Chemistry & Biology, Vol. 11, March, 2004, DOI 10.1016/j.chembiol.2004.03.002

The Modular Approach to Ligand Discovery

Identifying specific protein-ligand interactions is a long-standing problem in drug discovery and chemical biology, which is only exacerbated by the abundance of uncharacterized proteins revealed by genomics. Last month in *Chemistry Biology*, Sem et al. described a powerful technique for rapidly screening protein families for ligands [1].

Molecular targeting of specific proteins by small molecules lies at the very heart of the interface between chemistry and biology. If the protein in question is a validated therapeutic target, we call this process drug discovery. In the case of proteins of unknown function, finding a pharmacological tool or probe to investigate the target can be an excellent first step toward deciphering protein function in a larger, biological context. More often, small molecule tools provide orthogonal approaches to compliment the more biological methods for elucidating protein function (e.g., siRNA) [2]. The difficulty, of course, is faced in finding these tools in a rapid and efficient fashion.

Traditionally, the way researchers have tackled this problem has been to perform large, high throughput screens of structurally varied collections of small molecules to look for an effect on the protein of interest. Unfortunately, this strategy is resource intensive and relies a bit on serendipity; in short, it is neither rapid nor efficient. If detailed structural information of the protein target is available (e.g., X-ray structure or NMR structure), then more direct structure-based methods can be used. However, resolving the 3D structure of a protein is not a trivial matter; not all proteins crystallize and NMR approaches are size limited [3]. In last month's issue of *Chemistry & Biology*, Sem and colleagues demonstrated a modular, structure-guided approach for rapidly identifying new ligands for a family of proteins [1].

The modular nature of the approach stems from the fact that protein function and thus structure has been conserved evolutionarily. Presently, there are approximately 600 known protein structural domains and estimates based on genome sequencing projects predict that this number could rise to 8000 distinct folds [4]. By using protein domains as the fundamental unit of protein structure, the rationale employed by the Sem group argues that similar domains should have a propensity to bind similar pharmacophores, offering a shortcut to isolate ligand classes most likely to be useful across protein families. In this study, they chose the oxidoreductase family of proteins to test this hypothesis.

The process begins by distilling information from bioinformatic analysis of protein sequence databases to identify members of the oxidoreductase family of proteins, a nicotinamide adenine dinucleotide (NAD)-utilizing enzyme involved in primary and intermediary metabolism. After determining that approximately 4.5% of all the sequences found in the Swiss-Prot protein database are NAD(P)-utilizing enzymes, a cluster analysis on 288 oxidoreductase crystal structures helped the authors to further classify the oxidoreductases into pharmacofamilies [5]. Pharmacofamilies are defined as proteins related by sequence, NAD cofactor geometry, and protein fold. Interestingly, the two largest families only differ by the orientation of the nicotinamide ring relative to the ribose with antiorientation predominating (cofactor geometry, vide supra).

With the pharmacofamilies identified and by using the NAD cofactor as a starting point, a pharmacophore map of the cofactor is used to create a model to help identify commercially available reagents capable of binding to the enzyme. Competitive binding studies identified several common ligand mimics (CLM) that bind within the cofactor site of dihydrodipicolinate reductase (DHPR), an oxidoreductase essential for cell wall synthesis in *Mycobacterium tuberculosis*. Further screening of the CLMs against members of the same pharmacofamily revealed one compound that possessed crossreactivity across the family, in essence identifying a privileged ligand for this class of oxidoreductase enzymes.

This is where the structure-guided approach becomes important. By using the cofactor as a reference ligand, Sem's previously reported technique, NMR SOLVE (structurally oriented library valency engineering), is used to map key amino acid residues of the enzyme involved in cofactor binding [6]. Once the key sites are identified by changes in chemical shift, the residues can be used as reference points to determine the binding *and* orientation of the CLMs. The beauty of this method lies in the fact that no 3D structure of the overall enzyme is needed; only information illustrating key peaks at sites critical for small molecule binding is necessary. Furthermore, NMR is sensitive to changes in chemical shift, allowing detection of weak binders that could easily be missed by functional assays. In this study, Sem and colleagues focus on DHPR, a homotetrameric complex of ~170 kDa, which is far too large for resolving complete 3D structure by using current NMR techniques!

By using the "privileged" CLM as an anchor point, carefully chosen chemical building blocks are chemically ligated to the CLM core with the hope of identifying molecules that offer specificity within the pharmacofamily. This takes advantage of the fact that the cofactor binding site is adjacent to a specificity site usually occupied by the metabolite upon which the enzyme works. Any building blocks with affinity for the adjacent binding site should dramatically increase the overall affinity of the new ligand as a result of the "chelate effect," i.e., a reduced entropic penalty upon binding of a multidentate ligand [7]. In the case of DHPR, the current study revealed a novel structure with an affinity greater than two orders of magnitude better than the original CLM. Furthermore, this small molecule has a dramatically lower affinity for the close oxidoreductase relatives lac-

tate dehydrogenase (LDH) and 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DOXPR).

By using this technique as a foundation, it is interesting to speculate on what other classes of small molecule/protein pairs might be useful in this context. Without the advantage of an adjacent specificity site, is it possible to add new building blocks to privileged anchor molecules that broadly bind to other pharmacofamilies? One interesting test case might be the ATP-competitive small molecule genistein, which binds nonspecifically to many receptor tyrosine kinases (RTK), an active area of pharmaceutical research [8]. Using expressed kinase domains and genistein as a starting point might deliver interesting and novel chemotypes for this important family of proteins.

The ongoing effort to understand the correlation of protein sequence and function will only make this technique more valuable. It is not too much to think that in the future, starting from gene sequence, a multidomain protein structure will be predicted and all of the domains will be of known types. If ligands for any of the domains are known, they could serve as useful starting points for using the Sem approach to identifying new small molecules or drug leads.

Charles Karan and John A. Tallarico
Harvard Institute of Chemistry and Cell Biology
Harvard Medical School
250 Longwood Avenue, Room 622
Boston, Massachusetts 02115

Selected Reading

1. Sem, D.S., Bertolaet, B., Baker, B., Chang, E., Costache, A.D., Coutts, S., Dong, Q., Hansen, M., Hong, V., Huang, X., et al. (2004). Chem. Biol. *11*, 185–194.
2. Hardy, L.W., and Peet, N.P. (2004). Drug Discov. Today *9*, 117–126.
3. Waszkowycz, B. (2002). Curr. Opin. Drug Discov. Dev. *5*, 407–413.
4. Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). J. Mol. Biol. *247*, 536–540.
5. Kho, R., Newman, J.V., Jack, R.M., Villar, H.O., and Hansen, M.R. (2003). J. Proteome Res. *2*, 626–632.
6. Pellecchia, M., Meininger, D., Dong, Q., Chang, E., Jack, R., and Sem, D.S. (2002). J. Biomol. NMR *22*, 165–173.
7. Page, M.I., and Jencks, W.P. (1971). Proc. Natl. Acad. Sci. USA *97*, 9367–9372.
8. Kissau, L., Stahl, P., Mazitschek, R., Giannis, A., and Waldmann, H. (2003). J. Med. Chem. *46*, 2917–2931.