J Comput Aided Mol Des (2012) 26:13–14
DOI 10.1007/s10822-011-9518-x

PERSPECTIVE

A very short history of structure-based design: how did we get here and where do we need to go?

Jeff Blaney

Received: 28 November 2011 / Accepted: 5 December 2011 / Published online: 11 December 2011

Beddell et al. [1] published the first structure-based design paper in 1976 on hemoglobin ligands using Kendrew wireframe models.The first protein–ligand docking paper was in 1982 [2]. The PDB contained about 200 protein X-ray crystal structures in 1982, but very few were drug discovery targets. Many medicinal chemists believed that relevant macromolecule crystal structures were not likely to ever become available early enough during the lifetime of a drug discovery project to matter. Many also thought that designing an optimal ligand would be trivially obvious if they were lucky enough to have the high-resolution structure of their target.

The initial flurry of enthusiasm in the 80s around docking and de novo design yielded to the realization in the 90s that scoring is the hard problem, not design. It's relatively easy to design a molecule to optimize any given scoring function; the hard part is recognizing which apparently complementary molecules actually will bind to their target. This is the crux of the still-unsolved scoring problem. Predicting free energies of binding in aqueous solution has proven to be far more difficult than most of us realized. Despite 30 years of effort with many different approaches, many different investigators, and massive amounts of computer time, we still cannot reliably predict relative binding free energies with sufficient accuracy to drive organic synthesis during hit-to-lead optimization.

CASP (Critical Assessment of Protein Structure, http://predictioncenter.org) showed the protein folders in 1994 that true, blind prediction is much harder than retrospective "prediction" [3]. Their field has advanced steadily since then. Our own field just began to learn this recently, thanks to the CCDC (Cambridge Crystallographic Data Centre) and SAMPL (http://sampl.eyesopen.com) "contests", plus other critical analyses (e.g., [4]). Blind predictions for small molecule crystal structures began with the CCDC in 1999 [5]. Anthony Nicholls, Vijay Pande, and others started SAMPL0 in 2007 to predict small molecule solvation free energies. SAMPL3 was held this year and included blind prediction of host–guest complexes and trypsin-fragment binding. The initial CCDC contest results were disappointing, but helped stimulate dramatic progress. The most recent CCDC contests proved that high-resolution, accurate, blind prediction of small molecule crystal structures is possible for some small organic molecules, albeit with heroic amounts of computing [6, 7]. This is really exciting work: it proves that current theory and its implementation are sufficient to solve this problem in several non-trivial cases. SAMPL0–3 showed that true blind predictions of solvation free energies, host–guest complexes, protein-fragment structures and relative binding affinities are still remarkably difficult and typically have surprisingly large errors [8–10].

Literature in our field is still cluttered with work claiming to predict what's already known. It is not even clear where the major errors are (partial charges, dielectric model, fixed versus polarizable force fields, torsion terms, entropy changes due to torsional degrees of freedom, water model, etc.). We still don't understand why protein–ligand binding free energy doesn't continue to increase with increasing size of the binding interface [11]. Brute force computing has not helped. Major improvements in our methods are required.

The PDB now has over 75,000 structures, of which about 50,000 are protein–ligand complexes (http://www.pdb.org). High-resolution X-ray crystal structures are now routinely

J. Blaney (✉)
Computational Chemistry and Cheminformatics, Small Molecule Drug Discovery, Genentech, 1 DNA Way, South San Francisco, CA 94080, USA
e-mail: blaney.jeff@gene.com

available for most drug discovery targets, except for membrane-bound proteins. Even they are yielding and becoming tractable in many cases, though still very challenging. The noncovalent interactions observed within these structures provide a rich database to mine for qualitative structure-based design. Sadly, high-resolution X-ray crystal structures don't identify the strength of each interaction! Our ability to predict which molecules to synthesize next remains empirical, guided by these observations of preferred noncovalent interactions and supplemented with qualitative docking and other calculations [12].

Experienced medicinal chemists frequently argue that finding hits and optimizing them into potent and selective leads is easy compared with the later stage challenges faced in drug discovery. Maybe, but consider how much faster we could advance more series through those difficult later stages if we could reliably predict relative binding free energies within even a factor of 100 (<2 kcal/mol). This would allow us to prioritize which series to focus on and to identify those that are likely to hit an affinity or selectivity "wall". Major improvements in quantitative structure-based design will have a dramatic impact on our ability to advance drug candidates to the clinic.

The best design engine and scoring function is an expert human: we need fast, interactive, easily used tools to help us design compounds and test hypotheses. We have many useful tools, but they are seldom appropriately integrated: structure–activity relationships, physicochemical properties, DMPK, and safety data need to be tightly integrated with molecular modeling and design software. 2D and 3D structure searching of the CCDC, PDB, in-house, commercially available, synthetically accessible, and patented compounds also need to be tightly coupled.

Predicting solubility also remains an incredibly difficult, fundamental physical chemical problem. It is also critically important: poorly soluble compounds frequently slow series progression, often killing a series, and occasionally killing an entire project. Major advances here will require reliable, practical methods for calculating solvation free energy and small molecule crystal structures.

Which problems are likely to succumb to routine practice within the next 25 years? Accurate solubility prediction, DMPK and permeability prediction, automatic connection of patent and journal literature to current structures of interest, accurate 3D structures available for almost all drug discovery targets, prediction of relative binding free energies to within 2 logs, qualitative understanding of signal-transduction networks, delivery of large molecules across cell membranes and the blood–brain barrier.

Which problems are unlikely to succumb? Accurate prediction of absolute binding free energies, cost of custom synthesis, discovering small molecule drugs that inhibit protein–protein interactions (except for a small number of special cases).

What's needed for significant progress? Research in our area tends to be driven more by the investigator's interest and knowledge in a particular computational approach rather than a careful assessment of which approaches are likely to have the most impact. Our field needs to focus more effort on well-designed blind "contests" to better understand where the limitations are, prioritize improvements, and focus research in those areas. Blind prediction contests for ligand binding in aqueous solution (both pose and binding free energy), solvation free energies, and pKa predictions for protein sidechains and bound ligands will provide critical guidance for future improvements.

References

1. Beddell CR, Goodford PJ, Norrington FE, Wilkinson S, Wootton R (1976) Compounds designed to fit a site of known structure in human haemoglobin. Br J Pharmacol 57:201–209
2. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982) A geometric approach to macromolecule-ligand interactions. J Mol Biol 161:269–288
3. Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. Proteins 23, ii–v (1995)
4. Warren GL et al (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931
5. Lommerse JP et al (2000) A test of crystal structure prediction of small organic molecules. Acta Crystallogr B 56:697–714
6. Day GM et al (2009) Significant progress in predicting the crystal structures of small organic molecules—a report on the fourth blind test. Acta Cryst B65:107–125
7. Kazantsev AV et al. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. Intern J Pharm (2011)
8. Nicholls A, Mobley David L, Guthrie Peter J, Chodera John D, Bayly Christopher I, Cooper Matthew I, Pande Vijay S (2008) Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. J Med Chem 51:769–779
9. Nicholls A, Wlodek S, Grant JA (2009) The SAMPL1 solvation challenge: further lessons regarding the pitfalls of parametrization. J Phys Chem B 113:4521–4532
10. Skillman AG, Geballe MT, Nicholls A (2010) SAMPL2 challenge: prediction of solvation energies and tautomer ratios. J Comput Aided Mol Des 24:257–258
11. Kuntz I, Chen K, Sharp K, Kollman P. The maximal affinity of ligands. Proc Natl Acad Sci USA (1999)
12. Bissantz C, Kuhn B, Stahl M. A medicinal chemist's guide to molecular interactions. J Med Chem (2010)