Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Computational design of protein–protein interactions

Gideon Schreiber and Sarel J Fleishman

A long-term aim of computational design is to generate specific protein–protein interactions at desired affinity, specificity, and kinetics. The past three years have seen the first reports on atomically accurate *de novo* interactions. These were based on advances in design algorithms and the ability to harness high-throughput experimental characterization of design variants to optimize binding. Current state-of-the-art in computational design lacks precision, and therefore requires intensive experimental optimization to achieve parity with natural binders. Recent successes (and failures) point the way to future progress in design methodology that would enable routine and robust design of binders and inhibitors, while also shedding light on the essential features of biomolecular recognition.

Addresses
Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Corresponding authors: Schreiber, Gideon (sarel@weizmann.ac.il) and Fleishman, Sarel J (sarel@weizmann.ac.il)

Current Opinion in Structural Biology 2013, **23**:xx–yy

This review comes from a themed issue on **Protein–protein interactions**

Edited by **Joel Janin** and **Alexandre Bonvin**

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2013.08.003

Two fundamental features of natural protein interactions are specificity and affinity. Affinities of protein–protein interaction in biological systems span 10-orders of magnitude, from high μM to fM, with many interactions being exquisitely specific at their native concentrations and environments [1]. They form homo-complexes or hetero-complexes with one or many partners. Binding affinities may be influenced by temperature, pH, ionic strength or posttranslational modifications. The site of interaction can be anywhere on the surface, both in structured and unstructured regions; however, a tendency to interact at specific locations has been observed [1]. Finally, binding affinity is a function of the difference between the overall free energy of the unbound relative to the bound state of the proteins. This includes direct contributions (bonds between the two proteins) as well as changes in the internal energy of the proteins, the water structure, entropy and other factors. Partly because of these complicating factors the design of novel binding sites lagged behind the design of new folds [2] or enzymatic functions [3,4]. In this review we survey recent advances in computational design of protein–protein interactions, from engineering of altered affinity to *de novo* design of interfaces.

Predicting mutational effects

Predicting the energetic outcome of a set of point mutations provides a clear and unbiased benchmark for energy functions. However, constructing such a benchmark is complicated as the structure of the mutated protein is usually unknown and different experimental methods may lead to wide disparities in reported binding energies [5]. The ability to accurately predict the effects of mutations is still a formidable task [6,7]. Many new studies were presented in recent years to address this issue, both for predicting protein stability and protein–protein binding affinity. One of the potential problems is the sparse data available to calibrate energy functions. Moreover, the choice of mutations tested is biased by experimentalists' hypotheses, and usually does not provide a good coverage of the potential energy surface. In recent years, deep sequencing has become relatively cheap and accessible, providing for the first time complete data on all point mutations at the binding surface extracted from a natural pool of binders, or following random or focused mutagenesis and *in vitro* selection for binding. This allows for un-biased mapping of mutational effects on binding and stability. For example, deep sequencing of selected antibody clones against VEGF provided the sequences of higher-affinity variants that were used to calculate the probability of sequence space constituting the interface (using machine learning) [8]. The assumption is that a specific amino-acid preference is directly related to its contribution to the free energy of the selected trait. Accordingly, more favorable solutions will be selected and their propensity would increase in the selected population relative to the unselected, or naïve, population [9••]. Similarly, the availability of over 50 thousand HIV-1 protease sequences isolated from patients was used to explore the mutational tolerance of this protein and the space of stability versus function tradeoff that viruses explore during the accumulation of resistance mutations [10•].

Prediction and design of association kinetics

Binding affinity factors the rate of association (k_{on}) and dissociation (k_{off}). The dissociation rate measures the stability of the complex after it has formed (half-life), and is independent of protein concentration. The observed rate of association is a product of the protein concentration and the physical association-rate constant. The latter is mostly affected by electrostatic forces that can be manipulated by mutations [11]. Indeed, electrostatic

Please cite this article in press as: Schreiber G, Fleishman SJ. Computational design of protein–protein interactions, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.08.003

design of 'on-rates' through optimization of the electrostatic energy of interaction at the bound state by mutating amino-acids surrounding the physical interface but not making direct contact with the partner protein was an early success in protein-interface design [12]. Further developments in 'on-rate' predictions have successfully simulated the diffusion-driven binding rate. Combining these two factors reproduced observed rates of association for different cytokine-receptor interactions and mutants of these complexes [13]. The authors pointed out a correlation between the diverse (>1000-fold) association rate constants of the different pairs and the *in vivo* cytokine circulating concentrations, suggesting that the observed cytokine-receptor binding rates are close to the limits set by the half-lives of the receptors. These k_{on} calculations were successfully reproduced also in crowded environments, mimicking the intracellular milieu [14].

Two additional factors have to be taken into account when designing binding rates. One, is the need for conformational selection before binding [15], and second is that forming hydrogen bonds between buried donors/acceptors requires dehydration/rehydration, which involves an energetically penalized transition state. This results in a kinetic trap that may slow or weaken binding [16]. Together, these data point toward a reaction-limiting component in association, which has to be taken into account when designing binding sites.

Changing the electrostatic properties of the surface has the potential to stabilize proteins. Interestingly, using Rosetta to optimize the charge distribution at the surface of a single-chain antibody resulted, in addition to increased stability, also in a 30-fold enhancement in binding affinity, which apparently is not related to non-specific electrostatic attraction [17].

Designing affinity and specificity

Affinity and specificity of interactions are two sides of the same coin. To design binding while ignoring specificity can be achieved by making flexible hydrophobic surfaces that would bind one another or other similar hydrophobic surfaces. Imposing a mixture of bond types and shape complementarity between rugged surfaces encodes binding specificity. Using a structure-based energy score accounting for electrostatic forces and buried surface area of a residue and its conservation in known Regulators of G protein Signaling (RGS) proteins, Kosloff et al. [18•] identified a set of peripheral residues that encode interaction specificity toward G proteins. Mutation of those resulted in variations of RGS specificity toward G proteins. In another study the signaling cascade of the GTPase Cdc42 and its activator, Intersectin, was rewired. The altered interface designed between the two proteins resulted in GTPase activation exclusively by its engineered cognate partner, both *in vitro* and in mammalian cells [19]. Another example of applying predictive tools to alter biological activity through interface engineering was demonstrated for the binding of pro-apoptotic BH3 (Bcl-2 homology 3) proteins to Bcl-2 receptor proteins. BH3 promotes cell death by docking an α-helix into a hydrophobic groove on the surface of one or more of five pro-survival Bcl2-receptors. Mutations that were predicted to decrease binding resulted in increased apoptosis [20]. A different approach using peptide SPOT arrays and deep sequencing data from yeast display screening to determine the BH3 sequence space that binds to the Bcl-2 receptor proteins was undertaken by DeBartolo et al. [21]. The data were used to calibrate two algorithms (SPOT and DEEP) that predict binding affinity and specificity against five human Bcl-2 receptors. The experimental data were also used to evaluate a more general prediction algorithm, STATIUM, which is solely based on summing up the pairwise interactions within a given interface (scored according to their occurrence within a large database of protein structures). All three algorithms distinguished binders from non-binders, with STATIUM performing as well as the other two, suggesting that system specific data are not necessary in this case. On the same experimental system, Rosetta was used to design a constrained library, from which a variant with 1000-fold improved binding specificity for the BH3 region of Bad over the BH3 region of Bim was selected. Although negative design was applied only against the BH3 region of Bim, the best redesigned protein was specific against binding to other BH3 motifs [22•]. The design framework demonstrates an efficient route to design new Bcl-2 family complexes.

In a study aimed to achieve higher affinity of binding of MHC Class I Polypeptide-Related Sequence A (MICA) to the homodimeric immunoreceptor NKG2D the fixed-backbone RosettaDesign protocol was used to identify appropriate residues for mutation. The best-performing mutations were combined, resulting in a measured 50-fold increase in binding affinity [23]. Many of the multiple mutants were subadditive compared to the sum of the single mutants, suggesting cooperativity in the interface. In another study, Hawse et al. [24] aimed to explore whether T-cell signaling is dynamically controlled. As the binding between the DMF5 T-cell receptor (TCR) with a peptide presented by HLA-A2 was too weak to be detected by mass spectrometry, they used Rosetta to identify mutations that enhance binding affinity. Ten of the highest scoring mutants were experimentally analyzed, with two of them leading to enhanced binding from 11 μM to 40 nM. Rosetta, in combination with a range of experimental methods, was also used successfully to explain the molecular basis of a million-fold affinity maturation between T-cell receptor and bacterial antigen [25].

A method to introduce a binding site between two non-interacting proteins is to graft it from a different protein

Please cite this article in press as: Schreiber G, Fleishman SJ. Computational design of protein–protein interactions, Curr Opin Struct Biol (2013), http://dx.doi.org/10.1016/j.sbi.2013.08.003