Published in final edited form as:

Curr Opin Struct Biol. 2011 April ; 21(2): 189–199. doi:10.1016/j.sbi.2011.01.004.

Structure-based Systems Biology for Analyzing Off-target Binding

Lei Xie^{1,2,*}, Li Xie¹, and Philip E. Bourne^{1,*}

¹ Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego MC9743, 9500 Gilman Drive, La Jolla, CA 92093, USA

² Department of Computer Science, Hunter College, the City University of New York, 695 Park Avenue, New York City, NY 10065, USA

Abstract

Here off-target binding implies the binding of a small molecule of therapeutic interest to a protein target other than the primary target for which it was intended. Increasingly such off-targeting appears to be the norm rather than the exception, rational drug design notwithstanding, and can lead to detrimental side-effects, or opportunities to reposition a therapeutic agent to treat a different condition. Not surprisingly, there is significant interest in determining *a priori* what off-targets exist on a proteome-wide scale. Beyond determining putative off-targets is the need to understand the impact of such binding on the complete biological system, with the ultimate goal of being able to predict the phenotypic outcome. While a very ambitious goal, some progress is being made.

Introduction

The availability of fully sequenced bacterial genomes and then the human genome were expected to be the forerunner of a new era in drug discovery. As yet that era has not eventuated. What has become apparent is the problem is much more complicated than simply identifying new targets and designing a drug to affect an individual target in some way. Ehrlich's magic bullet of one drug, one target to treat one condition is likely the exception not the rule, which would explain why the attrition rate at the late stage of clinical trials is typically as high as 90% as a result of the lack of both efficacy and clinical safety [1]. These failures means a single new drug typically costs more than US\$800 million and may take 15 to 20 years to develop [2].

While genomics, and as a consequence the emergent field of systems biology, might not have provided the simple solution pharmaceutical scientists were seeking, it has illustrated the complexity of the problem and may point the way to a new approach to drug discovery. Simply stated, the disappointing outcome of current approaches to drug discovery emphasizes the need to assess and possibly employ a new polypharmacology paradigm

*To whom correspondence may be addressed: Lei Xie, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, MC0743, 9500 Gilman Drive, La Jolla, CA 92093, USA, Tel: (858)822-3686, lxie@ucsd.edu, Fax: (858)822-0873, Philip E. Bourne, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, MC0743, 9500 Gilman Drive, La Jolla, CA 92093, USA, Tel: (858)534-8301, pbourne@ucsd.edu, Fax: (858)822-0873.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

[3,4]. Polypharmacology focuses on searching for multi-target drugs to perturb disease-associated networks rather than designing selective ligands to target individual proteins. The efficacy of a multi-target therapy is supported by observations of the robustness and resilience of complex biological systems and hence failure of single target approaches. In a biological network, effecting multiple nodes is more likely to cause the system to fail than the removal of a single node; a result of diversity and redundancy in the biological system. From a causal perspective, multiple node failures have been called “fail-on” [3], and used to explain neurological disorders [5] and cancer [6,7] in recent genome-wide studies. Thus, in principle, from an effect perspective, multi-target therapeutics can exhibit greater efficacy and be less vulnerable to drug resistance by impacting multiple nodes at the system level.

Indeed, a large number of existing anti-bacterial, anti-viral and anti-cancer therapeutics are multi-target agents developed through drug combination (so-called cocktails) (for a detailed review, see ref. [8]) or discovered serendipitously but not rationally for a single molecule, for example, imatinib [9,10] and lipitor [11]. Although a multi-target therapeutic may enhance clinical efficacy, there is also the increased possibility of side-effects from a compound with promiscuous binding properties. Thus, it is a great challenge to rationally design appropriate effective multi-target therapies. To understand the impact of a ligand in targeting multiple proteins in a biological system we have to answer the following questions: What are the common features of the protein-ligand interaction that persists across gene families [12]? Hence, which specific proteins are inhibited/activated by the ligand? How are these proteins connected in biological pathway(s)? Given these pathways, how does inhibition/activation of these proteins affect the overall physiological process? Further, questions arise when considering an individual response to a given therapeutic or in the treatment of a fast evolving pathogen. Namely, how does single or multi-amino acid mutations alter ligand binding and consequently the physiological effect? Weighty questions indeed, but questions that are actively being addressed.

Polypharmacological drug design can be reformulated as a proteome wide off-target identification problem. Recent work on large-scale mapping of polypharmacology interactions by Paolini *et al.* revealed the extent of drug promiscuity in proteome-wide binding [13]. It was estimated that each existing drug binds to, on average, 6.3 protein receptors [14]. Identification of these off-targets will provide the molecular basis for a new kind of therapy as already indicated, but can also lead to a better understanding of potential drug side-effects, suggest drug repurposing in the treatment of different conditions than originally intended and suggest prioritization of multiple targets for polypharmacological drug design. Experimental identification of off-targets on a proteome-wide scale is still in its early stage, but *in silico* approaches can be used to guide a limited number of subsequent experiments. A number of computational methodologies have been developed where relate putative protein targets through their ligand chemistry [15,16] or biological profiles (e.g., binding activities [17], expression profiles [18,19], side-effects [20]). Application of these methods on a proteome-wide scale requires we further integrate methods for homology detection, structural bioinformatics analysis, protein-ligand docking, accurate free energy calculations and biological network reconstruction and simulation. This paper describes recent advances and results from the integration of such approaches for structural proteome-wide off-target prediction.

2. Existing Proteome and Ligand Coverage

Success of proteome-wide prediction of off-targets depends on the availability of sequences and structures for the whole proteome and ligands known to bind to at least one target. As of August 2010 there were 1,000 prokaryotic and more than 100 eukaryotic genomes, including human completely sequenced (<http://www.ncbi.nlm.nih.gov>). For a large number

of the associated encoded proteins their functions, co-factors, and substrates may be inferred from these protein sequences. However, 3-dimensional (3D) structure information is required to investigate the atomic details of protein-ligand interactions. Fortunately, in parallel with sequence growth, the structures deposited in the Protein Data Bank [21] have also been increasing at a very fast rate. For example, more than 6,000 unique experimentally determined human structures are currently in the PDB, covering around 30% of known functional classes [22]. The structural coverage of proteomes can be further increased using structure prediction and homology modeling techniques as embodied in software such as I-TASSER[23]. Reliable homology models built from known PDB structures using these techniques increase the human proteome structural coverage to about 50% [24]. Advances in structural genomics [25] and subsequent homology modeling [26] will continue to improve this coverage.

Currently the known drug target space (the part of the proteome to which known drugs bind) covers approximately 5% of the human proteome (red circle in Fig. 1). Drug-target coverage for pathogens (e.g., *M. tuberculosis* and *T. brucei*) is even more limited, with coverage of 0.01% to 0.2%. At the same time the overall structural coverage of these proteomes is relatively good. The poor proteome coverage by pharmaceutically investigated drug targets limits the application of ligand-based and activity-based methods. Conversely, the relatively good structural coverage provides opportunities for target-based methods for determining off-targets.

Finally, there are abundant structures of drug-protein complexes in the PDB. As shown in Figure 2, among 6,127 drugs or drug candidates in the SuperTarget database [27], around 15% and 60% of them have molecular fingerprint similarity scores larger than 90 and 70, respectively, to ligands deposited in PDB. Comprehensive coverage by structures and ligands provides an opportunity to apply structural proteome-wide methods to predict off-target binding.

3. Structural bioinformatics approaches for off-target prediction

In nature chemical space is infinite, yet only a small part of that space is used by endogenous ligands in binding to proteins. These endogenous ligands and associated protein sequences and structures have co-evolved [28,29] and thus sequence and structure similarities, both global and local to the binding site, can be used to infer binding to other targets by the same or similar ligand.

Tremendous efforts have been made in detecting global sequence and structure similar proteins to infer their functional relationship [30]. Consider examples relevant to drug discovery. The selection of anti-infectious drug targets in pathogens is focused on those targets that are essential for the growth and survival of the organism, and which are frequently mapped from the orthologous genes in model organisms [31]. Unwanted side-effects can be avoided by selecting only that subset of targets not having orthologs in the human genome. Moreover, clustering proteins into target families has been applied extensively in modern drug discovery [32] to facilitate lead discovery [33]. It is well known that protein structure is more conserved than sequence over large evolutionary time scales. Thus global structure comparison may detect functional relationships that, in many cases, imply similar ligand binding propensities. Protein structure similarity has been successfully applied in guiding combinatorial library design [34]. Existing drugs have been redesigned to bind off-targets that have similar structures to the primary targets. Recent successful stories include the discovery of nonsteroidal antagonists against the human androgen receptor [35] and the repurpose of cholesterol drugs to block *Staphylococcus aureus* virulence [36]. However, unlike the “twilight zone” that distinguishes homologous protein pairs from non-

homologous based on sequence alignment, there is not a clearly defined threshold for structural comparison to confidently determine the functional relationship between two structures. More confounding still, is that similar structures may be responsible for diverse biological functions. Recent efforts to combine sequence and structure features to improve the prediction of protein function [37], especially ligand binding cross-reactivity [38] have shown some success.

Increasing evidence suggests that cross-reactivity of ligands with proteins occurs beyond global sequence and structure homologs [12]. Evolutionary and functional linkage between proteins can exist across fold space at the level of functional sites [39–41]. It is expected that functional site comparison will detect more off-targets than sequence and structure similarity alone. By searching for similar ligand binding sites, it is possible to design novel lead compounds for the target of interest. Klebe *et al.* discovered two new privileged templates for HIV protease inhibitors based on a similarity between the unoccupied binding site region in HIV protease and an adenine binding pocket [42]. They also computationally predicted and experimentally validated that COX-2 specific inhibitors demonstrate nanomolar affinity to the totally unrelated carbonic anhydrase (CA) family due to their similar binding sites [43]. More recently, several pharmaceutically interesting off-targets have been predicted based on ligand binding site similarity, and subsequently validated by experiments.

Applying a pocket detection and comparison algorithm to model structures, it has been found that human antiviral APOBEC3 protein specifically binds 5.8S RNA [44]. Evolutionary linkage has been established among the ATP binding sites of protein kinases, ATP grasp, and SAICA synthetase-like superfamilies [39]. Independently Miller *et al.* discovered that protein kinase inhibitors can be repurposed to target the biotin carboxylase subunit of acetyl-CoA carboxylase, which is a member of the ATP-grasp superfamily [45]. In another study, the pharmaceutical Comtan (entacapone), which is a COMT methyltransferase inhibitor used to treat Parkinson's disease, has shown direct inhibition of InhA, one of the most used drug targets against tuberculosis [46]. In yet another study, Durrant *et al.* has identified a panel of human and *T. brucei* off-targets of NSC45208 [47], a novel drug lead to inhibit RNA-editing ligase [48]. *T. brucei* is the causative pathogen in the neglected, but deadly, African sleeping sickness. The identified off-targets have implications for developing new multi-target drugs with both the desired efficacy and safety profile to treat African sleeping sickness. Using a novel ligand binding site comparison method, Milletti *et al.* are able to predict the selectivity profile of kinase inhibitors and to detect sub-pocket similarities among proteins with different folds [49].

These success stories highlight the need for accurate ligand binding site prediction and comparison. A number of approaches have been developed for determining ligand binding pockets as summarized in recent reviews [50–52]. Notably, Skolnick and colleagues have developed a series of methods to extend ligand binding site characterization to predicted structures, including FINDSITE [37], FINDSITE^{LHM} [53], and Q-Dock^{LHM} [54]. FINDSITE identifies consensus binding residues from superimposed groups of threading templates, selects representative molecules bound to a particular binding site for use in ligand-based virtual screening and predicts molecular functions from the top-ranked templates. In FINDSITE^{LHM}, anchor substructures are detected from clusters of template-bound ligands that occupy top-ranked binding pockets predicted by FINDSITE and are used as references in virtual screen docking. The anchor-binding mode predicted by FINDSITE^{LHM} is further incorporated into the Q-Dock force field as position-specific anchor restraints to improve the sampling of native-like conformations. The integration of these techniques contributes to the success in the virtual profiling of cross-reactivity in the human kinome by X-React^{KIN} [38].

Several recent reviews have discussed the challenges and advances in ligand binding site comparison [52,55]. One aspect that has yet to be fully addressed is that most of the existing methods rely on knowing the location and boundary of predefined binding pockets either from a co-crystallized complex structure or computational prediction. Then the similarity between two pockets is typically measured by their geometric fitness or ratio of overlapping residues. However, there are no biologically meaningful criteria to clearly define the boundary of the binding pocket. Errors in the binding pocket definition may propagate to the subsequent similarity measure. Establishing similar ligand binding sites is further complicated by the intrinsic plasticity of such sites in proteins. For example, proteins may experience large conformation changes ($\text{RMSD} > 2.0 \text{ \AA}$) in the ligand binding site upon ligand binding [56]. To detect similar regions between two proteins, the ligand binding site comparison method needs to resemble a local sequence alignment where the similar sequence motif is not necessary predefined but can be directly detected. Moreover, the similarity measure requires the quantitative assessment of residue conservation or physiochemical compatibility beyond the geometric match, which alone may lead to high false positive rates or miss underlying functional relationships. Several algorithms fall into this category [39,49,57–59]. Another issue that affects the performance of functional site prediction is that the oligomeric state of potential off-targets may be unknown even if a single protein chain is deposited in the PDB, or can be predicted from homology modeling. This is important since there are a number of cases where the drug binds at the interface between multiple protein chains.

4. Molecular modeling of off-target binding on a structural proteome-wide scale

Although structurally, functionally and/or evolutionarily related proteins have the propensity to bind the same ligand, binding depends on the chemical nature of the ligand, as demonstrated in Figure 3. Further, theoretical and experimental work has suggested that even weak bindings to multiple targets may have profound impact on the biological system [60–62]. As such the physiological consequences of off-target binding may be determined by the relative difference in ligand binding affinity rather than the absolute value of binding of the primary and any off-targets [63–65]. All this suggests that a quantitative investigation of protein-ligand interactions is critical to understanding the phenotypic response to off-target binding.

The available computational tools used to study quantitative protein-ligand interactions are mostly based on protein-ligand docking and free energy calculations for the protein-ligand complexes [66,67]. Existing protein-ligand docking software is not only a notoriously weak predictor of binding affinity, but also designed to screen multiple small molecules against a single receptor. To apply docking to off-target prediction, the process is reversed, i.e., docking a single molecule to multiple receptors. The performance and applicability of reverse docking depends on the accuracy of the docking program, the reliability of ligand binding pocket identification, the quality of target structures [68–71], and the normalization of docking scores [72,73]; but success can be achieved. By combining protein-ligand docking and data mining techniques, Yang *et al.* identified candidate genes that may be responsible for serious adverse drug reactions (SADR) such as Stevens-Johnson syndrome [73]. Taking this a step further, genetic predisposition to drug side-effects can be rationalized by the interaction between the drug and the putative off-target predicted by structural modeling and protein-ligand docking. For example, Li *et al.* discovered that a nonsynonymous SNP in human cytosolic sialidase potentially causes an adverse drug reaction to oseltamivir (Tamiflu) [74].

Despite good structural proteome coverage, there are many important drug targets which are absent. In such cases, homology modeling can provide reasonable starting structures if the sequence identity between the target and template protein is 30% or higher [75]. The quality of homology models when used for docking has been evaluated for GPCRs [76–78], protein kinases [79–81], nuclear hormone receptors [82–84] and other families [85–88]. All these studies show that homology modeling can enrich the screening and provide reasonable results, but it is not as powerful as using experimental holo crystal structures. The limitation of homology modeling based docking comes not only from the quality of the models, but also from the conformational changes induced by different ligands. Fan *et. al* found that multiple models improved docking results for 38 targets in their benchmark with half providing results comparable to docking using holo crystal structures [89]. A similar conclusion was reached by Orozco *et. al* in their high-throughput docking study [90]. Typically in homology modeling associated small molecules are not considered, although methods now exist to consider them. Recently, Dalton *et. al* presented a “hybrid-template modeling” method which includes the ligand bound to the template at the beginning of homology modeling [91]. This method appears to be very promising when a suitable protein-ligand template can be identified.

Although docking is able to generate reasonable binding poses most of the time, the efficient and accurate estimation of the binding affinity is a challenging task [92]. A more formidable task is to explicitly take protein flexibility into account in the calculation of the binding free energy as scoring errors most often arise from the rigid protein conformation [93]. The modeling of protein flexibility requires computationally intensive molecular dynamic (MD) simulations, such as the relaxed complex schema [47,94,95]. Following MD simulation, MM-PBSA/GBSA [96] and absolute free energy calculations [93,97–99] can be applied to accurately estimate the binding affinity of the ligand to the putative off-targets. These computationally intensive calculations currently cannot be applied on a structural proteome-wide scale. However, they can serve as computational verification tools when putative off-targets have been predicted from structural bioinformatics, protein-ligand docking, ligand based methods, systems biology simulation (discussed subsequently), and other resources. This integrated strategy has been successfully applied to antimicrobial drug discovery [100].

5. Integration of off-target predictions with biological network analysis and simulation

The clinical efficacy and adverse effects of drugs begins at the molecular level, involves complex biological networks and is ultimately measured by clinical outcomes at the level of the whole organism. Thus, it is critical to model and analyze off-target binding in the context of biological networks [101]. Network analysis and simulation (the analysis of the dynamics inherent in the network) has proved to be a powerful tool in identifying novel drug targets [102], elucidating mechanisms of drug resistance [103], predicting drug response phenotypes [104–106], and identifying genetic risk factors [106]. The understanding of the drug response at the systems level is particularly important for understanding polypharmacology since there are many reported examples where the therapeutic efficacy is enhanced through synergistic relationships among multiple targets [107–111]. The question we address here is what has been done to date and what does the future hold?

The first structure-enabled metabolic network reconstruction of the thermophile *Thermotoga maritima* shed new light into the evolution of metabolic enzymes and pathways [112]. Although this work used *T. maritima* as a model system, which is not very relevant to drug discovery, it established a framework to integrate structural bioinformatics and metabolic modeling, and could be extended to other organisms more related to human health. As such a structural proteome-wide drug-target network – the TB-drugome - was constructed and

analyzed by combining structural bioinformatics, protein-ligand docking and flux balance analysis [113]. The TB-drugome links structurally characterized approved drugs and more than 1,000 solved protein structures and homology models from the *Mycobacterium tuberculosis* (*M.tb*) proteome. It revealed that a large number of drugs have the potential to be repositioned to treat tuberculosis and that many currently unexploited *M.tb* receptors may be chemically druggable and could serve as novel anti-tubercular targets. Further, a detailed analysis of the network topology of the TB-drugome supports the hypothesis that drug-target networks are inherently scale-free and modular i.e., a small numbers of targets are highly promiscuous, but most targets bind a small number of ligands. However, the drug-target network alone provides little information on the mode of drug action. To correlate drug off-target binding to clinical end-points such as observed side-effects and clinical efficacy, the drug-target network should be integrated using interconnected metabolic, signal transduction and gene regulation pathways. In this way, a cause-effect knowledge graph can be constructed, which links drugs, targets, pathways and clinical endpoints (see Figure 4 for an example) [72]. In the longer term, the development of the semantic web will facilitate the integration of such drug-target networks with heterogeneous data sets facilitating further analysis [114]. Although the data integration and knowledge representation of drug-target-pathway-disease offers a great opportunity to predict drug induced phenotype changes, this static picture does not capture the dynamic behavior of biological systems, which is critical for the predictive modeling of drug action. Recent advances in the proteome-wide reconstruction of human metabolic networks [115] and algorithms for context-specific modeling [116,117] make it possible to simulate drug induced perturbations in human. Chang *et al.* have developed a functional renal model to study the detailed impact of off-target binding of CETP inhibitors in hypertension [106]. One interesting finding from this study is that there exist cryptic genetic risk factors that are associated with disorder phenotypes during drug treatment but not under normal conditions. The identification of genetic risk factors associated with drug side-effects is a small step towards realizing the promise of personalized medicine and more immediately the optimal design of clinical trials.

In order to apply constraint-based modeling, which has been tremendously successful in studying microorganisms [118,119], to investigate the cellular response of drug perturbations in human, two big challenges are: (1) reconstruction of a functional sub-network; and (2) simulation of the state change of the protein (e.g., partial inhibition or over-activation, under/over expression, or gain/loss of function) through ligand binding or genetic modification. Continued progress in generating, mining and integrating omics-data [120–125], the active development of algorithms for network biology [126–128], and the emergence of executable biology [129,130] will provide rich resources and computational frameworks to address the network reconstruction issue. However, little attention has been paid to the second problem. Simple knock-out experiments may be too simple to account for complex drug actions and genetic alternations. Although computational techniques exist to simulate cellular responses to a broad range of genetic modifications, they are too computationally intensive to be applied on a proteome-wide scale [131]. There is much left to do.

Conclusions

Structural proteomics is complementary to other computational methods that have been developed for off-target prediction. There is a long history in utilizing small molecule similarity to identify off-targets [132]. A bioactivity-guided mapping of chemical space may improve the correlation between the chemical structure and its bioactivity [133–135]. Recently, chemogenomics has emerged as a new discipline to systematically establish target relationships based on the structural and biological similarity of their ligands [136]. Several studies have attempted to extend target-based methods through similar sequence motifs or

global structures [137]. Similarly, the scope of chemical genomics approaches can be further extended beyond sequence and fold similarity by directly searching for similar ligand binding sites. This leads to proteome-wide protein-ligand interaction networks. Taken together it is hoped that these approaches will eventually provide the foundation for an *in silico* simulation of the influence of small molecules on biological systems.

A structural proteomics approach to off-target binding has its limitations. Besides the issues inherent in the individual methodologies discussed here, a daunting task is the rigorous validation of the hypotheses arising from off-target prediction. Experimental validation involves two aspects: undertaking a ligand binding assay across a panel of off-targets and measuring the phenotype response resulting from off-target binding. Although multiple target assays are available for well-studied gene families such as protein kinases [138], it may not be feasible to develop assays that are able to detect weak bindings across a large panel of proteins from different gene families, especially for those families not well characterized. Recent development in proteomics methods, such as capture compound mass spectrometry [139], stability of proteins from rate of oxidation [140], small molecular microarrays [141], affinity SILAC [142], *in vivo* metabolites assay [143], and phosphoproteomics [144] may help address these issues. It is even more challenging to verify the predicted collective phenotypic off-target effect. The cellular response to drug treatment depends on many complex factors such as genetic variation [145–147], cell-cell communication [148], protein-protein interaction networks [149], and protein oligomerization and different conformation states [150]. Accumulated data and knowledge from chemical genomics and genetics of model organisms could lay the foundation for elucidating the cellular mechanisms for how biological systems respond to environmental perturbation and genetic alternation [151,152]. Increasing computer power will enable high-throughput hypothesis evaluation *in silico* [153]. Taking all these advances together offers some hope for a new approach in supporting early stage drug discovery.

Acknowledgments

We appreciate constructive suggestions from the editor and anonymous reviewer on improving the manuscript. This work was supported by the National Institutes of Health Grant GM078596.

References

1. Nwaka S, Hudson A. Innovative lead discovery strategies for tropical diseases. *Nat Rev Drug Discov.* 2006; 5:941–955. [PubMed: 17080030]
2. O'Connor KA, Roth BL. Finding new tricks for old drugs: an efficient route for public-sector drug discovery. *Nat Rev Drug Discov.* 2005; 4:1005–1014. [PubMed: 16341065]
3. Kitano H. A robustness-based approach to systems-oriented drug design. *Nat Rev Drug Discov.* 2007; 6:202–210. [PubMed: 17318209]
4. Hopkins AL. Network pharmacology. *Nature Biotechnology.* 2007; 25:1110–1111.
5. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandari A, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science.* 2008; 320:539–543. [PubMed: 18369103]
6. Parsons W, et al. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science.* 2008; 321:1807–1812. [PubMed: 18772396]
7. Jones S, et al. Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses. *Science.* 2008; 321:1801–1806. [PubMed: 18772397]
8. Jia J, Zhu F, Ma X, Cao Z, Li Y, Chen YZ. Mechanisms of drug combinations: interaction and network perspectives. *Nat Rev Drug Discov.* 2009; 8:111–128. [PubMed: 19180105]
9. Druker BJ, Lydon NB. Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest.* 2000; 105:3–7. [PubMed: 10619854]

10. Carroll M, Ohno-Jones S, Tamura S, Buchdunger E, Zimmermann J, Lydon NB, Gilliland DG, Druker BJ. CGP 57148, a tyrosine kinase inhibitor, inhibits the growth of cells expressing BCR-ABL, TEL-ABL, and TEL-PDGFR fusion proteins. *Blood*. 1997; 90:4947–4952. [PubMed: 9389713]
11. Jain MK, Ridker PM. Anti-inflammatory effects of statins: clinical evidence and basic mechanisms. *Nat Rev Drug Discov*. 2005; 4:977–987. [PubMed: 16341063]
- 12. Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol*. 2009; 27:157–167. An insightful review that highlights the importance of understanding protein promiscuity in synthetic biology and drug discovery. [PubMed: 19204698]
13. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nat Biotechnol*. 2006; 24:805–815. [PubMed: 16841068]
14. Mestres J, Gregori-Puigjane E, Valverde S, Sole RV. Data completeness--the Achilles heel of drug-target networks. *Nat Biotechnol*. 2008; 26:983–984. [PubMed: 18779805]
15. Mestres J. Computational chemogenomics approaches to systematic knowledge-based drug discovery. *Curr Opin Drug Discov Devel*. 2004; 7:304–313.
16. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol*. 2007; 25:197–206. [PubMed: 17287757]
17. Fliri AF, Loring WT, Thadeio PF, Volkmann RA. Biospectra analysis: model proteome characterizations for linking molecular structure and biological response. *J Med Chem*. 2005; 48:6918–6925. [PubMed: 16250650]
18. Lamb J. The Connectivity Map: a new tool for biomedical research. *Nat Rev Cancer*. 2007; 7:54–60. [PubMed: 17186018]
19. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*. 2006; 313:1929–1935. [PubMed: 17008526]
20. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science*. 2008; 321:263–266. [PubMed: 18621671]
21. Deshpande N, Address KJ, Bluhm WF, Merino-Ott JC, Townsend-Merino W, Zhang Q, Knezevich C, Xie L, Chen L, Feng Z, et al. The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res*. 2005; 33:D233–D237. [PubMed: 15608185]
22. Xie L, Bourne PE. Functional coverage of the human genome by existing structures, structural genomics targets, and homology models. *PLoS Comp Biol*. 2005; 1:e31.
23. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008; 9:40. [PubMed: 18215316]
24. Xie L, Wang J, Bourne PE. In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators. *PLoS Comp Biol*. 2007; 3:e217.
25. Todd AE, Marsden RL, Thornton JM, Orengo CA. Progress of structural genomics initiatives: an analysis of solved target structures. *J Mol Biol*. 2005; 348:1235–1260. [PubMed: 15854658]
26. Xiang Z. Advances in homology protein structure modeling. *Curr Protein Pept Sci*. 2006; 7:217–227. [PubMed: 16787261]
27. Gunther S, Kuhn M, Dunkel M, Campillos M, Senger C, Petsalaki E, Ahmed J, Urdiales EG, Gewiss A, Jensen LJ, et al. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res*. 2008; 36:D919–922. [PubMed: 17942422]
28. Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE. Co-evolution of proteins with their interaction partners. *J Mol Biol*. 2000; 299:283–293. [PubMed: 10860738]
29. Dupont CL, Yang S, Palenik B, Bourne PE. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proc Natl Acad Sci U S A*. 2006; 103:17822–17827. [PubMed: 17098870]
30. Whisstock JC, Lesk AM. Prediction of protein function from protein sequence and structure. *Q Rev Biophys*. 2003; 36:307–340. [PubMed: 15029827]
31. Aguero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, Carmona S, Carruthers IM, Chan AW, Chen F, et al. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat Rev Drug Discov*. 2008; 7:900–907. [PubMed: 18927591]

32. Muller G. Medicinal chemistry of target family-directed masterkeys. *Drug Discov Today*. 2003; 8:681–691. [PubMed: 12927511]
33. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002; 1:727–730. [PubMed: 12209152]
34. Koch MA, Breinbauer R, Waldmann H. Protein structure similarity as guiding principle for combinatorial library design. *Biol Chem*. 2003; 384:1265–1272. [PubMed: 14515987]
35. Bisson WH, Cheltsov AV, Bruey-Sedano N, Lin B, Chen J, Goldberger N, May LT, Christopoulos A, Dalton JT, Sexton PM, et al. Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proc Natl Acad Sci U S A*. 2007; 104:11927–11932. [PubMed: 17606915]
36. Liu CI, Liu GY, Song Y, Yin F, Hensler ME, Jeng WY, Nizet V, Wang AH, Oldfield E. A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science*. 2008; 319:1391–1394. [PubMed: 18276850]
- 37. Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:129–134. A novel approach that extends ligand binding site determination to predicted structures and accurately predicts protein functions. [PubMed: 18165317]
- 38. Brylinski M, Skolnick J. Cross-reactivity virtual profiling of the human kinome by X-react(KIN): a chemical systems biology approach. *Mol Pharm*. 2010; 7:2324–2333. A novel systematic approach to predict off-targets by integrating ligand homology modeling, ligand binding similarity, low-resolution docking and machine learning. [PubMed: 20958088]
- 39. Xie L, Bourne PE. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc Natl Acad Sci U S A*. 2008; 105:5441–5446. An algorithmic framework for non-sequential protein structure alignment that resembles local sequence alignment by incorporating evolutionary, physiochemical and geometric information into similarity measurements. [PubMed: 18385384]
40. Petrey D, Fischer M, Honig B. Structural relationships among proteins with different global topologies and their implications for function annotation strategies. *Proc Natl Acad Sci U S A*. 2009; 106:17377–17382. [PubMed: 19805138]
41. Zhang QC, Petrey D, Norel R, Honig BH. Protein interface conservation across structure space. *Proc Natl Acad Sci U S A*. 2010; 107:10896–10901. [PubMed: 20534496]
42. Specker E, Bottcher J, Lilie H, Heine A, Schoop A, Muller G, Griebenow N, Klebe G. An old target revisited: two new privileged skeletons and an unexpected binding mode for HIV-protease inhibitors. *Angew Chem Int Ed Engl*. 2005; 44:3140–3144. [PubMed: 15822136]
43. Weber A, Casini A, Heine A, Kuhn D, Supuran CT, Scozzafava A, Klebe G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. *J Med Chem*. 2004; 47:550–557. [PubMed: 14736236]
44. Stauch B, Hofmann H, Perkovic M, Weisel M, Kopietz F, Cichutek K, Munk C, Schneider G. Model structure of APOBEC3C reveals a binding pocket modulating ribonucleic acid interaction required for encapsidation. *Proc Natl Acad Sci U S A*. 2009; 106:12079–12084. [PubMed: 19581596]
45. Miller JR, Dunham S, Mochalkin I, Banotai C, Bowman M, Buist S, Dunkle B, Hanna D, Harwood HJ, Huband MD, et al. A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore. *Proc Natl Acad Sci U S A*. 2009; 106:1737–1742. [PubMed: 19164768]
46. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L, Bourne PE. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput Biol*. 2009; 5:e1000423. [PubMed: 19578428]
47. Amaro RE, Schnauffer A, Interthal H, Hol W, Stuart KD, McCammon JA. Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A*. 2008; 105:17278–17283. [PubMed: 18981420]
48. Durrant JD, Amaro RE, Xie L, Urbaniak MD, Ferguson MA, Haapalainen A, Chen Z, Di Guilmi AM, Wunder F, Bourne PE, et al. A multidimensional strategy to detect polypharmacological

- targets in the absence of structural and sequence homology. *PLoS Comput Biol*. 2010; 6:e1000648. [PubMed: 20098496]
49. Milletti F, Vulpetti A. Predicting polypharmacology by binding site similarity: from kinases to the protein universe. *J Chem Inf Model*. 2010; 50:1418–1431. [PubMed: 20666497]
 50. Leis S, Schneider S, Zacharias M. In Silico Prediction of Binding Sites on Proteins. *Current Medicinal Chemistry*. 2010; 17:1550–1562. [PubMed: 20166931]
 51. Laurie ATR, Jackson RM. Methods for the prediction of protein-ligand binding sites for Structure-Based Drug Design and virtual ligand screening. *Current Protein & Peptide Science*. 2006; 7:395–406. [PubMed: 17073692]
 52. Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov Today*. 2010; 15:656–667. [PubMed: 20685398]
 53. Brylinski M, Skolnick J. FINDSITE LHM: A Threading-Based Approach to Ligand Homology Modeling. *Plos Computational Biology*. 2009; 5:e1000405. [PubMed: 19503616]
 54. Brylinski M, Skolnick J. Q-Dock(LHM): Low-Resolution Refinement for Ligand Comparative Modeling. *Journal of Computational Chemistry*. 2010; 31:1093–1105. [PubMed: 19827144]
 55. Kellenberger E, Schalon C, Rognan D. How to measure the similarity between protein ligand-binding sites. *Current Computer-Aided Drug Design*. 2008; 4:209–220.
 56. Gunasekaran K, Nussinov R. How different are structurally flexible and rigid binding sites? sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J Mol Biol*. 2006; 365:257–273. [PubMed: 17059826]
 57. Binkowski TA, Adamian L, Liang J. Inferring functional relationships of proteins from local sequence and spatial surface patterns. *J Mol Biol*. 2003; 332:505–526. [PubMed: 12948498]
 58. Tseng YY, Dundas J, Liang J. Predicting Protein Function and Binding Profile via Matching of Local Evolutionary and Geometric Surface Patterns. *J Mol Biol*. 2009; 387:451–464. [PubMed: 19154742]
 59. Laskowski RA, Watson JD, Thornton JM. Protein function prediction using local 3D templates. *J Mol Biol*. 2005; 351:614–626. [PubMed: 16019027]
 60. Agoston V, Csermely P, Pongor S. Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2005; 71:051909. [PubMed: 16089573]
 61. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al. Diversity and complexity in DNA recognition by transcription factors. *Science*. 2009; 324:1720–1723. [PubMed: 19443739]
 62. Roy J, Cyert MS. Cracking the phosphatase code: docking interactions determine substrate specificity. *Sci Signal*. 2009; 2:re9. [PubMed: 19996458]
 63. Ferrell JE Jr. Signaling motifs and Weber's law. *Mol Cell*. 2009; 36:724–727. [PubMed: 20005833]
 64. Goentoro L, Kirschner MW. Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Mol Cell*. 2009; 36:872–884. [PubMed: 20005849]
 65. Goentoro L, Shoval O, Kirschner MW, Alon U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell*. 2009; 36:894–899. [PubMed: 20005851]
 66. Gilson MK, Zhou HX. Calculation of protein-ligand binding affinities. *Annu Rev Biophys Biomol Struct*. 2007; 36:21–42. [PubMed: 17201676]
 67. Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure*. 2009; 17:489–498. [PubMed: 19368882]
 68. Rockey WM, Elcock AH. Progress toward virtual screening for drug side effects. *Proteins: Struct Funct Bioinform*. 2002; 48:664–671.
 69. Ji ZL, Wang Y, Yu L, Han LY, Zheng CJ, Chen YZ. In silico search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction. *Toxicol Lett*. 2006; 164:104–112. [PubMed: 16563668]
 70. Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins-Structure Function and Genetics*. 2001; 43:217–226.

71. Bernard P, Dufresne-Favetta C, Favetta P, Do QT, Himbert F, Zubrzycki S, Scior T, Lugnier C. Application of drug repositioning strategy to TOFISOPAM. *Curr Med Chem*. 2008; 15:3196–3203. [PubMed: 19075663]
72. Xie L, Li J, Xie L, Bourne PE. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comp Biol*. 2009; 5:e1000387.
73. Yang L, Chen J, He L. Harvesting candidate genes responsible for serious adverse drug reactions from a chemical-protein interactome. *PLoS Comput Biol*. 2009; 5:e1000441. [PubMed: 19629158]
74. Li CY, Yu Q, Ye ZQ, Sun Y, He Q, Li XM, Zhang W, Luo J, Gu X, Zheng X, et al. A nonsynonymous SNP in human cytosolic sialidase in a small Asian population results in reduced enzyme activity: potential link with severe adverse reactions to oseltamivir. *Cell Res*. 2007; 17:357–362. [PubMed: 17426694]
75. Rost B. Twilight zone of protein sequence alignments. *Protein Engineering*. 1999; 12:85–94. [PubMed: 10195279]
76. Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins-Structure Function and Genetics*. 2003; 50:5–25.
77. Evers A, Klabunde T. Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the Alpha1A adrenergic receptor. *Journal of Medicinal Chemistry*. 2005; 48:1088–1097. [PubMed: 15715476]
78. Radestock S, Weil T, Renner S. Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring. *Journal of Chemical Information and Modeling*. 2008; 48:1104–1117. [PubMed: 18442221]
79. Diller DJ, Li RX. Kinases, homology models, and high throughput docking. *Journal of Medicinal Chemistry*. 2003; 46:4638–4647. [PubMed: 14561083]
80. Oshiro C, Bradley EK, Eksterowicz J, Evensen E, Lamb ML, Lancot JK, Putta S, Stanton R, Grootenhuys PDJ. Performance of 3D-database molecular docking studies into homology models. *Journal of Medicinal Chemistry*. 2004; 47:764–767. [PubMed: 14736258]
81. Rockey WM, Elcock AH. Structure selection for protein kinase docking and virtual screening: Homology models or crystal structures? *Current Protein & Peptide Science*. 2006; 7:437–457. [PubMed: 17073695]
82. Kasuya A, Sawada Y, Tsukamoto Y, Tanaka K, Toya T, Yanagi M. Binding mode of ecdysone agonists to the receptor: comparative modeling and docking studies. *Journal of Molecular Modeling*. 2003; 9:58–65. [PubMed: 12638012]
83. Marhefka CA, Moore BM, Bishop TC, Kirkovsky L, Mukherjee A, Dalton JT, Miller DD. Homology modeling using multiple molecular dynamics simulations and docking studies of the human androgen receptor ligand binding domain bound to testosterone and nonsteroidal ligands. *Journal of Medicinal Chemistry*. 2001; 44:1729–1740. [PubMed: 11356108]
84. Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. *Journal of Medicinal Chemistry*. 2003; 46:3045–3059. [PubMed: 12825943]
85. Ferrara P, Jacoby E. Evaluation of the utility of homology models in high throughput docking. *Journal of Molecular Modeling*. 2007; 13:897–905. [PubMed: 17487515]
86. McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *Journal of Medicinal Chemistry*. 2003; 46:2895–2907. [PubMed: 12825931]
87. Kairys V, Fernandes MX, Gilson MK. Screening drug-like compounds by docking to homology models: A systematic study. *Journal of Chemical Information and Modeling*. 2006; 46:365–379. [PubMed: 16426071]
88. Parrill AL, Echols U, Nguyen T, Pham TCT, Hoeglund A, Baker DL. Virtual screening approaches for the identification of non-lipid autotaxin inhibitors. *Bioorganic & Medicinal Chemistry*. 2008; 16:1784–1795. [PubMed: 18036821]
89. Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A. Molecular Docking Screens Using Comparative Models of Proteins. *Journal of Chemical Information and Modeling*. 2009; 49:2512–2527. [PubMed: 19845314]

90. Novoa EM, de Poupiana LR, Barril X, Orozco M. Ensemble Docking from Homology Models. *Journal of Chemical Theory and Computation*. 2010; 6:2547–2557.
- 91. Dalton JAR, Jackson RM. Homology-Modelling Protein-Ligand Interactions: Allowing for Ligand-Induced Conformational Change. *Journal of Molecular Biology*. 2010; 399:645–661. A new homology modeling method that includes the ligand during protein modeling. [PubMed: 20434455]
92. Guvench O, Mackerell AD Jr. Computational evaluation of protein-small molecule binding. *Curr Opin Struct Biol*. 2009; 19:56–61. [PubMed: 19162472]
93. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol*. 2007; 371:1118–1134. [PubMed: 17599350]
94. Amaro RE, Baron R, McCammon JA. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des*. 2008; 22:693–705. [PubMed: 18196463]
95. Cheng LS, Amaro RE, Xu D, Li WW, Arzberger PW, McCammon JA. Ensemble-based virtual screening reveals potential novel antiviral compounds for avian influenza neuraminidase. *J Med Chem*. 2008; 51:3878–3894. [PubMed: 18558668]
96. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*. 2000; 33:889–897. [PubMed: 11123888]
97. Luo H, Sharp K. On the calculation of absolute macromolecular binding free energies. *Proc Natl Acad Sci U S A*. 2002; 99:10399–10404. [PubMed: 12149474]
98. Rodinger T, Howell PL, Pomes R. Calculation of absolute protein-ligand binding free energy using distributed replica sampling. *J Chem Phys*. 2008; 129:155102. [PubMed: 19045232]
99. Woo HJ, Roux B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc Natl Acad Sci U S A*. 2005; 102:6825–6830. [PubMed: 15867154]
- 100. Shen Y, Liu J, Estiu G, Isin B, Ahn YY, Lee DS, Barabasi AL, Kapatral V, Wiest O, Oltvai ZN. Blueprint for antimicrobial hit discovery targeting metabolic networks. *Proc Natl Acad Sci U S A*. 2010; 107:1082–1087. Uses genome-wide metabolic modeling to identify drug targets and structural-based virtual screening to discover drug leads for antimicrobial drug discovery. An early example that integrates systems biology and structure-based drug design. [PubMed: 20080587]
101. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. *Nat Biotechnol*. 2004; 22:1253–1259. [PubMed: 15470465]
102. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. *Nat Biotechnol*. 2007; 25:1119–1126. [PubMed: 17921997]
103. Raman K, Chandra N. Mycobacterium tuberculosis interactome analysis unravels potential pathways to drug resistance. *BMC Microbiol*. 2008; 8:234. [PubMed: 19105810]
104. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*. 2009
105. Scheiber J, Chen B, Milik M, Sukuru SC, Bender A, Mikhailov D, Whitebread S, Hamon J, Azzaoui K, Urban L, et al. Gaining Insight into Off-Target Mediated Effects of Drug Candidates with a Comprehensive Systems Chemical Biology Analysis. *J Chem Inf Model*. 2009
- 106. Chang RL, Xie L, Xie L, Bourne PE, Palsson B. Drug Off-Target Effects Predicted Using Structural Analysis in the Context of a Metabolic Network Model. *PLoS Comp Biol*. 2010; 6:e1000938. As far as we are aware, the first effort to computationally predict human phenotype changes upon drug treatments and genetic alternations by bridging structural biology and systems biology.
107. Keith CT, Borisy AA, Stockwell BR. Multicomponent therapeutics for networked systems. *Nat Rev Drug Discov*. 2005; 4:71–78. [PubMed: 15688074]
108. Rasmussen K. Creating more effective antidepressants: clues from the clinic. *Drug Discov Today*. 2006; 11:623–631. [PubMed: 16793531]

109. Zimmermann GR, Lehar J, Keith CT. Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discovery Today*. 2007; 12:34–42. [PubMed: 17198971]
110. Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. *Nat Rev Cancer*. 2010; 10:130–137. [PubMed: 20094047]
111. Fitzgerald JB, Schoeberl B, Nielsen UB, Sorger PK. Systems biology and combination therapy in the quest for clinical efficacy. *Nat Chem Biol*. 2006; 2:458–466. [PubMed: 16921358]
112. Zhang Y, Thiele I, Weekes D, Li Z, Jaroszewski L, Ginalski K, Deacon AM, Wooley J, Lesley SA, Wilson IA, et al. Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science*. 2009; 325:1544–1549. [PubMed: 19762644]
113. Kinnings SL, Xie L, Fung K, Xie L, Bourne PE. The Mycobacterium tuberculosis Drugome and Its Polypharmacological Implications. *PLoS Comp Biol*. 2010; 6:e100976.
- 114. Chen B, Dong X, Jiao D, Wang H, Zhu Q, Ding Y, Wild DJ. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinformatics*. 2010; 11:255. An effort to link compounds, drugs, genes, proteins, and diseases using semantic web techniques. [PubMed: 20478034]
115. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*. 2007; 104:1777–1782. [PubMed: 17267599]
116. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Rupp E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*. 2008; 26:1003–1010. [PubMed: 18711341]
117. Becker SA, Palsson BO. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol*. 2005; 5:8. [PubMed: 15752426]
118. Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*. 2004; 2:886–897. [PubMed: 15494745]
119. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*. 2009; 7:129–143. [PubMed: 19116616]
120. Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science*. 2010; 329:439–443. [PubMed: 20558669]
121. Xu TR, Vyshemirsky V, Gormand A, von Kriegsheim A, Girolami M, Baillie GS, Ketley D, Dunlop AJ, Milligan G, Houslay MD, et al. Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci Signal*. 2010; 3:ra20.
122. Huang SS, Fraenkel E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci Signal*. 2009; 2:ra40. [PubMed: 19638617]
123. Yizhak K, Benyamini T, Liebermeister W, Rupp E, Shlomi T. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*. 2010; 26:i255–260. [PubMed: 20529914]
124. Komurov K, White MA, Ram PT. Use of data-biased random walks on graphs for the retrieval of context-specific networks from genomic data. *PLoS Comput Biol*. 2010; 6:e1000889. [PubMed: 20808879]
125. Nibbe RK, Koyuturk M, Chance MR. An integrative -omics approach to identify functional subnetworks in human colorectal cancer. *PLoS Comput Biol*. 2010; 6:e1000639. [PubMed: 20090827]
126. Koyuturk M. Algorithmic and analytical methods in network biology. *WIREs Systems Biology and Medicine*. 2009; 2:277–292. [PubMed: 20836029]
127. Dittrich MT, Klau GW, Rosenwald A, Dandekar T, Muller T. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics*. 2008; 24:i223–231. [PubMed: 18586718]
128. Yosef N, Ungar L, Zalckvar E, Kimchi A, Kupiec M, Rupp E, Sharan R. Toward accurate reconstruction of functional protein networks. *Mol Syst Biol*. 2009; 5:248. [PubMed: 19293828]
129. Fisher J, Henzinger TA. Executable cell biology. *Nat Biotechnol*. 2007; 25:1239–1249. [PubMed: 17989686]

130. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol.* 2009; 5:331. [PubMed: 19953085]
131. Zhao Q, Kurata H. Genetic modification of flux for flux prediction of mutants. *Bioinformatics.* 2009; 25:1702–1708. [PubMed: 19468056]
132. Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov Today.* 2007; 12:225–233. [PubMed: 17331887]
133. Koch MA, Schuffenhauer A, Scheck M, Wetzel S, Casaulta M, Odermatt A, Ertl P, Waldmann H. Charting biologically relevant chemical space: a structural classification of natural products (SCONP). *Proc Natl Acad Sci U S A.* 2005; 102:17272–17277. [PubMed: 16301544]
134. Renner S, van Otterlo WA, Dominguez Seoane M, Mocklinghoff S, Hofmann B, Wetzel S, Schuffenhauer A, Ertl P, Oprea TI, Steinhilber D, et al. Bioactivity-guided mapping and navigation of chemical space. *Nat Chem Biol.* 2009; 5:585–592. [PubMed: 19561619]
135. Wetzel S, Klein K, Renner S, Rauh D, Oprea TI, Mutzel P, Waldmann H. Interactive exploration of chemical space with Scaffold Hunter. *Nat Chem Biol.* 2009; 5:581–583. [PubMed: 19561620]
136. Bender A, Young DW, Jenkins JL, Serrano M, Mikhailov D, Clemons PA, Davies JW. Chemogenomic data analysis: prediction of small-molecule targets and the advent of biological fingerprint. *Comb Chem High Throughput Screen.* 2007; 10:719–731. [PubMed: 18045083]
137. Strömbergsson H, Kryshchovych A, Prusis P, Fidelis K, Wikberg JES, Komorowski J, Hvidsten TR. Generalized modeling of enzyme-ligand interactions using proteochemometrics and local protein substructures. *Proteins.* 2006; 65:568–579. [PubMed: 16948162]
138. Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, et al. A quantitative analysis of kinase inhibitor selectivity. *Nat Biotechnol.* 2008; 26:127–132. [PubMed: 18183025]
139. Fischer JJ, Michaelis S, Schrey AK, Graebner OG, Glinski M, Dreger M, Kroll F, Koester H. Capture compound mass spectrometry sheds light on the molecular mechanisms of liver toxicity of two Parkinson drugs. *Toxicol Sci.* 2010; 113:243–253. [PubMed: 19783845]
140. West GM, Tucker CL, Xu T, Park SK, Han X, Yates JR 3rd, Fitzgerald MC. Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. *Proc Natl Acad Sci U S A.* 2010; 107:9078–9082. [PubMed: 20439767]
141. Uttamchandani M, Wang J, Yao SQ. Protein and small molecule microarrays: powerful tools for high-throughput proteomics. *Mol Biosyst.* 2006; 2:58–68. [PubMed: 16880923]
142. Ong SE, Schenone M, Margolin AA, Li X, Do K, Doud MK, Mani DR, Kuai L, Wang X, Wood JL, et al. Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc Natl Acad Sci U S A.* 2009; 106:4617–4622. [PubMed: 19255428]
143. Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M. Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses. *Cell.* 2010; 143:639–650. [PubMed: 21035178]
144. Li J, Rix U, Fang B, Bai Y, Edwards A, Colinge J, Bennett KL, Gao J, Song L, Eschrich S, et al. A chemical and phosphoproteomic characterization of dasatinib action in lung cancer. *Nat Chem Biol.* 2010; 6:291–299. [PubMed: 20190765]
145. Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJ, Stokoe D, Gloor SL, Vigers G, et al. RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature.* 2010; 464:431–435. [PubMed: 20130576]
146. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature.* 2010; 464:427–430. [PubMed: 20179705]
147. Heidorn SJ, Milagre C, Whittaker S, Nourry A, Niculescu-Duvas I, Dhomen N, Hussain J, Reis-Filho JS, Springer CJ, Pritchard C, et al. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell.* 2010; 140:209–221. [PubMed: 20141835]
148. Wu M, Pastor-Pareja JC, Xu T. Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion. *Nature.* 2010; 463:545–548. [PubMed: 20072127]

149. Breitkreutz A, Choi H, Sharom JR, Boucher L, Neduva V, Larsen B, Lin ZY, Breitkreutz BJ, Stark C, Liu G, et al. A global protein kinase and phosphatase interaction network in yeast. *Science*. 2010; 328:1043–1046. [PubMed: 20489023]
150. Rajagopal S, Rajagopal K, Lefkowitz RJ. Teaching old receptors new tricks: biasing seven-transmembrane receptors. *Nat Rev Drug Discov*. 2010; 9:373–386. [PubMed: 20431569]
151. Lehar J, Krueger A, Zimmermann G, Borisy A. High-order combination effects and biological robustness. *Mol Syst Biol*. 2008; 4:215. [PubMed: 18682705]
152. Lehar J, Stockwell BR, Giaever G, Nislow C. Combination chemical genetics. *Nat Chem Biol*. 2008; 4:674–681. [PubMed: 18936752]
153. An G. Closing the scientific loop: bridging correlation and causality in the petaflop age. *Sci Transl Med*. 2010; 2:41ps34.

Figure 1. Proteome coverage of drug targets, PDB structures and homology models in human, virus, *M. tuberculosis*, and *T. brucei*.

Figure 2. Distribution of 2D molecular fingerprint similarities between 6,127 drugs and PDB ligands

Around 15% and 60% of drugs have similar molecular structures to PDB ligands with fingerprint similarity scores larger than 90 and 70, respectively.

Figure 3. Off-target binding depends on both the ligand binding propensity of receptors and the chemical nature of ligands

For two receptors A and B that share a similar ligand binding pocket, only ligand *L2* of receptor A and ligand *L3* of receptor B demonstrate cross-reactivity between receptors A and B. *L2* and *L3* may bind receptor C in a totally different pose than when binding receptors A and B. Existing structural bioinformatics methods may not detect that receptor C is an off-target of receptors A and B.

Figure 4. An example of a knowledge graph that links a drug-target network to clinical endpoints through biological pathways

The graph shows the effects of CETP inhibitors Torcetrapib, Anacetrapib and JTT705 on hypertension, inflammation and cancer through the combinational control of nuclear hormone receptors and fatty acid binding protein (FABP). The red, purple, and blue lines between inhibitors and off-targets indicate strong, relatively strong, and weak binding affinities, respectively. The brown and black lines between off-targets and pathways or clinical indications represent positive and negative regulation, respectively.