A Systematic Approach to Identifying Protein-Ligand Binding Profiles on a Proteome Scale

Li Xie[1], Nancy Buchmeier[2], Lei Xie[3]*, Philip E. Bourne[1,3]

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, [2]Department of Chemistry and Biochemistry and [3]San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

*To whom correspondence should be addressed: lxie@sdsc.edu

INTRODUCTION

Identification of protein-ligand interaction networks on a proteome scale is crucial to address a wide range of biological problems such as correlating molecular functions to physiological processes and designing safe and efficient therapeutics. We have developed a novel computational strategy to identify ligand binding profiles of proteins across gene families and applied it to predicting protein functions, elucidating molecular mechanisms of drug adverse effects, and repositioning safe pharmaceuticals to treat different diseases.

METHODS

Our cross-gene-family approach proceeds as follows:

1) The ligand binding site of the primary target is extracted or predicated from a 3D experimental structure or homology model of proteins and characterized by a geometric potential (1).
2) Off-target proteins with similar ligand binding sites to the primary target are identified across human structural genomes based on a new algorithm for fast and accurate estimation of statistics significance of Sequence Order Independent Profile-Profile Alignment (SOIPPA) (2).
3) The atomic details of interactions between the drugs and the putative off-targets from the step 2 are characterized using protein-ligand docking methods. The high ranking off-targets are further investigated using a normalized docking score.
4) The identified panel of off-targets is subject to structural and functional cluster analysis and connected into a network including multiple metabolic, signal transduction, and gene regulation pathways. Combinatorial controls of biological process by the off-target are analyzed based the above established network.

RESULTS

The above strategy has been successfully applied to reveal the molecular mechanism for the adverse drug effects of selective estrogen receptor modulators (SERMs) from experimental structures (3). Moreover, we have discovered that the marketed pharmaceuticals Entacapone (ENT) and Tolcapone (TOL) used for treating Parkinson's disease can potentially be repositioned to treat multi drug and extensively drug resistance tuberculosis (TB). The predication is based on the established evolutionary relationship between human SAM methyltransferases (including COMT) and *M. tuberculosis* enoyl-ACP reductase (InhA), both Rossmann fold proteins and the latter a major target of TB drugs (2). The inhibition by ENT and TOL of *M. tuberculosis* growth has been validated by microplate assay (Figure 1).

Appling our approach on a proteome scale we extended the methodology to include homology models as target receptors and identified a panel of off-targets of cholesteryl ester transfer protein (CETP) inhibitors (Table 1). The CETP inhibitor is developed as a new preventive therapy for cardiovascular disease by raising HDL cholesterol. Clinical studies have indicated that one of the CETP inhibitors, Torcetrapib, has deadly off-target effect that result in the excess induction of hypertension (4). Consequently, it was withdrawn from the phase III clinical trial. In contrast with Torcetrapib, another CETP inhibitor JTT-705 does not have unwanted side-effect that increases the blood pressure. The identified multiple off-targets of CETP inhibitors from our studies are involved in both positive and negative feedback controls of stress regulations and immune response through an interconnected metabolic, signal transduction, and gene regulation networks (Figure 2). Our predictions are strongly correlated to the clinical and *in vitro* observations, providing a molecular explanation on the difference in the side effect profile of the CETP

inhibitors. The finding further suggests that adverse drug reactions can be modulated by the fine-tuning of the off-target binding network.

CONCLUSION

In all of above cases, most of the identified off-targets belong to different protein superfamilies from the primary target. It indicates that complex protein-ligand interaction networks play key roles in physiological and pathological processes. Thus, a computational chemical genomics approach that systematically uses small molecules to probe biological systems will provide us with valuable clues as to the molecular basis of cellular functions and at the same time shift the conventional one-target-one-drug drug discovery process to a new multi-target-multi-drug paradigm.

FIGURES

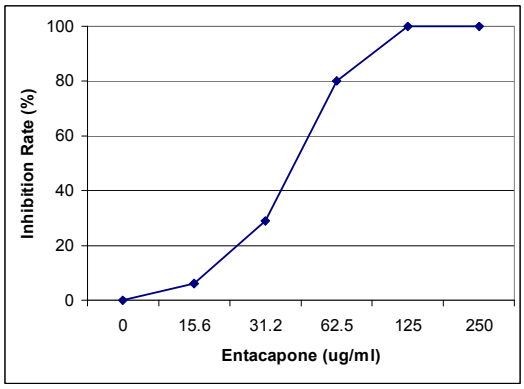

Figure 1. Inhibition rate of *M. tuberculosis* by Entacapone (ENT) from microplate assay.

The drugs ENT and TOL's primary target catechol-*O*-methyltransferase (COMT) belongs to a large superfamily of SAM methyltransferases. The drugs are used as adjuncts to treat Parkinson's disease by increasing the bioavailability of the primary drug levodopa. ENT and TOL may inhibit the *M.tuberculosis* InhA protein directly – a different mechanism from the first- and second-line drugs that result in MDR and XDR strains. ENT and TOL have excellent safety profiles with few side effects.

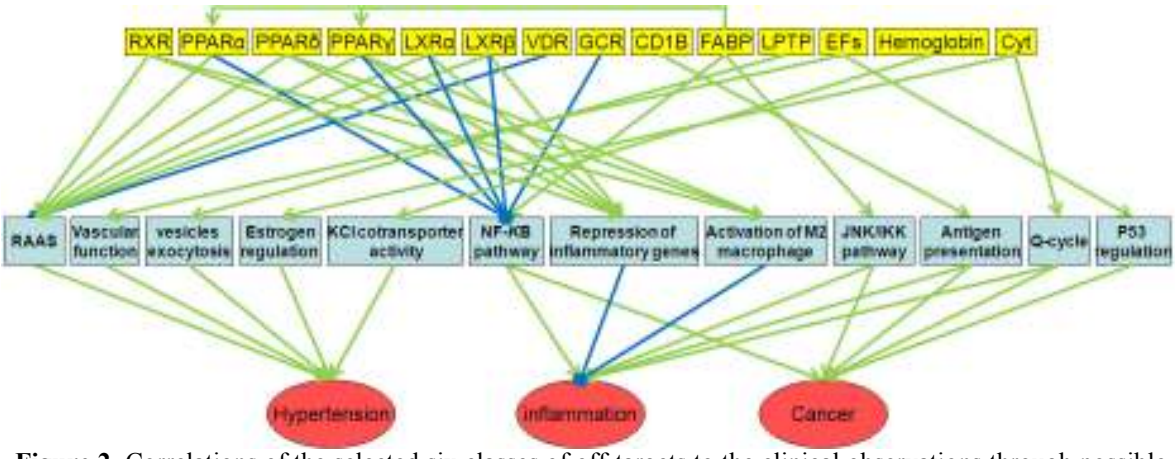

Figure 2. Correlations of the selected six classes of off-targets to the clinical observations through possible biological pathways. Green and blue lines indicate positive and negative regulation, respectively.

TABLES

Table 1. Binding site volumes and normalized docking scores (NDS) of CETP inhibitors for CETP and six classes of putative off-targets. The predicated binding affinities are colored red (strong, NDS< -2.0), purple (relatively strong, 0.0>NDS> -2.0), and blue (weak, NDS>0.0), respectively.

Protein	**PDB ID**	**Binding Site Volume (Å^3)**	**Normalized Docking Score**		
			Torcetrapib	*anacetrapib*	*JTT705*
CETP	2OBD	1084.2	**-5.6024**	**-4.6705**	**-1.9644**
Retinoid X receptor (agonist)	1YOW	1420.5	**-5.5803**	**-4.1922**	**-0.9344**
PPARδ (agonist)	1Y0S	1313.2	**-3.8703**	**-3.8384**	**-1.5662**
PPARα (agonist)	2P54	1059.4	**-4.0828**	**6.6785**	**-3.0660**
PPARγ (agonist)	1ZEO	726.5	**-3.9838**	**6.0096**	**-2.0316**
LXRα (agonist)	2ACL	1155.0	**5.7793**	**6.3052**	**-0.6900**
LXRβ (agonist)	1UPV	1553.5	**5.0882**	**5.5450**	**-1.7543**
Vitamin D receptor (agonist)	1IE8	879.7	**5.7622**	**6.1759**	**-1.1761**
Glucocorticoid receptor (agonist)	1P93	819.0	**5.5504**	**6.1432**	**-2.0131**
Glucocorticoid receptor (antagonist)	1NHZ	990.5	**-2.1235**	**-3.2125**	**-1.1673**
Glycolipid transfer protein	1TFJ	987.4	**-0.9839**	**-2.1587**	**-1.3249**
Phosphatidylcholine transfer protein	1LN1	1860.1	**-7.3050**	**-9.1032**	**-1.0794**
Phosphatidylinositol transfer protein	2A1L	2271.7	**-4.0881**	**-6.0708**	**-1.7366**
GM-2 activator	2AG9	955.0	**-4.0254**	**-3.8265**	**-3.6934**
Fatty acid binding protein	2NNQ	743.3	**3.3521**	**6.8334**	**-2.3356**
T-Cell CD1B receptor	1GZP	1056.4	**-2.0899**	**-6.1531**	**-1.3424**
EF hand-like	1DTL	1992.0	**-3.7771**	**-3.6403**	**-2.9955**
Cytochrome complex	1PP9	2963.0	**-3.8702**	**-7.0825**	**-2.2745**
Human cytoglobin	1V5H	1022.2	**-3.4827**	**-1.8246**	**-2.3848**

REFERENCES

1. Xie, L., Bourne, P.E. 2007. A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. *BMC Bioinformatics* 8: S9.
2. Xie, L., Bourne, P.E. 2008. Detecting evolutionary relationships across fold and functional space with sequence order independent profile-profile alignments. *Proc Natl Acad Sci USA* Accepted.
3. Xie, L., Wang, J., Bourne, P.E. 2007. In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators. *PLoS Comp Biol* 3: e217.
4. Cutler, D.M. 2007. The demise of the blockbuster? *The New England Journal of Medicine*. 356: 1292–1293.