Principles of Protein–Protein Interactions: What are the Preferred Ways For Proteins To Interact?

Ozlem Keskin,*,† Attila Gursoy,† Buyong Ma,‡ and Ruth Nussinov*,‡,§

Koc University, Center for Computational Biology and Bioinformatics and College of Engineering, Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey; Basic Research Program, SAIC–Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI—Frederick, Frederick, Maryland 21702; and Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Received July 13, 2007

Contents

* Correspondence should be addressed to R. Nussinov and O. Keskin at NCI—Frederick, Bldg. 469, Rm. 151, Frederick, MD 21702. Tel.: (301) 846-5579. Fax: (301) 846-5598. E-mail: okeskin@ku.edu.tr and ruthn@ncifcrf.gov.
† Center for Computational Biology and Bioinformatics and College of Engineering.
‡ NCI—Frederick.
§ Tel Aviv University.

1. Introduction

1.1. Protein–Protein Interactions: Toward Functional Prediction and Drug Design

Proteins are the working horse of the cellular machinery. They are responsible for diverse functions ranging from molecular motors to signaling. They catalyze reactions, transport, form the building blocks of viral capsids, traverse the membranes to yield regulated channels, and transmit the information from the DNA to the RNA. They synthesize new molecules, and they are responsible for their degradation. Proteins are the vehicles of the immune response and of viral entry into cells. The broad recognition of their involvement in all cellular processes has led to focused efforts to predict their functions from sequences, and if available, from their structures (e.g., refs 1−6). A practical way to predict protein function is through identification of the binding partners. Since the vast majority of protein chores in living cells are mediated by protein−protein interactions, if the function of at least one of the components with which the protein interacts is identified, it is expected to facilitate its functional and pathway assignment. Through the network of protein−protein interactions, we can map cellular pathways and their intricate cross-connectivity (e.g., refs 7−11). Since two protein partners cannot simultaneously bind at the same (or overlapping) site, discovery of the ways in which proteins associate should assist in inferring their dynamic regulation. Identification of protein−protein interactions is at the heart of functional genomics. Prediction of protein−protein interactions is also crucial for drug discovery. Knowledge of the pathway and its topology, length, and dynamics should provide useful information for forecasting side effects.

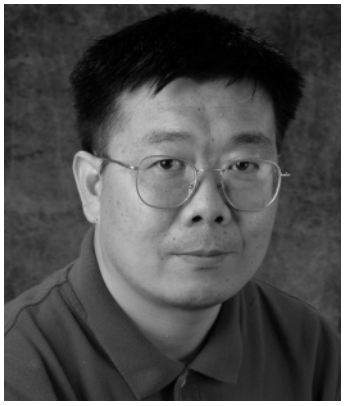
While it is important to predict protein associations, it is a daunting task. Some associations are obligatory, whereas others are transient, continuously forming and dissociating.[12−18] From the physical chemical standpoint, any two proteins can

10.1021/cr040409x CCC: $71.00
Published on Web 03/21/2008

Currently, Ozlem Keskin is an associate professor in the Chemical and Biological Engineering Department and Center for Computational Biology and Bioinformatics at Koc University, Istanbul. Her work focuses on computational biology and bioinformatics on understanding the physical principles and dynamics of macromolecular systems, basically the principles of protein–protein interactions and prediction of interactions. Before her career in Koc University, she was a postdoctoral fellow at the National Cancer Institute-National Institutes of Health, U.S.A., during 1999–2001. She received her Ph.D. degree in Chemical Engineering in 1999, at Bogazici University, Istanbul. She received UNESCO-L'OREAL Co-Sponsored Fellowship Award for Young Women in Life Sciences, 2005; Turkish Academy of Sciences (TUBA) Distinguished Young Investigator Award, 2006; best Ph.D. Dissertation Award, 1999, Bogazici University; and International Integrated Graduate Research Fellowship from Scientific and Technical Research Council of Turkey (TUBITAK) 1997–1999.

Attila Gursoy is an associate professor in the Computer Engineering Department, Koc University, Istanbul, Turkey. He received his Ph.D. degree from University of Illinois at Urbana–Champaign in Computer Science in 1994. He joined Theoretical Biophysics Group in Beckman Institute, University of Illinois, as a postdoctoral research associate, where he contributed significantly to the development of NAMD, a parallel molecular dynamics simulation program. Dr. Gursoy worked at Bilkent University, Ankara, Turkey, as an assistant professor, and since 2002, he is with Computer Engineering Department and Center for Computational Biology, Koc University. Dr. Gursoy's research interests are in the area of computational biology, particularly protein interactions and high-performance algorithms for computational biology.

interact. The question is under what conditions and at which strength. Protein−protein interactions are largely driven by the hydrophobic effect.[19−21] Hydrogen bonds and electrostatic interactions play crucial roles,[22−25] and covalent bonds are also important. The physical chemical principles of protein−protein interactions are general, and many of the interactions observed in vitro are the outcome of experimental overexpression or of crystal effects, complicating functional prediction. The Gibbs free energy upon complex formation (also called binding free energy) can be evaluated directly from the equilibrium constant of the reaction (usually denoted as



Buyong Ma received his B.Eng. degree in Polymer Chemistry from the Hefei University of Technology in 1984. He received his Ph.D. degree in Physical Chemistry from the University of Georgia in 1995. From 1995 to 1998, he was a postdoctoral researcher with Professor N. L. Allinger, working in the field of molecular mechanics. In 1998, he joined the National Cancer Institute (NCI) as a Research Fellow in Professor Ruth Nussinov's group. He was a research scientist in Locus Pharmaceuticals from 2002 to 2003. In 2003, he returned to NCI as a senior scientist at SAIC–Frederick, Inc. His research interests cover computational approaches to protein–protein interaction, protein-nucleic acid interaction, and protein aggregation.

Ruth Nussinov received her Ph.D. in 1977, from the Biochemistry Department at Rutgers University, and did postdoctoral work in the Structural Chemistry Department of the Weizmann Institute. Subsequently she was at the Chemistry Department at Berkeley, the Biochemistry Department at Harvard, and the NIH. In 1984 she joined Tel Aviv University. In 1990 she became a Professor in the Department of Human Genetics, at the Medical School. In 1985, she accepted a concurrent position at the National Cancer Institute of the NIH, where she is a Senior Principal Investigator heading the Computational Structural Biology Group. She has authored over 300 scientific papers. Her interests largely focus on protein folding, protein–protein interactions, amyloid conformations, and large multimolecular associations with the goal of understanding the protein structure–function relationship.

K_a and K_d, for association or dissociation constants) to assess how stable the interactions are. These constants are functions of the concentrations of the free protein and the complexed form at thermodynamic equilibrium. The K_d is wide (between Micromolar and Picomolar) in protein−protein interaction, resulting in free energy changes (ΔG_a) of −6 to −19 kcal/mol. Both enthalpic (ΔH) and entropic (ΔS) contributions are temperature dependent in the Gibbs free energy. The formation of the complex is said to be enthalpy driven if ΔH is negative (favoring association) and ΔS is negative (disfavoring association) and entropy driven otherwise.[26]

To be able to predict protein−protein interactions, there is a need to figure out the chemical aspects of their associations.[27−36] These range from shape complementarity

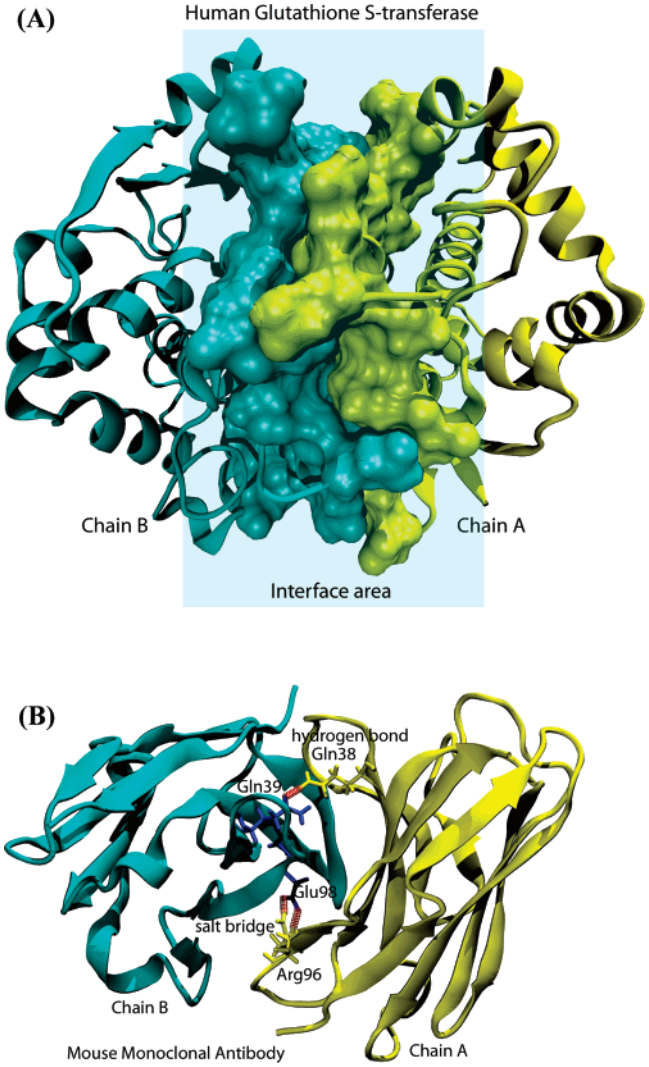


Figure 1. Illustration of protein−protein interfaces. (A) The figure represents two interacting proteins (human glutathione S-transferase, PDB ID: 10gs, Chains A and B). The two chains are colored yellow and cyan. Interacting residues from the two chains are shown with surface representation in order to emphasize the complementarity, while the rest of the proteins are illustrated with ribbon representations. (B) The details of the interface of mouse monoclonal antibody D1.3 (PDB ID: 1kir, Chains A (yellow) and B (cyan)). The H-bond between Gln38 in Chain A and Gln 39 in Chain B and the salt bridge between Arg96 in Chain A and Glu98 in Chain B are highlighted.

to the organization[37] and the relative contributions of the physical/chemical components to their stability. Proteins interact through their interfaces. Interfaces consist of interacting residues that belong to two different chains, along with residues in their spatial vicinity. Thus, interfaces consist of fragments of each of the chains and some isolated residues. Figure 1 illustrates some examples of protein−protein interfaces. To analyze protein−protein interactions, residues (or atoms) that are in contact across the two-chain interface are studied. In addition, residues in their vicinity are also inspected to explore the chemical effects of their supporting matrix.[32,38−42] At the same time, it behooves us to remember that proteins are flexible. Proteins that are free in solution exist in ensembles of interconverting conformations. Backbones and side-chains move. In addition, native proteins frequently populate distinct minima that are separated by low, yet not so easy to surmount, barriers. These conformers lie on the rugged bottom of the funnel, reflecting multiple conformational states and allosteric effects.[43] Conformational and dynamic allosteric effects are the outcome of binding to other molecules, proteins, small molecules, or nucleic acids, leading to population shifts. Such allosteric effects are the hallmarks of functional regulation. Depending on the extent of the conformational change in the binding site, they may mislead predictions of protein−protein interactions. In viewing proteins as static structures, the properties of a particular population are explored. Yet, if we consider hub proteins, proteins with shared binding sites, or proteins involved in regulation, different populations may preferentially associate with different partners.

A large fraction of cellular proteins are estimated to be "natively disordered", i.e., unstable in solution.[44−46] The structures of disordered proteins are not "random". Rather, the disordered state has a significant residual structure.[47−50] In the "disordered" state, a protein exists in an ensemble of conformers. In many cases, these regions constitute only certain parts or domains of the whole protein. Disordered proteins are believed to account for a large fraction of all cellular proteins and to play roles in cell-cycle control, signal transduction, transcriptional and translational regulation, and large macromolecular complexes.[51] While disordered on their own, their native conformation is stabilized upon binding. The global fold of disordered proteins does not change upon binding to different partners; however, local conformational variability can be observed, inevitably complicating the predictions of protein interactions.

The overriding reasons for the heightened interest in protein−protein interactions are that better understanding and better quantization of the key features controlling the interactions should lead to higher success in the prediction of protein associations.[28,52,53] This would assist in the elucidation of cellular pathways and in drug design. It will also assist in figuring out the effects of crucial mutations, which are often clustered in binding sites, as in p53.[54,55]

Below, we aim to provide an overview of the principles of protein−protein interactions. Within this framework, we highlight what we consider are key components in the question of "what are the preferred ways for proteins to interact". The goal is to be able to predict *how* the proteins will interact. Our assumption is that the structures are available and that there are experimental data that the proteins do interact. In the absence of such data, docking the structures of any pair of proteins will always find a matching patch of surface that may appear favorable.[56−58]

1.2. Proteins are Flexible Molecules Even Though We Frequently Treat Them as Rigid

When carrying out an analysis of protein−protein binding interfaces, the routine procedure is to examine the complexes as they are available in their crystal structures. Hence, the protein is treated as a rigid molecule in that crystal conformation. Yet, the conformation observed in the complex is not necessarily the one that prevails in solution.[59−62] Moreover, depending on its binding state, i.e., whether it is already bound to another protein (or ligand) although at another binding site, different prevailing conformational states may be populated.[63−66] Figure 2A illustrates the free energy landscape and the shift in the populations and, consequently, in the prevailing binding-site shape upon binding to another protein at another site.

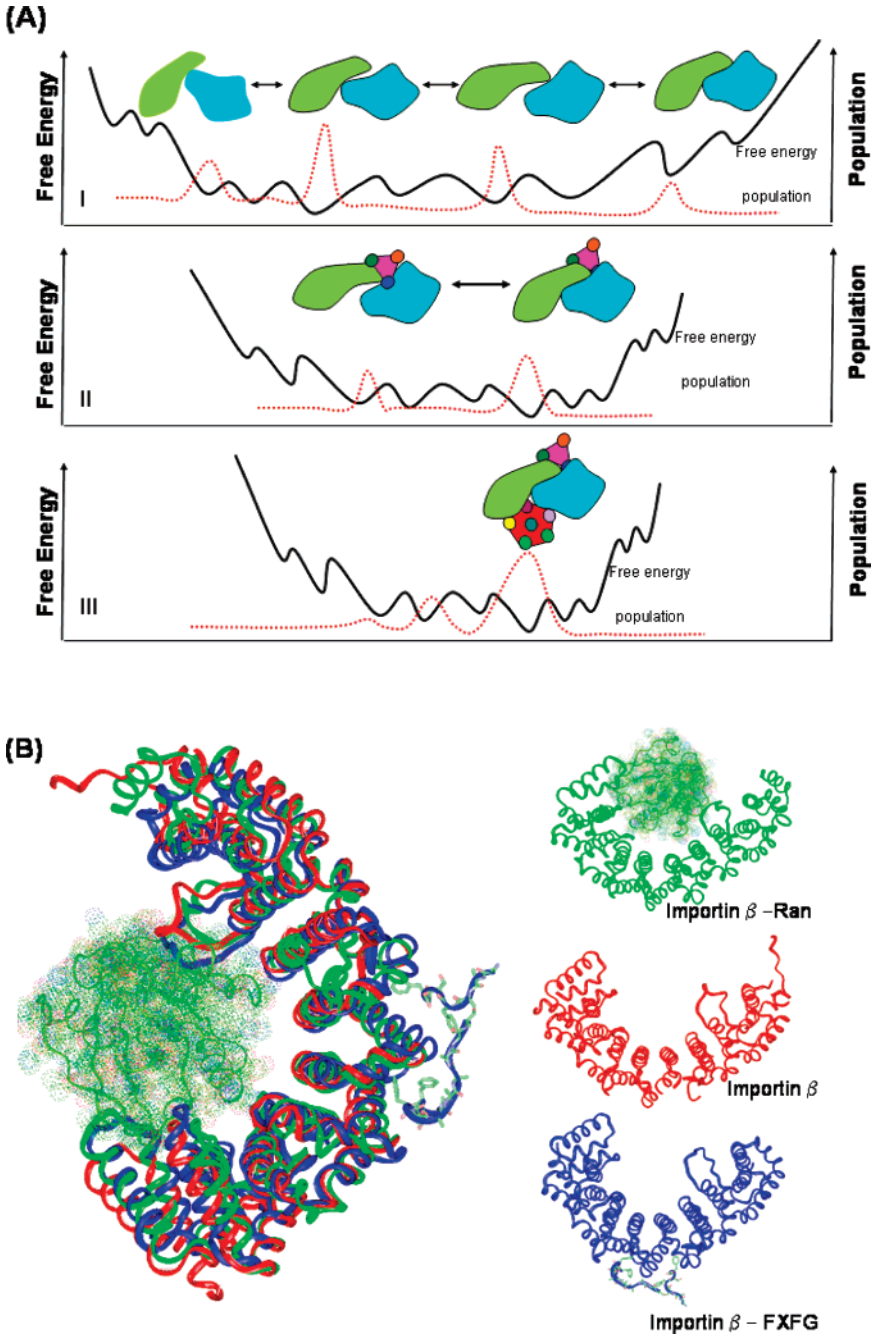


Figure 2. (A) The free energy landscape of a protein may change upon binding to another protein. Binding may induce a shift in the distribution of the populations of the conformational states of the protein; consequently, the relative population of the conformer with an altered binding site shape at another location on the protein surface may increase. The solid black line refers to the free energy landscape, and the dashed red line refers to the relative populations. (I) Distribution of the substates of the protein conformations, presenting several binding possibilities. (II) When a ligand binds at the first binding site, it shifts the conformational energy landscape and the distribution of the populations to favor selective binding at a second, allosteric site. (III) The final dominant conformer recognizes both ligands. (B) Conformational variability is very important for importin to mediate nucleo-cytoplasmic transportation. Shown here are the superimposition (left panel) of three crystal structures of importin in the free state (red ribbon, left panel, PDB ID: 1gcj), bound to RanGTP (green ribbon, left panel, PDB ID: 1ibr; RanGTP is represented by ribbon and surface dots), and bound to nucleoporin (blue ribbon, left panel, PDB ID: 1f59). The bound/unbound conformational states are coupled with the importin functions of cargo binding and release by RanGTP binding. The importin conformations in the three crystal structures differ significantly in their binding sites with an overall rmsd around 3.5 Å. In solution, SAXS revealed much larger conformational variations.[224]

Further, the crystal structure used in the prediction of the protein−protein interaction is likely to also be affected by the crystallization conditions.[61] The crystal structure presents a homogeneous population of one conformer, whereas other conformers are not accounted for. For example, importin has different conformations in different complexes (Figure 2B). The existence of populations of such conformers is reflected in the crystallization time scales. Molecular dynamics simulations assist in the sampling; however, the sampling is a function of the barrier heights between the different populations and of the simulation time scales. Hence, the small backbone and the side-chain movements are likely to be sampled; however, distinct conformers even with a limited conformational change may not be visited in the simulations, presenting a problem in the analysis and prediction of the preferred

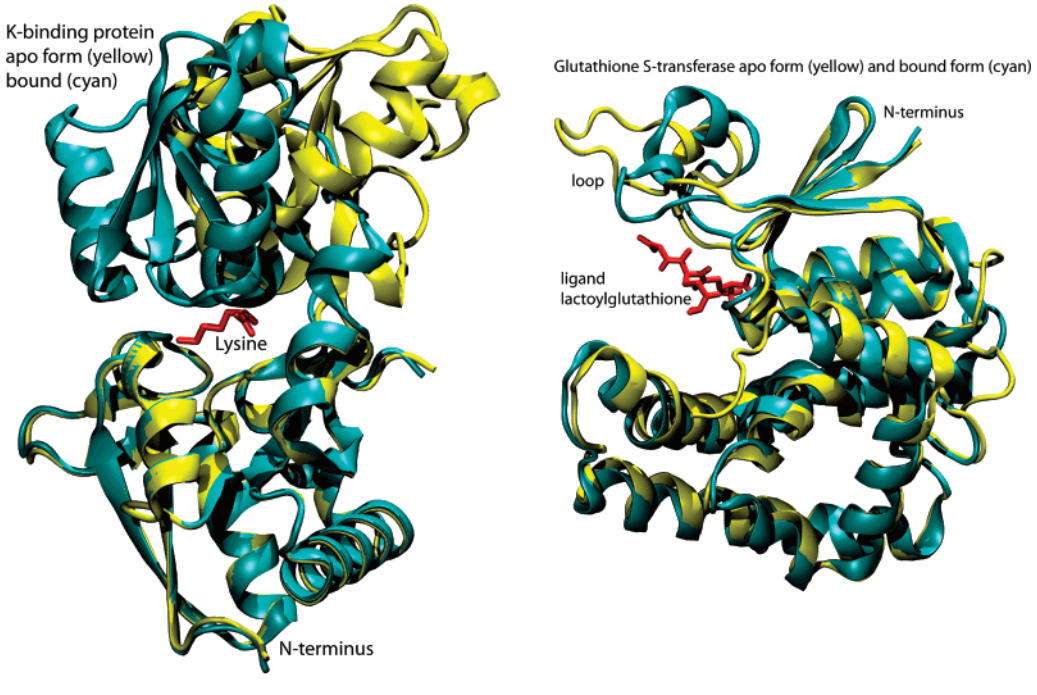


Figure 3. Comparisons of the proteins when they are in the bound, complexed states versus in the free (apo) states. (A) The conformational changes undertaken by K-binding protein (PDB IDs: 2lao (yellow) and 1lst (cyan)). The free structure (yellow) closes up and becomes stabilized when it is bound (cyan structure) to its ligand. The ligand, shown in red, belongs to the cyan structure. This is a domain motion example. (B) Glutathione S-transferase-I in free and bound forms (PDB IDs: 1aw9 (shown in cyan) and 1axd (yellow), respectively). The ligand introduces a conformational change in the loop.

interactions. Figure 3 presents a few examples of complexed versus free protein molecules.

1.3. Proteins Interact through Their Surfaces

Proteins interact through their surfaces. Consequently, analyses usually focus on protein surfaces. To identify the residues and atom groups that line the surfaces, it is essential to have the structures of the proteins. The determination of which residues and atoms are on the surface is usually carried out through calculations of the surface area that is accessible to the solvent.[32,39,67,68] Figure 4 illustrates some binary protein complexes. In each complex, one protein is colored purple and the accompanying protein is in yellow. Both side-chain and backbone atoms can be on the surface, interacting with solvent molecules. If the molecule interacts with another protein molecule, atoms on the surface of one molecule will interact with atoms on the surface of the partner protein. To understand the nature of the intermolecular interaction, various properties of the protein−protein interface are examined, for example, the surface area that is buried by the interacting molecules and what fraction is nonpolar; the hydrogen bonds across the interface and the salt bridges; buried water molecules; the composition of the interface; residue conservation; the strength of the interaction; residues that contribute significantly to the free energy of binding; the shape of the binding interface; and the types of secondary structures.[17,37,40,69−72] Figure 1 presents a few examples of protein−protein interaction interfaces, highlighting some of these features. Yet, while all of these properties are essential, they provide insufficient description of the binding. This can be best judged by the difficulties in the correct prediction of protein−protein associations and in accounting for mutational effects.

The major features of the interaction vary substantially among proteins. These depend on the protein surface at the binding site, on protein stability, and on the distribution of the protein conformational substates, as well as the types and locations of the conformational changes that are involved. This is the major reason why the availability of the protein structures and the description of their surfaces are insufficient for an accurate prediction of protein−protein interactions. Despite the detailed chemical description of the protein molecular surface, our ability to correctly assess a possible association is limited. Hence, while the availability of the protein structures is essential for prediction of the protein−protein complex and an estimation of its stability, in the absence of additional biochemical data, the problem is still extremely difficult and predictions cannot be considered reliable. A potential exception that increases the confidence level is when the proteins present complementary surface patches similar to those shown to interact (e.g., ref 1). However, here too computational predictions are mere candidates for the experiment to test.

2. Cooperativity in Protein Folding and in Protein–Protein Associations

Cooperativity is nonindependence. It is generally accepted that proteins fold cooperatively. If proteins were to fold noncooperatively, in order to reach the global minimum they would need to perform an exhaustive search of the conformational space. However, the time scales that are involved in an exhaustive search are not physiologically relevant. This challenging question of the physical basis of cooperativity through which proteins would avoid an exhaustive search has been the focus of considerable research (e.g., refs 73 and 74). Cooperativity derives from the hydrophobic effect,

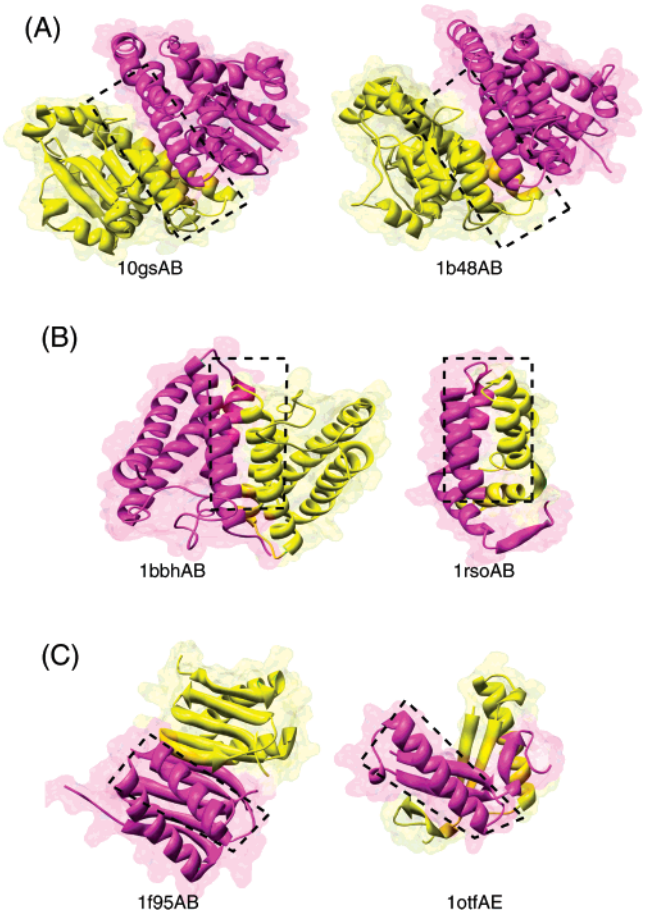


Figure 4. Several examples of crystal structures of binary protein complexes. The interfaces are highlighted with boxes. In part A, the two glutathione S-transferase complexes (PDB IDs: 10gs and 1b48) are homologous; they use similar interfaces to bind each other. In part B, the two complexes, cytochrome C and neuropeptide/membrane protein (PDB IDs: 1bbh and 1rso) are not related evolutionarily, yet the interface architecture is similar. Part C represents two complexes (dynein light chain 8 (PDB ID: 1f95AB) and 4-oxalocrotonate tautomerase (PDB ID: 1otfAE)) where only one side of the interface has similar architectures, the accompanying sides are unrelated. The similar side belongs to the magenta chains.

the driving force in a single-chain protein folding.[73] Proteins that are approximated by a two-state transition correspond to an *all-or-none* description of protein folding displayed by cooperatively folding hydrophobic folding units. Such a behavior is typically observed in small globular proteins consisting of one hydrophobic unit; on the other hand, larger chains do not fold cooperatively into a single hydrophobic unit. The hydrophobic folding units that are observed at the interfaces of *two-state complexes* similarly suggest the cooperative nature of the two-chain protein folding, also the outcome of the hydrophobic effect.[75] Thus, cooperativity implies preferred protein folding pathways.

To understand cooperativity, we need to think of the system as a cohesive unit, where the parts do not behave independently of each other. The behavior of the system is the outcome of the properties of the system as a whole, rather than the sum of the properties of the individual components. In our case, the thermodynamic stability of the protein−protein complex is not a simple summation of the individual contributions of each of the residues or of the pairs of residues; rather, residues that are in direct spatial contact, or in close contact through a few tightly packed intermediate residues, impact the stability of the association in a nonadditive manner. Substitution of a tightly packed residue would inevitably affect the interactions of its neighboring residues. Thus, a mutation affects the stability of the complex since the interactions will change; however, at the same time, since the residue is tightly packed in the native complex, its substitution will also impact the stability of the complex indirectly, through the changes of the interactions of its neighbors. This may occur if a large residue is substituted by a smaller residue leading to side-chain (and backbone) movements to fill the "hole" that is created; by contrast, if a smaller residue is substituted by a larger one, the neighboring residues' contacts will change to allow accommodation of the inserted residue in the tight environment.[76,77] The extent and direction of the impact depends on the type and environment of the substitutions. Either way, this would affect the stability of the complex beyond the direct altered interactions of the mutated residue. This implies that, if we simultaneously mutate two contacting or spatially nearby residues in a tightly packed environment, the change in the stability would not be the sum of the measured changes of each one separately. The measured change in the thermodynamic stability upon a mutation of a *single* residue already implicitly takes into account changes in the interactions of its closely packed neighboring residues. Hence, a summation of the substitutions of two residues that are in spatial proximity may overestimate (or underestimate) the total contribution. On the other hand, if the protein−protein interface can be separated into units, the impact of mutations in each of these is independent and these can be summed. That is, these contributions are noncooperative. Such effects have been shown in a range of systems.[42,78−82] The affinity maturation process through which proteins evolve to bind with increased affinity has been shown to be a particularly useful system for studies of cooperative effects at the residue level.[81] Cooperative effects complicate the estimation of the stability of the interactions, since the free energy change upon a mutation already implicitly accounts for some of the effects of the neighboring residues as well, making the accuracy of the per residue (or per chemical group) parametrization less accurate.

3. Protein–Protein Interfaces Have Preferred Organization

3.1. Description of Protein–Protein Interfaces

Above, we have discussed attributes that hamper predictions of protein associations. Among these, we highlighted protein flexibility, the existence of ensembles with distinct conformations separated by barriers, the difficulties encountered by the presence of even partial disorder, and the cooperativity in protein−protein association. Are there any attributes of protein−protein interactions that may assist in the prediction? For example, is there a property that distinguishes interfaces from the rest of the protein surface? If there were such a property, it could a priori be used toward a prediction, allowing us to focus on the binding sites, thus reducing the conformational search. Toward this aim, various data sets of protein−protein interfaces have been derived, divided into groups, and analyzed.[83−88] Homodimers, which are frequently *permanent complexes*, were mostly analyzed separately from heterodimers. Homodimeric interfaces resemble protein cores.[19,20] They are typically large, are hydrophobic as measured by high values of nonpolar buried surface areas, and show good complementarity between the

two chains. These interfaces can often be distinguished from the remainder of the protein surface. In contrast, this is not the case for heterocomplexes, where the chains differ from each other. Yet these largely nonpermanent complexes are the interfaces we would, in particular, like to be able to predict, since the structures of homodimeric proteins are usually obtained in the complex state. Heterocomplexes cannot be distinguished by the extent of their hydrophobicity.[89−93] Jones and Thornton[94] have compared the residue types weighted by their accessible surface areas. They have observed that large hydrophobic and uncharged polar residues were more frequent in the interfaces of heterocomplexes as compared to the rest of the surface. Charged residues were more frequent on the exposed, noninterface surface. They have further divided the surface into patches. Analysis of these has illustrated that interface patches are more planar, and their residues have larger accessible surface areas. For some interfaces, the geometric and electrostatic complementarity is important, and a small fraction of the interface residues may make a large contribution to the binding energy.[89] Thus, no single physicochemical property distinguishes sufficiently well interfaces from the remainder of the surface; on the other hand, all hydrophobicities, solvation energies, and relative solvent accessible areas and residue compositions show trends that differ in the interfaces versus the rest of the protein surface.[52]

Residue conservation was also observed to be higher in interfaces as compared to the rest of the surface.[71] Quantification of the conservation through calculation of sequence entropies complements existing methods.[90] It was further found that central interface residues were more conserved than peripheral ones.[89] Li et al.[95] examined the hydrophobicity in the center of the interface versus its periphery. To measure the hydrophobicity at the center, they replaced buried phenylalanine by smaller hydrophobic residues in structures of antibody−antigen complexes, obtaining an estimated energy change of 46 cal/mol per Å.[2] Ofran and Rost[96] observed that six types of protein−protein interfaces differed significantly from each other in their residue composition and interaction preferences. Janin and coworkers have suggested dividing the interface into cores and their surrounded rims and have used it to differentiate between biological interfaces and nonspecific crystal packing ones.[27,97] Nevertheless, while these trends may assist in the prediction, they too are insufficient.

3.2. Some Amino Acids at the Interface Are Hot Spots Since They Contribute Significantly to the Stability of the Protein–Protein Association

Are there residues in the interface that contribute dominantly to the binding free energy of the protein−protein complex or do all residues contribute roughly equally? In folding, some residues in the protein core have been shown to be important for the stability of the protein. Does the same hold for protein−protein association? To address this question, Wells and his colleagues have carried out alanine scanning.[98] Residues in the interface were systematically replaced by alanine, and the difference in the binding free energy ($\Delta\Delta G$) between the wild type and each mutant was measured. They have defined a hot spot as a residue whose substitution by alanine leads to a significant ($\Delta\Delta G \geq 2$ kcal/mol) drop in the binding free energy.[76] Clackson et al.[99] provided structural data coupled with binding and kinetic analysis of these mutants and proposed that hot spots are "assembled cooperatively" and that many residues contribute indirectly to binding. They suggested that several hydrophobic residues serve to orient key tryptophan residues and that the electrostatic contacts (receptor Arg43 to human growth hormone) were less important than the intramolecular packing of its alkyl chain with Trp169. Sundberg et al.[100] have correlated the detailed structural effects of hot spot substitution with the energetics of binding.

While identification of hot spots is crucial, exhaustive screening is still very expensive. Thus, to date, only a limited number of interfaces have been screened for residue hot spots. Thorn and Bogan[101] compiled experimentally assessed hot spots from the literature. This compilation facilitated the development of computational strategies to screen protein−protein interfaces with the goal of identifying the hot spots.[102,103] Since structure conservation is expected to positively correlate with the stability constraints acting on a position in a protein, hot spots are expected to correlate with structurally conserved residues. Consistently,[104,105] it has been shown that the alanine scanning mutagenesis data assembled by Bogan and Thorn[101,106] correlate well with residue conservation. Thus, "computational hot spots" correlate with experimental ones, suggesting that hot spots may be identified based on their structural conservation and sequence identity. Residue conservation, particularly if it is a methionine or a tryptophan, suggests that it is likely to be a hot spot.

Bogan and Thorn postulated that it is the burial of a hot spot in a hydrophobic environment that leads to its major stabilizing contribution.[106] Further investigation has illustrated that packing along the interface is not homogeneous and that the hot spots are located within the densely packed areas.[32] This explains why these residues contribute dominantly to the stability of the complex and why they are conserved. A replacement of a residue under such circumstances is difficult: substitution by a smaller residue would create holes, while substitution by a larger residue would lead to steric clashes. It is striking that, in the complexes where both protein partners were alanine-scanned, the $\Delta\Delta G$ of a hot spot correlates remarkably well with the local packing density.[29]

Analysis of structurally conserved and experimental hot spot residues illustrates that they tend to be coupled across the interface more than expected by random distribution. Charge−charge couples are disfavored, and the total number of hydrogen bonds and salt bridges contributed by hot spots is as expected. At first glance this appears surprising, since electrostatic interactions and hydrogen bonds are well-known to be crucial to the stability of protein−protein complexes. Further, the high success rate of the simple physical models in the prediction of the hot spots binding energy contribution clearly illustrates the important role of electrostatic interactions and hydrogen bonds in the hot spots contributions. This suggests that the charged/polar residues may act through a water-exclusion mechanism. Since the hot spots are located within highly packed regions, water molecules are easily removed upon binding, leading to strengthened electrostatic contributions of charge−charge interactions. This explanation is consistent with the insightful Bogan and Thorn[106] proposition of a hydrophobic "O-ring" around the hot spots.

Thus, to conclude, as we noted above, estimation of the stability of a candidate complex by computationally scanning its interface with the goal of quantifying the association may be inaccurate, given potential hot spot cooperativity.[32,33,42] Summation of $\Delta\Delta G$ for hot spots may overestimate the

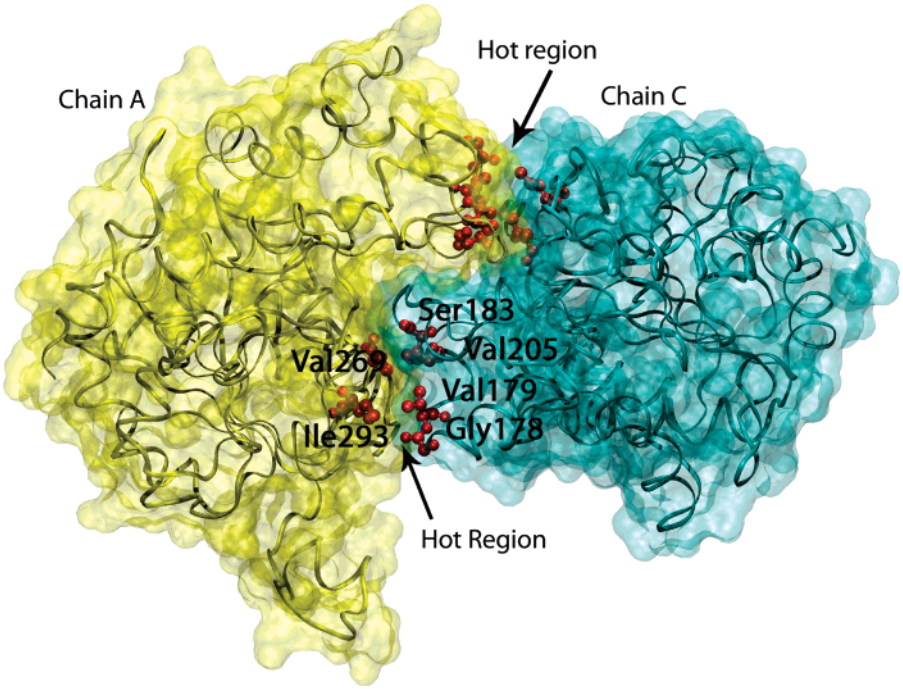


Figure 5. Crystal structure of a complex displaying the hot regions between two M chains of the human muscle L-lactate dehydrogenase (PDB ID: 1i10). Two interacting chains are shown in yellow and cyan. The hot spot residues (red) are shown in ball and stick representation. There are two hot regions in this interface of the homodimer. The figure illustrates that hot spots are in contact with each other and form a network of interactions forming *hot regions*. The bottom hot region is composed of residues Ser183, Val205, Val179, Gly178 from Chain C and Val269 and Ile293 from Chain A of the complex.

binding free energy. Nevertheless, since hot spots correlate with residue conservation and they tend to be coupled across the two sides of the interface, these measures can assist in the prediction. Furthermore, their properties, as described in the next two sections below, make them potentially useful attributes in the prediction, although to date these properties have not been used in prediction strategies.

3.3. Protein Binding Sites Can Be Described as Consisting of a Combination of Self-Contained Modules, or Hot Regions

Hot spots tend to occur in clusters. Within the cluster, the tightly packed hot spots are in contact with each other and form a network of interactions (Figure 5) constituting *hot regions*.[32,33] This organization implies that, within a cluster, the contributions of the hot spots to the stability of the complex are cooperative; however, the contributions of independent clusters are additive. Such a conclusion is further supported by the double mutant cycle analysis.[42,107] For the barnase−barstar interface, it was observed that the coupling energy between two residues decreases with the distance between them. Residues within a distance of 10 Å are defined as modules. Residues located within a module may be cooperative, while residues located in different modules are additive.[41,42]

At greater distances, the effects of mutations are additive, and the energetics of the interactions are independent of each other. This organization reenforces our conclusion above: the binding free energy is not a simple summation of the single hot spot residue contributions; however, that is the case for hot spots within the same *hot region*.

Protein binding sites have been described either in terms of the residues that take part in the interaction with the binding partner or in terms of the binding area patch. Here we describe protein binding sites as a combination of "*hot regions*". This description is not merely semantic; rather, it represents a new view of macromolecular binding. A "classical" description that employs single amino acids that interact across the interface implies that the contributions of single residues to the stability of the protein−protein association are additive. At the other extreme, a "patch" definition usually refers to the area over which the intermolecular interactions extend. In contrast to both views, we view the binding interface as consisting of independent regions. Each region is tightly packed. The amino acids that contribute dominantly to the stability are clustered within these regions. Their tightly packed environment rationalizes their high contributions and the observation that they are strongly conserved by evolution. The clustered hot spot residues form a network of conserved interactions. The implications of such a description are that, within a hot region, the contributions of the hot spot residues to the stability of the complex are *cooperative*. On the other hand, since the regions are independent of each other, the contributions of the hot regions are *additive*.

Such a description suggests that, in between the tightly packed hot regions, packing is not optimal, allowing binding-site flexibility. One clear advantage of such a model is that it highlights the similarity between protein folding and protein binding. The cooperative contributions of conserved residues in the tightly packed protein cores have long been known to be a hallmark of protein folding. Thus, here we argue that protein−protein interactions might be understood in terms of hot-region organization. We stress that a hot region includes residues from both chains, which form a network of interactions (Figure 5).

3.4. Hot Spots Tend to Occur in Preorganized (Complemented) Pockets That Disappear Upon Binding

The protein surface is not flat. It is studded with pockets, crevices, and indentations.[68] In the unbound state, depending on their sizes and shapes, these imperfections of the protein surface may be occupied by water.[108] In the bound state, the water may or may not be replaced by the partner protein molecule. Unfilled pockets are those that remain unfilled by the protein partner. Complemented pockets are pockets that disappear upon binding, representing tightly fit regions.[39] The question arises as to whether there is a preference for the hot spot residues to occur in a specific geometry. Since the hot spots are tightly packed, they are strongly favored to be located in complemented pockets and are disfavored in unfilled pockets. Interestingly, however, complemented pockets often pre-exist binding. In 16 of 18 protein−protein complexes with complemented pockets whose unbound structures were available, the pockets were identified to pre-exist in the unbound structures.[39] Figure 6 presents such an example. The root-mean-squared deviations of the atoms lining the pockets between the bound and unbound states were observed to be as small as 0.9 Å, suggesting that such pockets constitute features of the populated native state. Thus, these pockets are usually already *preorganized* in the unbound state, prior to the protein complexation. The finding that key residues have preferred states is in agreement with the observations of Rajamani et al.[109] that some key residues act as "ready-made" recognition motifs by acquiring native-like conformation prior to binding. The conferred rigidity in the unbound state minimizes the entropic cost on binding, whereas the surrounding residues form a flexible cushion. The studies of Smith et al.[110] further reinforce these conclusions: the fluctuations that they observed in a set of 41 proteins that form binary complexes took parts of the molecules into regions of conformational space close to the bound state; however, at no point in their simulations does each protein as whole sample the complete bound state. As in Rajamani et al., in simulations in the absence of the binding partner, the core interface residues presented a tendency to be less mobile (either measured by the size of the fluctuation or by its entropy) than the rest of the surface, while the peripheral interface residues were more mobile. This result, obtained across 40 of the 41 proteins, suggests different roles for these regions in protein recognition and binding. In a recent study, we compared the mobility of conserved and nonconserved residues in 17 protein−protein interfaces by performing molecular dynamics simulations.[111] Figure 7 presents the results from our simulations illustrating this interesting hallmark of protein−protein interactions. The results further suggest that docking algorithms may treat these regions differently in the docking process and substantiate the feasibility of targeting hot spots in drug design.[112]

3.5. There Are Favorable Organizations in Protein–Protein Interactions

The molecular architecture of protein−protein binding sites, which can be defined as the secondary structural organization, have been reviewed recently.[37] While the interfaces are heterogeneous in terms of size, shape, and chemical composition, amino acid sequence order-independent structural alignment procedures are able to cluster the large set of interfaces (>20 000) from different protein

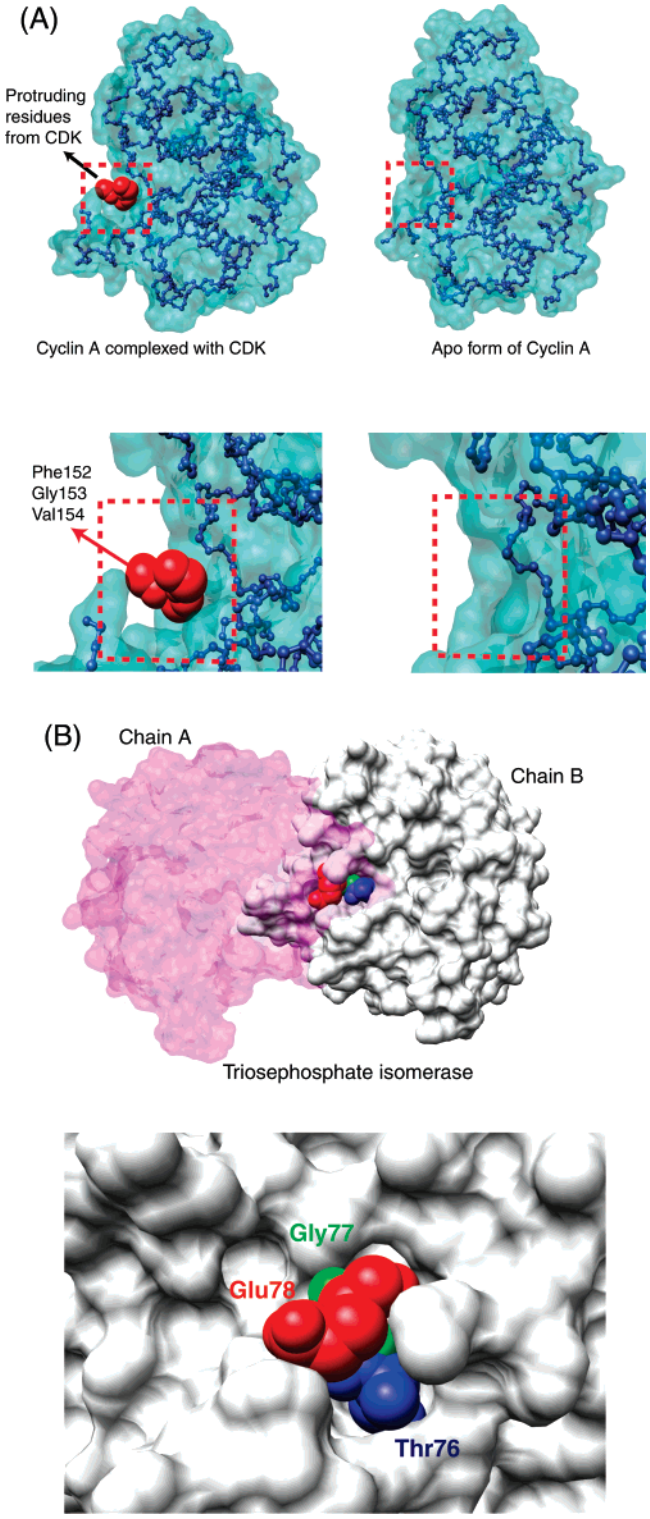


Figure 6. Illustration of pockets in protein interfaces. (A) The upper panel shows the Cyclin A protein in bound (left) and free (right) forms. For clarity, the residues except the protruding ones from the accompanying protein (cyclin dependent kinase) are not shown in the complexed form. (PDB IDs of the complex and monomer are 1fin and 1vin, respectively). The bottom figure shows the details of the pocket. The red residues (belonging to CDK) protrude into the pocket. The same pocket exists in the free form as shown in the boxed rectangular region of the apo form. (B) The top figure shows triosephosphate isomerase in complex form (PDB ID: 1b9b). Red, blue, and green residues are the protruding residues belonging to the left (pink) protein. The bottom figure displays the pocket and the protruding residues in detail.

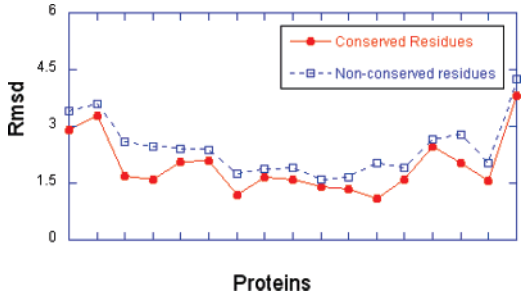


Figure 7. Flexibility of conserved and nonconserved residues in the interfaces. Each point represents a different complex. Seventeen complexes are shown (the first eight and last five points are for homodimers and enzyme−inhibitor complexes, respectively; the middle points correspond to antibodies). The flexibility of residues over 5 ns molecular dynamics simulations of the complexes[111] are compared to determine the difference in the dynamic behavior of conserved and nonconserved interface residues. First, the average rmsd of each residue in the interface is calculated over the entire simulation time. Before calculating the residue side-chain rmsd values, all heavy backbone atoms (N, Cα, C, O) of the interface residues are aligned with the initial structure at the beginning of the simulations to avoid systematic errors caused by translational motions. Side-chain rmsd values are obtained by comparing each frame during the simulations with the structure at the beginning of the simulations after the equilibration step. The red and blue lines represent the flexibility of conserved and nonconserved residues, respectively. RMSD units in Angstroms.

families into a small set of groups (~3 500 clusters),[113] with similar architectures. Studies of these clusters have shown that interfaces sharing similar scaffolds may derive from globally different structures and belong to functionally different protein families.[114] This, however, is not surprising, as it is well-known that proteins with similar structures can have different functions.[115] Different structures whose associations lead to similar interface architectural motifs are particularly interesting: these similar-interfaces, dissimilar-protein folds fall into different families (according to the SCOP classification).[116] In Figure 8a, the interfaces and the global protein architectures are similar; in Figure 8b, the 3-dimensional structures of the monomers are different, yet their interfaces have similar architectures. A real case is given in Figure 4B. Two complexes, cytochrome C and neuropeptide/membrane protein, are not related evolutionarily, yet their interface architectures are similar. Thus, as in monomer structures, evolution has reutilized "good" favorable motifs, leading to preferred architectures. These interface motifs resemble those of protein chains. Despite the absence of chain connections, global features of the architectural motifs that are present in monomers recur in the interfaces, reflecting the limited set of the folding patterns. However, the details of the architectural motifs may vary. In particular, the extent of the similarity correlates with the consideration of how the interface has been formed: whether the proteins cofold (two-state folders) or fold separately (three-state folders).[20] Architectures of interfaces derived from two-state complexes, i.e., where the chains fold cooperatively, are similar to those in protein cores, as judged by the quality of their geometric superposition. On the other hand, three-state interfaces, representing binding of already folded molecules, manifest a larger variability and resemble the monomer architecture only in general outline.[20,75] The origin of the difference between the monomers and the three-state interfaces can be understood in terms of the different nature of the folding and the binding that are involved. Whereas in the former all degrees of freedom are available to the backbone to

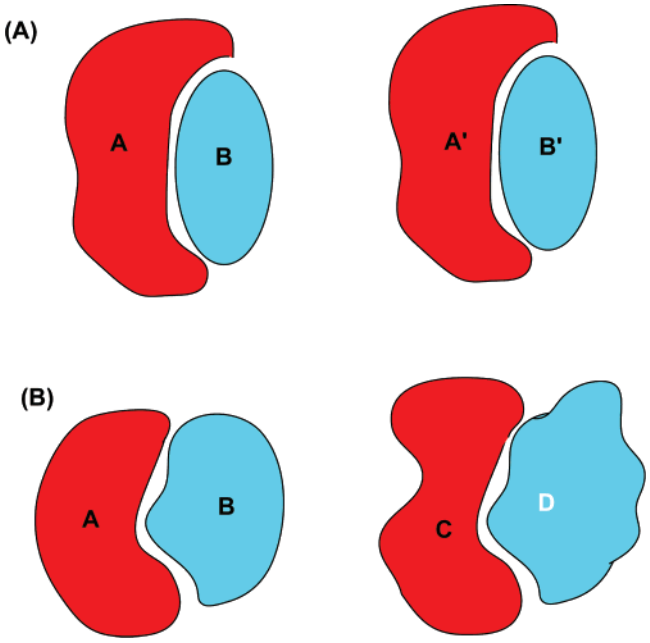


Figure 8. Schematic representation of the interfaces and the global architectures of protein complexes. Part (A) shows cases where both the interfaces and the global architectures are similar; Protein A is homologous to Protein A′ and B is homologous to B′. In Part (B), the three-dimensional structures of the monomers are different, yet their interfaces have similar architectures. Proteins A and C are non-homologous, as are proteins B and D.

maximize favorable interactions, in rigid body three-state binding, only six degrees of freedom are allowed.[20] Examples include four-helix bundles, extensions of β-sheets across the interface, two-helices packed against each other, β-sandwiches, etc.[13,113,114]

Thus, like protein folds, protein−protein interfaces have preferred architectures. Since the number of secondary structure organizations is limited because of the restricted freedom upon secondary structure formation,[117] these preorganized secondary structure motifs may be important in limiting the conformational space, key to protein association. On the practical side, similar to schemes for predictions of protein structures by threading through available folds, a library of protein−protein interaction architectures may provide patterns for modeling protein−protein associations, assisting in docking predictions. However, a large portion of protein−protein interfaces are formed by disordered loops presenting a difficulty in such modeling strategies.

4. Different Protein Partners May Share Similar Binding Sites

Preferred organization is further observed in the reutilization of given binding sites by different partners.[13] The recent increase in the number of protein structures, the additional experimental results of protein−protein interactions, and the construction of maps of protein interactions for some organisms all consistently indicate that some proteins are centrally connected, whereas others are at the edges of the map. The centrally connected hub proteins may interact with a large number of proteins.[118,119] Genomic maps indicate that some proteins have as many as tens of connections. While this may be an overestimate, nonetheless it does suggest multiple interactions, beyond the possibility of the surface providing as many separate, isolated sites. Thus, whereas some binding sites are distinct, it may be expected that others

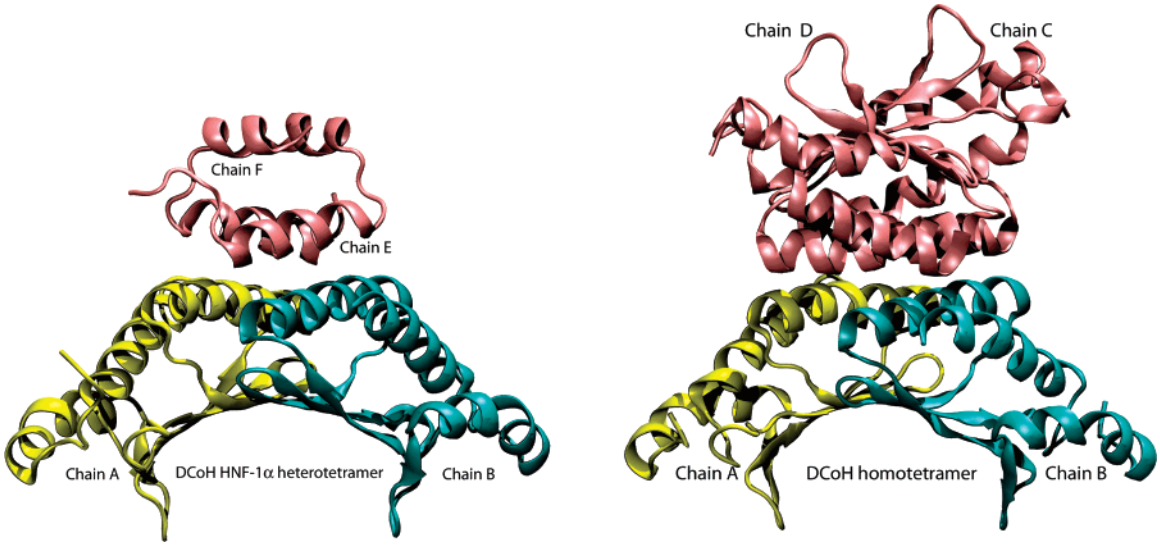


Figure 9. Example of multiple proteins binding at the same site on the protein surface, dimerization cofactor of hepatocyte nuclear factor (DCoH). DCoH serves as an enzyme and a transcription coactivator. The left figure is the crystal structure of hepatocyte nuclear factor dimerization domain, HNF-1α, bound to a DCoH dimer (PDB ID: 1F93, Chains A, B of DCoH, and Chains E, F of HNF-1α). In order to act as a coactivator, DCoH binds to HNF 1α. The figure on the right displays the enzymatic form of the protein DCoH forming dimers of dimers (shown Chains A, B, C, and D, PDB ID: 1DCH).

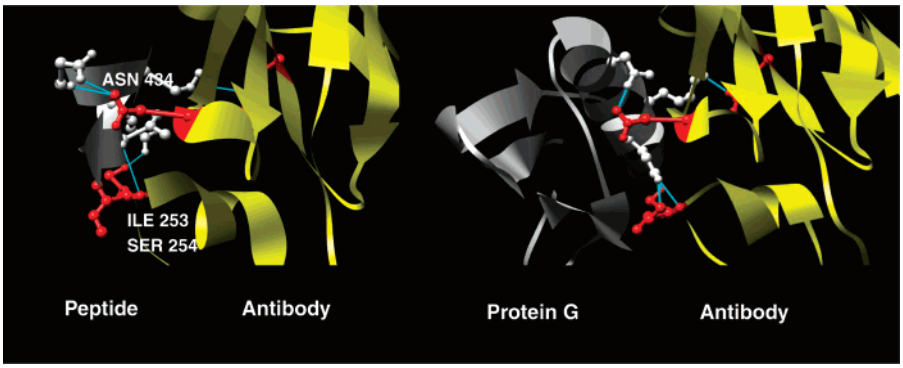


Figure 10. Shared binding sites. The figure highlights the conserved interactions of a given site when interacting with multiple partners. The yellow protein is the antibody interacting with a peptide and protein G (PDB IDs: 1dn2 and 1fcc). The residues shown in red belong to the antibody and they are utilized to form H-bonds with both partners.

may bind different molecules at the same location. This suggests that there are binding sites that are multiply reutilized, albeit with different affinities. Furthermore, for a few cases, there are documented examples with crystal structures, like the Elongin B/Elongin C/VHL and Elongin B/Elongin C/SOCS2.[120,121] Beckett has recently highlighted "functional switches" in transcriptional regulation,[122] focusing on the ability of proteins to bind alternative proteins at the same binding site. Figure 9 presents one such example. A table illustrating similar binding sites among proteins with globally different structures and with different functions was provided in our previous study.[13] To create this table, we have used the data set of structurally and sequentially nonredundant protein−protein interfaces.[113] The clustered binding sites in this table provide a set of structurally similar sites that bind different partners. For the β-catenin, Beckett[122,123] observed that similar interactions are responsible for binding to the different partners. This is expected, since the hot spots are those residues conserved in the protein families. Our analysis of the data set validates this observation. Figure 10 highlights conserved interactions of a given site when interacting with multiple partners. This observation suggests that these optimized local interactions involve the preorganized conserved hot spots. On the other hand, their actual contributions are likely to be functionally modulated. Thus, while the patterns of the local interactions are similar in multipartners and in single partners, the multipartners have been optimized by evolution to accommodate different ligand shapes, sizes, and composition.

5. Obligatory and Transient Complexes

Protein complexes have been classified into obligatory, or permanent, and transient.[17,30,124,125] Obligatory protein−protein complexes are formed by proteins that only function when associated in the complex. Homodimers provide a nice example for obligatory complexes; however, many other proteins consisting of heteromultimers may also fall into this category. By contrast, formation of transient complexes depends on the functional state of the partners. Examples include enzyme−inhibitor, hormone−receptor, and signaling−effector types of interactions. In recent years, considerable attention has focused on the distinction between the two types of complexes.[42] The relative contributions of the physical interactions differ between the two. Obligatory associations are in general tighter, with a stronger hydrophobic effect, better packing, and fewer structural water molecules trapped between the monomers, and they manifest better shape complementarity. In contrast, the interfaces in the transient complexes are generally less extensive and more

polar/charged, and the surfaces of the interacting proteins at their interface are not as optimized, leading to weaker associations with the exception of some enzyme−inhibitor complexes.[30,124,125] Quantifying these differences is important since many predictive protein−protein schemes use knowledge-based scoring parameters derived from the combined data set of complexes.

Interestingly, analysis of the interfaces of both types of complexes illustrates that residues in the interfaces of obligate complexes tend to evolve at a relatively slower rate, which allows the protein-partners to coevolve. By contrast, the less tight transient partners illustrate increased rate of mutations at the interface and no evidence of correlated mutations.[15]

6. Disordered Proteins: A Major Component of Protein–Protein Interactions

While the presence of "disordered" proteins has been recognized for a long time, in recent years they have drawn increasing attention. Disordered proteins (or "intrinsically unstructured" proteins) lack a stable, well-defined structure under physiological conditions, existing in a continuum of conformations from the less to the more structured states.[47−50] Natively unstructured proteins undergoing a disorder-to-order transition upon binding their partner, and stable monomeric proteins, which exist as multimers in their crystal form but not in solution, provide examples of two vastly different scenarios. There are two major reasons for the recent heightened interest in the disordered protein state: First, a large number of proteins have now been identified to belong to this category, with a diverse functional spectrum. Second, the disordered state is analogous to the denatured state. Comprehension of the protein-folding reaction necessitates knowledge of the ensembles of the folded and the denatured states under different conditions. The lack of understanding of the denatured state impedes understanding of the folding process.

Natively unstructured proteins have a broad range of functions,[44−46,51] including regulation of transcription and translation, cellular signaling, phosphorylation, regulation of large multimolecular self-assemblies, and small molecule storage.[45] Analysis of the structural characteristics of complexes of natively unstructured proteins, ribosomal proteins, two-state and three-state complexes, and crystal-packing dimers has suggested that ordered monomers can be distinguished from disordered monomers on the basis of the per-residue surface and interface areas, which are significantly smaller for ordered proteins.[126] With this scale, two-state dimers (where the monomers unfold upon dimer separation) and ribosomal proteins resemble disordered proteins. On the other hand, crystal-packing dimers, whose monomers are stable in solution, fall into the ordered protein category. While there is a continuum in the distributions, nevertheless, the per-residue scale measures the confidence in the determination of whether a protein can exist as a stable monomer. Disordered proteins lack a strong hydrophobic core and are composed of highly polar surface area.

Molecules or regions displaying disorder have been considered inherently unstructured. Yet prevailing conformations still exist, with population times higher than those of other conformations.[47−49] Disordered molecules are the outcome of rugged energy landscapes away from the native state. Ruggedness has a biological function, creating a distribution of conformers that bind via conformational selection, driving association, and multimolecular complex formation. A rugged energy landscape modulates the lifetimes of different conformers, depending on the biological function.

Disordered functional proteins provide evidence that the function of a protein and its properties are not only decided by its static folded three-dimensional structure; they are determined by the distribution and redistribution of the conformational substates. Enumeration[127] of all sterically allowed conformations for short polyalanine chains consistently shows that, in the denatured state, not all conformations are accessible. Even for alanines, local steric effects beyond nearest neighbors already restrict significantly the conformational space. For variable-sequence chains with bulkier side-chains, this effect is likely to be enhanced, biasing the local conformations.[127−131] Preferred conformation implies that there is no need to search for the favored binding partner over broad space in time-scales not biologically relevant. Hence, the fact that binding is fast implies selection: the conformation is already there. With its binding, the equilibrium shifts in its favor, further driving the reaction. As binding and folding are similar processes with similar underlying principles, this principle applies to disordered molecules in binding and to unstable, conformationally fluctuating building blocks in folding. Folding and binding imply selection, rationalizing rugged energy landscapes away from the native conformations. However, local conformational diversity can be expected, allowing latitude in the associations, depending on the binding partner.

7. Systems Biology and the Chemistry of Protein–Protein Interactions

Proteins function in cellular processes. Unfortunately, for the vast majority of the proteins participating in these, there are no structural data; only databases citing experiments that infer which proteins interact and sequence information allow for prediction of protein−protein interactions based on various schemes, such as coevolution,[132] orthologous relationship,[133] or, for example, based on domain combinations,[134] to name a few. In the absence of structures, it is not possible to address the chemistry of the interactions. Nevertheless, by crossing the structural data available in the PDB[135] with the connectivity data for the yeast map,[136−139] we have obtained a data set of proteins that have complete structures and interactivity data. The problem is, however, that, even for these, the interfaces are largely unknown. Even if interfaces are available, they are not necessarily those which play transient regulatory roles in the network; rather, they may belong to the protein multimeric permanent interactions. Bearing these caveats in mind, it is nevertheless interesting to look into the structural/chemical properties of the central versus the edge proteins. By definition, a central protein has a large number of interacting partners, whereas a loner has one or very few. Developed organisms typically have a more centralized network topology. Topologies consisting of highly connected proteins are functionally advantageous, leading to higher efficiency and inherently superior regulation. In this respect, it is interesting to note that the human genome has a fewer number of genes as compared to some lower organisms, implying that our genome is more flexible and functionally more complex. It is now clear that one gene can specify more than one protein, with gene expression regulated by different factors at the different levels of control. A more highly connected map implies more proteins binding at a shared site. This leads to

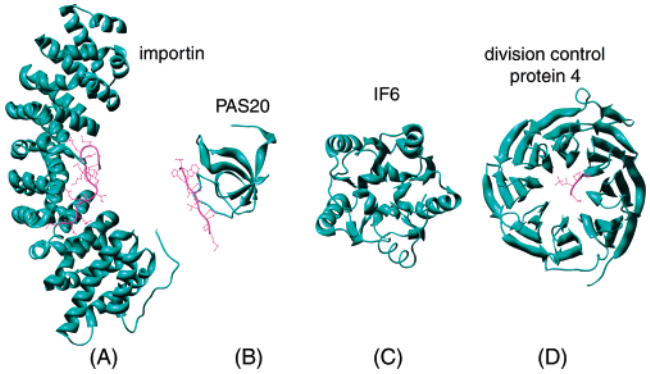


Figure 11. (A) Structure of importin alpha (connectivity from DIP[225] is 197; DIP ID: 728N, PDB ID: 1un0A) is shown here complexed with a peptide containing the nuclear localization sequences. The peptide is in a wire frame. (B) The structure of SH3 domain fragment of peroxisomal membrane protein PAS20 (connectivity 21, DIP ID: 2473N, PDB ID: 1n5zA) in complex with a peptide substrate in a wire frame. (C) The structure of Eukaryotic translation initiation factor 6 (connectivity is 48, DIP ID: 5395N; PDB ID: 1g62A). (D) The structure of cell division control protein 4 (connectivity 7, DIP ID: 1625N, PDB ID: 1nexB). The length of the full protein is 779, while the PDB file contains 444 residues. For clearer images, other chains are removed.

the question of whether there are any structural features that characterize such proteins and binding sites, making them increasingly central in the network as compared to highly specific ones.

7.1. Are There Any Structural Features That Distinguish Highly Interactive Proteins from Loners?

7.1.1. Interface Size and Binding Modes

Highly connected proteins have interfaces of very different sizes. For example, a highly connected protein is importin, whose structure consists of 10 Armadillo repeats forming a superhelical structure.[140,141] Importin is a karyopherin that transports molecules into the nucleus through the pores in the nuclear envelope. In Figure 11a, we see that the binding site is very extensive, running along the inner groove of the superhelix. This binding site is thought to be negatively charged, forming numerous electrostatic contacts with basic residues in the nuclear localization sequence.[142,143] By contrast, SH3, which is one of the most highly interactive domains, recognizes a short peptide PxxP as in Figure 11b.[144−146] The two prolines fit very snugly in two especially designated pockets on the SH3 surface. By contrast to the importin, this binding site is quite small.[144] Importin uses numerous electrostatic contacts for the import of proteins into the nucleus,[143] while the calmodulin binding site consists of a "mat of methionines",[147] hydrophobic and highly flexible side-chains. The two hydrophobic pockets of CaM can accommodate a variety of bulky aromatic rings, providing a plausible structural basis for the diversity in CaM-mediated molecular recognition.[148,149] Histone is yet another example: the tail of the histone can be extensively modified by the addition of acetyl and methyl groups to lysine residues,[150,151] and the modification is thought to at least partially recruit other partners, such as chromatin remodeling complexes.

7.1.2. Protein Fold

The fold of the protein is also not an indication of the protein interactivity. For example, WD40 domain, which forms a β-propeller (Figure 11d), has a wide range of connectivity distribution.[152,153] Yet a similar β-propeller structure, with the fourth strand of the propeller blade replaced with a helix (Figure 11c), has a connectivity which is quite high, 48.[154] An even more extreme example is importin and the regulatory subunit H of V-type ATP-synthase,[155] both of which are Armadillo repeats but whose connectivities are 197 (Figure 11a) and 5, respectively.

7.1.3. Structural and/or Sequence Repeats

Many of the most highly connected proteins are structural and/or sequence repeats. The importance of repeats for protein−protein interactions is not unexpected.[156] In the example of importin, repeats provide a structural context where multiple interactions are made by combinatorial contacts, and different contacts can be used by different partners. Repeats are easy to make (by duplication) and offer an opportunity to divergently evolve the particular parts.[142,157] An example is the proteasome, which in all kingdoms consists of 4 heptameric rings, but which in bacteria is made of a single protein, while in eukaryotes this protein diverged into 7 related paralogs of similar structures but different compositions.[158] An additional example relates to the BRCT domains. BRCT domains from breast cancer-susceptibility gene product BRCA1 and the 53BP1 protein have similar structures yet different binding behaviors with the p53 core domain. 53BP1-BRCT domain forms a stable complex with p53. In contrast, BRCA1−p53 interaction is weak or other mechanisms (differing from an 53BP1-BRCT domain interaction) may operate.[159]

7.1.4. Function

On the other hand, it is no surprise that the most interactive proteins are those that perform the same function for many partners. For example, importin performs the same function (transport into nucleus) for all proteins destined for the nucleus; cell-cycle proteins phosphorylate a slew of proteins in order to advance the cell cycle to the next stage; and proteasome proteins recognize proteins for degradation (we note, however, that it is the regulatory unit that recognizes ubiquitin).[142,153,158] Importin can bind to all these different proteins, as they have similar nuclear association sequences; thus, the binding is specific to a certain protein domain that is found in all these proteins. In the case of the kinases,[160] the underlying principle revealed by the structural organization of Src—the use of protein interaction domains to regulate catalytic activity—couples targeting with catalytic activation. This principle applies quite frequently in modular signaling proteins, where a substrate needs to be carefully positioned in order to be accessible for phosphotransfer.

7.1.5. Residue Propensities and Conservation

The relative frequencies of different types of amino acids in the interfaces of protein−protein complexes are used to derive their propensities. Further, different types of complexes possess different residue propensities. Figure 12 displays the logarithmic propensities of the 20 residues for the different interface types as bars. Thus, overall, current data suggest that there is no universal mechanism nor recurring chemical features differentiating between highly connected proteins versus loners; rather, optimization appears to have occurred through evolution where central proteins increasingly became more centralized. Repeat proteins and variations of specific protein interaction domains are recur-

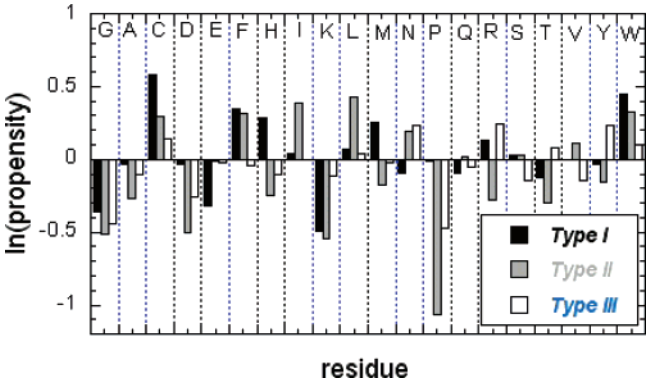


Figure 12. Logarithmic propensities of the contacting residues in the different interface types. A positive value indicates a favorable propensity in the interfaces as compared to the rest of the protein, whereas a negative propensity indicates that it is less likely to find the particular residue in the interfaces compared to the rest of the protein. Here, Types 1, 2, and 3 refer to different types of complexes according to our definition.[13,32,113,114] In order to separate interfaces into different types, we used the data set of structurally and sequentially nonredundant protein−protein interfaces.[113] The data set was created by extracting all existing interfaces between two protein chains obtained from higher complexes of proteins. These interfaces are compared structurally with a sequence- and order-independent algorithm. Interfaces sharing similar architectures are clustered. We divided the 103 clusters into 3 types: In Type 1 clusters, the global folds of the parent chains are similar and the functions of the members of the cluster are also similar (Figure 4A). In Type 2 clusters, members often do not share similar functions and do not have globally similar structures (Figure 4B). Members of Type 3 clusters only have one side of their chains aligned. Thus, members of a Type 3 cluster have similar binding sites on one side of the interface, but the partner proteins are different (Figure 4C). Here, all member interfaces have dissimilar functions. The listings of the three types were given previously.[13,32,113,114] The data was obtained from 358 Type 1, 94 Type 2, and 367 Type 3 complexes.

ring themes. Nevertheless, as described in the next section, by studying a data set of shared binding sites and its comparison with specific interacting pairs, some trends are observed.

7.2. Interfaces of Shared Proteins

For a protein to be a hub, it must be involved in more than a single complex;[136,161,162] therefore, hub proteins are shared proteins that can act as linkers of cellular processes, joining complexes into higher order networks. Dandekar and co-workers[162] investigated the properties of shared protein components in six sets of protein complexes. They concluded that many of the shared proteins appear to be primarily regulatory links in cellular processes acting as peripheral components of protein interaction networks.[162]

Different properties of intermolecular interfaces can have a strong effect in modulating binding affinity and specificity of molecular recognition. Comparison of the flexibilities of homologous proteins across species suggested that, as the species gets more complex, its proteins become more flexible.[163] Ekman et al.[164] observed that multiple and repeat domains are enriched in hub proteins. At the same time, there is evidence that proteins whose function requires a number of specific interactions evolve slowly.[15,165−168] Thus, binding regions with high specificity evolve more slowly than those with lower specificity; this in turn may suggest that additional central links evolve faster than the unique links of loners.

Understanding how a given site binds to different binding sites may shed light on identifying the mechanism of protein

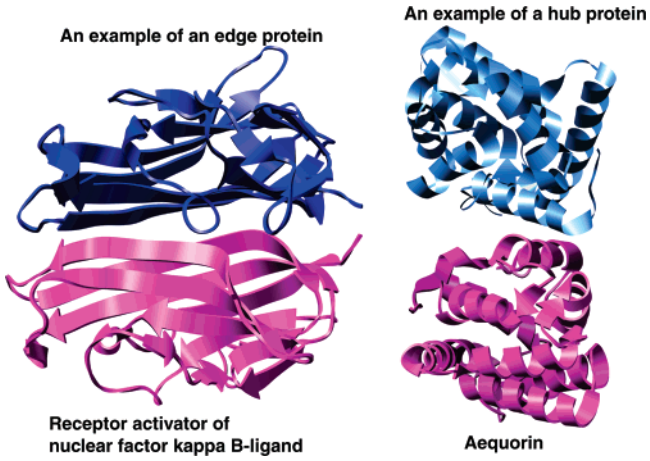


Figure 13. Crystal structures of two proteins: one edge and one hub. The left figure is a receptor activator of nuclear factor kappa B-ligand (PDB ID: 1iqa, Chains A and B). This (edge) protein has 3 connections according to MINT database. The right figure is an aequorin (PDB ID: 1ej3, Chains A and B) with 57 interactions with other proteins.

interactions. To look into this question, we have assembled a data set of protein−protein interfaces from the PDB.[87] We clustered interfaces where one side of the interface is similar but the second, complementary, side is different.[13] Such similar interfaces interacting with different binding sites can be defined as shared binding sites. Inspection of the connectivity of these clusters confirms that the proteins with shared binding sites have higher numbers of interactions with other proteins (∼13)[13] compared with the average connectivity number in yeast interactome (∼5).[169] We find that proteins with common binding-site motifs preferentially use conserved interactions at similar interface locations, despite the different partners. Our analysis of multipartner interfaces further indicates that proteins that use common interface motifs to bind to other proteins have smaller interfaces than complexes with specific partners. The average accessible surface area (ASA) of multiprotein interfaces is 1235 Å,^{2} compared to the 1967 Å^2 ASA of the other types. It appears that, with a large interface, it would be more difficult to bind to different complementary sites. Multipartner interfaces are not as well packed and organized as other proteins. The geometrical matching is not as optimized, and there are water molecules, allowing variability in the interactions. In particular, we observe that multipartner interfaces preferentially consist of α helices. Helices appear as the major vehicle through which similar binding sites are able to bind different partners. Helices at multipartner binding sites allow alternate variable ways to achieve favorable binding, depending on the side-chain identities. They allow more dynamics in the optimization of the helical associations as compared to extension of β-sheets. It will be of interest to examine whether centrally located proteins with multiple proteins binding at the same sites are enriched in α-helical folds as compared to the edge proteins. Figure 13 displays two proteins: one edge and one hub protein. The left figure is a receptor activator of nuclear factor kappa B-ligand (PDB ID: 1iqa). This protein has 3 connections according to MINT database. The right figure is an aequorin (PDB ID: 1ej3) with 57 interactions with other proteins.

7.3. Chemistry of the Interactions: How Are Subtle Differences Distinguished?

Given the similarities between features of protein−protein interfaces, the question arises: how does nature nevertheless distinguish subtle differences, and what happens if nature's choice fails.[159] An insight into these questions should assist in figuring out the principles of protein−protein interactions and in predicting the preferred ways in which proteins interact.

The BRCT domain from the breast cancer-susceptibility gene product BRCA1 noted above is a good example. BRCA1 relates to 45% of the families with inherited breast cancers and 90% of the families with inherited breast and ovarian cancers.[170,171] BRCA1 encodes a large protein of 1863 amino acids, with a zinc-finger RING domain N-terminal and tandem BRCT (BRCA1 C-terminal) domains. BRCT was first identified in BRCA1 as ~95 amino acid tandem repeats[172] and has been found in many proteins, such as p53-binding protein, 53BP1,[173,174] the base excision response scaffold protein, XRCC1, and DNA ligase IV,[175] many of which appear to participate in cell-cycle checkpoints or DNA repair in many species.[176] BRCA1 stimulates p53 transcriptional activity.[177−180] It was reported to associate with p53 with two interaction domains: the central disordered region of BRCA1 interacting with C-terminal domain of p53,[181] and there are some in vitro studies suggesting that BRCT domain of BRCA1 binds to the core domain of p53.[177] 53BP1−p53 interactions were observed directly by X-ray crystallography of the 53BP1−p53 complex. The 53BP1−p53 binding site partially overlaps the p53 DNA-binding site, thus inhibiting the DNA-binding activities of p53.[182]

Both 53BP1 and human BRCA1 have two BRCT repeats, with high structural similarities, even though the sequence identity is only 19%. Each repeat consists of four β-strands and four α-helices, with the exception that one of the α-helices is disordered in the C-terminal repeat of BRCA1. The BRCT region of 53BP1 (taken from the 53BP1−p53 complex, PDB ID: 1kzy) superimposed (by Swiss-Pdb-Viewer http://www.expasy.org/spdbv/ on the crystal structure of BRCA1 BRCT, PDB ID: 1jnx), gives a root-mean-squared deviation of 1.44 Å for 133 out of 211 BRCA1 Cα atoms, including all eight β-strands and seven of eight α-helices. The N-terminal repeat (repeat 1) of 53BP1 and BRCA1 gives an rmsd of 1.38 Å (for 69 out of 88 Cα atoms), and the C-terminal repeat (repeat 2) has an rmsd of 1.25 Å (for 60 out of 94 Cα atoms). The sequence identities of repeats 1 and 2 are 24% and 17%, respectively. The least conserved region is the linker between repeats 1 and 2, with a low 10% identity. Except for the linker, the region involved in 53BP1 bound to p53, including α3A through α4A, has a striking structural conservation with the corresponding region of BRCA1, with an rmsd of 0.58 Å for all 23 Cα atoms. The sequence identity of this region (26%) is also higher than that in the other regions. Yet despite the structural conservation, the p53 core domain interacts with the BRCT domains of 53BP1 and BRCA1 proteins to different extents. Isothermal titration calorimetry, analytical ultracentrifugation, and analytical size-exclusion chromatography confirmed the p53 core domain interactions with the BRCT domain of 53BP1 protein but not with the BRCA1 BRCT domain.[183] While it is possible that these biophysical methods are not sensitive enough, it does imply that, if there is an interaction between BRCA1 BRCT domain and p53 core domain, it is very weak, or that there is no interaction with this BRCT domain repeat and the interaction is with the second repeat. Hence, within the global similarity, the complex details of the structure and the chemistry lead to such selective differentiation. On the other hand, in silico mutations in the first repeat may stabilize the interactions, possibly leading to a non-native, diseased state.[159]

8. Allostery

Allostery is a key in regulation; it has a crucial role in practically all proteins: in hubs and loners. Allostery involves coupling of conformational and dynamic changes between two—nearby or widely separated—binding sites. Proteins are not rigid as it appears when looking at crystal or averaged NMR structures.[62,63] Hydrogen/deuterium (H/D) exchange clearly indicates that native proteins exist as statistical ensembles[184−186] distinguished by locally unfolded region-(s) in the binding sites or elsewhere. Elber and Karplus have demonstrated that the potential energy surface of myoglobin is characterized by a large number of thermally accessible minima around the native structure.[187,188] These observations suggest that the Gibbs energy of stabilization is not equally distributed in the structure. Since local unfolding occurs in the functional state, its significance is beyond protein folding per se. There is experimental and theoretical support that binding at one site effectively can shift the population, showing conformational and dynamic changes at some other sites. Structural perturbation at any site leads to a redistribution of the populations. One source of structural perturbation is the binding of inhibitors (or effectors). Other sources include mutations, binding to sister molecules, binding to nucleic acids or to small molecules, changes in pH, ionic strength, temperature, and covalent modification such as phosphorylation and acetylation, discussed above. Redistributed conformations are not a manifestation unique to allostery. Rather, they are physical attributes of proteins. Allostery derives from populations. Thus, there is no well-defined path, nor a distinct series of steps that molecules follow. Rather than every single molecule undergoing a series of steps to reach the conformational change observed in the snapshot of a shape of a site that is far away, *what we observe is the outcome of the ensemble*. The perturbations at one site do not yield a homogeneous distribution. Since some portions of the molecule are less stable than others, these parts will manifest larger variability. When thought of in these terms, allosteric activation should not produce an alternate rigid binding site shape fits the ligand (substrate or protein). Rather, the perturbation upon effector binding leads to a redistribution of the ensemble, which would be largely reflected in binding sites which are a priori less stable. Nevertheless, the "active" conformer is also present in the presumably "inactive" ensemble, albeit at a lower concentration. Upon binding, there is an equilibrium shift in its direction, further driving the binding reaction.

p53 presents a relevant example of protein allostery. The last 30 residues of the C-terminal domain were proposed to negatively regulate DNA binding by an allosteric mechanism (reviewed in ref 189). This hypothesis was based on the observation that the interaction of p53 with a short oligonucleotide containing a consensus p53-binding site is greatly enhanced either by the deletion of the C-terminal basic region (30 residues) or by binding of the antibody PAb421 to the same region.[190] This was confirmed by a study showing that p53 transcriptional activity is activated by PAb421 in cells.[191] Recent studies have demonstrated that, within the context

of chromatin or supercoiled DNA, the C-terminal domain may actually facilitate binding of the core to its target DNA sequence by providing an additional anchorage to specific DNA sites via nonspecific DNA binding.[192−194] Cross-talk between the different p53 domains has also been indicated in earlier studies, showing that destabilization of the core by mutation (R273H) inhibits the transactivation activity of the N-terminal domain.[195] NMR studies confirm that the N- and C-terminal domains have an impact on the thermodynamic instability of the p53 tetramer.[196] However, how the cross-talk between p53 domains occurs is still unclear. A similar situation is observed in other tumor suppressors that serve as hub proteins, such as pVHL or suppressors of the cytokine signaling (SOCS) family,[120] which function as key regulators at all levels of cellular pathways. Binding of the pVHL to the elongin B-elongin C complex leads to a conformational change that allows it to bind to the HIF;[197] in contrast, without the pVHL binding to the elongin C/elongin B, the pVHL has not been observed to bind to HIF.

Although allostery plays a role in protein−protein interactions in general, it is likely to play a particularly important role in central shared proteins. The conformational change may or may not be large. However, even if small, it may lead to distinct minima at the bottom of the protein folding funnel; the low barrier heights may nevertheless be physiologically sufficiently high to necessitate a change in the conditions to allow surpassing them. Such a change may be the binding of another protein, leading to the population shift. Distinct minima with small conformational changes may explain the more centralized nature of the cellular network and how central regulatory proteins are able to bind an astonishingly large number of different partners. In many cases, such as in p53 and pVHL, the allosteric communication is transmitted via cross-talk between domains.

9. Large Assemblies

Accurate determination of the structures of protein molecules and their complexes constitute major challenges in the biological sciences. Availability of the structures would facilitate drug design, identification of the origin of misfunction, and disease. It would provide crucial assistance to the prediction of protein function. Yet despite the broad recognition of the importance of the knowledge of the structures, experimental structure determination and computational prediction still face immense hurdles. Proteins never function when they exist isolated in solution. Function dictates molecular associations. The sizes of the assemblies can be very large. This arrangement effectively increases the local concentrations of the reactants/products in enzyme pathways, provides effective regulation and control of cellular processes, and leads to structural scaffolds and regulated molecular machines. Large assemblies are an advantage of complex and robust systems. To understand how they work at the molecular level, it is essential to have the interactions between the components and their spatial organization. The structure is also crucial if we are to figure out how the machine performs its work and how the regulation is achieved. The structure is crucial in order to understand the conformational switches and alternate potential binding modes.

The main reaction pathways in the living cell are carried out by functional modules, namely, macromolecular machines with compact structure or the multimolecular ensembles that change their composition and spatial organization during function.[198] Function relies on spatial sequestration, chemical specificity, and time series of the dynamic ensembles. The addition of the individual components results in systemic properties that could not be predicted by considering the components individually.[199]

The molecular ensembles constitute compact, specific, and transient functional modules. Type III secretion systems (TTSSs) constitute one example. TTSSs are multiprotein macromolecular "machines" that have a central function in the virulence of many Gram-negative pathogens.[200] The TTSSs directly mediate the secretion and translocation of bacterial effector proteins into the cytoplasm of eukaryotic cells. The 20 unique structural components constituting this secretion apparatus are largely conserved among animal and plant pathogens and are evolutionarily related to proteins in the flagellar-specific export system. Electron microscopy revealed a "needle-shaped" morphology of the TTSS. The 1.8 Å crystal structure of EscJ from enteropathogenic *Escherichia coli* (EPEC), a member of the YscJ/PrgK family whose oligomerization represents one of the earliest events in TTSS assembly, provided the detailed structural characteristics and the organization of these protein components.[200] Molecular modeling has indicated that EscJ could form a large 24-subunit "ring" superstructure with extensive grooves, ridges, and electrostatic features. Electron microscopy, labeling, and mass spectrometry studies on the orthologous *Salmonella typhimurium* PrgK within the context of the assembled TTSS support the stoichiometry, membrane association, and surface accessibility of the modeled ring.

Another example is the nuclear pore complex.[200−202] The nuclear pore complex (NPC) consists of multiple copies of ~30 different proteins (nucleoporins, nups). They form a channel in the nuclear envelope that mediates macromolecular transport between the cytosol and the nucleus. Only <5% of the nup residues are currently available in experimentally determined structures, and consequently, very little is known about the detailed structure of the NPC. Nevertheless, a combined computational and biochemical approach was used to assign folds for ~95% of the residues in the yeast and vertebrate nups. The assigned folds suggest simplicity in the composition and modularity in the architecture of all eukaryotic NPCs, reflected in the presence of only 8-fold types; three of the most frequent folds account for ~85% of the residues. The modularity in architecture is reflected in the hierarchical and symmetrical organization. These partition the predicted nup folds into three groups: the transmembrane group with transmembrane helices and a cadherin fold; the central scaffold group consisting of β-propeller and α-solenoid folds; and the peripheral FG group containing predominantly the FG repeats and the coiled-coil fold. These led to the suggestion that the small number of fold types in the NPC and their internal symmetries evolved through extensive motif and gene duplication from a simple precursor set of only a few proteins.

10. Crystal Interfaces

A considerable effort has been directed at differentiating between the "real" biological interfaces and interfaces that are the outcome of crystal effects.[30] This is an important issue since potential functions and other chemical and functional attributes are derived from the protein−protein complexes in the PDB. Analysis of a manually curated crystal interface data set has illustrated that the average area of crystal-packing

interfaces is only 570 Å^2 per interface. Nonetheless, some crystals contain pairwise interfaces comparable in size to those of protein−protein complexes. Large packing interfaces are often associated with twofold symmetry elements forming "crystal dimers" that may be mistaken for real dimers. To identify structural features other than size that distinguish between the two, Janin et al. selected from a set of crystals of monomeric proteins 188 packing interfaces with >800 Å^2 including 105 with twofold symmetry.[27] Their results showed that, on average, these large crystal-packing interfaces are standard size and with similar nonpolar fraction as in complexes. Because homodimers have a larger average fraction of nonpolar buried surface area, the chemical composition of the interface may distinguish between real and crystal dimers; however, the distributions of individual values overlap. The amino acid compositions reflect similar trends; however, these are not sufficiently distinct to remove ambiguities between the crystal and the biological interfaces.[27,203,204] Our results also show that crystal interfaces can either be unique or share similar patterns with biological interfaces. However, for the majority of the interfaces, there are no details in the literature to elucidate their real tertiary structure. Therefore, it is not clear whether some of the interfaces that share the same chemical and structural features with crystal interfaces are indeed "real" interfaces, or perhaps they, too, are crystal interfaces. A conclusive solution for this problem has not yet been found.

A relevant example of the complexity and relevance of the problem can be gauged from examination of the p53 crystal structures.[205−207] The first crystal structure obtained by Pavletich et al.[205] has presented a trimer structure for the p53 proteins with interfaces differing from the expected symmetric associations of the p53 dimer binding to the DNA. Crystal structures with symmetric associations were only published in 2006,[206,207] 12 years after the nonsymmetric trimeric crystal organization. Detailed high-resolution structure of the full native p53 tetramer−DNA has yet to be determined experimentally, probably in a supramolecular association given its disordered regions.

11. Concluding Remarks: Preferred Organization in Protein Interactions

Considerable statistics have accumulated over the years on protein binding sites and protein−protein interactions. Studies have been carried out on data sets of structures, focusing on particular complexes and on their dynamics. Protein−protein interactions have been studied in binary associations and within the framework of the cellular networks. Much progress has been made in our understanding of the types of associations, permanent and transient; the conformational transitions; and the ordered and disordered states.

Currently we know that protein−protein interactions are largely driven by the hydrophobic effect. Nevertheless, although the hydrophobic effect plays a dominant role in protein−protein binding, it is not as strong as that observed in the interior of protein monomers, and its extent is variable. The binding site is not necessarily at the largest patch of hydrophobic surface. At the interface, there are higher proportions of buried charged and polar residues as compared to protein cores, suggesting that hydrogen bonds and ion pairs contribute more to the stability of protein binding than to that of protein folding. Residue conservation has also been observed to tend to be higher at binding sites as compared to other protein surface areas. However, such observations are insufficient to assist in predicting protein−protein interactions. Protein binding sites have neither the largest total buried surface area nor the most extensive nonpolar buried surface area. They cannot be uniquely distinguished by their electrostatic characteristics, as observed by parameters such as unsatisfied buried charges, or the number of hydrogen bonds. Although the geometry of molecular surfaces has provided clues to binding sites on enzyme surfaces, which are often shaped as the largest or deepest clefts on the surface, none were found for protein−protein binding sites. On the other hand, a description of binding sites in terms of preferred residue, and particularly *region* and *architectural organizations*, may lead to classification strategies assisting in predictions of the preferred ways for proteins to interact. Within the recurring favorable architectures, there are preferred cooperative hot spot organizations. A combination of all the mentioned features can be used to distinguish the location of interfaces with an average success rate of 75%.[40,87,208,209]

Preferred organization is a key in chemistry and in protein science, whether in amyloid microfilaments or in globular protein−protein associations. Evolution reutilizes favorable patterns and modulates these toward different functions. The well-recognized fact that protein architectures do not span the entire conformational space and certain topologies are disallowed has led to the imaginative proposition of using the limited repertoire in folding strategies.[117,210] Despite the absence of the chain connectivity between the interacting partners, nature appears to similarly follow these preferred organizations in protein−protein associations. Within these scaffolds, functional hot spot residues are conserved. The energetic contributions of the hot spots derive from their local networked organization in tightly packed "hot" regions. Between these, packing is less optimal, allowing flexibility and binding to multiple—different—partners. A self-contained hot region organization offers many advantages and may also play a role in binding sites to other molecules such as DNA, RNA, and small molecules.

The proposition that an interface can be divided into parts or patches is not new. Jones and Thornton analyzed protein−protein interaction sites using surface patches, defined in terms of solvation potential, residue propensity, hydrophobicity, planarity, protrusion, and accessible surface area.[31,94,211] Shanahan and Thornton analyzed the conservation of surface patch polarity.[72] Surface complementarity in complexes has been estimated using patches. The so-called Ile-44 surface patch of ubiquitin binds to the alpha3 helix of the GAT domain, which is responsible for ubiquitin binding and ubiquitination.[212,213] Surface patches containing basic and aromatic residues were detected in domains of the La protein that interacts with RNA. These account for the cooperative binding of short oligonucleotides. A surface patch with two exposed tryptophan residues that interface with lipid bilayers was noted for the GM2-activator protein. Using a patch analysis, side-chain conformational entropy at protein−protein interfaces has also been performed.[214] These constitute only a few examples. Thus, while patch definitions vary, it has been recognized that a binding site or an interface can be divided into parts. Here, however, our definition is in terms of continuous paths of interacting residues within densely packed neighborhoods, leading to cooperative effects.

A recent review[215] summarized the challenges in modeling structures and protein interactions by sequence and structure.

Sali and co-workers[216] discussed the localization of protein binding sites within families of proteins. They observed that 72% of the 1847 SCOP domains have binding sites at similar positions, that is, members of that domain family have their binding regions at or around the same positions. Their finding can assist in describing the functional diversity of protein−protein interactions, as well as introducing spatial constraints in modeling protein assemblies. Similarly, Aloy et al.[217] analyzed the relationship between sequence similarity and binding orientation and showed that the geometry of the interactions tends to be conserved between highly similar pairs. On the other hand, Henschel et al.[218] investigated binding at equivalent sites between nonhomologous proteins when interacting with a common partner. They found that, of all nonhomologous domains that bind with a common interaction partner, 4.2% use the same interface of the same common interaction partner (excluding immunoglobulins and proteases). Aytuna et al.[1] employed a bottom-up approach, combining structure and sequence conservation in protein interfaces to predict protein−protein interactions. Running the algorithm on a template data set of 67 known interfaces and a sequentially nonredundant data set of 6170 protein structures, they found a number of potential interactions, which they further verified with experimental data.[87] These indicate that there is progress toward this profound problem of predicting protein−protein interaction.

So, *what is the preferred way for proteins to interact*? In a thought-provoking comment already some years ago, van Regenmortel[219] argued that analyzing the interactions between biological molecules cannot be reduced to the description of (static) molecular structures. Integrated functional approaches need to consider the binding partner and the time component of the interaction. The function of a protein and its properties are decided not only by the static folded three-dimensional structure but also by the distribution and redistributions of the populations of its conformational and dynamic substates under different (physical or binding) environments.[65] Such mechanisms provide multiple pathways and allow a single molecular surface to interact with numerous structurally distinct binding partners, accommodating mutations through shifts in the dynamic energy landscape, and, as such, are evolutionarily advantageous. Yet the distribution of the conformations is not homogeneous, and the protein topology dictates the more dynamic regions.[220−223] Future work along integrative lines will provide insight into this profound protein−protein interaction problem.

12. Acknowledgments

We thank members of the group for discussions and suggestions. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. N01-CO-12400 and TUBITAK (Research Grant No. 104T504). O.K. has been granted with Turkish Academy of Sciences Young Investigator Programme (TUBA-GEBIP). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

13. References

(1) (1) Aytuna, A. S.; Gursoy, A.; Keskin, O. *Bioinformatics* **2005**, *21*, 2850.
(2) Malmstrom, L.; Riffle, M.; Strauss, C. E.; Chivian, D.; Davis, T. N.; Bonneau, R.; Baker, D. *PLoS Biol.* **2007**, *5*, e76.
(3) Nariai, N.; Kolaczyk, E. D.; Kasif, S. *PLoS ONE* **2007**, *2*, e337.
(4) Punta, M.; Forrest, L. R.; Bigelow, H.; Kernytsky, A.; Liu, J.; Rost, B. *Methods* **2007**, *41*, 460.
(5) Sharan, R.; Ulitsky, I.; Shamir, R. *Mol. Syst. Biol.* **2007**, *3*, 88.
(6) Watson, J. D.; Sanderson, S.; Ezersky, A.; Savchenko, A.; Edwards, A.; Orengo, C.; Joachimiak, A.; Laskowski, R. A.; Thornton, J. M. *J. Mol. Biol.* **2007**, *367*, 1511.
(7) Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; McBroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M.; Taylor, R.; Dharsee, M.; Ho, Y.; Heilbut, A.; Moore, L.; Zhang, S.; Ornatsky, O.; Bukhman, Y. V.; Ethier, M.; Sheng, Y.; Vasilescu, J.; Abu-Farha, M.; Lambert, J. P.; Duewel, H. S.; Stewart, II; Kuehl, B.; Hogue, K.; Colwill, K.; Gladwish, K.; Muskat, B.; Kinach, R.; Adams, S. L.; Moran, M. F.; Morin, G. B.; Topaloglou, T.; Figeys, D. *Mol. Syst. Biol.* **2007**, *3*, 89.
(8) Hart, G. T.; Ramani, A. K.; Marcotte, E. M. *Genome Biol.* **2006**, *7*, 120.
(9) Kim, P. M.; Lu, L. J.; Xia, Y.; Gerstein, M. B. *Science* **2006**, *314*, 1938.
(10) Rachlin, J.; Cohen, D. D.; Cantor, C.; Kasif, S. *Mol. Syst. Biol.* **2006**, *2*, 66.
(11) Yip, A. M.; Horvath, S. *BMC Bioinformatics* **2007**, *8*, 22.
(12) Cho, K. I.; Lee, K.; Lee, K. H.; Kim, D.; Lee, D. *Proteins* **2006**, *65*, 593.
(13) Keskin, O.; Nussinov, R. *Structure* **2007**, *15*, 341.
(14) Maerkl, S. J.; Quake, S. R. *Science* **2007**, *315*, 233.
(15) Mintseris, J.; Weng, Z. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10930.
(16) Morell, M.; Espargaro, A.; Aviles, F. X.; Ventura, S. *Proteomics* **2007**, *7*, 1023.
(17) Nooren, I. M.; Thornton, J. M. *J. Mol. Biol.* **2003**, *325*, 991.
(18) Sprinzak, E.; Altuvia, Y.; Margalit, H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14718.
(19) Tsai, C. J.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *Protein Sci.* **1997**, *6*, 53.
(20) Tsai, C. J.; Xu, D.; Nussinov, R. *Protein Sci.* **1997**, *6*, 1793.
(21) Young, L.; Jernigan, R. L.; Covell, D. G. *Protein Sci.* **1994**, *3*, 717.
(22) Norel, R.; Sheinerman, F.; Petrey, D.; Honig, B. *Protein Sci.* **2001**, *10*, 2147.
(23) Sheinerman, F. B.; Honig, B. *J. Mol. Biol.* **2002**, *318*, 161.
(24) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
(25) Xu, D.; Lin, S. L.; Nussinov, R. *J. Mol. Biol.* **1997**, *265*, 68.
(26) Kleanthous, C. *Protein−protein recognition*; Oxford University Press: Oxford, New York, 2000.
(27) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. *J. Mol. Biol.* **2004**, *336*, 943.
(28) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409.
(29) Halperin, I.; Wolfson, H.; Nussinov, R. *Structure* **2004**, *12*, 1027.
(30) Janin, J.; Rodier, F.; Chakrabarti, P.; Bahadur, R. P. *Acta Crystallogr., Sect. D* **2007**, *63*, 1.
(31) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 133.
(32) Keskin, O.; Ma, B.; Nussinov, R. *J. Mol. Biol.* **2005**, *345*, 1281.
(33) Keskin, O.; Ma, B.; Rogale, K.; Gunasekaran, K.; Nussinov, R. *Phys. Biol.* **2005**, *2*, S24.
(34) Laskowski, R. A.; Luscombe, N. M.; Swindells, M. B.; Thornton, J. M. *Protein Sci.* **1996**, *5*, 2438.
(35) Mintz, S.; Shulman-Peleg, A.; Wolfson, H. J.; Nussinov, R. *Proteins* **2005**, *61*, 6.
(36) Nooren, I. M.; Thornton, J. M. *EMBO J.* **2003**, *22*, 3486.
(37) Reichmann, D.; Rahat, O.; Cohen, M.; Neuvirth, H.; Schreiber, G. *Curr. Opin. Struct. Biol.* **2007**, *17*, 67.
(38) Haliloglu, T.; Keskin, O.; Ma, B.; Nussinov, R. *Biophys. J.* **2005**, *88*, 1552.
(39) Li, X.; Keskin, O.; Ma, B.; Nussinov, R.; Liang, J. *J. Mol. Biol.* **2004**, *344*, 781.
(40) Neuvirth, H.; Raz, R.; Schreiber, G. *J. Mol. Biol.* **2004**, *338*, 181.
(41) Reichmann, D.; Cohen, M.; Abramovich, R.; Dym, O.; Lim, D.; Strynadka, N. C.; Schreiber, G. *J. Mol. Biol.* **2007**, *365*, 663.
(42) Reichmann, D.; Rahat, O.; Albeck, S.; Meged, R.; Dym, O.; Schreiber, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 57.
(43) Gunasekaran, K.; Ma, B.; Nussinov, R. *Proteins* **2004**, *57*, 433.
(44) Chong, P. A.; Ozdamar, B.; Wrana, J. L.; Forman-Kay, J. D. *J. Biol. Chem.* **2004**, *279*, 40707.
(45) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197.
(46) Dyson, H. J.; Wright, P. E. *IUBMB Life* **2006**, *58*, 107.

(47) Gunasekaran, K.; Tsai, C. J.; Kumar, S.; Zanuy, D.; Nussinov, R. *Trends Biochem. Sci.* **2003**, *28*, 81.
(48) Kortemme, T.; Kelly, M. J.; Kay, L. E.; Forman-Kay, J.; Serrano, L. *J. Mol. Biol.* **2000**, *297*, 1217.
(49) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3.
(50) Shortle, D.; Ackerman, M. S. *Science* **2001**, *293*, 487.
(51) Dosztanyi, Z.; Chen, J.; Dunker, A. K.; Simon, I.; Tompa, P. *J. Proteome Res.* **2006**, *5*, 2985.
(52) Camacho, C. J.; Ma, H.; Champ, P. C. *Proteins* **2006**, *63*, 868.
(53) Camacho, C. J.; Vajda, S. *Curr. Opin. Struct. Biol.* **2002**, *12*, 36.
(54) Joerger, A. C.; Fersht, A. R. *Oncogene* **2007**, *26*, 2226.
(55) Kim, E.; Deppert, W. *Oncogene* **2007**, *26*, 2185.
(56) Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. *Phys. Biol.* **2005**, *2*, S156.
(57) Inbar, Y.; Benyamini, H.; Nussinov, R.; Wolfson, H. J. *J. Mol. Biol.* **2005**, *349*, 435.
(58) Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. J. *Curr. Med. Chem.* **2004**, *11*, 91.
(59) Fuentes, E. J.; Gilmore, S. A.; Mauldin, R. V.; Lee, A. L. *J. Mol. Biol.* **2006**, *364*, 337.
(60) James, L. C.; Roversi, P.; Tawfik, D. S. *Science* **2003**, *299*, 1362.
(61) Lindner, A. B.; Eshhar, Z.; Tawfik, D. S. *J. Mol. Biol.* **1999**, *285*, 421.
(62) Ma, B.; Kumar, S.; Tsai, C. J.; Nussinov, R. *Protein Eng.* **1999**, *12*, 713.
(63) Demchenko, A. P. *J. Mol. Recognit.* **2001**, *14*, 42.
(64) Fetler, L.; Kantrowitz, E. R.; Vachette, P. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 495.
(65) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10.
(66) Popovych, N.; Sun, S.; Ebright, R. H.; Kalodimos, C. G. *Nat. Struct. Mol. Biol.* **2006**, *13*, 831.
(67) Binkowski, T. A.; Joachimiak, A.; Liang, J. *Protein Sci.* **2005**, *14*, 2972.
(68) Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. *Nucleic Acids Res.* **2006**, *34*, W116.
(69) Glaser, F.; Rosenberg, Y.; Kessel, A.; Pupko, T.; Ben-Tal, N. *Proteins.* **2005**, *58*, 610.
(70) Jones, S.; Marin, A.; Thornton, J. M. *Protein Eng.* **2000**, *13*, 77.
(71) Nimrod, G.; Glaser, F.; Steinberg, D.; Ben-Tal, N.; Pupko, T. *Bioinformatics* **2005**, *21* (Suppl 1), i328.
(72) Shanahan, H. P.; Thornton, J. M. *Biopolymers* **2005**, *78*, 318.
(73) Dill, K. A.; Fiebig, K. M.; Chan, H. S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1942.
(74) Kolinski, A.; Galazka, W.; Skolnick, J. *Proteins* **1996**, *26*, 271.
(75) Tsai, C. J.; Nussinov, R. *Protein Sci.* **1997**, *6*, 1426.
(76) Clackson, T.; Wells, J. A. *Science* **1995**, *267*, 383.
(77) Karpusas, M.; Baase, W. A.; Matsumura, M.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8237.
(78) Albeck, S.; Unger, R.; Schreiber, G. *J. Mol. Biol.* **2000**, *298*, 503.
(79) Pal, G.; Ultsch, M. H.; Clark, K. P.; Currell, B.; Kossiakoff, A. A.; Sidhu, S. S. *J. Mol. Biol.* **2005**, *347*, 489.
(80) Teufel, D. P.; Kao, R. Y.; Acharya, K. R.; Shapiro, R. *Biochemistry* **2003**, *42*, 1451.
(81) Yang, J.; Swaminathan, C. P.; Huang, Y.; Guan, R.; Cho, S.; Kieke, M. C.; Kranz, D. M.; Mariuzza, R. A.; Sundberg, E. J. *J. Biol. Chem.* **2003**, *278*, 50412.
(82) Yang, P. L.; Schultz, P. G. *J. Mol. Biol.* **1999**, *294*, 1191.
(83) De, S.; Krishnadev, O.; Srinivasan, N.; Rekha, N. *BMC Struct. Biol.* **2005**, *5*, 15.
(84) Gao, Y.; Wang, R.; Lai, L. *J. Mol. Model.* **2004**, *10*, 44.
(85) Kim, W. K.; Henschel, A.; Winter, C.; Schroeder, M. *PLoS Comput. Biol.* **2006**, *2*, e124.
(86) Lu, H.; Lu, L.; Skolnick, J. *Biophys. J.* **2003**, *84*, 1895.
(87) Ogmen, U.; Keskin, O.; Aytuna, A. S.; Nussinov, R.; Gursoy, A. *Nucleic Acids Res.* **2005**, *33*, W331.
(88) Saha, R. P.; Bahadur, R. P.; Chakrabarti, P. *J. Proteome Res.* **2005**, *4*, 1600.
(89) Bordner, A. J.; Abagyan, R. *Proteins* **2005**, *60*, 353.
(90) Elcock, A. H.; McCammon, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2990.
(91) Korn, A. P.; Burnett, R. M. *Proteins* **1991**, *9*, 37.
(92) Kufareva, I.; Budagyan, L.; Raush, E.; Totrov, M.; Abagyan, R. *Proteins* **2007**, *67*, 400.
(93) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177.
(94) Jones, S.; Thornton, J. M. *J. Mol. Biol.* **1997**, *272*, 121.
(95) Li, Y.; Huang, Y.; Swaminathan, C. P.; Smith-Gill, S. J.; Mariuzza, R. A. *Structure* **2005**, *13*, 297.
(96) Ofran, Y.; Rost, B. *J. Mol. Biol.* **2003**, *325*, 377.
(97) Bahadur, R. P.; Chakrabarti, P.; Rodier, F.; Janin, J. *Proteins* **2003**, *53*, 708.
(98) Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081.
(99) Clackson, T.; Ultsch, M. H.; Wells, J. A.; de Vos, A. M. *J. Mol. Biol.* **1998**, *277*, 1111.
(100) Cho, S.; Swaminathan, C. P.; Yang, J.; Kerzic, M. C.; Guan, R.; Kieke, M. C.; Kranz, D. M.; Mariuzza, R. A.; Sundberg, E. J. *Structure* **2005**, *13*, 1775.
(101) Thorn, K. S.; Bogan, A. A. *Bioinformatics* **2001**, *17*, 284.
(102) Guerois, R.; Nielsen, J. E.; Serrano, L. *J. Mol. Biol.* **2002**, *320*, 369.
(103) Kortemme, T.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14116.
(104) Hu, Z.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins* **2000**, *39*, 331.
(105) Ma, B.; Elkayam, T.; Wolfson, H.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5772.
(106) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1.
(107) Schreiber, G.; Fersht, A. R. *J. Mol. Biol.* **1995**, *248*, 478.
(108) Clodfelter, K. H.; Waxman, D. J.; Vajda, S. *Biochemistry* **2006**, *45*, 9393.
(109) Rajamani, D.; Thiel, S.; Vajda, S.; Camacho, C. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 11287.
(110) Smith, G. R.; Sternberg, M. J.; Bates, P. A. *J. Mol. Biol.* **2005**, *347*, 1077.
(111) Erdemli, S. B.; Yogurtcu, O. N.; Turkay, M.; Nussinov, R.; Keskin, O. *Biophys. J.* **2008**.
(112) Landon, M. R.; Lancia, D. R., Jr.; Yu, J.; Thiel, S. C.; Vajda, S. *J. Med. Chem.* **2007**, *50*, 1231.
(113) Keskin, O.; Tsai, C. J.; Wolfson, H.; Nussinov, R. *Protein Sci.* **2004**, *13*, 1043.
(114) Keskin, O.; Nussinov, R. *Protein Eng. Des. Sel.* **2005**, *18*, 11.
(115) Moult, J.; Melamud, E. *Curr. Opin. Struct. Biol.* **2000**, *10*, 384.
(116) Andreeva, A.; Howorth, D.; Brenner, S. E.; Hubbard, T. J.; Chothia, C.; Murzin, A. G. *Nucleic Acids Res.* **2004**, *32*, D226.
(117) Finkelstein, A. V.; Ptitsyn, O. B. *Prog. Biophys. Mol. Biol.* **1987**, *50*, 171.
(118) Kohn, K. W.; Aladjem, M. I. *Mol. Syst. Biol.* **2006**, *2*, 2006 0002.
(119) Kohn, K. W.; Aladjem, M. I.; Weinstein, J. N.; Pommier, Y. *Mol. Biol. Cell* **2006**, *17*, 1.
(120) Bullock, A. N.; Debreczeni, J. E.; Edwards, A. M.; Sundstrom, M.; Knapp, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7637.
(121) Kamura, T.; Maenaka, K.; Kotoshiba, S.; Matsumoto, M.; Kohda, D.; Conaway, R. C.; Conaway, J. W.; Nakayama, K. I. *Genes Dev.* **2004**, *18*, 3055.
(122) Beckett, D. *Biochemistry* **2004**, *43*, 7983.
(123) Beckett, D. *Phys. Biol.* **2005**, *2*, S67.
(124) Block, P.; Paern, J.; Hullermeier, E.; Sanschagrin, P.; Sotriffer, C. A.; Klebe, G. *Proteins* **2006**, *65*, 607.
(125) Winter, C.; Henschel, A.; Kim, W. K.; Schroeder, M. *Nucleic Acids Res.* **2006**, *34*, D310.
(126) Gunasekaran, K.; Tsai, C. J.; Nussinov, R. *J. Mol. Biol.* **2004**, *341*, 1327.
(127) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565.
(128) Baldwin, R. L.; Zimm, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12391.
(129) Tran, H. T.; Wang, X.; Pappu, R. V. *Biochemistry* **2005**, *44*, 11369.
(130) Wang, X.; Vitalis, A.; Wyczalkowski, M. A.; Pappu, R. V. *Proteins* **2006**, *63*, 297.
(131) Keskin, O.; Yuret, D.; Gursoy, A.; Turkay, M.; Erman, B. *Proteins* **2004**, *55*, 992.
(132) Bowers, P. M.; Pellegrini, M.; Thompson, M. J.; Fierro, J.; Yeates, T. O.; Eisenberg, D. *Genome Biol.* **2004**, *5*, R35.
(133) Huang, T. W.; Tien, A. C.; Huang, W. S.; Lee, Y. C.; Peng, C. L.; Tseng, H. H.; Kao, C. Y.; Huang, C. Y. *Bioinformatics* **2004**, *20*, 3273.
(134) Han, D. S.; Kim, H. S.; Jang, W. H.; Lee, S. D.; Suh, J. K. *Nucleic Acids Res.* **2004**, *32*, 6312.
(135) Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. *Nucleic Acids Res.* **2007**, *35*, D301.
(136) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141.
(137) Ho, Y.; Gruhler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutilier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson,

F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M. *Nature* **2002**, *415*, 180.
(138) Salwinski, L.; Miller, C. S.; Smith, A. J.; Pettit, F. K.; Bowie, J. U.; Eisenberg, D. *Nucleic Acids Res.* **2004**, *32*, D449.
(139) Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamodar, G.; Yang, M.; Johnston, M.; Fields, S.; Rothberg, J. M. *Nature* **2000**, *403*, 623.
(140) Fontes, M. R.; Teh, T.; Kobe, B. *J. Mol. Biol.* **2000**, *297*, 1183.
(141) Kobe, B. *Nat. Struct. Biol.* **1999**, *6*, 388.
(142) Bayliss, R.; Littlewood, T.; Strawn, L. A.; Wente, S. R.; Stewart, M. *J. Biol. Chem.* **2002**, *277*, 50597.
(143) Cingolani, G.; Bednenko, J.; Gillespie, M. T.; Gerace, L. *Mol. Cell* **2002**, *10*, 1345.
(144) Fazi, B.; Cope, M. J.; Douangamath, A.; Ferracuti, S.; Schirwitz, K.; Zucconi, A.; Drubin, D. G.; Wilmanns, M.; Cesareni, G.; Castagnoli, L. *J. Biol. Chem.* **2002**, *277*, 5290.
(145) Gao, Y. G.; Yan, X. Z.; Song, A. X.; Chang, Y. G.; Gao, X. C.; Jiang, N.; Zhang, Q.; Hu, H. Y. *Structure* **2006**, *14*, 1755.
(146) Lewitzky, M.; Harkiolaki, M.; Domart, M. C.; Jones, E. Y.; Feller, S. M. *J. Biol. Chem.* **2004**, *279*, 28724.
(147) Yuan, T.; Vogel, H. J. *J. Biol. Chem.* **1998**, *273*, 30328.
(148) Ishida, H.; Nakashima, K.; Kumaki, Y.; Nakata, M.; Hikichi, K.; Yazawa, M. *Biochemistry* **2002**, *41*, 15536.
(149) Zhang, M.; Yuan, T. *Biochem. Cell. Biol.* **1998**, *76*, 313.
(150) Garcia, B. A.; Hake, S. B.; Diaz, R. L.; Kauer, M.; Morris, S. A.; Recht, J.; Shabanowitz, J.; Mishra, N.; Strahl, B. D.; Allis, C. D.; Hunt, D. F. *J. Biol. Chem.* **2007**, *282*, 7641.
(151) Taverna, S. D.; Ueberheide, B. M.; Liu, Y.; Tackett, A. J.; Diaz, R. L.; Shabanowitz, J.; Chait, B. T.; Hunt, D. F.; Allis, C. D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 2086.
(152) Orlicky, S.; Tang, X.; Willems, A.; Tyers, M.; Sicheri, F. *Cell* **2003**, *112*, 243.
(153) Smith, T. F.; Gaitatzes, C.; Saxena, K.; Neer, E. J. *Trends Biochem. Sci.* **1999**, *24*, 181.
(154) Groft, C. M.; Beckmann, R.; Sali, A.; Burley, S. K. *Nat. Struct. Biol.* **2000**, *7*, 1156.
(155) Sagermann, M.; Stevens, T. H.; Matthews, B. W. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7134.
(156) Main, E. R.; Lowe, A. R.; Mochrie, S. G.; Jackson, S. E.; Regan, L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 464.
(157) Magliery, T. J.; Regan, L. *BMC Bioinformatics* **2005**, *6*, 240.
(158) Voges, D.; Zwickl, P.; Baumeister, W. *Annu. Rev. Biochem.* **1999**, *68*, 1015.
(159) Liu, J.; Pan, Y.; Ma, B.; Nussinov, R. *Structure* **2006**, *14*, 1811.
(160) Boggon, T. J.; Eck, M. J. *Oncogene* **2004**, *23*, 7918.
(161) Krogan, N. J.; Peng, W. T.; Cagney, G.; Robinson, M. D.; Haw, R.; Zhong, G.; Guo, X.; Zhang, X.; Canadien, V.; Richards, D. P.; Beattie, B. K.; Lalev, A.; Zhang, W.; Davierwala, A. P.; Mnaimneh, S.; Starostine, A.; Tikuisis, A. P.; Grigull, J.; Datta, N.; Bray, J. E.; Hughes, T. R.; Emili, A.; Greenblatt, J. F. *Mol. Cell* **2004**, *13*, 225.
(162) Krause, R.; von Mering, C.; Bork, P.; Dandekar, T. *Bioessays* **2004**, *26*, 1333.
(163) Demirel, M. C.; Keskin, O. *J. Biomol. Struct. Dyn.* **2005**, *22*, 381.
(164) Ekman, D.; Light, S.; Bjorklund, A. K.; Elofsson, A. *Genome Biol.* **2006**, *7*, R45.
(165) Bloom, J. D.; Adami, C. *BMC Evol. Biol.* **2003**, *3*, 21.
(166) Caffrey, D. R.; Somaroo, S.; Hughes, J. D.; Mintseris, J.; Huang, E. S. *Protein Sci.* **2004**, *13*, 190.
(167) Fraser, H. B.; Hirsh, A. E.; Steinmetz, L. M.; Scharfe, C.; Feldman, M. W. *Science* **2002**, *296*, 750.
(168) Jordan, I. K.; Wolf, Y. I.; Koonin, E. V. *BMC Evol. Biol.* **2003**, *3*, 1.
(169) Grigoriev, A. *Nucleic Acids Res.* **2003**, *31*, 4157.
(170) Futreal, P. A.; Liu, Q.; Shattuck-Eidens, D.; Cochran, C.; Harshman, K.; Tavtigian, S.; Bennett, L. M.; Haugen-Strano, A.; Swensen, J.; Miki, Y., et al. *Science* **1994**, *266*, 120.
(171) Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P. A.; Harshman, K.; Tavtigian, S.; Liu, Q.; Cochran, C.; Bennett, L. M.; Ding, W., et al. *Science* **1994**, *266*, 66.
(172) Koonin, E. V.; Altschul, S. F.; Bork, P. *Nat. Genet.* **1996**, *13*, 266.
(173) Derbyshire, D. J.; Basu, B. P.; Serpell, L. C.; Joo, W. S.; Date, T.; Iwabuchi, K.; Doherty, A. J. *EMBO J.* **2002**, *21*, 3863.
(174) Joo, W. S.; Jeffrey, P. D.; Cantor, S. B.; Finnin, M. S.; Livingston, D. M.; Pavletich, N. P. *Genes Dev.* **2002**, *16*, 583.
(175) Sibanda, B. L.; Critchlow, S. E.; Begun, J.; Pei, X. Y.; Jackson, S. P.; Blundell, T. L.; Pellegrini, L. *Nat. Struct. Biol.* **2001**, *8*, 1015.
(176) Glover, J. N.; Williams, R. S.; Lee, M. S. *Trends Biochem. Sci.* **2004**, *29*, 579.
(177) Chai, Y. L.; Cui, J.; Shao, N.; Shyam, E.; Reddy, P.; Rao, V. N. *Oncogene* **1999**, *18*, 263.
(178) Hartman, A. R.; Ford, J. M. *J. Mol. Med.* **2003**, *81*, 700.
(179) MacLachlan, T. K.; Takimoto, R.; El-Deiry, W. S. *Mol. Cell. Biol.* **2002**, *22*, 4280.
(180) Navaraj, A.; Mori, T.; El-Deiry, W. S. *Cancer Biol. Ther.* **2005**, *4*, 1409.
(181) Mark, W. Y.; Liao, J. C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabartty, A.; Arrowsmith, C. H. *J. Mol. Biol.* **2005**, *345*, 275.
(182) Iwabuchi, K.; Bartel, P. L.; Li, B.; Marraccino, R.; Fields, S. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 6098.
(183) Ekblad, C. M.; Friedler, A.; Veprintsev, D.; Weinberg, R. L.; Itzhaki, L. S. *Protein Sci.* **2004**, *13*, 617.
(184) Bai, Y.; Sosnick, T. R.; Mayne, L.; Englander, S. W. *Science* **1995**, *269*, 192.
(185) Kim, K. S.; Woodward, C. *Biochemistry* **1993**, *32*, 9609.
(186) Woodward, C. *Trends Biochem. Sci.* **1993**, *18*, 359.
(187) Elber, R. *Biophys. J.* **2007**, *92*, L85.
(188) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318.
(189) Selivanova, G.; Wiman, K. G. *Oncogene* **2007**, *26*, 2243.
(190) Hupp, T. R.; Lane, D. P. *Curr. Biol.* **1994**, *4*, 865.
(191) Lu, X.; Lane, D. P. *Cell* **1993**, *75*, 765.
(192) Espinosa, J. M.; Emerson, B. M. *Mol. Cell* **2001**, *8*, 57.
(193) McKinney, K.; Mattia, M.; Gottifredi, V.; Prives, C. *Mol. Cell* **2004**, *16*, 413.
(194) McKinney, K.; Prives, C. *Mol. Cell. Biol.* **2002**, *22*, 6797.
(195) Fields, S.; Jang, S. K. *Science* **1990**, *249*, 1046.
(196) Bell, S.; Klein, C.; Muller, L.; Hansen, S.; Buchner, J. *J. Mol. Biol.* **2002**, *322*, 917.
(197) Paltoglou, S.; Roberts, B. J. *Oncogene* **2007**, *26*, 604.
(198) Hofmann, K. P.; Spahn, C. M.; Heinrich, R.; Heinemann, U. *Trends Biochem. Sci.* **2006**, *31*, 497.
(199) Bork, P.; Serrano, L. *Cell* **2005**, *121*, 507.
(200) Yip, C. K.; Kimbrough, T. G.; Felise, H. B.; Vuckovic, M.; Thomas, N. A.; Pfuetzner, R. A.; Frey, E. A.; Finlay, B. B.; Miller, S. I.; Strynadka, N. C. *Nature* **2005**, *435*, 702.
(201) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Rout, M. P.; Sali, A. *Nature* **2007**, *450*, 683.
(202) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Sali, A.; Rout, M. P. *Nature* **2007**, *450*, 695.
(203) Mintseris, J.; Weng, Z. *Proteins* **2003**, *53*, 629.
(204) Ponstingl, H.; Henrick, K.; Thornton, J. M. *Proteins* **2000**, *41*, 47.
(205) Cho, Y.; Gorina, S.; Jeffrey, P. D.; Pavletich, N. P. *Science* **1994**, *265*, 346.
(206) Ho, W. C.; Fitzgerald, M. X.; Marmorstein, R. *J. Biol. Chem.* **2006**, *281*, 20494.
(207) Kitayner, M.; Rozenberg, H.; Kessler, N.; Rabinovich, D.; Shaulov, L.; Haran, T. E.; Shakked, Z. *Mol. Cell* **2006**, *22*, 741.
(208) Bradford, J. R.; Needham, C. J.; Bulpitt, A. J.; Westhead, D. R. *J. Mol. Biol.* **2006**, *362*, 365.
(209) Liang, S.; Zhang, C.; Liu, S.; Zhou, Y. *Nucleic Acids Res.* **2006**, *34*, 3698.
(210) Bowie, J. U.; Luthy, R.; Eisenberg, D. *Science* **1991**, *253*, 164.
(211) Jones, S.; Thornton, J. M. *Curr. Opin. Chem. Biol.* **2004**, *8*, 3.
(212) Bilodeau, P. S.; Winistorfer, S. C.; Allaman, M. M.; Surendhran, K.; Kearney, W. R.; Robertson, A. D.; Piper, R. C. *J. Biol. Chem.* **2004**, *279*, 54808.
(213) Zhu, G.; Zhai, P.; He, X.; Wakeham, N.; Rodgers, K.; Li, G.; Tang, J.; Zhang, X. C. *EMBO J.* **2004**, *23*, 3909.
(214) Wright, C. S.; Zhao, Q.; Rastinejad, F. *J. Mol. Biol.* **2003**, *331*, 951.
(215) Schueler-Furman, O.; Wang, C.; Bradley, P.; Misura, K.; Baker, D. *Science* **2005**, *310*, 638.
(216) Korkin, D.; Davis, F. P.; Alber, F.; Luong, T.; Shen, M. Y.; Lucic, V.; Kennedy, M. B.; Sali, A. *PLoS Comput. Biol.* **2006**, *2*, e153.
(217) Aloy, P.; Ceulemans, H.; Stark, A.; Russell, R. B. *J. Mol. Biol.* **2003**, *332*, 989.
(218) Henschel, A.; Kim, W. K.; Schroeder, M. *Bioinformatics* **2006**, *22*, 550.
(219) Van Regenmortel, M. H. *J. Mol. Recognit.* **1999**, *12*, 1.
(220) Keskin, O.; Jernigan, R. L.; Bahar, I. *Biophys. J.* **2000**, *78*, 2093.
(221) Ma, B.; Shatsky, M.; Wolfson, H. J.; Nussinov, R. *Protein Sci.* **2002**, *11*, 184.
(222) Ma, B.; Wolfson, H. J.; Nussinov, R. *Curr. Opin. Struct. Biol.* **2001**, *11*, 364.
(223) Sinha, N.; Nussinov, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3139.
(224) Fukuhara, N.; Fernandez, E.; Ebert, J.; Conti, E.; Svergun, D. *J. Biol. Chem.* **2004**, *279*, 2176.
(225) Xenarios, I.; Rice, D. W.; Salwinski, L.; Baron, M. K.; Marcotte, E. M.; Eisenberg, D. *Nucleic Acids Res.* **2000**, *28*, 289.

CR040409X