–R80.
Tsai,C.J., Kumar,S., Ma,B. and Nussinov,R. (1999) *Protein Sci.*, **6**, 1181–1190.
Waldburger,C.D., Jonsson,T. and Sauer,R.T. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 2629–2634.
Wallon,G., Lovett,S.T., Magyar,C., Svingor,A., Szilagyi,A., Zavodszky,P., Ringe,D., Patsko,G.A. (1997) *Protein Engng*, **10**, 665–672.
Wang,Y. and Shortle,D. (1996) *Protein Sci.*, **5**, 1898–1906.
Wedemayer,G.J., Patten,P.A., Wang,L.H., Schultz,P.G. and Stevens R.C. (1997) *Nature Struct. Biol*., **276**, 1665–1669.
Wilderspin,A.F. and Sugrue,R.J. (1994) *J. Mol. Biol.*, **239**, 97–103.
Wolynes,P.G., Onuchic,J.N. and Thirumalai,D. (1995) *Science*, **267**, 1619–1620.
Yip,K.S.P. *et al.* (1995) *Structure*, **3**, 1147–1158.
Zavodszky,P., Kardos,J., Svingor,A. and Petsko,G.A. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 7406–7411.

Received March 25, 1999; revised May 28, 1999; accepted June 4, 1999