Novel Computational Approaches to Polypharmacology as a Means to Define Responses to Individual Drugs

Lei Xie,[1,*] Li Xie,[2] Sarah L. Kinnings,[3] and Philip E. Bourne[2,*]

[1]Department of Computer Science, Hunter College, The City University of New York, New York, New York 10065; email: lei.xie@hunter.cuny.edu

[2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093; email: pbourne@ucsd.edu

[3]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093

Annu. Rev. Pharmacol. Toxicol. 2012. 52:361–79

First published online as a Review in Advance on October 17, 2011

The *Annual Review of Pharmacology and Toxicology* is online at pharmtox.annualreviews.org

This article's doi: 10.1146/annurev-pharmtox-010611-134630

0362-1642/12/0210-0361$20.00

*Corresponding authors.

Keywords

protein-ligand interaction, drug-target network, in vivo efficacy, dynamic simulation, metabolic modeling

Abstract

Polypharmacology, which focuses on designing therapeutics to target multiple receptors, has emerged as a new paradigm in drug discovery. Polypharmacological effects are an attribute of most, if not all, drug molecules. The efficacy and toxicity of drugs, whether designed as single- or multitarget therapeutics, result from complex interactions between pharmacodynamic, pharmacokinetic, genetic, epigenetic, and environmental factors. Ultimately, to predict a drug response phenotype, it is necessary to understand the change in information flow through cellular networks resulting from dynamic drug-target interactions and the impact that this has on the complete biological system. Although such is a future objective, we review recent progress and challenges in computational techniques that enable the prediction and analysis of in vitro and in vivo drug-response phenotypes.

INTRODUCTION

The molecular mechanisms of action of a large number of preclinical and clinical drugs still remain elusive. Unexpected drug efficacy, toxicity, and resistance are often observed in treating complex multigenic diseases such as cardiovascular disease, central nervous system disorders, and cancer, as well as in anti-infectious drug discovery. In recent years, it has been discovered that multiple on-target and off-target binding is essential for efficacy and is responsible for the side effects of many drugs from a wide range of therapeutic areas. For example, both statins (intended target, HMG-COA reductase) and torcetrapib (intended target, cholesterol ester transfer protein) can lower the level of LDL cholesterol efficiently by inhibiting, with high affinity, the receptors that are critical for cholesterol biosynthesis and transport. However, these two drugs produce completely different clinical endpoints. Whereas statins demonstrate cardiovascular benefits to patients, torcetrapib, unfortunately, increases the risk of heart disease. Several studies have shown that unintended off-target binding may contribute to the therapeutic efficacy of statins (1) and the unwanted side effect of torcetrapib (2). Similarly, although originally designed for specificity, a number of different protein kinase inhibitors mediate their anticancer effects through the modulation of several different targets synergistically (3, 4). Besides protein kinase inhibitors, antipsychotic drugs, such as clozapine, mediate their effects through binding entire families of serotonin and dopamine receptors. In fact, it has been proposed that the clinical failures of many antipsychotic drugs can be attributed to their being too selective in nature (3). It has recently been revealed that the anticancer drug lenalidomide inhibits angiogenesis and metastasis through disrupting the association of adherens junction proteins such as cadherin 5, β-catenin, and CD31, and by inhibiting the VEGF-induced PI3K-Akt signaling pathway (5). Several successful multitarget anti-infective therapeutics have also been discovered serendipitously (6).

The efficacy of multitarget therapy is supported by observations concerning the robustness and resilience of complex biological systems. Indeed, when faced with a drug-induced node failure in the cell, a disease network may be able to resort to systems robustness to ensure its survival. This robustness results from either the inherent diversity of genetics and signaling pathways or the resilient architecture of the complete cellular system (7). Because compensatory signaling pathways that bypass the inhibition of individual proteins frequently exist, it is often necessary to modulate several nodes simultaneously to affect the disease phenotype. Furthermore, large-scale functional genomics studies in model organisms have shown that the vast majority of single-gene knockouts exhibit little or no effect on phenotype (4). The concept of synthetic lethality—genes that are essential, not individually, but in combination—uncovers a whole new plethora of drug targets that may have been overlooked owing to their nonessentiality in individual-gene knockout studies (8). Moreover, dynamic system control of multiple drug-target interactions plays a key role in mediating therapeutic efficacy, side effects, and resistance to drug molecules. For example, it has been proposed that the efficiency of receptor tyrosine kinase inhibitors requires combined inhibition of the MAPK and PI3K pathways because the cross talk between PI3K and tyrosine kinase may contribute to clinical resistance of tyrosine kinase inhibitors. Furthermore, there exists a negative feedback loop between the PI3K and mTOR pathways (9). Indeed, the inhibition of mTOR increases the activation of PI3K, thus reducing the antiproliferative effect of mTOR inhibitors. In another example, the serious cardiovascular side effects of selective cyclooxygenase-2 inhibitors may result from the unbalanced production of prostacyclin and thromboxane A2 in the disease state, prompting platelet-dependent thrombosis (10).

The serendipitous discovery of the multitarget effects of drug action and the theoretical consideration of biological network properties have prompted efforts to assess a new polypharmacological paradigm in drug discovery (4). In contrast to designing selective ligands to target

individual proteins, polypharmacology focuses on searching for multitarget drugs to perturb disease-associated networks. The immediate question to address is how we can select the correct combination of therapeutic targets and rationally design polypharmacological drugs. Although most current efforts focus on the targeted polypharmacology of selected pathways or gene families (9), we argue that the polypharmacological drug phenotype response can be defined and predicted only on the basis of the quantitative and systematic analysis of drug-protein interactions on a proteome-wide scale. For this purpose, novel computational tools are needed to model drug actions on a multiscale, from the atomic details of drug-target interactions to the dynamic behavior of single cells, specific tissues, and whole organisms. In this review, we first discuss the nature of proteome-wide polypharmacology, and then highlight recent computational endeavors that facilitate multiscale modeling of polypharmacology on a proteome-wide scale.

THE NATURE OF POLYPHARMACOLOGICAL EFFECTS

Proteome-Wide Drug-Target Interactions Are not Rare but Common

The serendipitously discovered multitarget drug interactions mentioned in the introduction may be more than a coincidence. A number of recent studies have revealed that protein and ligand promiscuity is a phenomenon much more common than previously thought. Yildirim et al. (11) applied network analysis to drugs and their targets by integrating publicly available drug data. Their resulting network of polypharmacological interactions was dense in nature and revealed not only that a single target can often bind multiple drugs, but also that it is far more common than expected for a single drug to modulate several different molecular targets, which may be involved in multiple diseases. Similarly, Paolini et al. (12) discovered that approximately 35% of their database of 276,122 active compounds had observed activity for more than one target. Although the majority of these promiscuous compounds were active against targets within the same gene family, a significant proportion (approximately one-quarter) had demonstrable activity across different gene families. In fact, it has been estimated that, on average, each drug can bind to approximately six receptors (13).

Although most multitarget drug effects are identified from pharmaceutically investigated targets, it is likely that drugs can bind to multiple receptors on a proteome-wide scale, including proteins that are less functionally characterized (14). The promiscuity of drug-target interactions is rooted in the underlying promiscuity of protein function (15). Protein-functional promiscuity plays a critical role in the evolution and survival of organisms. Such promiscuity may be found in either a single member of the family or the complete family, and it arises from changed intracellular (e.g., differential expression, substrate concentration) and environmental (e.g., temperature, pH) conditions as well as through the adoption of a wide spectrum of mechanisms (15). Moreover, many endogenous ligands such as nucleotides are highly promiscuous (16). Recently, an unbiased and systematic analysis of in vivo metabolite-protein interactions discovered the unexpected binding of sterols to protein kinases and the interaction of pentaporphyrin with ergosterol biosynthesis proteins (17). Hydrophobic metabolites are thought to bind to more than 1,200 soluble proteins in yeast (17). Metabolite-protein interactions are undoubtedly more extensive in humans. Because successful drug molecules are likely to share chemical structural similarities with metabolites (18), multitarget drug effects are expected to extend beyond characterized drug targets.

At the level of the organism, the phenotypic response to a drug perturbation is a consequence of complex interactions between multiple intracellular and extracellular components. Besides the drug binding directly to multiple targets involved in drug actions, interactions include both specific and promiscuous binding to other proteins responsible for pharmacokinetics (transporters,

drug-metabolizing enzymes, etc.), as well as genetic, epigenetic, and environmental factors. Transport proteins play a critical role in drug development. Defining drug action requires a detailed understanding of how drugs specifically interact with and regulate various transport protein subtypes (19). Likewise, drug-metabolizing enzymes contribute significantly to the effective concentration and lifetime of the active form of the drug molecule. Drug-binding selectivity of well-known drug-metabolizing enzymes such as the cytochrome P450 (CYP) family has been the subject of extensive study (20). However, a large number of drug-metabolizing enzymes that affect the efficacy and safety of drugs remain unknown. For example, only in the past year was paraoxonase 1 identified as a key determinant in the variable efficacy of the antiplatelet drug clopidogrel (21). In addition, the effect of drugs can be altered through their binding to proteins that regulate the expression of transporters or drug-metabolizing enzymes. For example, various drugs act as agonists or antagonists of nuclear receptors that up- or downregulate the expression of different CYP subtypes. A diverse set of molecules can enhance the expression of multidrug-resistance pumps (22–24). However, in many cases, the molecular targets of regulatory networks are still unclear. Besides regulating multidrug-resistance proteins, off-target binding can be directly responsible for drug resistance (25). Furthermore, the phenotypic response of a drug depends on the state of the multivariate cellular network, which, in turn, is a function of genetic, epigenetic, and environmental variation. Although progress in pharmacogenetics and pharmacogenomics has linked an increasing number of genetic variations to individual drug responses (26), many genetic risk factors are off the radar. A recent discovery demonstrated the complexity of drug action under the influence of genetic variation. A selective *RAF* inhibitor that blocks the *RAS-RAF-MEK-ERK* signaling pathway activates the same pathway for mutant *RAS* (27–29), thereby turning an anticancer therapy into a carcinogen.

Besides intracellular factors, cell-cell communication plays an important role in the development of tumors and other diseases. For example, two cells—one of which harbors an oncogene, the other a loss-of-function mutation in a tumor suppressor gene—can cooperate to activate tumorigenesis (30). Variations in drug response also result from cell-to-cell variability with either genetic or nongenetic origins (31). Microenvironment also has a significant impact on drug phenotype response (32). It is well known that the in vitro potency of a drug may be dramatically different from its in vivo activity (33). Moreover, a drug may cause different responses in a disease versus a normal state. For example, the side effect of selective COX-2 inhibitors is more severe in the inflammation state than in the healthy state.

In summary, to define and predict a drug response, it is essential to identify drug-target interactions on a proteome-wide scale, including the molecular components involved in both pharmacodynamics and pharmacokinetics. In addition, it is necessary to account for genetic and epigenetic variations, cell-to-cell communication, cell-to-cell variability, and other environmental factors. Stated another way, we believe the future of successful drug discovery lies in a systems-based approach whereby we identify the cellular connectivity that links all of these aforementioned components, and where we can simulate the dynamic behavior upon drug perturbation for a given cellular state.

Multiple Weak Interactions Cannot be Ignored in Polypharmacological Actions

For the past two decades, the paradigm of drug discovery has been to develop highly selective ligands that interact with individual target proteins with high binding affinity. However, a highly potent lead compound usually yields a drug candidate with high molecular weight, which is often linked to a higher risk of failure during drug development (34). It has now been accepted that an optimal lead should balance binding potency and molecular size (35). Analysis of the binding

affinities of marketed drugs and natural products indicates that therapeutic efficacy is not necessarily associated with high binding affinity (34, 36). The lack of correlation between binding affinity and efficacy is well illustrated by the failure of sorafenib in treating melanoma, a type of skin cancer. Sorafenib was developed as a potent nanomolar inhibitor of *BRAF*, a protein implicated in the survival of melanoma cells. Disappointingly, the drug failed during clinical trials owing to its low antimelanoma efficacy. In contrast, the moderate micromolar *RAF* inhibitor PLX4720 is a potent inhibitor of downstream signaling and proliferation in cells harboring *BRAF*; as such, it is active in treating melanoma cell lines (see Reference 37 and references cited therein).

At the molecular level, it is estimated that more than 80% of the interactions in cellular networks are weak (38). Weak interactions play critical roles in molecular recognition in biological systems, from the classic example of protein folding to recent discoveries in metabolism, gene regulation, and signal transduction. There exist "underground" metabolic reactions that use endogenous metabolites as alternative substrates (39). The binding affinity between the enzyme and the alternative substrate is usually low. In the normal state, underground metabolism may have little impact on the phenotype. However, genetic or drug perturbation can generate or boost underground metabolic reactions, leading to complex phenotypic changes. An example of genetic perturbation is the on/off switch of transcription and epigenetic modification that starts from a long-range all-or-none conformational change in DNA, induced by regulatory protein binding. For example, weak protein-DNA interactions cause the large-scale, all-or-none conformational switch, whereas strong interactions lead to a gradual conformational change in DNA (40). Analysis of data from ChIP-chip experiments indicates that low-affinity protein-DNA binding is common in yeast (41). A highly accurate computational model of the expression patterns in *Drosophila* development suggests that weak DNA binding is as important as strong DNA binding. Cooperative weak binding is necessary to regulate accurately and robustly temporal and spatial expression (42). More recently, analysis of genome-scale DNA-binding sites suggested that nearly half of all transcriptional factors possess secondary low-affinity DNA-binding sites. DNA-binding specificity of transcriptional factors may arise from their differing preferences toward low-affinity sites (43). Weak binding is extensive in protein-protein interactions. For example, a small molecular drug with moderate micromolar-binding affinity to a protein-protein interface may cause observed phenotypic changes (44). The substrate specificity of phosphatases has been determined by the combination of several low-affinity docking interactions (45). Furthermore, weak protein-protein interactions play key roles in regulating integrin signal transduction (46). Micromolar, and even millimolar, binding has also been observed in the transient protein-protein interactions of intrinsically disordered proteins, which are often associated with abnormal cellular regulation and signaling (47).

From the point of view of biological networks, collective weak interactions may have more profound effects on biological systems than a single, strong interaction (48). Weak interactions contribute toward the robustness and diversity of biological networks (49). However, a robust network is fragile to the malfunction of multiple nodes, a phenomenon known as "fail-on" failure (7). Such fail-on failure has been observed in many disease states, such as neurological disorders (50) and cancer (51, 52), where subtle but multiple genetic variations cause significant phenotypic changes. Although their interactions are seemingly minor, micronutrients may have a significant impact on disease progression (53). It has been suggested that multigenic diseases should be treated through the multiple intervention of abnormal components to restore a normal state, rather than knocking out only one or two components (54). Indeed, low-affinity, multitarget noncompetitive NMDA receptor antagonists, which were developed for the treatment of Alzheimer's disease, may have fewer side effects than do high-affinity, single-target drugs (55, 56). In a bistable cellular process regulated by a ligand, switch-like activation can result from

cooperative low-affinity binding (57). The transition threshold can be tuned by intracellular signaling in response to environmental cues. For example, the GSK-3β inhibitor 7AIPM can modulate the progesterone threshold for the maturation of amphibian oocytes (58). In another study, increasing the activation threshold of the Wnt pathway reduced the chance of embryonic abnormalities and diseases in mouse models (59). In addition to targeting disease genes, efficient therapeutics may, thus, be designed to regulate the threshold of biological networks (59).

Moreover, drug-target interactions in vivo are different from those in vitro. In target or cell-based assays, the concentrations of both drug and target are fixed, and binding affinity is measured by thermodynamic equilibrium constants such as IC_{50} values, which reflect binding potency. However, in a living organism, the concentration of the drug, the target, and the endogenous ligand of the target constantly changes with time, rarely reaching equilibrium. Thus, the drug-binding affinity is not an appropriate indicator of drug efficacy in vivo (36). An increasing body of evidence suggests that the drug-target residence time correlates with drug efficacy more strongly than with binding affinity (60, 61). For example, efavirenz, a highly active therapy for the treatment of HIV type 1 infection, has a long residence time of 4.1 h (62). Lu et al. (63) observed a strong correlation between drug-target residence time and drug efficacy in vivo. Although high-affinity binders usually exhibit a long residence time, binding affinity is not necessarily correlated with residence time.

Because most molecular interactions in cellular networks are weak, low-affinity polypharmacological agents may significantly modify the system. Thus, it is time to reconsider the role of low-affinity binding events in polypharmacology. To investigate quantitatively both high- and low-affinity binding on a proteome-wide scale, new computational methods are needed to accurately and efficiently calculate free-energy landscapes in the association and dissociation of protein-ligand complexes. Moreover, translating drug-binding events in vitro to drug phenotypic responses in vivo remains a challenge. The proteome-scale modeling of multiple target interventions of context-specific biological networks may begin to fill this gap.

Quantitative, Proteome-Wide, and Multiscale Modeling of Polypharmacology

The process of defining and predicting polypharmacological effects requires a quantitative understanding of the structure and function of a protein, as well as an understanding of the protein's interaction with small molecules in the context of biological networks (**Figure 1**). In an ideal situation, modeling starts with characterization of the thermodynamics and kinetics of protein-drug interactions at the atomic level, followed by determination of the conformational and chemical states of proteins upon drug binding through allosteric or orthosteric interactions. Molecular dynamics simulation is one of the most powerful tools used for this purpose. Whenever possible, such atomic-level dynamic analyses should be extended to a proteome-wide scale. Bioinformatics and chemoinformatics analysis may help to narrow down the candidates for computationally intensive, molecular dynamics simulations, or to explore the solution space when the three-dimensional (3D) protein structure is not available. Eventually, the physiological or pathological consequences of protein conformational changes upon drug binding should be not only mapped to biological pathways, but also integrated into cellular or tissue-specific networks. The multivariable state of biological networks could be simulated either stoichiometrically using constraint-based modeling or dynamically using mass-action modeling. Network reconstruction and static analysis would provide a solid foundation for dynamic analysis through the identification of cellular circuits. Over the past few years, these fields have developed independently. However, to explore new frontiers in polypharmacology, it is necessary to integrate these disjointed computational

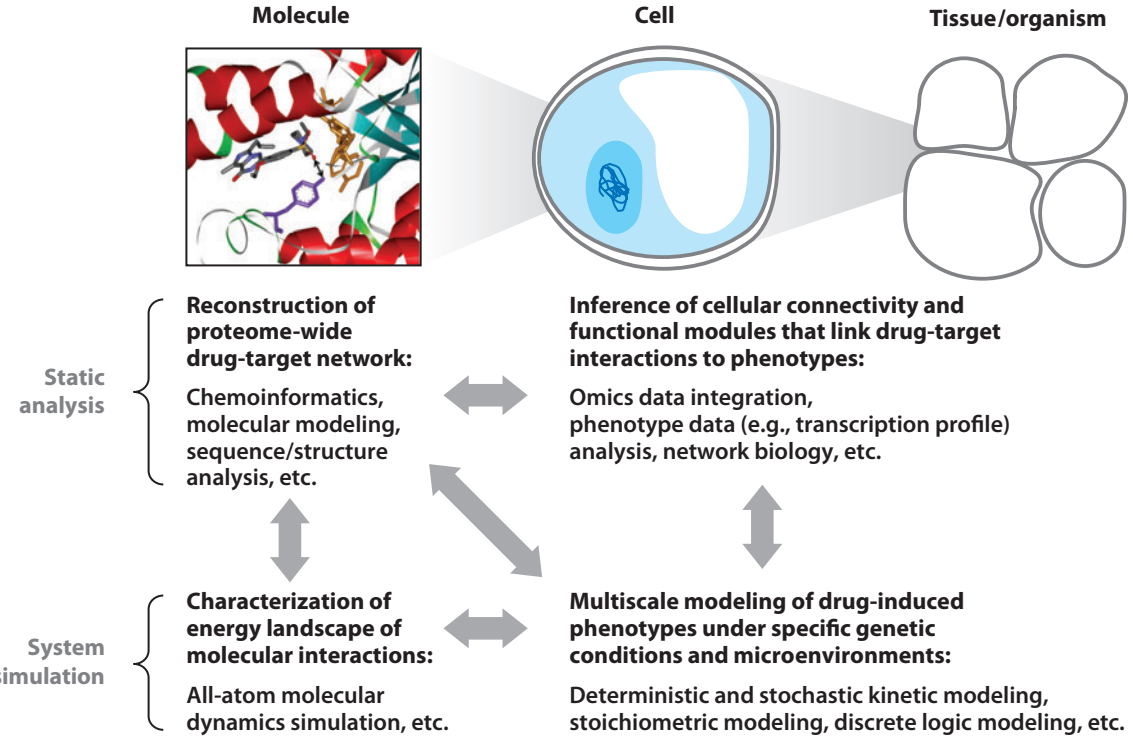

Figure 1

The definition of polypharmacological effects requires multiscale modeling from the atomic details of molecular interactions to dynamic modeling at the level of the organism. The listed computational techniques are the focus of this paper.

techniques into a unified framework. In the remainder of this review, we focus on four computational areas that are critical to polypharmacology: the proteome-wide prediction of drug-target interactions, the quantitative modeling of protein-ligand interactions, the integrated analysis of biological networks, and the dynamic and stoichiometric simulation of biological networks. Related software and Web servers are listed in **Supplemental Table 1** (for all **Supplemental Material**, including **Sections 1–3, Table 1**, and associated terms and references, follow the respective link from the Annual Reviews home page at **http://www.annualreviews.org**). Covered elsewhere are several important topics such as the chemical aspects of polypharmacological drug design (3, 64, 65) and combination therapies (66, 67).

Supplemental Material

TOWARD A PROTEOME-WIDE DRUG-TARGET NETWORK

Overview

The majority of existing work focusing on the prediction of binary drug-drug associations begins with defined signatures of drug molecules (**Figure 2**). The molecular signature could be a topological fingerprint of the chemical compound, or it could be derived from phenotypic responses such as gene expression profiles. The molecular signature allows for the use of statistical analyses, including clustering and classification, to quantify the similarity between drugs. If two drugs share similar structures or phenotypes, then their targets may be related. The relationship between

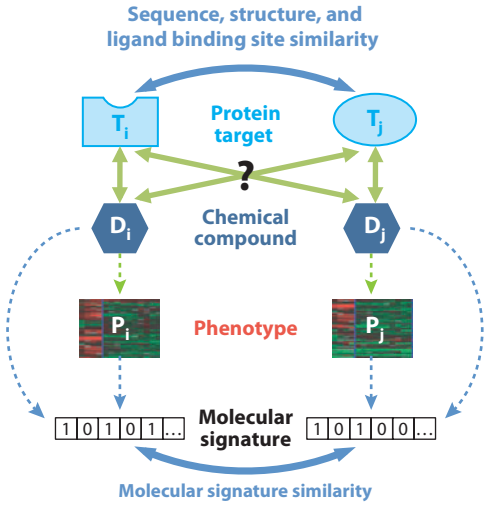

Figure 2

Molecular signatures of drugs can be derived mathematically from their chemical structures or their phenotypic responses. Similarity between molecular signatures implies similarity between drug modes of action. If Drug i (D_i) is known to interact with Target i (T_i), and the signature of Drug j (D_j) is similar to that of D_i, it can be inferred that D_j binds to T_i. Drug-target pairs can be related by comparing targets directly on the basis of similarity between sequence, structure, or ligand binding site: Depicted here are the known drug-target interactions (*green solid arrows*), predicted drug-target interactions (*green solid arrows with question mark*), mathematically derived relationships (*blue dashed arrows*), and experimentally determined properties (*green dashed arrows*).

two targets can be inferred directly using a bioinformatics approach. If two proteins have similar molecular functions, they have the propensity to bind similar ligands. Computational strategies for predicting drug-target associations can, thus, be roughly classified as ligand, phenotype, or target based. The mathematical and computational formulization of these methods is fundamentally similar. The challenge lies in addressing domain-specific issues.

Ligand-Based Prediction of Drug-Drug and Drug-Target Interaction Networks

The hypothesis underlying ligand-based methods states that two ligands with similar chemical structures may exhibit similar bioactivity. Although this hypothesis does not always hold (68), there has been a long history of using small-molecule similarity to establish quantitative structure-activity relationships and virtual screening in drug discovery (69). Small molecules are usually represented as two-dimensional (2D) topological fingerprints encoding atom types and their bond connectivity. Molecular fingerprints can then be used as feature vectors to compare (e.g., using the Tanimoto coefficient) and cluster molecules, or to train statistical regression or classification models. Self-organizing maps, an unsupervised-machine learning technique, has been used to cluster drug molecules (70). Nidhi et al. (71) trained a Bayesian model to classify chemical compounds on the basis of their targets. A notable recent advance in relating proteins by their ligand similarity was the development of the similarity ensemble approach (SEA) (72). Based on an empirical extreme value distribution model, the SEA normalizes the sum of similarity scores between two sets of ligands known to bind to their targets. Several predictions by the SEA have been experimentally confirmed (72, 73). Despite the success of the topological representation of

small molecules, it has been argued that more informative molecular descriptors that are able to capture the essential features of protein-ligand interactions are required (74). Considerable efforts have been made to develop 3D representations of molecules (75–79). Pharmacophore models (80), inferred from active compounds or the structure of protein-ligand complexes, could be used to predict off-targets (81). The success of generalized 3D shape descriptors in predicting drug-target interactions remains to be seen.

Pharmacophore: a description of the spatial and other properties of molecules that are essential for their biological activity

Phenotype-Based Prediction of Drug-Drug and Drug-Target Association Networks

Drugs and their targets can be related by their phenotypic responses, such as gene expression profiles (82, 83), although similar drug phenotypic responses are not always associated with similar drug actions (84). Several reviews have comprehensively assessed the recent advances in applying chemical genomics to establish drug-target relationships (85). One of the central computational challenges in using genomics- and proteomics-based studies to generate molecular signatures is to separate molecular targets and essential genes or protein markers from the noise resulting from cell variations, dosages, and other effects. Once a robust molecular signature is derived, a wide spectrum of computational techniques such as machine learning, statistical analysis, and network analysis can be applied. Iorio et al. (86) developed an automatic and robust method to rank up- and downregulated genes from drug-response transcriptional-profile data. Using this approach, they predicted and confirmed that the Rho-kinase inhibitor fasudil demonstrated strong activities in cellular autophagy (87). Iskar et al. (88) developed a pipeline to filter and normalize drug-response gene expression profiles. The robust assessment of gene expression profiles allowed them to identify target expression changes resulting from drug perturbation. The drug-induced gene expression profile can be directly compared with the gene expression profile from a disease state to reveal disease mechanisms. This strategy was applied to identify the mTOR inhibitor rapamycin as a modulator of glucocorticoid resistance in acute lymphoblastic leukemia (89), to construct a global drug-disease network for the inference of drug mechanisms (90), and to identify common disease modules and pluripotent targets (91).

Low-dimensional phenotypic information such as cell imaging and side effects can be represented as a high-dimensional molecular signature. For example, Young et al. (92) developed a factor analysis technique to profile chemical compounds from image-based cellular phenotypic screening. Campillos et al. (93) used English terms related to drug side effects as a molecular signature to associate a drug with its unknown targets using text mining. Using PubChem bioassay data, Chen et al. (94) constructed a bioassay network that was then mapped to a drug-target network, a protein-protein interaction network, and biological pathways through similar nodes or edges to identify compound-specific profiles and biologically important target pairs. Recent advances in RNAi technology have introduced a simple, yet powerful, approach to investigate the cellular mechanisms of drug action in higher organisms (95). Similar computational techniques developed for chemical genomics and gene expression data analysis can be applied to RNAi-based drug screening. It is expected that additional novel, drug-target, pathway-disease associations will arise from future phenotype screening experiments.

Target-Based Prediction of Drug-Drug and Drug-Target Association Networks

The structural, functional, and evolutionary relationships between proteins provide a natural constraint to reduce the complexity of drug-target association predictions given that proteins and

ligands may have coevolved (96, 97). Thus, protein promiscuity in ligand binding can be inferred from the divergent or convergent evolutionary relationships between proteins by detecting sequence and structure homologs (98, 99). Recently, Chikina et al. (100) showed that the combination of gene expression data and sequence similarity information can improve the prediction of functional homology. Melvin et al. (101) introduced a semantic embedding approach to detect remote evolutionary relationships by exploring the global sequence and structure space of proteins. Several recent publications have combined sequence features of the receptor with the fingerprint of the ligand to train models based on statistical machine learning to predict drug-target interactions (102–105). Unlike sequence comparison, in which statistical models have been developed to rigorously evaluate similarity, there is no practical mathematical framework for protein structure comparison. To address this problem, Poleksic (106) formulized an algorithm that guarantees optimal rigid superposition between protein structures. Liu et al. (107) proposed the use of geodesic distance to rigorously model flexible structure comparisons. Consolidation of these efforts may improve the detection of drug-target associations from protein sequences and structures.

Ligand binding site comparison that is independent of sequence order may detect more drug-target associations than sequence and structure similarity alone because protein cross-reactivity occurs beyond sequence and structure homologs (15). It has recently become a subject of tremendous interest to develop 3D ligand binding site characterization and analysis algorithms (99, 108, 109). The ligand binding site can be represented as a one-dimensional (1D) fingerprint of atomic spatial distributions (110), a 2D graph in which atoms or residues in the structure are nodes and their spatial relationships are encoded as edges (111), or a cloud of atoms or residues in 3D space (112). Although many algorithms developed for 1D vector, 2D graph, or 3D object analysis can be utilized for sequence-order-independent ligand binding site comparison, several challenges remain for protein structures. First, because proteins are intrinsically flexible, the algorithm should be robust to conformational changes in the binding site, yet sensitive to different interaction patterns that contribute toward binding specificity (113). Second, if the algorithm is to be applied on a proteome-wide scale, it should be tolerant to the uncertainty in homology models. Finally, most existing work focuses on small-molecule binding pockets or catalytic active sites. It will be interesting to extend binding site comparison to protein-protein interaction interfaces, antibody-antigen recognition surfaces, and other irregular protein functional sites (114), as they contribute to less-characterized polypharmacology effects yet can play important roles in drug discovery, antibody design, and synthetic biology. Although several recent efforts have explored these less-characterized binding sites (115–118), much more work remains to be done.

Data Integration and Knowledge Discovery

Literature and patent searches are straightforward ways to mine information for protein-ligand interactions. Several public databases focusing on drug-target relationships have been developed to annotate drug-target interactions. DrugBank (119), Therapeutic Target Database (120), SuperTarget, and Matador (121) mine the literature and collect experimental data that are related to drug-target interactions. BindingDB (122), ChEMBL (**https://www.ebi.ac.uk/chembldb**), and PDSP *Ki* (**http://pdsp.med.unc.edu/**) are valuable resources that contain binding-affinity data. The PubChem Bioassay database (123) is one of the largest publicly available repositories of bioactivity data. ChemBank (124) is another public resource of small-molecule screens. PharmGKB collects and disseminates pharmacogenomics data and knowledge (125). A number of Web portals have been developed to integrate multiple drug-target databases. STITCH (126) is a public tool that integrates experimental data, databases, text mining, and homology to infer protein-ligand relationships. PROMISCUOUS (127) accumulates 25,000 drug molecules and

annotates 21,050 drug-target and 104,000 protein-protein associations. ChemProt (128) integrates ChEMBL, WOMBAT & WOMBAT-PK, DrugBank, PubChem, PDSP *K*i, BindingDB, PharmGKB, CTD, and STITCH. It includes more than 700,000 chemicals with bioactivities, 30,578 protein targets, and more than 2 million interactions.

Gene expression profiles are considered as one of the most general signatures of drug activity because no information about the compound is required a priori. The development of the Connectivity Map (CMAP) database for drug-response gene expression (82, 83) and rich expression data from other model systems makes gene expression profile analysis a powerful tool for predicting drug modes of action. However, CMAP is based on a limited number of cell lines and does not provide information about drug-response phenotypes at the level of the tissue and the organism. A notable effort is the use of semantic Web techniques and ontologies to integrate multiple domain knowledge and heterogeneous databases, thereby linking drugs to genes, targets, pathways, and disease phenotypes. Such efforts may facilitate knowledge discovery for polypharmacology (129, 130).

In spite of these advances, proteome-wide drug-target binding networks are a long way from realization. Each of the computational techniques discussed in this section has its own limitations in terms of both predictive power and target coverage. It is important to integrate ligand- and phenotype-based approaches with target-based methodologies, as they are complementary in nature (99). In addition, most existing drug-target databases not only cover limited and biased target space, but also lack the capability to support quantitative data analysis. Finally, although proteomics has generated an increasing amount of information in proteome-wide protein-ligand interactions, there is no central repository or mechanism to store, disseminate, and mine weak protein-ligand interaction data and to integrate it with other resources.

OVERVIEW OF OTHER COMPUTATIONAL TECHNIQUES

Owing to limited space, other critical computational techniques for studying polypharmacology—e.g., quantitative modeling of protein-ligand interactions (**Supplemental Section 1**), statistical and network analysis to link disease- and drug-associated genotypes to phenotypes (**Supplemental Section 2**), and dynamic and stoichiometric simulation of phenotypic changes upon genetic and drug perturbation (**Supplemental Section 3**)—cannot be extensively discussed in this review. Below, we provide a brief overview of these techniques.

Supplemental Material

Quantitative Modeling of Protein-Ligand Interactions

The binary relationship of protein-ligand interactions is not sufficient to infer the phenotypic changes resulting from drug perturbation, because the affinity and specificity of ligand binding are critical to correlate molecular interactions with clinical outcomes. As discussed previously, multiple weak bindings may have a more profound impact on the biological system than a single strong binding. Moreover, the drug response in signal transduction and gene-regulation networks may depend on the relative fold change in drug concentration of the simulation, rather than the absolute concentration (131). For example, the dose response to a drug concentration of 10 nM relative to 1 nM may be the same response as that from 5 nM to 50 nM, but different from 5 nM to 10 nM. Thus, it is necessary to simulate both dynamically and quantitatively the drug-target complex formation in vivo. As a result of next-generation sequencing, personalized medicine is becoming a realistic possibility. Drug efficacy and side effects may strongly depend on individual genetic disposition, as already shown, for example, in warfarin treatment (132). If a nonsynonymous single-nucleotide polymorphism occurs in the ligand binding site or allosteric site, the change in drug

response may result from the variation in drug-target or drug–off-target interactions. Thus, the computational challenge is to accurately and efficiently characterize the change in the free-energy landscape of binding and folding resulting from a single residue mutation. Quantitative structure-activity relationships, protein-ligand docking, and all-atom molecular dynamics simulation are several important techniques to quantitatively assess protein-ligand interactions. Recent advances in algorithm development and tremendous improvement in computer power make them powerful tools in polypharmacology. More details are given in **Supplemental Section 1**.

Supplemental Material

Statistical and Network Analysis to Link Disease- and Drug-Associated Genotypes to Phenotypes

Revealing causal relationships between genotypes and phenotypes is critical to polypharmacology. Not only are disease-associated genes potential drug targets, but the mode of action of drugs also depends on multiple genetic factors. Advances in next-generation sequencing, proteomics, chemical genomics, and functional genomics have generated abundant genotypic and phenotypic data from genome-wide association studies as well as genome-scale chemical perturbation screens. Parallel to the development of new genotyping and phenotyping techniques, a number of novel computational tools have been developed to integrate and analyze genetic, transcriptomic, proteomic, and interactomic data. Such research provides unprecedented opportunities for studying polypharmacology. It is now possible to reveal potential drug targets, genetic and epigenetic factors, and their associated interaction networks, all of which impact our understanding not only of possible drug therapies, but also of side effects.

As discussed in the previous section, there are two types of phenotypes from genome-wide screening: a high-dimensional intermediate phenotype such as a gene expression profile, and a low-dimensional endpoint phenotype such as a side effect or cell-growth rate. An intermediate phenotype can be treated either as a molecular signature of the drug or genetic perturbation, or as a feature vector in the language of machine learning. An endpoint phenotype can be considered as the label of the molecular signature. The computational task is to establish correlations and causal relationships between genetic markers, intermediate phenotypes, and endpoint phenotypes (for a see review, see Reference 133 and references cited therein). For example, Chen et al. (134) developed the CAMELOT (causal modeling expression linkage for complex traits) algorithm that uses both genetic markers and gene expression data to show causal relationships with phenotype as features to predict drug resistance in yeast. Two areas that are of importance to polypharmacology are the prediction and analysis of genetic interactions as well as the identification of functional modules related to drug actions (see **Supplemental Section 2**).

Dynamic and Stoichiometric Simulation of Phenotypic Changes Upon Genetic and Drug Perturbation

Cellular processes can be modeled as an integrated biochemical reaction network, which consists of gene regulatory, signaling transduction, and metabolism modules. The reactions in these network modules span different timescales. For example, most signal cascades and metabolic reactions take a matter of seconds, whereas regulatory events occur within minutes to hours. Traditionally, the dynamic characteristics of these modules were investigated independently. Regulatory networks are mainly studied using discrete logical modeling, ordinary differential equations, piecewise-linear differential equations, or stochastic modeling. Signal transduction networks are typically investigated using discrete logical modeling and stochastic optimization coupled with ordinary differential equations. Computational techniques for metabolic networks

range from classic enzyme kinetics for well-characterized pathways, to constraint-based modeling such as flux balance analysis for genome-scale networks. A dynamic model of a biological network consists of three main components: inputs, outputs, and network structures that include a defined cellular wiring diagram and dynamic parameters (the initial concentration of species and reaction rates). In the context of polypharmacology, there are three basic questions: What is the phenotypic response (network output) given the input and network structure? What is the optimal drug intervention (network input) given the desirable output and network structure? What is the network structure given defined inputs and observed outputs? A wide array of computational techniques has been developed to address these questions, as discussed in **Supplemental Section 3**.

Supplemental Material

CONCLUSIONS

Identifying a single disease-causing gene as a drug target, along with screening high-affinity chemical compounds against that single target, has been the core of drug discovery for the past decade. However, this reductionist view of drug modes of action will be less fruitful in tackling most diseases given the common phenomenon of the multitarget interactions of drugs as well as the stochastic and multivariable nature of biological systems. Drug efficacy, toxicity, and resistance can be rationalized only by taking into account proteome-wide drug-target interactions and their collective effects in the context of biological networks. Pharmacology is traditionally studied at the organism level with a renewed interest in phenotype-based screening. Although the potential therapeutic and side effects of drugs can be observed in phenotypic screening, drug responses in vitro or in model organisms may not be directly related to those in humans. Moreover, the molecular mechanisms of drug actions may not be identified from phenotypic screening in a straightforward manner. Efficient and effective in silico approaches to polypharmacology, which enable an understanding of the information flow of dynamic interactions in cellular processes upon drug and genetic perturbations, may leverage investment in target-based and phenotype-based screening campaigns, thereby bridging the reductionist view of modern molecular biology and the holistic view of in vivo pharmacology. Although recent progress in both computational and experimental techniques has greatly enriched our knowledge of protein-ligand interactions in the context of biological systems, no single technique is sufficient to uncover the whole spectrum of drug-target interactions on either a proteome-wide or multiscale level. New methodologies are required to integrate diverse tools from chemoinformatics, bioinformatics, molecular modeling, and systems biology (135–137) as well as heterogeneous omics data (138–141). The ultimate goal is to predict the polypharmacological effects of drug molecules at the level of the organism.

SUMMARY POINTS

1. Multitarget interactions of drugs are not rare but common. To define and predict drug phenotypic responses, it is necessary to investigate drug-target interactions on a proteome-wide scale.
2. Multiple weak interactions, as opposed to a single strong binding, may play critical roles in drug action. It is important to develop efficient and accurate computational techniques to explore the energy landscape of protein-ligand binding and unbinding events.
3. A drug response is a consequence of complex interactions between multiple intracellular and environmental components. Network analysis and multiscale simulation are key to revealing systems behavior at the level of the cell, tissue, and organism.

FUTURE ISSUES

1. Techniques that reconcile microscopic molecular interactions with the macroscopic dynamic behavior of complete biological systems need to be developed.
2. Methodologies that correlate drug response phenotypes in vitro to those in vivo also need to be developed.
3. Data and knowledge discovery must be integrated to link drugs to targets, genes, pathways, and phenotypes.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health grant number GM078596.

LITERATURE CITED

1. Jain MK, Ridker PM. 2005. Anti-inflammatory effects of statins: clinical evidence and basic mechanisms. *Nat. Rev. Drug Discov.* 4:977–87
2. Xie L, Li J, Xie L, Bourne PE. 2009. Drug discovery using chemical systems biology: identification of the protein-ligand binding network to explain the side effects of CETP inhibitors. *PLoS Comput. Biol.* 5:e1000387
3. Hopkins AL, Mason JS, Overington JP. 2006. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* 16:127–36
4. Hopkins AL. 2007. Network pharmacology. *Nat. Biotechnol.* 25:1110–11
5. Lu L, Payvandi F, Wu L, Zhang LH, Hariri RJ, et al. 2009. The anti-cancer drug lenalidomide inhibits angiogenesis and metastasis via multiple inhibitory effects on endothelial cell function in normoxic and hypoxic conditions. *Microvasc. Res.* 77:78–86
6. Lange RP, Locher HH, Wyss PC, Then RL. 2007. The targets of currently used antibacterial agents: lessons for drug discovery. *Curr. Pharm. Des.* 13:3140–54
7. Kitano H. 2007. A robustness-based approach to systems-oriented drug design. *Nat. Rev. Drug Discov.* 6:202–10
8. Hopkins AL. 2008. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* 4:682–90
9. Knight ZA, Lin H, Shokat KM. 2010. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* 10:130–37
10. Antman EM, DeMets D, Loscalzo J. 2005. Cyclooxygenase inhibition and cardiovascular risk. *Circulation* 112:759–70
11. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. 2007. Drug-target network. *Nat. Biotechnol.* 25:1119–26
12. Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. 2006. Global mapping of pharmacological space. *Nat. Biotechnol.* 24:805–15
13. Mestres J, Gregori-Puigjane E, Valverde S, Sole RV. 2008. Data completeness–the Achilles heel of drug-target networks. *Nat. Biotechnol.* 26:983–84
14. Kinnings SL, Xie L, Fung K, Xie L, Bourne PE. 2010. The *Mycobacterium tuberculosis* drugome and its polypharmacological implications. *PLoS Comp. Biol.* 6:e100976

15. Nobeli I, Favia AD, Thornton JM. 2009. Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* 27:157–67
16. Ji HF, Kong DX, Shen L, Chen LL, Ma BG, Zhang HY. 2007. Distribution patterns of small-molecule ligands in the protein universe and implications for origin of life and drug discovery. *Genome Biol.* 8:R176
17. Li X, Gianoulis TA, Yip KY, Gerstein M, Snyder M. 2010. Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses. *Cell* 143:639–50
18. Dobson PD, Patel Y, Kell DB. 2009. 'Metabolite-likeness' as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Discov. Today* 14:31–40
19. Yan Q. 2010. Membrane transporters and drug development: relevance to pharmacogenomics, nutrigenomics, epigenetics, and systems biology. *Methods Mol. Biol.* 637:1–21
20. Veith H, Southall N, Huang R, James T, Fayne D, et al. 2009. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat. Biotechnol.* 27:1050–55
21. Bouman HJ, Schomig E, van Werkum JW, Velder J, Hackeng CM, et al. 2010. Paraoxonase-1 is a major determinant of clopidogrel efficacy. *Nat. Med.* 17:110–16
22. Piddock LJ. 2006. Multidrug-resistance efflux pumps—not just for resistance. *Nat. Rev. Microbiol.* 4:629–36
23. Lewis K. 1994. Multidrug resistance pumps in bacteria: variations on a theme. *Trends Biochem. Sci.* 19:119–23
24. Fletcher JI, Haber M, Henderson MJ, Norris MD. 2010. ABC transporters in cancer: more than just drug efflux pumps. *Nat. Rev. Cancer* 10:147–56
25. Baysarowich J, Koteva K, Hughes DW, Ejim L, Griffiths E, et al. 2008. Rifamycin antibiotic resistance by ADP-ribosylation: structure and diversity of Arr. *Proc. Natl. Acad. Sci. USA* 105:4886–91
26. Weinshilboum RM, Wang L. 2006. Pharmacogenetics and pharmacogenomics: development, science, and translation. *Annu. Rev. Genomics Hum. Genet.* 7:223–45
27. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N. 2010. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. *Nature* 464:427–30
28. Heidorn SJ, Milagre C, Whittaker S, Nourry A, Niculescu-Duvas I, et al. 2010. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell* 140:209–21
29. Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, et al. 2010. RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. *Nature* 464:431–35
30. Wu M, Pastor-Pareja JC, Xu T. 2010. Interaction between Ras^{V12} and scribbled clones induces tumour growth and invasion. *Nature* 463:545–48
31. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. 2009. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459:428–32
32. Kong HJ, Mooney DJ. 2007. Microenvironmental regulation of biomacromolecular therapies. *Nat. Rev. Drug Discov.* 6:455–63
33. Koschny R, Walczak H, Ganten TM. 2007. The promise of TRAIL—potential and risks of a novel anticancer therapy. *J. Mol. Med.* 85:923–35
34. Mestres J, Gregori-Puigjane E. 2009. Conciliating binding efficiency and polypharmacology. *Trends Pharmacol. Sci.* 30:470–74
35. Rees DC, Congreve M, Murray CW, Carr R. 2004. Fragment-based lead discovery. *Nat. Rev. Drug Discov.* 3:660–72
36. Gleeson MP, Hersey A, Montanari D, Overington J. 2011. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat. Rev. Drug Discov.* 10:197–208
37. Garraway LA, Hahn WC. 2010. On or off target: mutations, models, and predictions. *Sci. Transl. Med.* 2:35ps28
38. Korcsmaros T, Szalay MS, Bode C, Kovacs IA, Csermely P. 2007. How to design multi-target drugs: Target search options in cellular networks. *Expert Opin. Drug Discov.* 2:799–808
39. D'Ari R, Casadesus J. 1998. Underground metabolism. *BioEssays* 20:181–86
40. Akitaya T, Seno A, Nakai T, Hazemoto N, Murata S, Yoshikawa K. 2007. Weak interaction induces an ON/OFF switch, whereas strong interaction causes gradual change: folding transition of a long duplex DNA chain by poly-L-lysine. *Biomacromolecules* 8:273–78

41. Tanay A. 2006. Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* 16:962–72
42. Segal E, Raveh-Sadka T, Schroeder M, Unnerstall U, Gaul U. 2008. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451:535–40
43. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, et al. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science* 324:1720–23
44. Zhai D, Jin C, Satterthwait AC, Reed JC. 2006. Comparison of chemical inhibitors of antiapoptotic Bcl-2-family proteins. *Cell Death Differ.* 13:1419–21
45. Roy J, Cyert MS. 2009. Cracking the phosphatase code: docking interactions determine substrate specificity. *Sci. Signal.* 2:re9
46. Legate KR, Wickstrom SA, Fassler R. 2009. Genetic and cell biological analysis of integrin outside-in signaling. *Genes Dev.* 23:397–418
47. Cheng Y, LeGall T, Oldfield CJ, Mueller JP, Van YY, et al. 2006. Rational drug design via intrinsically disordered protein. *Trends Biotechnol.* 24:435–42
48. Agoston V, Csermely P, Pongor S. 2005. Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example. *Phys. Rev. E* 71:051909
49. Csermely P. 2004. Strong links are important, but weak links stabilize them. *Trends Biochem. Sci.* 29:331–34
50. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, et al. 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–43
51. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, et al. 2008. An integrated genomic analysis of human glioblastoma multiforme. *Science* 321:1807–12
52. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, et al. 2008. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321:1801–6
53. Ames BN. 2006. Low micronutrient intake may accelerate the degenerative diseases of aging through allocation of scarce micronutrients by triage. *Proc. Natl. Acad. Sci. USA* 103:17589–94
54. Xie L, Evangelidis T, Xie L, Bourne PE. 2011. Drug discovery using chemical systems biology: Weak inhibition of multiple kinases may contribute to the anti-cancer effect of Nelfinavir. *PLoS Comp. Biol.* 7:e1002037
55. Lipton SA. 2004. Turning down, but not off. *Nature* 428:473
56. Youdim MB, Buccafusco JJ. 2005. Multi-functional drugs for various CNS targets in the treatment of neurodegenerative disorders. *Trends Pharmacol. Sci.* 26:27–35
57. Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, et al. 2001. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 414:514–21
58. Justman QA, Serber Z, Ferrell JE Jr, El-Samad H, Shokat KM. 2009. Tuning the activation threshold of a kinase network by nested feedback loops. *Science* 324:509–12
59. Buchert M, Athineos D, Abud HE, Burke ZD, Faux MC, et al. 2010. Genetic dissection of differential signaling threshold requirements for the Wnt/β-catenin pathway in vivo. *PLoS Genet.* 6:e1000816
60. Copeland RA, Pompliano DL, Meek TD. 2006. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov.* 5:730–39
61. Lu H, Tonge PJ. 2010. Drug-target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.* 14:467–74
62. Braz VA, Holladay LA, Barkley MD. 2010. Efavirenz binding to HIV-1 reverse transcriptase monomers and dimers. *Biochemistry* 49:601–10
63. Lu H, England K, am Ende C, Truglio JJ, Luckner S, et al. 2009. Slow-onset inhibition of the FabI enoyl reductase from *Francisella tularensis*: residence time and in vivo activity. *ACS Chem. Biol.* 4:221–31
64. Apsel B, Blair JA, Gonzalez B, Nazif TM, Feldman ME, et al. 2008. Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat. Chem. Biol.* 4:691–99
65. Morphy R, Rankovic Z. 2007. Fragments, network biology and designing multiple ligands. *Drug Discov. Today* 12:156–60
66. Jia J, Zhu F, Ma X, Cao Z, Li Y, Chen YZ. 2009. Mechanisms of drug combinations: interaction and network perspectives. *Nat. Rev. Drug Discov.* 8:111–28

67. Calzolari D, Bruschi S, Coquin L, Schofield J, Feala JD, et al. 2008. Search algorithms as a framework for the optimization of drug combinations. *PLoS Comput. Biol.* 4:e1000249
68. Brown RD, Martin YC. 1996. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. *J. Chem. Inf. Comput. Sci.* 36:572–84
69. Varnek A, Tropsha A. 2008. *Chemoinformatics: An Approach to Virtual Screening*. Cambridge, UK: R. Soc. Chem.
70. Schneider P, Tanrikulu Y, Schneider G. 2009. Self-organizing maps in drug discovery: compound library design, scaffold-hopping, repurposing. *Curr. Med. Chem.* 16:258–66
71. Nidhi, Glick M, Davies JW, Jenkins JL. 2006. Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. *J. Chem. Inf. Model.* 46:1124–33
72. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. 2007. Relating protein pharmacology by ligand chemistry. *Nat. Biotechnol.* 25:197–206
73. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, et al. 2009. Predicting new molecular targets for known drugs. *Nature* 462:175–81
74. Schneider G. 2010. Virtual screening: an endless staircase? *Nat. Rev. Drug Discov.* 9:273–76
75. Ballester PJ, Finn PW, Richards WG. 2009. Ultrafast shape recognition: evaluating a new ligand-based virtual screening technology. *J. Mol. Graph. Model.* 27:836–45
76. Ballester PJ, Westwood I, Laurieri N, Sim E, Richards WG. 2010. Prospective virtual screening with Ultrafast Shape Recognition: the identification of novel inhibitors of arylamine *N*-acetyltransferases. *J. R. Soc. Interface* 7:335–42
77. Rush TS III, Grant JA, Mosyak L, Nicholls A. 2005. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* 48:1489–95
78. Jenkins JL, Glick M, Davies JW. 2004. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J. Med. Chem.* 47:6144–59
79. Jain AN. 2000. Morphological similarity: a 3D molecular similarity method correlated with protein-ligand recognition. *J. Comput. Aided Mol. Des.* 14:199–213
80. Guner OFE. 2000. *Pharmacophore Perception, Development, and Use in Drug Design*. La Jolla, CA: Int. Univ. Line
81. Klabunde T, Evers A. 2005. GPCR antitarget modeling: pharmacophore models for biogenic amine binding GPCRs to avoid GPCR-mediated side effects. *ChemBioChem* 6:876–89
82. Lamb J. 2007. The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer* 7:54–60
83. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–35
84. Fliri AF, Loging WT, Volkmann RA. 2009. Drug effects viewed from a signal transduction network perspective. *J. Med. Chem.* 52:8038–46
85. Bender A, Young DW, Jenkins JL, Serrano M, Mikhailov D, et al. 2007. Chemogenomic data analysis: prediction of small-molecule targets and the advent of biological fingerprint. *Comb. Chem. High Throughput Screen.* 10:719–31
86. Iorio F, Tagliaferri R, di Bernardo D. 2009. Identifying network of drug mode of action by gene expression profiling. *J. Comput. Biol.* 16:241–51
87. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaokar P, et al. 2010. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl. Acad. Sci. USA* 107:14621–26
88. Iskar M, Campillos M, Kuhn M, Jensen LJ, van Noort V, Bork P. 2010. Drug-induced regulation of target expression. *PLoS Comput. Biol.* 6:e1000925
89. Wei G, Twomey D, Lamb J, Schlis K, Agarwal J, et al. 2006. Gene expression–based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. *Cancer Cell* 10:331–42
90. Hu G, Agarwal P. 2009. Human disease-drug network based on genomic expression profiles. *PLoS ONE* 4:e6536
91. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. 2010. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput. Biol.* 6:e1000662

92. Young DW, Bender A, Hoyt J, McWhinnie E, Chirn GW, et al. 2008. Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat. Chem. Biol.* 4:59–68
93. Campillos M, Kuhn M, Gavin AC, Jensen LJ, Bork P. 2008. Drug target identification using side-effect similarity. *Science* 321:263–66
94. Chen B, Wild D, Guha R. 2009. PubChem as a source of polypharmacology. *J. Chem. Inf. Model.* 49:2044–55
95. Jiang H, Pritchard JR, Williams RT, Lauffenburger DA, Hemann MT. 2010. A mammalian functional-genetic approach to characterizing cancer therapeutics. *Nat. Chem. Biol.* 7:92–100
96. Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE. 2000. Co-evolution of proteins with their interaction partners. *J. Mol. Biol.* 299:283–93
97. Dupont CL, Yang S, Palenik B, Bourne PE. 2006. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry. *Proc. Natl. Acad. Sci. USA* 103:17822–27
98. Whisstock JC, Lesk AM. 2003. Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.* 36:307–40
99. Xie L, Xie L, Bourne PE. 2011. Structure-based systems biology for analyzing off-target binding. *Curr. Opin. Struct. Biol.* 21:189–99
100. Chikina MD, Troyanskaya OG. 2011. Accurate quantification of functional analogy among close homologs. *PLoS Comput. Biol.* 7:e1001074
101. Melvin I, Weston J, Noble WS, Leslie C. 2011. Detecting remote evolutionary relationships among proteins by large-scale semantic embedding. *PLoS Comput. Biol.* 7:e1001047
102. Yamanishi Y, Araki M, Gutteridge A, Honda W, Kanehisa M. 2008. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24:i232–40
103. Nagamine N, Shirakawa T, Minato Y, Torii K, Kobayashi H, et al. 2009. Integrating statistical predictions and experimental verifications for enhancing protein-chemical interaction predictions in virtual screening. *PLoS Comput. Biol.* 5:e1000397
104. Vina D, Uriarte E, Orallo F, Gonzalez-Diaz H. 2009. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Mol. Pharm.* 6:825–35
105. He Z, Zhang J, Shi XH, Hu LL, Kong X, et al. 2010. Predicting drug-target interaction networks based on functional groups and biological features. *PLoS ONE* 5:e9603
106. Poleksic A. 2009. Algorithms for optimal protein structure alignment. *Bioinformatics* 25:2751–56
107. Liu W, Srivastava A, Zhang J. 2011. A mathematical framework for protein structure comparison. *PLoS Comput. Biol.* 7:e1001075
108. Kellenberger E, Schalon C, Rognan D. 2008. How to measure the similarity between protein ligand-binding sites. *Curr. Comput. Aided Drug Des.* 4:209–20
109. Perot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. 2010. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* 15:656–67
110. Xiong B, Wu J, Burk DL, Xue M, Jiang H, Shen J. 2010. BSSF: a fingerprint based ultrafast binding site similarity search and function analysis server. *BMC Bioinforma.* 11:47
111. Mernberger M, Klebe G, Hullermeier E. 2011. SEGA: Semi-Global Graph Alignment for structure-based protein comparison. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 8(5):1330–43
112. Hoffmann B, Zaslavskiy M, Vert JP, Stoven V. 2010. A new protein binding pocket similarity measure based on comparison of clouds of atoms in 3D: application to ligand prediction. *BMC Bioinform.* 11:99
113. Chen BY, Honig B. 2010. VASP: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. *PLoS Comput. Biol.* 6:e1000881
114. Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009. Protein sectors: evolutionary units of three-dimensional structure. *Cell* 138:774–86
115. Gao M, Skolnick J. 2010. iAlign: a method for the structural comparison of protein-protein interfaces. *Bioinformatics* 26:2259–65
116. Gao M, Skolnick J. 2010. Structural space of protein-protein interfaces is degenerate, close to complete, and highly connected. *Proc. Natl. Acad. Sci. USA* 107:22517–22
117. Bouvier B, Grunberg R, Nilges M, Cazals F. 2009. Shelling the Voronoi interface of protein-protein complexes reveals patterns of residue conservation, dynamics, and composition. *Proteins* 76:677–92

118. Zhang QC, Petrey D, Norel R, Honig BH. 2010. Protein interface conservation across structure space. *Proc. Natl. Acad. Sci. USA* 107:10896–901
119. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. 2008. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 36:D901–6
120. Chen X, Ji ZL, Chen YZ. 2002. TTD: Therapeutic Target Database. *Nucleic Acids Res.* 30:412–15
121. Gunther S, Kuhn M, Dunkel M, Campillos M, Senger C, et al. 2008. SuperTarget and Matador: resources for exploring drug-target relationships. *Nucleic Acids Res.* 36:D919–22
122. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. 2007. BindingDB: a Web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 35:D198–201
123. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. 2009. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 37:W623–33
124. Seiler KP, George GA, Happ MP, Bodycombe NE, Carrinski HA, et al. 2008. ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res.* 36:D351–59
125. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, et al. 2001. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. *Pharmacogenomics J.* 1:167–70
126. Kuhn M, Szklarczyk D, Franceschini A, Campillos M, von Mering C, et al. 2009. STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res.* 38:D552–56
127. von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R. 2010. PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res.* 39:D1060–66
128. Taboureau O, Nielsen SK, Audouze K, Weinhold N, Edsgard D, et al. 2011. ChemProt: a disease chemical biology database. *Nucleic Acids Res.* 39:D367–72
129. Chen B, Dong X, Jiao D, Wang H, Zhu Q, et al. 2010. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinform.* 11:255
130. Qu XA, Gudivada RC, Jegga AG, Neumann EK, Aronow BJ. 2009. Inferring novel disease indications for known drugs by semantically linking drug action and disease mechanism relationships. *BMC Bioinform.* 10(Suppl. 5):S4
131. Ferrell JE Jr. 2009. Signaling motifs and Weber's law. *Mol. Cell* 36:724–27
132. Klein TE, Altman RB, Eriksson N, Gage BF, Kimmel SE, et al. 2009. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* 360:753–64
133. Li Y, Tesson BM, Churchill GA, Jansen RC. 2010. Critical reasoning on causal inference in genome-wide linkage and association studies. *Trends Genet.* 26:493–98
134. Chen BJ, Causton HC, Mancenido D, Goddard NL, Perlstein EO, Pe'er D. 2009. Harnessing gene expression to identify the genetic basis of drug resistance. *Mol. Syst. Biol.* 5:310
135. Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ. 2010. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comp. Biol.* 6:e1000938
136. Obiol-Pardo C, Gomis-Tena J, Sanz F, Saiz J, Pastor M. 2011. A multiscale simulation system for the prediction of drug-induced cardiotoxicity. *J. Chem. Inf. Model.* 51:483–92
137. Khalili S, Monaco JM, Armaou A. 2010. Development of a stochastic model for the efficacy of NRTIs using known mechanisms of action. *J. Theor. Biol.* 265:704–17
138. Schmid A, Blank LM. 2010. Systems biology: hypothesis-driven omics integration. *Nat. Chem. Biol.* 6:485–87
139. Joyce AR, Palsson BØ. 2006. The model organism as a system: integrating 'omics' data sets. *Nat. Rev. Mol. Cell Biol.* 7:198–210
140. Palsson B, Zengler K. 2010. The challenges of integrating multi-omic data sets. *Nat. Chem. Biol.* 6:787–89
141. Wiley HS. 2011. Integrating multiple types of data for signaling research: challenges and opportunities. *Sci. Signal.* 4:pe9

Annual Review of Pharmacology and Toxicology

Volume 52, 2012

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Pharmacology and Toxicology* articles may be found at http://pharmtox.annualreviews.org/errata.shtml