Report on the research conducted under the funding of the Sloan foundation postdoctoral fellowship in Computational Molecular Biology

A central question of my research proposal was what structural features are common at protein-protein interfaces and what energetic factors define the affinity and specificity of protein-protein association. Analysis of structural and mutational data on protein-protein interfaces revealed that protein-protein interfaces of different functional classes contain many more energetically important charged and polar residues that was previously thought. Since, in the context of protein folding studies, polar interactions are believed to mostly destabilize the folded proteins, this observation raised the question as to the forces that determine the stability of protein complexes [1]. To investigate this issue in detail we developed a number of partitioning schemes that allowed us to investigate the role of selected residues, ion pairs and networks of polar interactions in protein-protein association.

We applied the developed methods to the analysis of four different protein-protein interfaces [2]. The complexes examined are: the ribonuclease barnase and its inhibitor barstar, the human growth hormone and its receptor, subtype N9 influenza virus neuraminidase and NC41 antibody, and the Ras Binding Domain (RBD) of kinase cRaf and a Ras homologue Rap1A. The calculations revealed a surprising variability in how polar interactions affect the stability of different complexes. We found that the overall impact of electrostatic interactions on binding affinity ranges from destabilizing (hormone-receptor and neuraminidase-antibody complexes) to neutral (barnase-barstar complex) to stabilizing (RBD-Rap1A complex). Two principal conclusions were reached. (1) Residues with stabilizing electrostatic contributions to binding often participate in the formation of networks of ionic and hydrogen-bonding interactions involving several closely positioned interfacial residues. Formation of such clusters of interacting residues is significantly favored over the formation of isolated salt bridges by the balance of the electrostatic effects. In addition, interfacial residues that are partially buried on the surface of a monomer can make strong stabilizing contributions to the free energy of association since the penalty for their desolvation upon formation of a complex is reduced. (2) The balance of all polar interactions is most favorable for the stability of complexes, where significant stabilization due to other physical mechanisms (i.e. hydrophobic effect, reduced penalty for the loss of conformational entropy) is unlikely. Thus, optimization of the placement of charged and polar residues on protein surfaces may improve dramatically the impact of polar interactions on the stability of a protein-protein complex and, along with other factors, defines the affinity of protein-protein association.

The finding that positions of charged and polar residues on protein-protein interfaces are optimized with respect to electrostatic interactions, suggests that this property can be employed for the discrimination between native conformations and trial complexes generated by a docking algorithm. We found that a computationally inexpensive scoring function that penalized interfaces burying many unsatisfied charges and rewarding them for the formation of networks of ionic and hydrogen-bonding interactions correlates well with the exact calculations on the impact of electrostatics on protein complex stability.
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
The incorporation of this scoring function into protein docking algorithm [3] was found to improve dramatically the discrimination of native protein-protein complexes among many trial conformations generated with a docking algorithm, including those that are indistinguishable from native in terms of geometric complementarity and overall physico-chemical properties (number of charges at the interface, number of hydrogen bonds and salt bridges, buried polar and non-polar area).

The ability to predict the ligand-binding function of members of a protein family that bind to different ligands provides an important test of our understanding of the affinity and specificity of association. We analyzed binding of phosphotyrosine-containing peptides by SH2 domains, and developed an algorithm for the prediction of an SH2 domain's binding preferences based solely on its sequence (ref). The strategy for the analysis of the binding preferences of SH2 domains comprises two major steps. First we developed a highly accurate algorithm to generate a structure-guided alignment of sequences of all SH2 domains currently available in the SWISSPROT sequence database. We demonstrate that the produced sequence alignment is of higher quality than that available in sequence databases, or that can be produced by widely used methods, such as PSI-BLAST or HMM. Then, in cases where 3D structures of complexes of SH2 domains with high affinity peptides are known, we analyze what binding site residues are most important energetically for the protein-peptide association. The partitioning scheme developed for the analysis of protein-protein complexes [2] is employed. The identity and the location of all residues predicted to be important for the recognition of a high affinity target form a signature of the peptide-binding site for a given SH2 domain. Alignment of sequences of all SH2 domains is then used to analyze similarities between subsets of residues lining these energetically important positions and to judge on possible similarities in ligand-binding function between a domain of known structure and a prediction target, the structure and binding preference of which are unknown. Our analysis demonstrates that the same binding preferences are not necessarily shared between proteins with the highest overall sequence similarity. Comparison with the experimental data indicates that the developed approach allows one to make accurate predictions on the binding specificity of an SH2 domain of unknown binding function. In fact, in cases where experimental data is available, our prediction agrees with experiment in 23 out of 25 cases [4].

Our analysis indicated the presence of SH2 domains in Janus family of non-receptor protein tyrosine kinases (JAKs). The presence of SH2 domains in this family of kinases is a debated issue in the literature. We thus performed additional analysis, including secondary structure prediction, threading calculations and homology modeling to establish, whether JAKs indeed contain SH2 domains. All analyses performed supported our initial conclusion. An SH2 domain located in one of the JAKs, human Tyk2 kinase, was found to be unique among all analyzed SH2 domains in containing a Histidine instead of Arginine at the key phosphotyrosine binding position PBS, suggesting that the affinity of Tyk2-SH2 for phosphopeptides should depend critically on the pH of the local environment. If confirmed to take place, such recognition would represent a novel mechanism of how phosphorylation of a tyrosine occurring within a particular cellular context may regulate signal transmission. The lack of binding to phosphotyrosine-
containing ligands in a biologically relevant range of conditions, on the other hand, would identify the first example of a protein that apparently adopts an SH2 fold, yet has lost during the course of evolution its ability to associate with phosphorylated targets [5].


[4]. Sheinerman, F. B., Al-Lazikani, B., & Honig, B., Predicting structure and phosphopeptide selectivity of SH2 domains. (manuscript in preparation)


**Presentations based on the research supported by the Sloan fellowship**

