No free energy lunch

Brian K Shoichet

Prediction of a protein's affinity for a ligand has been improved through fundamental physical modeling.

One of the great unmet promises of molecular biology is the calculation of protein-ligand binding affinities. Although tens of thousands of protein structures have been determined to atomic resolution, interpreting these structures to understand their function, and modulate it by designed ligands, remains elusive. In this issue, Lippow *et al.*[1] take a step towards predicting binding affinities using physics-based simulations of protein structures, swelling a small tide of related studies[2–4].

In principle, nothing stands between a protein structure and a calculation of the affinities of its ligands. The fundamental forces driving molecular association—enthalpies and entropies of interaction, desolvation, internal energy and conformational change[5]—are well understood. The energies may be calculated at an atomic level, consonant with the atomic-resolution structures available for proteins, in a series of terms that includes ionic, ion-dipole, dipole-dipole, dipole-induced dipole and instantaneous dipole-induced dipole interactions. Each may be calculated with some accuracy for two molecules interacting with each other and nothing else.

The challenge is that we are concerned not with the association of proteins and ligands in a vacuum but rather with their association in aqueous solution[6], where water competes for groups that form the ligand-protein interface. Almost all the driving forces that we associate with a strong ligand-protein complex—hydrogen bonds, salt bridges, van der Waals complementarity—also drive the association between the molecules and water. Indeed, it is difficult for protein and ligand groups to improve in conjunction what they can achieve with water alone. The energy terms favoring association or solvation are both high in magnitude and almost equal. Thus, calculating the interaction energy between a protein and a ligand involves subtracting two sets of large-magnitude energies to find a small net free energy of binding. Even so, the problem would be soluble if we could calculate each term accurately. But the interaction energies are context dependent[7], and a hydrogen bond in one structural environment, say one exposed to bulk solvent, contributes differently to net interaction energy than the same hydrogen bond in a different structural context, say one buried from bulk solvent (**Fig. 1**).

Lippow *et al.* tackled an especially challenging instance of protein-ligand association, namely, that in which the ligand is itself a protein. One must now model two complicated, highly functionalized surfaces rather than one, when the ligand is a small organic molecule (**Fig. 1**). Using a physics-based energy function, the authors calculated and tested the effects of tens of single and several multi-site residue substitutions on the net free energy of binding of antibodies for protein ligands. These included the affinities of the D44.1 antibody for lysozyme and of the anti-cancer antibody cetuximab (Erbitux) for epidermal growth factor receptor (EGFR). Retrospective calculations were also conducted for the effects of residue substitutions on the association of bevacizumab (Avastin) and the 4-4-20 antibodies with their ligands. Substitutions that amounted to a tenfold (–1.4 kcal/mol at room temperature) increase in affinity between cetuximab and EGFR and to a 140-fold (–2.9 kcal/mol) increase in affinity between lysozyme and D44.1 were correctly predicted. A linear correlation between the predicted electrostatic and the measured binding free energies was observed over multiple residue substitutions. The ability to calculate the effects on affinity of such multiple substitutions makes this a pragmatic approach for identifying residues to modify for improved ligand-protein affinity and, by extension, protein stability or solubility.

Intriguingly, it was the electrostatic component of the interaction and solvation energies that correlated well with experiment, whereas the full energy function, including nonpolar and hydrophobic effects, for instance, did not. This largely reflects the identification of residues in

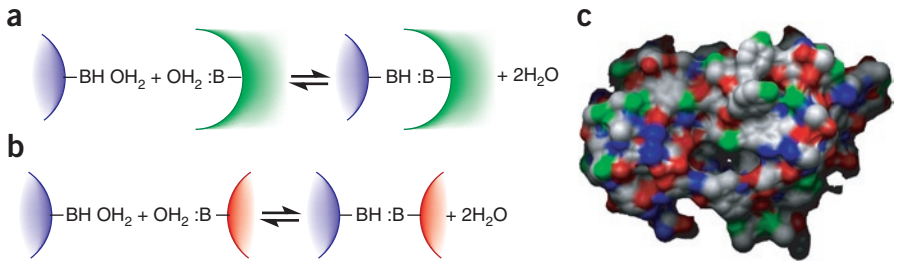

Figure 1 Debit and credit in the hydrogen-bond economy. No net hydrogen bonds are gained in going from the solvated state (left side) to the bound state (right side) for the hydrogen-bond donor (BH) and acceptor (:B) pair. (**a**,**b**) The context of the buried interaction in **a** is more likely to lead to a smaller net interaction energy than that of **b**, where the hydrogen bond is more exposed to bulk water. (**c**) The EGFR binding surface of the Fab fragment of cetuximab[8]. The complicated surface of this interface and the many functional groups involved in it can confound accurate calculation. Apolar regions are in gray, hydrogen-bond donors in blue, hydrogen-bond acceptors in red and hydrogen-bond donors/acceptors in green.

Brian K. Shoichet is in the Dept. of Pharmaceutical Chemistry, University of California, San Francisco, 1700 4th Street, Suite 508D, San Francisco, California 94158-2330, USA.
e-mail: shoichet@cgl.ucsf.edu

the wild-type complexes that, on burial from solvent, pay an electrostatic desolvation penalty that is uncompensated by the interactions that they make in the complex. Such a distinction is something to which the physics-based electrostatic model deployed in the calculations is well suited. Thus, most successful predictions involved the substitution of polar with apolar residues, which will be less penalized for desolvation on complex formation. Where new polar interactions were also introduced, for instance, by replacing a wild-type threonine with an asparate, these occurred on the periphery of the complex where detailed complementarity is less important. A pragmatic lesson may be that one can often improve affinity more by reducing desolvation penalties than by optimizing interaction energies.

By the same standard, the lack of correlation between the full energy function and the experimental interaction energies makes clear that this work is not a general solution to the problem of calculating ligand-protein binding affinities. As the authors acknowledge, the geometries of the complexes were refined by calculation using the full energy function, which was then discarded in favor of the electrostatic terms alone. This cannot be right, and may be expected to fail in many cases. Indeed, whereas single-site substitutions were well predicted in this study, calculations on multi-site substitutions correlated poorly with experiment, as did predictions on the D1.3 antibody–lysozyme complex.

Notwithstanding these caveats, the results reported by Lippow *et al.* are encouraging. Their predictions derive from a general physics-based model that should be widely applicable to the electrostatic component of binding affinity, and the calculations were rapid enough to guide an experimental program of affinity maturation. The problems encountered here may be addressable by greater sampling of the states and configurations open to proteins and their ligands; methods to do so have been developed, although they remain slow enough to prohibit application to a large number of possible ligands. The calculations in Lippow *et al.* balance physical rigor with practicality and may be useful for optimizing other protein-protein complexes.

1. Lippow, S.M., Wittrup, K.D. & Tidor, B. *Nat. Biotechnol.* **25**, 1171–1176 (2007).
2. Chang, C.E. & Gilson, M.K. *J. Am. Chem. Soc.* **126**, 13156–13164 (2004).
3. Woo, H.J. & Roux, B. *Proc. Natl. Acad. Sci. USA* **102**, 6825–6830 (2005).
4. Mobley, D.L. *et al. J. Mol. Biol.* **371**, 1118–1134 (2007).
5. Tirado-Rives, J. & Jorgensen, W.L. *J. Med. Chem.* **49**, 5880–5884 (2006).
6. van Gunsteren, W.F. & Berendsen, H.J.C. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023 (1990).
7. Matthews, B.W. *Annu. Rev. Biochem.* **62**, 139–160 (1993).
8. Li, S. *et al. Cancer cell* **7**, 301–311 (2005).

Network pharmacology

Andrew L Hopkins

Network biology illuminates our understanding of drug action.

The common analogy of drug action is that of a lock and key, with a drug acting as a selective 'key' that fits into the 'lock' of a specific drug target. Over the past two decades, the concept of designing exquisitely selective ligands to avoid unwanted side effects has become the predominant paradigm in drug discovery. However, a growing body of post-genomic biology is revealing a far more complex picture of drug action. An elegant new study by Yıldırım *et al.*[1] in this issue illustrates not only that there are many keys for each lock but also that it is far more common than expected for a single key to fit multiple locks. Viewing drug action through the lens of network biology may provide insights into how we can improve drug discovery for complex diseases.

In recent years it has been appreciated that many effective drugs in therapeutic areas as diverse as oncology, psychiatry and anti-infectives act on multiple rather than single targets[2]—a phenomenon known as 'polypharmacology'[3]. For example, the antibacterial action of β-lactams depends on inhibition of at least two of the multiple penicillin-binding proteins to induce cell death. Similarly, fluoroquinolone antibiotics inhibit both ParC and GyrA. Anti-psychotic drugs commonly exhibit a wide spectrum of activities across entire families of serotonin and dopamine receptors[2]. Extensive studies on several new protein kinase inhibitors, including sunitinib (Sutent) and imatinib (Gleevec), have demonstrated that their anti-cancer effects are most likely due to their action on multiple signaling kinases. Chemogenomics and large-scale ligand profiling have also provided evidence that many drugs act on multiple targets.

Polypharmacology has traditionally been viewed by drug designers as an undesirable property that needs to be removed or reduced to produce 'clean' drugs that act on single targets. The assumption of 'one drug for one target for one disease' has influenced many aspects of drug-discovery strategy, including disease classification, target identification, screening, drug design and clinical-trial design. However, advances in systems biology suggest that complex diseases may not be effectively treatable by interventions at single nodes.

Large-scale gene knock-out experiments in model organisms have shown that biological systems are remarkably resilient to attack and perturbation. The robust phenotypes of biological systems often result from compensatory signaling routes that bypass the inhibition of individual proteins. Network biology theory predicts that modulating multiple nodes simultaneously is often required for modifying phenotypes[4]. Taken together, observations of phenotypic robustness after gene deletion and network biology theory indicate that in several instances exquisitely selective compounds may exhibit a lower-than-desired efficacy for the treatment of disease. Thus, compounds that selectively act on two or more targets of interest in theory should be more efficacious than single-target agents.

Yıldırım *et al.* applied network analysis to drugs and drug targets. Integrating publicly available drug data with genetic-disease associations, gene-expression information and protein-protein interaction data, they investigated the relationships between approved drugs. If drugs acted selectively on single targets, we should not expect a network but rather isolated, bipartite nodes. Although their analysis was limited by the relatively scarce amount of public drug-screening data[5] compared with proprietary screening and literature databases, and by the incomplete mapping of the human protein interactome, the authors nevertheless observed a rich network of polypharmacology interactions between drugs and their targets. Indeed, drugs acting on single targets appear to be the exception.

Yıldırım *et al.*'s findings add to our growing understanding of the role of polypharmacology in drug action. Using network analysis of integrated data sets and a network distance metric, they were able to distinguish between palliative drugs, which relieve symptoms,

Andrew L. Hopkins is in the Division of Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, United Kingdom.
e-mail: a.hopkins@dundee.ac.uk