Published in final edited form as:

Prog Biophys Mol Biol. 2014 ; 116(0): 194–202. doi:10.1016/j.pbiomolbio.2014.05.004.

Prediction and redesign of protein–protein interactions

Rhonald C. Lua^a, David C. Marciano^a, Panagiotis Katsonis^a, Anbu K. Adikesavan^a, Angela D. Wilkins^{a,c}, and Olivier Lichtarge^{a,b,c,*}

^aDepartment of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

^bVerna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

^cComputational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, TX 77030, USA

Abstract

Understanding the molecular basis of protein function remains a central goal of biology, with the hope to elucidate the role of human genes in health and in disease, and to rationally design therapies through targeted molecular perturbations. We review here some of the computational techniques and resources available for characterizing a critical aspect of protein function – those mediated by protein–protein interactions (PPI). We describe several applications and recent successes of the Evolutionary Trace (ET) in identifying molecular events and shapes that underlie protein function and specificity in both eukaryotes and prokaryotes. ET is a part of analytical approaches based on the successes and failures of evolution that enable the rational control of PPI.

Keywords

Evolutionary trace; Protein–protein interaction networks; Functional sites; Molecular evolution; Functional annotation; Protein design

1. Introduction

Protein–protein interactions (PPI) are ubiquitous and underlie the assembly of macromolecular machines, mediate signaling pathways in cellular networks (Jones and Thornton, 1996; Schadt, 2009; Yamada and Bork, 2009) and control cell-to-cell communication (Sattentau and Moore, 1993). Thus, characterizing PPI sites is an essential step towards: 1) identifying drug targets and mimicking them to create potential therapeutics (Moellering et al., 2009; Wells and McClendon, 2007), 2) engineering and modifying protein activity (Zorn and Wells, 2010), and 3) interpreting the impact of allelic variations.

To achieve these goals, a key step is to identify which proteins bind together and which amino acids mediate binding affinity and interaction specificity.

PPIs are diverse but they can be classified broadly into permanent (obligate, oligomeric, tight, more stable) and transient (relatively weak) interactions, which present different levels of co-expression and co-evolution (Jones and Thornton, 1996; Nooren and Thornton, 2003a,b; Mintseris and Weng, 2005; La et al., 2013). The interacting proteins of permanent complexes display higher tendencies to be co-expressed and co-localized than transiently-interacting proteins (Brown and Jurisica, 2007). Moreover, the amino acid residues that mediate permanent interactions evolve at a slower rate (they are more conserved) than residues involved in transient interactions (Mintseris and Weng, 2005). Therefore, it is expected to be more challenging to recognize the transient than the permanent interactions, because of the weaker signs of evolutionary pressure across the interface. However, this difficulty is mitigated by the fact that even the interface residues of the weak transient homodimers are generally more conserved than surface residues (Nooren and Thornton, 2003a,b).

A possible third category of PPI (or a special subcategory of transient interactions) is one mediated by disordered regions or unstructured segments of proteins (Babu et al., 2012; Tompa and Fuxreiter, 2008; Brown et al., 2011). Most eukaryotic proteins have both structured and disordered regions (Babu et al., 2012), and some proteins, such as the tumor suppressor p53, use disorder to achieve conformational heterogeneity and bind to a multitude of different protein partners (Hsu et al., 2012). Alternative splicing events and posttranslational modifications, which commonly occur in disordered regions, also contribute to distinct partner binding (Hsu et al., 2013). Disordered proteins generally evolve more rapidly than ordered proteins (Brown et al., 2011), although functional residues have been detected in disordered regions using evolutionary conservation (Nguyen Ba et al., 2012). The structured protein partner of the disordered protein showed substantially higher conservation in binding regions as compared to non-binding regions (Hsu et al., 2012).

Direct elucidation of protein–protein interfaces comes from crystallographic and NMR complexes (Berman et al., 2000) and are compiled into catalogs of interfaces (Schlessinger et al., 2006; Huang and Honda, 2006; de Beer et al., 2014; Krissinel and Henrick, 2007). Mutational studies can probe the energetic, catalytic or dynamic role of interface residues (Jackson and Fersht, 1993), but only a small fraction of them has been studied, due to the high cost, length of time required, and the growing number of interfaces (Hall, 2007; Liolios et al., 2006). Therefore, it is imperative to invent, refine, and update computational techniques that bypass experimental mutational assays by analyzing sequence and structure information.

In this review, we discuss how computational approaches can help in identifying PPI (See Fig. 1). Specifically, Section 2.1 focuses on predicting which proteins associate functionally or interact physically. Section 2.2 shows how to predict binding sites on protein surfaces, by using protein sequences and evolution data. Section 2.3 moves on to predicting pairs of interacting residues using new promising methods that rely on coevolution principles. Section 2.4 discusses *ab initio* and template-based docking, and how it can benefit from

predictions of binding sites and interacting residues described in Sections 2.2 and 2.3. Section 2.5, in turn, focuses on narrowing predictions of specificity and binding determinants to the ones relevant to the studied interaction. Section 3 illustrates recent applications of using such computational methods to identify, modulate and inhibit PPIs. The main application case focuses on the attempts to solve the puzzle of the long sought RecA–LexA PPI sites.

2. Current methods for PPIs

2.1. Finding and establishing links between proteins (“Which proteins interact?”)

In order to characterize protein–protein interfaces, the knowledge of which proteins physically interact is critical. Computational biology often transfers functional information from well-understood proteins to lesser-known ones using the concept of homology (Tatusov et al., 1997). Similarity searches (Mount, 2007) or shared domains (Aloy and Russell, 2006) can point to proteins in which the query of interest likely shares similar binding partners. This has become a common practice and has been applied in organizing PPI networks (Brown and Jurisica, 2007; Huang et al., 2004; Persico et al., 2005). However, homology transfer can be unreliable for interactions in phylogenetically distant species and should be used carefully (Lewis et al., 2012).

A complementary approach is to identify the proteins that are concurrently present or absent across large numbers of species. This co-occurrence inferred from phylogenetic profiling suggests a biological connection (Pellegrini et al., 1999; Tatusov et al., 2000; Schneider et al., 2013). The similarity of phylogenetic profiles can be assessed by assigning to each protein a vector encoding the patterns of presence or absence of that protein throughout many species. By finding matching or similar vectors, we can hypothesize which proteins interact. The resolution is expected to be low because disentangling physical and functional associations can be problematic (Kensche et al., 2008) but, in conjunction with multiple types of data, this approach can be useful (Snel and Huynen, 2004; Kim and Subramaniam, 2006). Gene co-expression is used in a similar fashion to identify proteins that likely interact (Ge et al., 2001; Taylor et al., 2008).

Several databases are already available that collect experimental knowledge of interactions and functional associations, and then aggregate this information to potentially predict new interactions. STRING (Franceschini et al., 2013) (<http://string-db.org/>) is particularly notable because it attempts to integrate these many sources of data into a network of physical and functional associations. It merges co-expression, co-occurrence, and homology with databases of protein–protein interactions and associations. These databases are derived from sources such as genomic context, high-throughput experiments (e.g., immunoprecipitation, yeast-two-hybrid, co-expression), PPI database imports, and literature cooccurrence. STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable (forming a “supergenomic” network). Currently STRING does not incorporate the structures of the proteins or close homologs in its predictions, but in the future may integrate this knowledge as well.

Unfortunately, the number of experimentally known interactions is growing at a much faster rate than the structurally characterized ones (Mosca et al., 2013), whose availability would allow the mapping of allelic variations and disease related mutations, rationalization of their mechanism of action, and the identification of drug targets. The situation mirrors the large gap between the number of known proteins and those that are characterized functionally with experimental annotations (only about 1%) (Erdin et al., 2011). As in the case of protein function annotation, we would have to infer a majority of PPIs using computational methods to dramatically expand the coverage of interaction space. Several databases (Tuncbag et al., 2011a,b; Zhang et al., 2012, 2013; Shoemaker et al., 2012; Hosur et al., 2012; Singh et al., 2010; Tuncbag et al., 2011a,b) show how the growing availability of structural data makes large-scale prediction of PPIs possible, including predictions of PPIs down to atomistic details, by knowledge-based approaches and structural homology to transfer interologs. These approaches and databases were recently reviewed by Mosca et al., (2013).

2.2. Finding protein interface residues (Meaningful sequence variations and ET)

Functionally important residues are often evolutionarily conserved, and therefore serve as good candidates when searching for possible protein–protein interfaces. Structural alignments showed that conservation of Tyr, Phe and Met on the protein surface suggests a likely binding site (Ma et al., 2003; Ma and Nussinov, 2007). In addition, methods using machine-learning algorithms, such as ISIS, were developed to identify binding interfaces from protein sequence (Ofra and Rost, 2007a, b; Res et al., 2005). However, conservation does not always seem a clear guide to interfaces (Caffrey et al., 2004). Previous studies also show that only a small subset of interface residues (mostly in the core) is essential to protein–protein recognition or binding (Ofra and Rost, 2007a, b; Clackson and Wells, 1995; Bogan and Thorn, 1998; Valdar and Thornton, 2001; Mihalek et al., 2007), although the conservation signal can be enhanced by splitting the interfaces into core and rim regions (Guharoy and Chakrabarti, 2005; Janin et al., 2008), and by considering the presence of multiple interfaces on a single protein (Choi et al., 2009).

Going beyond simple models of evolutionary conservation to identify key binding sites is the earlier incarnation of the Evolutionary Trace (ET) (Wilkins et al., 2012a, b; Mihalek et al., 2004; Lichtarge et al., 1996). ET uses a more natural model of evolution that accounts not only for absolute conservation of key functional residues, but also for systematic variation of specificity determinants as they adapt to different evolutionary constraints (Res and Lichtarge, 2005). Residues that are invariant among the widely divergent branches of evolution are expected to have a larger functional impact than other residues that vary among closely related species (Lichtarge et al., 1996). The distribution of evolutionarily important residues is far from random when mapped onto the protein structure (Lichtarge et al., 1996; Baranski et al., 1999; Madabushi et al., 2002), forming clusters that overlap with known functional sites (Madabushi et al., 2002; Baameur et al., 2010; Lichtarge et al., 2003; Yao et al., 2003; Sowa et al., 2000; Onrust et al., 1997; Chiba et al., 2006; Madabushi et al., 2004; Wilkins et al., 2013; Wilkins et al., 2010) from which one may then automatically extract 3D motifs useful for transferring functional annotation between proteins on the scale of the entire structural proteome (Venner et al., 2010; Amin et al., 2013; Erdin et al., 2013). These 3D motifs consists of five or six residues identified by ET, they do not rely on prior

knowledge of the molecular basis of any function, and a recent study shows that they can be accurate enough for *bona fide* prediction of both catalytic activity and substrates (Amin et al., 2013). Inclusion of evolutionary importance by ET rankings in machine learning or other methods improves the classification of protein residues into those that are part of protein interfaces and those that are not (Res et al., 2005; Engelen et al., 2009).

ET typically needs at least 15-20 sequences (Wilkins et al., 2012a, b) from homologous proteins that are well distributed across evolution to extract meaningful sequence variations for functional site predictions. ET requires many fewer sequences than methods that depend on residue covariation, consistent with the difference between first order correlations (between a sequence position and the ensemble of all others positions) and second order correlations (between two separate residue positions). The sequences may come from sources such as BLAST (Altschul et al., 1990) searches, then filtered for acceptable ranges of sequence identity and lengths relative to the query protein, or from HSSP (Joosten et al., 2011) alignments. The quality of the functional site predictions can then be assessed by the degree of clustering of the top-ranked ET residues, or more generally by the degree of smoothness in the spatial variation of ET ranks, as described previously. If the sequences given to ET are too few, or are too similar to each other (or too divergent from each other), the degree of clustering (quantified by a z-score) tends to be small. ET servers and viewers are available at <http://mammoth.bcm.tmc.edu>.

2.3. Key interacting and contacting residues in between proteins

In proteins known to interact, an additional analysis to find interaction determinants is to study the covariation and the complementarity of amino acids across the interface. Indeed, when the proteins of a complex are treated as a single, concatenated sequence, it becomes possible to search for common patterns of residue variations at the interfaces of interacting proteins (Mintseris and Weng, 2005; Pazos et al., 1997; Gouldson et al., 2001; Wang and DeLisi, 2006; Pazos and Valencia, 2008; Halperin et al., 2006) or to study patterns of complementarity. (Madaoui and Guerois, 2008) However, these relationships may be subtle and not especially specific for identifying binding determinants (Hakes et al., 2007). For example, better coevolutionary relationships may be found between residues that are more distant from the interface (Chakrabarti and Panchenko, 2009). Thus, although there are evolutionary constraints on either side of a protein–protein interface, (Guharoy and Chakrabarti, 2005) the gain in resolution may be modest. Nevertheless, by examining patterns of amino acid co-evolution in large, multiple sequence alignments of cognate histidine kinase and response regulator pairs in bacteria, Skerker et al.(2008) were able to demonstrate that a subset of the coevolving residues (1) map to the interface, and (2) when mutated, completely switched the substrate specificity of the kinase EnvZ, enabling a rational rewiring of two-component signaling pathways. This work suggested that analyses of coevolution could facilitate the reprogramming of other PPIs.

Recently, improved methods relying on amino acid coevolution/covariation/correlated mutations were developed that more accurately captured a large number of native contacts across many protein families (Morcos et al., 2011) and used to predict 3D structures/models of several proteins (de Juan et al., 2013; Cocco et al., 2013; Sulkowska et al., 2012),

successfully reproducing known structures and offering several novel predictions, including proteins in the pharmacologically important class of G Protein Coupled Receptors (GPCRs) (Hopf et al., 2012). Crucially, these methods have the ability to disentangle direct and indirect correlations that confounded previous attempts to identify contacts from sequence correlation analysis, by using entropy maximization to identify an essential set of correlations between pairs of positions in the sequence that best explains all the other correlations and are therefore likely to be causative. This statistical inference approach is characterized as ‘global’, in contrast to ‘local’ approaches such as mutual information. Because coevolution between residue positions on different proteins is no different from that between residue positions within the same protein chain, these methods can potentially be used to predict protein–protein interfaces (de Juan et al., 2013) (inter-protein contacts (Weigt et al., 2009) as well as intra-protein and intra-domain (Dago et al., 2012)), but may require large numbers of homologous yet variable paired sequences (~1000) (Morcos et al., 2011).

2.4. Predicting protein–protein interaction surface when the unbound structures are available (Docking)

The availability of structure for one or both interacting protein partners opens up a whole new avenue for predicting the binding surfaces. First, the structures reveal the exposed surface residues that are available for interaction. Second, the protein surfaces may be inspected for signs of binding sites, such as clusters of evolutionarily important residues (Madabushi et al., 2002; Yao et al., 2003; Wilkins et al., 2012). Third, the proposed interacting surfaces can be tested for shape (geometric, steric) and physicochemical (e.g. charge) complementarity (Madaoui and Guerois, 2008; Wass et al., 2011), binding free energy and affinity (Kastritis and Bonvin, 2013; Kortemme et al., 2004; Moal et al., 2013), coupling of structurally conserved residues and experimental and computational hotspots (Cukuroglu et al., 2012; Halperin et al., 2004).

Predicting the binding mode of two protein monomers falls under the realm of docking. Traditionally, the first step in docking is the generation of candidate complexes by shifting and rotating one monomer relative to the other until a satisfactory fit is obtained between the two (Janin et al., 2008; Mashich et al., 2010). This step, which is typically very fast because the interactions can be treated at a coarse-grained level, treats the monomers as rigid but allows steric clashes in the complex. In addition, the docking search space can be reduced given knowledge of binding sites that cross-link or reduce binding affinity upon mutagenesis. The second step is ranking the candidates using a scoring function that may include molecular mechanics energy, residue pair potentials, proximity of evolutionarily important or structurally conserved residues, geometric and physicochemical complementarity, as mentioned above. The third step is the refinement of the structures (backbone or side chain) of top-ranked complexes to reduce steric clashes and improve (minimize) some energy function, followed by rescoring and reranking.

The scoring and ranking of docking models can take advantage of the spatial proximity at protein–protein interfaces of structurally conserved (Halperin et al., 2004), evolutionarily conserved (Aloy et al., 2001), or complementary residues (Madaoui and Guerois, 2008) and

coevolving residues (Weigt et al., 2009), and are valuable enough to enrich the accuracy of docking solutions. For example, Madaoui and Guerois (2008) systematically linked contacts across the protein–protein interface to their physicochemical complementarity, whereby they formed hydrophobic–hydrophobic, polar–polar, or oppositely charged contacts in 95% of their pairings in a multiple sequence alignment. The physicochemical complementarity between sites was used as a component of the SCOTCH score, which was shown to be highly sensitive in detecting near–native complexes from other models produced by computational docking experiments when starting from the (albeit) bound protein partners (Madaoui and Guerois, 2008). This raises the possibility that a proper accounting of the evolutionary forces operating jointly across the interface may reveal a useful correlation, much as was seen at protein–DNA interfaces (Raviscioni et al., 2005). Weigt et al. improved upon this approach by the addition of a statistical inference step aimed at disentangling direct from indirect correlations (Weigt et al., 2009) (as described above using entropy maximization). The top residue pairs identified by their algorithm were found to be in contact at the protein interface between sensor kinase (SK) and response regulator (RR) proteins, and for homointeractions between RR proteins. Though limited to a single example, this demonstrated the large-scale feasibility of this application on protein complexes.

Another way to perform docking without generating complexes *ab initio* is to search for similarity of the two target protein structures among co-crystallized proteins (templates) in biological assemblies, followed by superposition of the target proteins onto the matching templates (Kundrotas et al., 2012). Template-based modeling can be more reliable and efficient than *ab initio* model building, extrapolating from the results of structure prediction of single proteins (Moult et al., 2009, 2011). The first and simplest of the knowledge/information/template-based approaches search for (global) sequence or structure homology between target and template proteins (Zhang et al., 2013; Haberle et al., 2011). Using only interface architectures as templates, independent of global sequence or structure homology, is a significant improvement on this scheme, because it was found that similar protein interface architectures are reused by structurally and functionally different protein pairs (Tuncbag et al., 2011a,b; Gunther et al., 2007; Kundrotas and Vakser, 2010; Sinha et al., 2010, 2012). In the PRISM protocol (Tuncbag et al., 2011a,b) the target proteins, after being transformed onto the template complex, additionally undergo flexible refinement.

The performance of various docking and scoring methods, servers and groups are evaluated in a community-wide experiment, the Critical Assessment of Predicted Interactions (CAPRI) (Lensink and Wodak, 2013; Janin, 2013). The results of this experiment have demonstrated the success of information-driven docking (Rodrigues et al., 2013).

2.5. Finding key residues, binding specificity determinants, hotspots

One approach to identify the binding determinants in PPIs is to measure the contribution of interfacial amino acids to the free energy of binding (Schreiber et al., 1997). Experimentally, alanine scanning mutagenesis makes this possible (Clackson and Wells, 1995), in which the change in free energy of binding is measured when residues are systematically mutated to alanine. Such experiments demonstrated that only a few residues contribute significantly to

the binding free energy at protein interfaces (Clackson and Wells, 1995; Bogan and Thorn, 1998; Moreira et al., 2007). In general, experiments showed that Trp, Tyr and Arg contributed most often to the binding free energy (Bogan and Thorn, 1998). Spatial clusters of amino acids with large free energy contributions can also be suggested computationally with either all-atom molecular dynamics simulations (Moreira et al., 2007; Gilson et al., 1997; Simonson et al., 2002; Swanson et al., 2004) or with simpler models (Guerois et al., 2002; Kortemme and Baker, 2002) and structural analysis (Nussinov and Wolfson, 1991; Tsai et al., 1996; Hu et al., 2000). One may also simply exploit “anchor” residues, i.e. amino acid side-chains deeply buried at protein–protein interfaces, which also indicate possible druggable pockets to be targeted by small molecules (Meireles et al., 2010; Rajamani et al., 2004). Although energetic aspects of these residues were extensively studied in the literature (Moreira et al., 2007; Gilson et al., 1997; McCammon, 1998; Sheinerman et al., 2000), the interplay between evolutionary conservation and the extent of free-energy contribution of interfacial amino acids remains unclear.

The binding determinants can be designed into peptides or small molecule inhibitors of PPIs. In a simple strategy, these key residues mimic the interaction with the native protein partner, while the other residues can be tuned to optimize other useful properties such as solubility or helical propensity. Some applications of this strategy are described in the next section.

3. Applications specific to ET

3.1. Rational design guided by ET

The analysis of evolutionary patterns in proteins was a useful step towards identifying PPIs. ET has been extensively tested to identify both protein–protein interfaces as well ligand binding specificity in Gα protein subunits (Onrust et al., 1997), regulators of G protein signaling (Madabushi et al., 2002; Sowa et al., 2000; Sheikh et al., 1996), Ku heterodimer (Ribes-Zamora et al., 2007) and prestin (Rajagopalan et al., 2006) and as well many types of receptors: beta adrenergic receptor (Shenoy et al., 2006), C5a receptor (Geva et al., 2000), oxytocin receptor (Yang et al., 2002), dopamine receptor (Rodriguez et al., 2010) and rhodopsin (Baranski et al., 1999; Madabushi et al., 2004), and nuclear hormone receptors (Baranski et al., 1999; Innis, 2007). These studies identified key amino acids that made up spatial sites. In turn, these sites were modulated to block or separate functions, swap interactions, and design peptides that inhibited function.

The Ku70/80 is a well-known heterodimer essential for the nonhomologous end joining (NHEJ) pathway of DNA repair. This complex binds and travels along the DNA searching for double-strand break ends and recruits fellow proteins to facilitate the nonhomologous end joining process. A mutational study of predicted functional sites produced separation of function in yeast orthologs (Ribes-Zamora et al., 2007) segregating double-strand break DNA repair and telomere maintenance to opposite ends of the Ku structure. These studies narrowed the focus of experiments to help clarify how antagonistic functions exist in the same complex.

Evolutionary patterns not only identify the binding sites, but they can suggest residue substitutions that switch functions between homologous proteins. By swapping residues, we rewired the transcriptional specificity of the bHLH proneural transcription factors from the frog homologue to the fly homologue, and vice-versa (Quan et al., 2004). In a more recent application, we demonstrated the ability to rewire the allosteric pathway in the Dopamine receptor, D2, to increase response to serotonin (Rodriguez et al., 2010).

These evolutionarily important sites have also been used in the design of peptides that inhibit protein interactions. In GRK, which regulates the activity of GPCRs through phosphorylation of the intracellular loops, we predicted multiple functional sites (Baameur et al., 2010) that involved mostly surface exposed helices. These helices were ideal for peptide design, but when isolated from the rest of the protein, the native sequence region had low helical propensity (Munoz and Serrano, 1994). In order for the sequence fragment to keep its helical fold, we generated new peptides that maintained the identity of the evolutionarily important residues and allowed the rest to vary. The helical propensity (Munoz and Serrano, 1994) was used to select for the best target peptides. Experimental studies found that the peptides disrupted GPCR phosphorylation (Baameur et al., 2010).

In a more recent application with broad therapeutic implications in the treatment of human diseases, a cell-permeable autophagy-inducing peptide, Tat-Beclin 1, was designed such that a set of amino acids were altered to enhance solubility in the native peptide sequence. (Shoji-Kawata et al., 2013) The peptide sequence was derived from a functionally important region of the autophagy protein, Beclin-1, that binds the HIV-1 pathogenicity protein, Nef. ET analysis of Beclin 1 identified the three residue positions as unlikely to be functionally important but occupied by charged residues in Beclin 1 orthologues. These cognate residues were then substituted into the peptide (H275E, S279D and Q281E) to increase solubility without disrupting its mimicry of Beclin 1 function. Aside from decreasing the replication of several pathogens (including HIV-1) *in vitro* and improving clinical outcomes in mice infected with chikungunya or West Nile virus, the peptide was also used to screen and discover a previously unknown inhibitor of autophagy, GAPR-1 (also called GLIPR2).

3.2. Find interaction sites in RecA and LexA

The long sought protein–protein interaction between the RecA and LexA proteins controls the bacterial SOS response to DNA damage. Although RecA and LexA have been extensively studied with targeted mutagenesis and their unbound crystallographic structures are available, the interaction between RecA and LexA is still poorly understood. Here we briefly explain how an ET study provides insight into this protein–protein interaction.

Under conditions of cell stress, an inducible DNA repair system (SOS response) increases the rate of mutation in bacteria, (Michel, 2005; Al Mamun et al., 2012) conferring antibiotic resistance. The DNA recombinase, RecA, and the self-repressor LexA are the two key regulators, the latter also repressing over 40 downstream SOS response genes (Courcelle et al., 2001; Adikesavan et al., 2011). RecA acts as a co-protease to promote the autocatalytic cleavage of LexA, thus initiating the SOS response. SOS derepresses error-prone polymerases that introduce mutations, leading to the development of antibiotic resistance.

The RecA crystal structure (Story et al., 1992; Chen et al., 2008) gave atomic resolution for the helical oligomerization of RecA monomers around DNA (Fig. 2A). Scanning transmission electron microscopy images reconstruct a low-resolution picture of the RecA–LexA structure suggesting that LexA binds within the deep helical groove of the RecA filament (Yu and Egelman, 1993). Genetics studies had identified several RecA mutants that alter LexA cleavage (McGrew and Knight, 2003) and, with the RecA crystal structure, led to the proposal of a LexA binding site that extends across two adjacent RecA monomers (Story et al., 1992).

ET analysis identified key functional RecA residues that were in contact with DNA, ATP/ADP, and other RecA monomers, but it also identified residues that, upon mutation, separate LexA cleavage from SOS induction (Adikesavan et al., 2011). Many of the residues under strongest selection are within or adjacent to the helical groove (Fig. 2B). The ET-guided mutational study revealed two novel sites (residues G87, K88, G108 and residues G22, K23, G24) adjacent to the helical groove of RecA, which could potentially interact with LexA. The location of these mutations on the active RecA filament narrowed down the space for docking analysis and led us to propose a model for LexA binding across the helical groove (Fig. 2C). A subsequent, independent docking study placed a LexA monomer in a similar location but in a different orientation and deeper within the helical groove (Kovacic et al., 2013).

These findings illustrate how applying the available computational tools can lead to better understanding on the molecular basis of protein function, RecA and LexA in this example, while a future direction of designing peptides to inhibit the RecA–LexA interaction will open up a new window for manipulation of bacterial stress response and provide a new drug target to combat the evolution of antibiotic resistance.

4. Concluding remarks

Due to the diversity and regulatory impact of PPIs, there is tremendous interest in targeting the interaction surfaces between proteins for therapeutic purposes (Wells and McClendon, 2007). A rational design approach would use a resolved 3D structure of the protein–protein or protein–ligand complex, from which the contact surfaces can be derived. It will be exciting if the success of coevolution-based methods that predict intra-protein contacts can be extended to also predict the interaction sites between transient protein–protein partners such as RecA and LexA, and perhaps confirm one of the docking models described previously. This success will depend on the quality and diversity of the paired sequence alignments. Once the RecA–LexA interaction sites are known and characterized (e.g. as a sequentially continuous or discontinuous binding site (Eichler, 2008; Jubb et al., 2012)), the peptide design approach described in this review (Shoji-Kawata et al., 2013) or one of the recent technological advances to tackle PPIs as drug targets (Eichler, 2008; Higuieruelo et al., 2013) can be selected to potentially discover a new drug in the fight against antibiotic resistance. The ET-based approach could perhaps be combined with the new techniques such as stapled peptides that improve cell permeability and resistance to proteolysis and help usher in this new class of therapeutics.

Acknowledgments

Funding: National Institutes of Health (NIH-GM079656, NIHGM066099, NLM 5T15LM07093); National Science Foundation (NSF CCF-0905536, NSF DBI-1062455).

References

- Adikesavan AK, et al. Separation of recombination and SOS response in *Escherichia coli* RecA suggests LexA interaction sites. *PLoS Genet.* 2011; 7:e1002244. [PubMed: 21912525]
- Al Mamun AA, et al. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science.* 2012; 338:1344–1348. [PubMed: 23224554]
- Aloy P, Russell RB. Structural systems biology: modelling protein interactions. *Nat. Rev. Mol. Cell. Biol.* 2006; 7:188–197. [PubMed: 16496021]
- Aloy P, Querol E, Aviles FX, Sternberg MJ. Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation and to protein docking. *J. Mol. Biol.* 2001; 311:395–408. [PubMed: 11478868]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol.* 1990; 215:403–410. [PubMed: 2231712]
- Amin SR, Erdin S, Ward RM, Lua RC, Lichtarge O. Prediction and experimental validation of enzyme substrate specificity in protein structures. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:E4195–E4202. [PubMed: 24145433]
- Baameur F, et al. Role for the regulator of G-protein signaling homology domain of G protein-coupled receptor kinases 5 and 6 in beta 2-adrenergic receptor and rhodopsin phosphorylation. *Mol. Pharmacol.* 2010; 77:405–415. [PubMed: 20038610]
- Babu MM, Kriwacki RW, Pappu RV. Structural biology. Versatility from protein disorder. *Science.* 2012; 337:1460–1461. [PubMed: 22997313]
- Baranski TJ, et al. C5a receptor activation. Genetic identification of critical residues in four transmembrane helices. *J. Biol. Chem.* 1999; 274:15757–15765. [PubMed: 10336477]
- Berman HM, et al. The protein data bank. *Nucleic Acids Res.* 2000; 28:235–242. [PubMed: 10592235]
- Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 1998; 280:1–9. [PubMed: 9653027]
- Brown KR, Jurisica I. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.* 2007; 8:R95. [PubMed: 17535438]
- Brown CJ, Johnson AK, Dunker AK, Daughdrill GW. Evolution and disorder. *Curr. Opin. Struct. Biol.* 2011; 21:441–446. [PubMed: 21482101]
- Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci.* 2004; 13:190–202. [PubMed: 14691234]
- Chakrabarti S, Panchenko AR. Coevolution in defining the functional specificity. *Proteins.* 2009; 75:231–240. [PubMed: 18831050]
- Chen Z, Yang H, Pavletich NP. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature.* 2008; 453:484–489.
- Chiba P, Mihalek I, Ecker GF, Kopp S, Lichtarge O. Role of transmembrane domain/transmembrane domain interfaces of P-glycoprotein (ABCB1) in solute transport. Convergent information from photoaffinity labeling, site directed mutagenesis and in silico importance prediction. *Curr. Med. Chem.* 2006; 13:793–805. [PubMed: 16611068]
- Choi YS, Yang JS, Choi Y, Ryu SH, Kim S. Evolutionary conservation in multiple faces of protein interaction. *Proteins.* 2009; 77:14–25. [PubMed: 19350617]
- Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science.* 1995; 267:383–386. [PubMed: 7529940]
- Cocco S, Monasson R, Weigt M. From principal component to direct coupling analysis of coevolution in proteins: low-eigenvalue modes are needed for structure prediction. *PLoS Comput Biol.* 2013; 9:e1003176. [PubMed: 23990764]

- Courcelle J, Khodursky A, Peter B, Brown PO, Hanawalt PC. Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*. *Genetics*. 2001; 158:41–64. [PubMed: 11333217]
- Cukuroglu E, Gursoy A, Keskin O. HotRegion: a database of predicted hot spot clusters. *Nucleic Acids Res*. 2012; 40:D829–D833. [PubMed: 22080558]
- Dago AE, et al. Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:E1733–E1742. [PubMed: 22670053]
- de Beer TA, Berka K, Thornton JM, Laskowski RA. PDBsum additions. *Nucleic Acids Res*. 2014; 42:D292–D296. [PubMed: 24153109]
- de Juan D, Pazos F, Valencia A. Emerging methods in protein co-evolution. *Nat. Rev. Genet.* 2013; 14:249–261. [PubMed: 23458856]
- Eichler J. Peptides as protein binding site mimetics. *Curr. Opin. Chem. Biol.* 2008; 12:707–713. [PubMed: 18935974]
- Engelen S, Trojan LA, Sacquin-Mora S, Lavery R, Carbone A. Joint evolutionary trees: a large-scale method to predict protein interfaces based on sequence sampling. *PLoS Comput Biol*. 2009; 5:e1000267. [PubMed: 19165315]
- Erdin S, Lisewski AM, Lichtarge O. Protein function prediction: towards integration of similarity metrics. *Curr. Opin. Struct. Biol.* 2011; 21:180–188. [PubMed: 21353529]
- Erdin S, Venner E, Lisewski AM, Lichtarge O. Function prediction from networks of local evolutionary similarity in protein structure. *BMC Bioinforma.* 2013; 14(Suppl 3):S6.
- Franceschini A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41:D808–D815. [PubMed: 23203871]
- Ge H, Liu Z, Church GM, Vidal M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* 2001; 29:482–486. [PubMed: 11694880]
- Geva A, Lassere TB, Lichtarge O, Pollitt SK, Baranski TJ. Genetic mapping of the human C5a receptor. Identification of transmembrane amino acids critical for receptor function. *J. Biol. Chem.* 2000; 275:35393–35401. [PubMed: 10952985]
- Gilson MK, Given JA, Bush BL, McCammon JA. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 1997; 72:1047–1069. [PubMed: 9138555]
- Gouldson PR, et al. Lipid-facing correlated mutations and dimerization in G-protein coupled receptors. *Protein Eng.* 2001; 14:759–767. [PubMed: 11739894]
- Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* 2002; 320:369–387. [PubMed: 12079393]
- Guharoy M, Chakrabarti P. Conservation and relative importance of residues across protein-protein interfaces. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:15447–15452. [PubMed: 16221766]
- Gunther S, May P, Hoppe A, Frommel C, Preissner R. Docking without docking: ISEARCH–prediction of interactions using known interfaces. *Proteins*. 2007; 69:839–844. [PubMed: 17803236]
- Haberle J, et al. Molecular defects in human carbamoyl phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations. *Hum. Mutat.* 2011; 32:579–589. [PubMed: 21120950]
- Hakes L, Lovell SC, Oliver SG, Robertson DL. Specificity in protein interactions and its relationship with sequence diversity and coevolution. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:7999–8004. [PubMed: 17468399]
- Hall N. Advanced sequencing technologies and their wider impact in microbiology. *J. Exp. Biol.* 2007; 210:1518–1525. [PubMed: 17449817]
- Halperin I, Wolfson H, Nussinov R. Protein-protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure*. 2004; 12:1027–1038. [PubMed: 15274922]
- Halperin I, Wolfson H, Nussinov R. Correlated mutations: advances and limitations. A study on fusion proteins and on the Cohesin–Dockerin families. *Proteins*. 2006; 63:832–845. [PubMed: 16508975]

- Higueruelo AP, Jubb H, Blundell TL. Protein-protein interactions as druggable targets: recent technological advances. *Curr. Opin. Pharmacol.* 2013; 13:791–796. [PubMed: 23735579]
- Hopf TA, et al. Three-dimensional structures of membrane proteins from genomic sequencing. *Cell.* 2012; 149:1607–1621. [PubMed: 22579045]
- Hosur R, et al. A computational framework for boosting confidence in high-throughput protein-protein interaction datasets. *Genome Biol.* 2012; 13:R76. [PubMed: 22937800]
- Hsu, WL., et al. Intrinsic protein disorder and protein-protein interactions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*; 2012. p. 116-127.
- Hsu WL, et al. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* 2013; 22:258–273. [PubMed: 23233352]
- Hu Z, Ma B, Wolfson H, Nussinov R. Conservation of polar residues as hot spots at protein interfaces. *Proteins.* 2000; 39:331–342. [PubMed: 10813815]
- Huang TW, et al. POINT: a database for the prediction of protein-protein interactions based on the orthologous interactome. *Bioinformatics.* 2004; 20:3273–3276. [PubMed: 15217821]
- Huang J, Honda W. CED: a conformational epitope database. *BMC Immunol.* 2006; 7:7. [PubMed: 16603068]
- Innis CA. siteFiNDER|3D: a web-based tool for predicting the location of functional sites in proteins. *Nucleic Acids Res.* 2007; 35:W489–W494. [PubMed: 17553829]
- Jackson SE, Fersht AR. Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin BPN'. *Biochemistry.* 1993; 32:13909–13916. [PubMed: 8268166]
- Janin J, Bahadur RP, Chakrabarti P. Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.* 2008; 41:133–180. [PubMed: 18812015]
- Janin J. The targets of CAPRI rounds 20–27. *Proteins.* 2013; 81:2075–2081. [PubMed: 23900782]
- Jones S, Thornton JM. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:13–20. [PubMed: 8552589]
- Joosten RP, et al. A series of PDB related databases for everyday needs. *Nucleic Acids Res.* 2011; 39:D411–D419. [PubMed: 21071423]
- Jubb H, Higueruelo AP, Winter A, Blundell TL. Structural biology and drug discovery for protein-protein interactions. *Trends Pharmacol. Sci.* 2012; 33:241–248. [PubMed: 22503442]
- Kastritis PL, Bonvin AM. Molecular origins of binding affinity: seeking the Archimedean point. *Curr. Opin. Struct. Biol.* 2013; 23:868–877. [PubMed: 23876790]
- Kensche PR, van Noort V, Dutilh BE, Huynen MA. Practical and theoretical advances in predicting the function of a protein by its phylogenetic distribution. *J. Royal Soc. Interface/Royal Soc.* 2008; 5:151–170.
- Kim Y, Subramaniam S. Locally defined protein phylogenetic profiles reveal previously missed protein interactions and functional relationships. *Proteins.* 2006; 62:1115–1124. [PubMed: 16385560]
- Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* 2002; 99:14116–14121. [PubMed: 12381794]
- Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE: signal Transduct. Knowl. Environ.* 2004; 2004:pl2.
- Kovacic L, et al. Structural insight into LexA-RecA* interaction. *Nucleic Acids Res.* 2013; 41:9901–9910. [PubMed: 23965307]
- Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 2007; 372:774–797. [PubMed: 17681537]
- Kundrotas PJ, Vakser IA. Accuracy of protein-protein binding sites in high-throughput template-based modeling. *PLoS Comput Biol.* 2010; 6:e1000727. [PubMed: 20369011]
- Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:9438–9441. [PubMed: 22645367]
- La D, Kong M, Hoffman W, Choi YI, Kihara D. Predicting permanent and transient protein-protein interfaces. *Proteins.* 2013; 81:805–818. [PubMed: 23239312]

- Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. *Proteins*. 2013; 81:2082–2095. [PubMed: 24115211]
- Lewis AC, Jones NS, Porter MA, Deane CM. What evidence is there for the homology of protein-protein interactions? *PLoS Comput Biol*. 2012; 8:e1002645. [PubMed: 23028270]
- Lichtarge O, Bourne HR, Cohen FE. Evolutionarily conserved Galphabetagamma binding surfaces support a model of the G protein-receptor complex. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:7507–7511. [PubMed: 8755504]
- Lichtarge O, Yao H, Kristensen DM, Madabushi S, Mihalek I. Accurate and scalable identification of functional sites by evolutionary tracing. *J. Struct. Funct. Genomics*. 2003; 4:159–166. [PubMed: 14649300]
- Liolios K, Tavernarakis N, Hugenholtz P, Kyripides NC. The genomes on line database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res*. 2006; 34:D332–D334. [PubMed: 16381880]
- Ma B, Nussinov R. Trp/Met/Phe hot spots in protein-protein interactions: potential targets in drug design. *Curr. Top. Med. Chem*. 2007; 7:999–1005. [PubMed: 17508933]
- Ma B, Elkayam T, Wolfson H, Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. U. S. A.* 2003; 100:5772–5777. [PubMed: 12730379]
- Madabushi S, et al. Structural clusters of evolutionary trace residues are statistically significant and common in proteins. *J. Mol. Biol*. 2002; 316:139–154. [PubMed: 11829509]
- Madabushi S, et al. Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *J. Biol. Chem*. 2004; 279:8126–8132. [PubMed: 14660595]
- Madaoui H, Guerois R. Coevolution at protein complex interfaces can be detected by the complementarity trace with important impact for predictive docking. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:7708–7713. [PubMed: 18511568]
- Mashiach E, et al. An integrated suite of fast docking algorithms. *Proteins*. 2010; 78:3197–3204. [PubMed: 20607855]
- McCammon JA. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol*. 1998; 8:245–249. [PubMed: 9631300]
- McGrew DA, Knight KL. Molecular design and functional organization of the RecA protein. *Critical Rev. Biochem. Mol. Biol*. 2003; 38:385–432. [PubMed: 14693725]
- Meireles LM, Domling AS, Camacho CJ. ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res*. 2010; 38:W407–W411. [PubMed: 20525787]
- Michel B. After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biol*. 2005; 3:e255. [PubMed: 16000023]
- Mihalek I, Res I, Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J. Mol. Biol*. 2004; 336:1265–1282. [PubMed: 15037084]
- Mihalek I, Res I, Lichtarge O. On itinerant water molecules and detectability of protein-protein interfaces through comparative analysis of homo-logues. *J. Mol. Biol*. 2007; 369:584–595. [PubMed: 17434530]
- Mintseris J, Weng Z. Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 2005; 102:10930–10935. [PubMed: 16043700]
- Moal IH, Moretti R, Baker D, Fernandez-Recio J. Scoring functions for protein-protein interactions. *Curr. Opin. Struct. Biol*. 2013; 23:862–867. [PubMed: 23871100]
- Moeller RE, et al. Direct inhibition of the NOTCH transcription factor complex. *Nature*. 2009; 462:182–188. [PubMed: 19907488]
- Morcos F, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:E1293–E1301. [PubMed: 22106262]
- Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis—an improved methodological approach. *J. Comput Chem*. 2007; 28:644–654. [PubMed: 17195156]

- Mosca R, Pons T, Ceol A, Valencia A, Aloy P. Towards a detailed atlas of protein-protein interactions. *Curr. Opin. Struct. Biol.* 2013; 23:929–940. [PubMed: 23896349]
- Moult J, Fidelis K, Kryshtafovych A, Rost B, Tramontano A. Critical assessment of methods of protein structure prediction – round VIII. *Proteins.* 2009; 77(Suppl 9):1–4. [PubMed: 19774620]
- Moult J, Fidelis K, Kryshtafovych A, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)–round IX. *Proteins.* 2011; 79(Suppl 10):1–5. [PubMed: 21997831]
- Mount DW. Using the basic local alignment search tool (BLAST). *CSH Protoc.* 2007 2007 pdb top.17.
- Munoz V, Serrano L. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. *Proteins.* 1994; 20:301–311. [PubMed: 7731949]
- Nguyen Ba AN, et al. Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci. Signal.* 2012; 5:rs1. [PubMed: 22416277]
- Nooren IM, Thornton JM. Diversity of protein-protein interactions. *EMBO J.* 2003a; 22:3486–3492. [PubMed: 12853464]
- Nooren IM, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.* 2003b; 325:991–1018. [PubMed: 12527304]
- Nussinov R, Wolfson HJ. Efficient detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques. *Proc. Natl. Acad. Sci. U. S. A.* 1991; 88:10495–10499. [PubMed: 1961713]
- Ofran Y, Rost B. ISIS: interaction sites identified from sequence. *Bioinformatics.* 2007a; 23:e13e16. [PubMed: 17237081]
- Ofran Y, Rost B. Protein-protein interaction hotspots carved into sequences. *PLoS Comput Biol.* 2007b; 3:e119. [PubMed: 17630824]
- Onrust R, et al. Receptor and betagamma binding sites in the alpha subunit of the retinal G protein transducin. *Science.* 1997; 275:381–384. [PubMed: 8994033]
- Pazos F, Valencia A. Protein co-evolution, co-adaptation and interactions. *Embo J.* 2008; 27:2648–2655. [PubMed: 18818697]
- Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.* 1997; 271:511–523. [PubMed: 9281423]
- Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:4285–4288. [PubMed: 10200254]
- Persico M, et al. HomoMINT: an inferred human network based on orthology mapping of protein interactions discovered in model organisms. *BMC Bioinforma.* 2005; 6(Suppl. 4):S21.
- Quan XJ, et al. Evolution of neural precursor selection: functional divergence of proneural proteins. *Development.* 2004; 131:1679–1689. [PubMed: 15084454]
- Rajagopalan L, et al. Essential helix interactions in the anion transporter domain of prestin revealed by evolutionary trace analysis. *J. Neurosci.* 2006; 26:12727–12734. [PubMed: 17151276]
- Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:11287–11292. [PubMed: 15269345]
- Raviscioni M, Gu P, Sattar M, Cooney AJ, Lichtarge O. Correlated evolutionary pressure at interacting transcription factors and DNA response elements can guide the rational engineering of DNA binding specificity. *J. Mol. Biol.* 2005; 350:402–415. [PubMed: 15946684]
- Res I, Lichtarge O. Character and evolution of protein-protein interfaces. *Phys. Biol.* 2005; 2:S36–S43. [PubMed: 16204847]
- Res I, Mihalek I, Lichtarge O. An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics.* 2005; 21:2496–2501. [PubMed: 15728113]
- Ribes-Zamora A, Mihalek I, Lichtarge O, Bertuch AA. Distinct faces of the Ku heterodimer mediate DNA repair and telomeric functions. *Nat. Struct. Mol. Biol.* 2007; 14:301–307. [PubMed: 17351632]
- Rodrigues JP, et al. Defining the limits of homology modeling in information-driven protein docking. *Proteins.* 2013; 81:2119–2128. [PubMed: 23913867]

- Rodriguez GJ, Yao R, Lichtarge O, Wensel TG. Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:7787–7792. [PubMed: 20385837]
- Sattentau QJ, Moore JP. The role of CD4 in HIV binding and entry. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 1993; 342:59–66. [PubMed: 7904348]
- Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature.* 2009; 461:218–223. [PubMed: 19741703]
- Schlessinger A, Ofra Y, Yachdav G, Rost B. Epitome: database of structure-inferred antigenic epitopes. *Nucleic Acids Res.* 2006; 34:D777–D780. [PubMed: 16381978]
- Schneider A, Seidl MF, Snel B. Shared protein complex subunits contribute to explaining disrupted co-occurrence. *PLoS Comput Biol.* 2013; 9:e1003124. [PubMed: 23874172]
- Schreiber G, Frisch C, Fersht AR. The role of Glu73 of barnase in catalysis and the binding of barstar. *J. Mol. Biol.* 1997; 270:111–122. [PubMed: 9231905]
- Sheikh SP, Zvyaga TA, Lichtarge O, Sakmar TP, Bourne HR. Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F. *Nature.* 1996; 383:347–350. [PubMed: 8848049]
- Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.* 2000; 10:153–159. [PubMed: 10753808]
- Shenoy SK, et al. beta-arrestin-dependent, G protein-independent ERK1/2 activation by the beta2 adrenergic receptor. *J. Biol. Chem.* 2006; 281:1261–1273. [PubMed: 16280323]
- Shoemaker BA, et al. IBIS (Inferred Biomolecular Interaction Server) reports, predicts and integrates multiple types of conserved interactions for proteins. *Nucleic Acids Res.* 2012; 40:D834–D840. [PubMed: 22102591]
- Shoji-Kawata S, et al. Identification of a candidate therapeutic autophagy-inducing peptide. *Nature.* 2013; 494:201–206. [PubMed: 23364696]
- Simonson T, Archontis G, Karplus M. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* 2002; 35:430–437. [PubMed: 12069628]
- Singh R, Park D, Xu J, Hosur R, Berger B. Struct2Net: a web service to predict protein-protein interactions using a structure-based approach. *Nucleic Acids Res.* 2010; 38:W508–W515. [PubMed: 20513650]
- Sinha R, Kundrotas PJ, Vakser IA. Docking by structural similarity at protein-protein interfaces. *Proteins.* 2010; 78:3235–3241. [PubMed: 20715056]
- Sinha R, Kundrotas PJ, Vakser IA. Protein docking by the interface structure similarity: how much structure is needed? *PloS one.* 2012; 7:e31349. [PubMed: 22348074]
- Skerker JM, et al. Rewiring the specificity of two-component signal transduction systems. *Cell.* 2008; 133:1043–1054. [PubMed: 18555780]
- Snel B, Huynen MA. Quantifying modularity in the evolution of biomolecular systems. *Genome Res.* 2004; 14:391–397. [PubMed: 14993205]
- Sowa ME, He W, Wensel TG, Lichtarge O. A regulator of G protein signaling interaction surface linked to effector specificity. *Proc. Natl. Acad. Sci. U. S. A.* 2000; 97:1483–1488. [PubMed: 10677488]
- Story RM, Weber IT, Steitz TA. The structure of the E. coli recA protein monomer and polymer. *Nature.* 1992; 355:318–325. [PubMed: 1731246]
- Sulkowska JI, Morcos F, Weigt M, Hwa T, Onuchic JN. Genomics-aided structure prediction. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:10340–10345. [PubMed: 22691493]
- Swanson JM, Henchman RH, McCammon JA. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* 2004; 86:67–74. [PubMed: 14695250]
- Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. *Science.* 1997; 278:631–637. [PubMed: 9381173]
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000; 28:33–36. [PubMed: 10592175]

- Taylor RC, Acquah-Mensah G, Singhal M, Malhotra D, Biswal S. Network inference algorithms elucidate Nrf2 regulation of mouse lung oxidative stress. *PLoS Comput Biol.* 2008; 4:e1000166. [PubMed: 18769717]
- Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 2008; 33:2–8. [PubMed: 18054235]
- Tsai CJ, Lin SL, Wolfson HJ, Nussinov R. A dataset of protein-protein interfaces generated with a sequence-order-independent comparison technique. *J. Mol. Biol.* 1996; 260:604–620. [PubMed: 8759323]
- Tuncbag N, Gursoy A, Keskin O. Prediction of protein-protein interactions: unifying evolution and structure at protein interfaces. *Phys. Biol.* 2011a; 8:035006. [PubMed: 21572173]
- Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat. Protoc.* 2011b; 6:1341–1354. [PubMed: 21886100]
- Valdar WS, Thornton JM. Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins.* 2001; 42:108–124. [PubMed: 11093265]
- Venner E, et al. Accurate protein structure annotation through competitive diffusion of enzymatic functions over a network of local evolutionary similarities. *PloS one.* 2010; 5:e14286. [PubMed: 21179190]
- Wang YE, DeLisi C. Inferring protein-protein interactions in viral proteins by co-evolution of conserved side chains. *Genome Inf.* 2006; 17:23–35.
- Wass MN, Fuentes G, Pons C, Pazos F, Valencia A. Towards the prediction of protein interaction partners using physical docking. *Mol. Syst. Biol.* 2011; 7:469. [PubMed: 21326236]
- Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. U. S. A.* 2009; 106:67–72. [PubMed: 19116270]
- Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature.* 2007; 450:1001–1009. [PubMed: 18075579]
- Wilkins AD, et al. Accounting for epistatic interactions improves the functional analysis of protein structures. *Bioinformatics.* 2013; 29:2714–2721. [PubMed: 24021383]
- Wilkins AD, Lua R, Erdin S, Ward RM, Lichtarge O. Sequence and structure continuity of evolutionary importance improves protein functional site discovery and annotation. *Protein Sci.* 2010; 19:1296–1311. [PubMed: 20506260]
- Wilkins A, Erdin S, Lua R, Lichtarge O. Evolutionary trace for prediction and redesign of protein functional sites. *Methods Mol. Biol.* 2012a; 819:29–42. [PubMed: 22183528]
- Wilkins AD, Bachman BJ, Erdin S, Lichtarge O. The use of evolutionary patterns in protein annotation. *Curr. Opin. Struct. Biol.* 2012b; 22:316–325. [PubMed: 22633559]
- Yamada T, Bork P. Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nat. Rev. Mol. Cell. Biol.* 2009; 10:791–803. [PubMed: 19851337]
- Yang M, et al. Lysine 270 in the third intracellular domain of the oxytocin receptor is an important determinant for G alpha(q) coupling specificity. *Mol. Endocrinol.* 2002; 16:814–823. [PubMed: 11923477]
- Yao H, et al. An accurate, sensitive, and scalable method to identify functional sites in protein structures. *J. Mol. Biol.* 2003; 326:255–261. [PubMed: 12547207]
- Yu X, Egelman EH. The LexA repressor binds within the deep helical groove of the activated RecA filament. *J. Mol. Biol.* 1993; 231:29–40. [PubMed: 8496964]
- Zhang QC, et al. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature.* 2012; 490:556–560. [PubMed: 23023127]
- Zhang QC, Petrey D, Garzon JI, Deng L, Honig B. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res.* 2013; 41:D828–D833. [PubMed: 23193263]
- Zorn JA, Wells JA. Turning enzymes ON with small molecules. *Nat. Chem. Biol.* 2010; 6:179–188. [PubMed: 20154666]

Fig. 1.

Computational characterization of PPI that also serves as an outline for much of this article.

A. Databases of PPI networks allow us to answer the question “Which proteins interact?” directly or functionally. **B.** Computational predictions of interacting/binding sites on one protein surface from sequence analysis using ET (yellow circles) can complement experimental data such as alanine scanning mutagenesis. **C.** Coevolution from simultaneous analysis of two protein partners known to directly interact can complement biochemical and cross-linking data. Dashed lines indicate coevolving or significantly correlated residues. **D.** Prediction of PPI interfaces by docking. **E.** Knowledge of the protein–protein complex enables refinement of computational predictions of key residues (red circles). **F.** Design of peptide to mimic and inhibit native interaction. Key binding and specificity determinants (yellow circles) are built into a scaffold that enhances other desirable features such as solubility or helical propensity. Note that characterization of PPI is not necessarily followed in the sequential order **A–F**.

Fig. 2.

ET-guided mutagenesis leads to a protein–protein interaction model. **A.** The helical oligomerization of RecA (alternating green and yellow monomers) around DNA. The positions of RecA mutations that alter LexA cleavage (Adikesavan et al., 2011; McGrew and Knight, 2003) are colored in magenta. **B.** ET shows the RecA phylogenetic variation pattern as a color spectrum ranging from violet (most variation) to red (least variation). **C.** Overlapping docking solutions for the LexA–RecA interaction are shown for a LexA C-terminal dimer (dark blue (Adikesavan et al., 2011)) and a LexA monomer (light blue (Kovacac et al., 2013)). The positions used to guide docking in Adikesavan et al. (G22 & G108) are shown in red.