Chemistry & Biology
Minireview

Systems Biology and Systems Chemistry: New Directions for Drug Discovery

J.B. Brown[1] and Yasushi Okuno[1,*]
[1]Department of Systems Bioscience for Drug Discovery, Graduate School of Pharmaceutical Sciences, Kyoto University,
Kyoto 606-8501, Japan
*Correspondence: okuno@pharm.kyoto-u.ac.jp

DOI 10.1016/j.chembiol.2011.12.012

Improvements in drug design have historically been centered around structure-based optimization of molecule specificity for a targeted protein, in an effort to reduce unintentional binding to other proteins and off-target effects. Although the "one-to-one" drug design strategy has been successful in impairing function of targets associated with a number of diseases, recent reports of drug promiscuity, which are a potential source of adverse reactions in patients, make a case to refine the drug design strategy such that it includes an awareness of multiple interactions from both ligand and protein perspectives. Polypharmacology and chemical biology studies are amassing interaction data at rapid rates, and the integration of such data into an interpretable model requires zooming our perspective out from the single ligand-target level to the larger network-wide level. We review some of the recent developments in systems-level research for drug design and discovery, and discuss the directions that some drug design efforts are heading toward.

Single-Target Drug Design and Side Effects

The information obtained from analysis of protein three-dimensional structure often kick-starts the modern drug design process, providing information for designing small molecules that are complementary to the shape of an investigated protein. Thus, the basic idea here is to take advantage of complementarity built into designed small molecules to inactivate the protein function and stop downstream cellular processes. Subsequent design revisions are typically aimed at optimizing the binding, reducing molecular weight, or increasing lipophilicity, which all improve the molecule's drug-like properties. With the assistance of computational techniques, a number of drugs have been refined with this mindset, and a number of reviews have discussed the impact made by computational drug design (Schneider and Fechner, 2005; Jacoby, 2011).

Although an appreciable number of drugs have been designed with a particular protein as the target, and although the desired effects of drugs are observable in a statistically significant manner during early in vitro and in vivo development stages, unexpected side effects continue to be a problem. This is frequently a contributor to failure in clinical trials, where the drug that showed success in vitro fails to work as intended when placed in a more complex in vivo environment. The fact that drug side effects exist indicates that there is something in the underlying assumptions about drug design that needs to be questioned and refined. Immediately, several questions come to mind. Even though a drug has been designed to target a single protein, it is important to question whether such a design is truly optimal for inhibiting a disease process which involves the orchestration of multiple receptors and proteins passing signals to each other. Additionally, is the assumption of a simple linear signaling cascade from ligand binding to the effect a valid one? Finally, is the designed drug molecule truly binding to only the intended protein?

In recent drug design, we have adjusted our mindset from the traditional one-protein-one-ligand model to incorporate the view that the underlying response mechanisms activated by the ligand stimulation are a network of processes, and that it is this network or systems biology level of understanding that needs to be considered for advancing our knowledge of the consequences of single-target drug design, including side effects. Much like the Internet that is a dynamic, multiple connection network system to transmit information, we believe the next generation of molecule design will need to consider a more dynamic, multipathway system that a drug can take to exert an effect.

Here, we will introduce ligand multi-targeting properties and how the properties can be used to develop a different kind of small molecule effectors that intentionally interact at the network level rather than at the single target level. We will illustrate our ideas with recent examples that highlight the issue and provide potential solutions, by drawing from both computational and chemical biology efforts.

Multiple Interactions between Ligands and Target Proteins

As mentioned, thus far drug design strategies by and large employed the one-protein-one-ligand model. However, in the past decade, the emergence of chemical biology is increasingly influencing the directions that drug design proceeds in, because it provides a platform for improving our understanding of how small molecules impact the underlying functional frameworks that connect intracellular macromolecules (Schreiber, 2005). Chemical biology approaches are being applied in a wide variety of research model organisms including yeast (Hübel, 2009), chickens (Yamamoto et al., 2011), and mice (Chen et al., 2011). In addition to chemical biology, polypharmacology (MacDonald et al., 2006) has also gained attention in recent years as efforts to understand biological network signaling cascades that are perturbed by drug stimulation have increased. Such studies demonstrate the "many-to-many" nature of compound-protein interactions, as opposed to the traditional "one-to-one" model, and it is believed that drug side effects partially result from unintentional binding. The reasons for these unexpected reactions are complex; molecular shape and properties, proteins' allosteric behavior, pathway perturbation, and pharmaco-kinetics/dynamics are all influential.

First, we need to consider the extent to which a target protein will bind to a wide variety of ligands. Hopkins (2007) performed a large-scale systematic study of polypharmacology from a protein perspective, finding that many G protein-coupled receptors (GPCRs) were able to bind a large number of ligands. In addition to GPCRs such as histamine H2 and the α2B adrenergic receptor, three types of cytochrome P450 enzymes and protein kinase C-delta, among many other proteins, have also been reported to bind to over 100 ligands at an activity threshold of 10 μM (Paolini et al., 2006). We briefly evaluated target promiscuity by investigating ligand specificity of proteins found in the ChEMBL SARfari databases (http://www.ebi.ac.uk/chembl), identifying the targets in which at least 100 ligands were reported with exact nanomolar inhibitory values (GPCRs: Ki; kinases: IC_{50}). The numbers of proteins matching this criterion were 129 for GPCRs and 90 for kinases. This calculation is an underestimate of target promiscuity, as we excluded ligands with inexact nM or uM inhibitory values and set a rather high threshold for the number of protein ligands, which resulted in excluding a number of less promiscuous yet still functionally important proteins such as GPCRs EDG 1/2/3/4/7 involved in endothelial differentiation and CCR 1/2/4/5/8 chemokine receptors, and apoptosis pathway kinases DAPK 1/2/3 and DYRK 1A/2/3.

One should also consider the extent to which drug molecules designed for a particular protein are actually promiscuous and bind to many proteins. In previous research, we developed the GPCR-ligand database (GLIDA) (Okuno et al., 2008), a GPCR-specific resource that has cataloged GPCR-ligand interactions, containing visualization maps that both demonstrate polypharmacology and allow one to rapidly identify promiscuous compounds such as clozapine, an example of a well-known antipsychotic drug that is considered "effectively promiscuous" (Hopkins et al., 2006). Using ChEMBL, we scanned both GPCR and kinase interaction data to get a first-hand glimpse into small molecule and ligand promiscuity. Based on the proteins from the above analysis, we found that 19,528 of 33,237 ligands (59%) in the kinase database have exact nanomolar inhibitory values. When we eliminated those ligands that were bound to only a single target, there were still 6,942 promiscuous molecules (36%) left. For ChEMBL's GPCR database, 35,090 of 118,013 (30%) molecules bind to the reduced set of multi-ligand receptors. Even after filtration of single-target ligands, the set of GPCR ligands still contained 18,564 (53%) promiscuous, multi-target molecules. These values should be considered as an underestimate of the situation as the search conditions were highly constrained.

Thus, the databases accumulating chemical biology data reveal a more realistic view of many-to-many interactions between proteins and ligands, suggesting that consideration of multiple interactions could be a key to improving drug design strategies.

Widening the Vision of Scope to the Network Level

These target and ligand promiscuities can be regarded as a very reasonable mechanism for subtle control of complicated biological systems because the existence of multiple combinations of the limited number of endogenous ligands and target proteins is a potential reason for explaining the diversity of input patterns into the downstream signaling pathways, suggesting that understanding and controlling of the promiscuities is a critical issue for drug design. Up to the present, ligand promiscuity has not been the explicit intent of drug design. Hence, the idea of polypharmacology suggests that principles that govern drug design should be reconsidered, and that drug design should be undertaken with a broader perspective. Another way of thinking about the opportunity presented by evidence of polypharmacology is that we should attempt to design "network-oriented" drugs. Considering entire networks, even social or logistical, allows one to derive a "context" for a subset of the network, where the context is often derived dynamically from the neighboring nodes. For network-oriented drugs, the context is the signaling network with multiple entry points, a design philosophy in sharp contrast to the "network-less" (single molecular target) designs that are intuitive when limited to a one-to-one local scope. Similar arguments, questioning whether the use of the reductionist approach typically applied in chemistry and physics is appropriate for drug design, given that reduction to network pieces does not provide directions for network reassembly, were made recently (Maggiora, 2011).

Once we accept that drug design needs to incorporate a "many-to-many" network approach, the next issue to solve is how to create and validate such a ligand-protein network. Next, we discuss several related computational drug discovery methods, including a network-oriented approach with experimental validation.

A New Drug Design Approach Based on Machine Learning of Network-Wide Interaction Space between Chemistry and Biology

GPCRs are involved in vision, smell, immune system activity, and many other high-level physiological functions. Binding of extracellular ligands to GPCRs affects cellular internal downstream signaling, which has a crucial impact on the function of an organism. The downstream signal processes are complex nonlinear relationships, reaffirming our need to shift the ligand design mindset from one-to-one to a larger network perspective. Although at least 300 GPCRs are of therapeutic interest, drugs currently available on the market target <10% of them (Okuno et al., 2008) and much of the GPCR-ligand interaction space remains to be explored. Using the 39,000 interactions available in GLIDA for exploration of new regions in GPCR interaction space not only uncovers new polypharmacological interactions that could be contributing to drug side effects, but equally important, represents the potential for identification of starting points to develop new drugs with alternative scaffolds and binding modes.

In a recent report (Yabuuchi et al., 2011), we proposed a drug design concept, "Chemical Genomics-Based Drug Design" (CGBDD) for incorporating system-level multi-interaction networks connecting chemistry and biology. To better understand GPCRs from a systems perspective, we implemented a new method for the computational prediction of novel GPCR-ligand interactions, so called polypharmacological interactions.

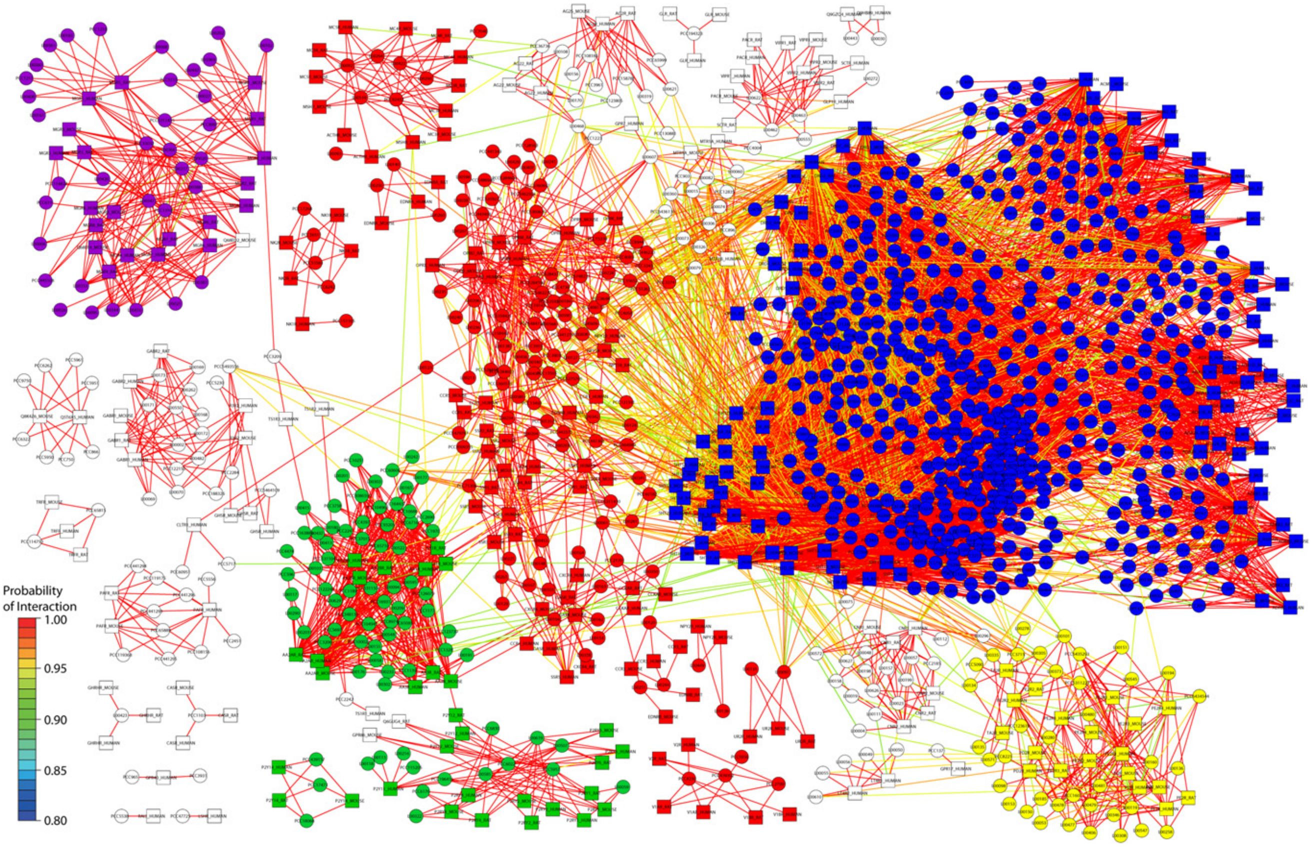

Figure 1. Complexity of Interaction Networks
Using known GPCR-ligand interactions, chemical genomics-based virtual screening was applied to discover novel interactions. When a GPCR-ligand interaction is predicted at a bioactivity of 10 uM, a line between the GPCR and ligand is drawn in the interaction network. The resulting network, including known interactions, demonstrates the complexity of interactions between chemical and biological spaces, and reinforces the need to shift the drug design strategy from "one-ligand-one-protein" to "many-to-many." The node color indicates the classes that compounds and GPCRs belong to (blue, amines; red, peptides; yellow, prostanoids; green, nucleotides). The links colored from green to yellow to red indicate increasing confidence in the GPCR-ligand interaction, with a number of interclass GPCR-ligand interactions exhibiting high predictive confidence.

Generalizing beyond GPCRs, our overall goal was to create a technique that leverages an interaction database for construction of a model that accurately and compactly expresses interaction patterns, such that the model has sufficient predictive performance in translational testing. Although the idea of multi-interaction data mining has been explored by several other groups (Jacob and Vert, 2008; Wassermann et al., 2009), only a very limited number of computational results have actually been tested in experimental assays.

Although massive HTS studies are now possible, it is important that drug discovery costs do not inflate simply because it is easy to perform all possible assays. Thus, virtual screening (VS) will become essential for efficiently reducing the number of possible chemical candidates and bioactivities to assay. However, the existing widely-used VS approaches are based only on a one-to-one mindset, such as structure-based VS (SBVS) and ligand-based VS (LBVS) methods. The CGBDD is a drug design concept for leveraging multiple compound-protein interaction networks (Figure 1), and has been implemented to develop a virtual screening method, called CGBVS (Yabuuchi et al., 2011). The CGBVS approach requires less computational time and complexity than existing SBVS methods, and overcomes problems encountered when performing LBVS for orphaned receptors. Its goal is to efficiently mine the broad, global chemical data space before using SBVS or LBVS with a reduced number of drug scaffolds.

As mentioned at the beginning, recent in silico development of drug leads and pharmaceuticals is being aided by the use of computational techniques, and in the CGBVS approach we employ machine learning, an active research field that develops computational algorithms to extract statistically meaningful information from large data sets. For those unfamiliar with machine learning, it is easy to think of how a human child learns to distinguish colors or shapes, after which they can cluster new, unseen objects of "similar" color or shape together. Thus, machine learning is a critical tool for the extraction and representation of patterns existing in protein-ligand interactions that can then be subsequently applied to drug lead discovery and optimization.

Below, we briefly describe the CGBVS interaction prediction procedure (Figure 2). For proteins in known and hypothetical query interaction pairs, an analysis technique counting the frequency of all dipeptide sub-sequences is applied; no structural information is required. We apply vector representation to

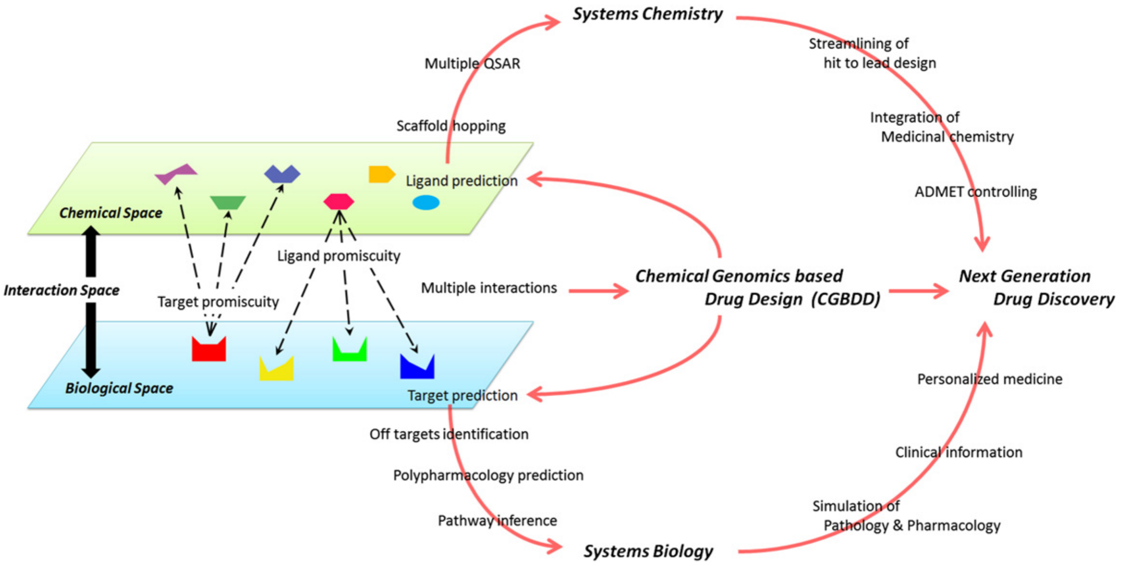

Figure 2. Roadmap for Systems-Level Drug Design
Similar to the CGBVS method that can utilize protein and ligand promiscuity, the incorporation of multiple interactions (polypharmacology and chemical biology) will be a key for advancements in drug design. Systems-level awareness of chemical and biological processes and the feedback resulting from clinical application will provide constraints to guide future generations of molecule design for producing medicines that are more personal and contain fewer side effects.

describe a ligand's chemical structure and physicochemical properties, such as atom counts, topology patterns, connectivity frequencies, electrostatic properties, lipophilic character, and other measurable types of information. For known (non-)interactions, we also create a piece of information (“bind”, “non-bind”) representing bioactivity at a modifiable threshold of 10 μM. The various pieces of information are concatenated per interaction example, and all examples are combined to form the dataset used for predictive model construction. We tested CGBVS against pure SBVS and LBVS methods using statistical validation techniques and obtained improved prediction performance. This result indicates that this representation of chemical-protein interactions is effective in identifying frequent patterns in physicochemical properties in order to model complex interactions, and the machine learning approach provides a way to analyze the nonlinear interaction data in such a way that different scaffolds can still be analyzed and clustered because of their similar physicochemical characteristics. It provides a new opportunity to perform reverse design from desired physicochemical characteristics to potential scaffolds, which the medicinal chemist can then optimize.

Using the full set of interactions in the GLIDA database, we used the above procedure to derive a model for prediction of novel GPCR ligands. Using the external Bionet chemical library (Key Organics Ltd., Cornwall, UK), the top 30 *novel* β2AR interactions predicted by CGBVS were tested in calcium mobilization assays. Of those candidates, 9 of 30 compounds had EC_{50} or IC_{50} values in the nM–μM range. With a similar procedure, cAMP assays confirmed novel interactions in 3 of 20 neuropeptide type 1 (NPY1R) candidates that we predicted. Changing the focus from GPCRs to kinases, we performed similar experiments for epidermal growth factor receptor (EGFR) and cyclin-dependent kinase 2 (CDK2), which are being considered as targets for anticancer therapy. Off-chip mobility shift assays resulted in novel molecule interaction hit rates of 25% (EGFR, 5/20) and 10% (CDK2, 2/20). These hit rates are improvements relative to the typical success rates encountered when screening entire chemical libraries, although even more importantly, many of the assay hits were compounds with scaffolds different from the known ligands for each of the targets. All of the interactions and novel scaffolds uncovered are published (Yabuuchi et al., 2011). As the same technique was applicable to GPCRs, kinases, ion channels, and other types of proteins, we anticipate its concept will be repeatedly applied in the paradigm shift from one-protein-one-ligand to a more complete, systems-level multi-interaction network.

A recent related study described a ligand-based approach to predict ligand-protein binding propensities by using no protein sequence information but rather only a database of existing ligands (Keiser et al., 2009). The success of this method reiterates the importance of the many-to-many drug design strategy. One major strength of the CGBDD philosophy is that it can characterize multiple compound-protein interactions to not only explore biological space and find new targets for existing drugs (polypharmacology), but also explore chemical space and find new drugs for existing targets (chemical biology). Equally important, the CGBVS method has the ability to discover novel scaffolds in unexplored chemical regions through effective utilization of existing compound-protein interaction patterns, which can link systems biology and systems chemistry.

Into the Future

A vast amount of bioassay data is now accumulating in databases such as PubChem. The massive screening data could expand our knowledge regarding chemical-biological interaction

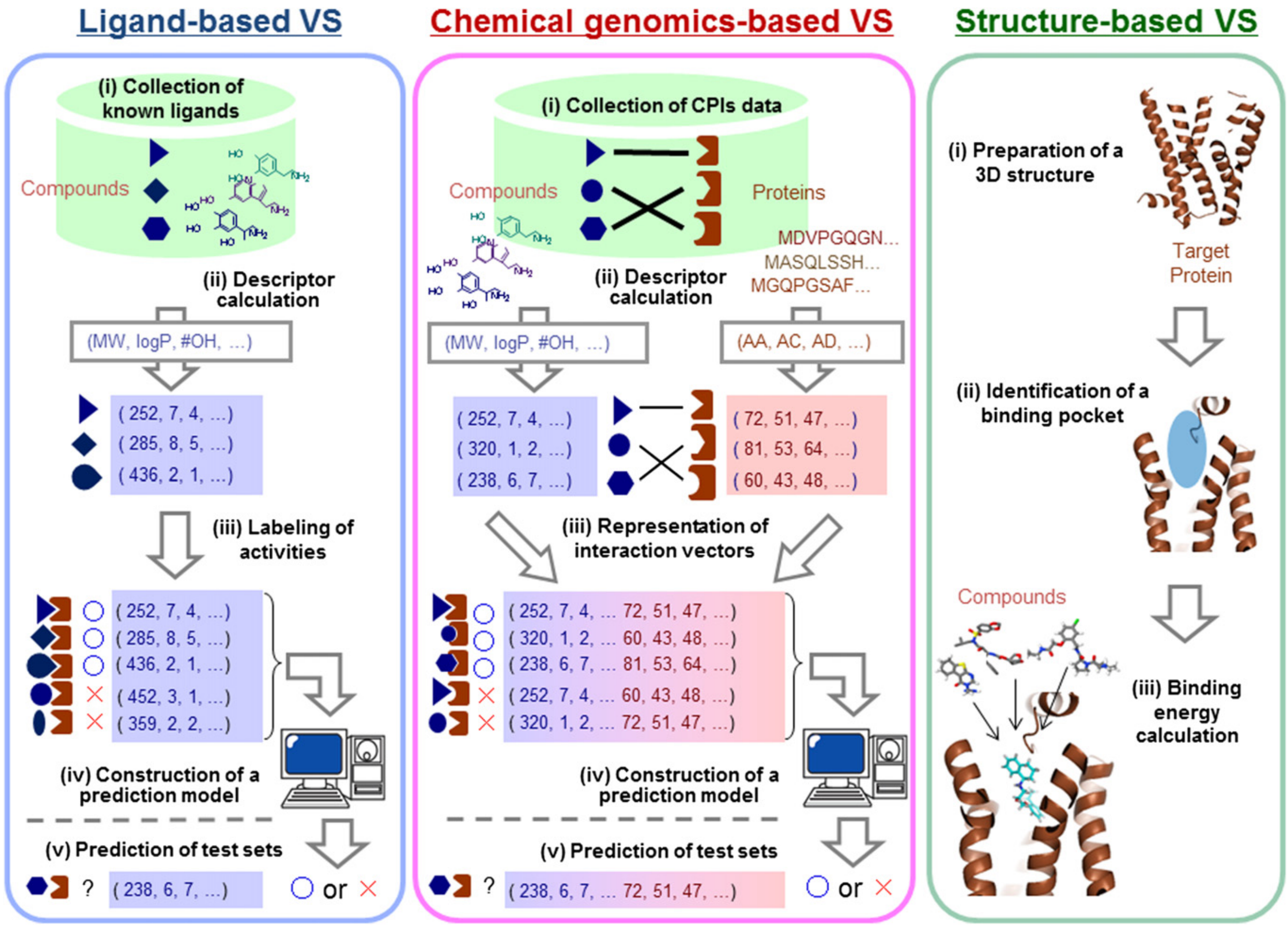

Figure 3. Compound-Prediction Interaction Scheme

The CGBVS ligand-protein interaction discovery method (center) is a change in exploring the interface between chemistry and biology, not requiring the protein three-dimensional structure required in traditional SBVS (right), nor limited in scope to a single protein as is the case in LBVS (left). In CGBVS, protein subsequences and chemical descriptions of topology and other physicochemical properties are combined for each known interaction and non-interaction. The set of (non-)interactions is used to build a predictive model that can rank novel ligand-protein interactions for prioritization in bioassay experiments.

space to a size and complexity we have yet to see. Although the concept of one-ligand-one-target is simple and easy to understand, it risks narrowing our interpretation of interaction space to the point where it prohibits us from understanding interaction space as a whole. In fact, the study with our CGBVS method successfully uncovered new ligands for multiple protein families by leveraging existing experimental assays with a representation appropriate for interaction mining. By incorporating more polypharmacology and chemical biology data, interaction prediction programs should continue to become more accurate as time progresses.

The present CGBVS starts with contexts of protein and chemical descriptors (Figure 2). The emerging tools along with the CGBDD concept will serve as the foundation for a new generation of drug discovery tools that take extended contexts from the dual viewpoints of systems biology and systems chemistry (Figure 3). Systems biology contexts might incorporate pathway knowledge to model the rules governing a biological network topology and dynamics. For systems chemistry contexts, we could utilize building blocks for chemical synthesis as chemical descriptors, a strategy of fragment-based drug design.

Furthermore, it is critical to address how to handle side effects that come about after molecule design, experiment, optimization, and mass deployment. Efforts to electronically accumulate drug adverse event reports have begun only in the past decade. These data might be beneficial in refining models about polypharmacology and chemical biology. Although the ability to perfectly coordinate ligation to receptors and all downstream signaling events in order to eliminate side effects is a major challenge, inroads are being made to create a positive feedback loop for driving signal mechanism model refinement. Standardized pharmacovigilance methods such as reporting odds ratios (Hauben and Bate, 2010) are now employed by governments in analyzing the safety of drugs after they have reached the market, and we also have made sure that the clinical information is statistically informative (Kadoyama et al., 2011). The upcoming integration of heterogeneous knowledge of chemical biology, systems chemistry, polypharmacology, systems biology, and clinical information, among others, is an exciting and critical advancement for intensive acceleration of drug discovery and pharmaceutical development.

ACKNOWLEDGMENTS

The work described herein was supported by grants from the Funding Program for Next Generation World-Leading Researchers. Additional support was provided by the New Energy and Industrial Technology Development Organization (NEDO) and the Grant-in-Aid for Scientific Research on Innovative Areas "HD Physiology" of the Ministry of Education, Culture, Sports, Science, and Technology, Japan. Financial support from Ono Pharmaceutical Co., Ltd. and the Takeda Science Foundation is also gratefully acknowledged. The Chemical Structure Association Trust has provided support through the Jacques-Émile Dubois Grant for research on chemical information.

REFERENCES

Chen, T., Ozel, D., Qiao, Y., Harbinski, F., Chen, L., Denoyelle, S., He, X., Zvereva, N., Supko, J.G., Chorev, M., et al. (2011). Chemical genetics identify eIF2α kinase heme-regulated inhibitor as an anticancer target. Nat. Chem. Biol. *7*, 610–616.

Hauben, M., and Bate, A. (2010). Data mining in pharmacovigilance. In Pharmaceutical Data Mining: Approaches and Applications for Drug Discovery, K.V. Balakin, ed. (New York: John Wiley & Sons), pp. 341–377.

Hopkins, A.L. (2007). Network pharmacology. Nat. Biotechnol. *25*, 1110–1111.

Hopkins, A.L., Mason, J.S., and Overington, J.P. (2006). Can we rationally design promiscuous drugs? Curr. Opin. Struct. Biol. *16*, 127–136.

Hübel, K. (2009). Yeast-based chemical genomic approaches. In Chemical Biology: Learning through Case Studies, H. Waldmann and P. Janning, eds. (Weinham: Wiley-VCH), pp. 1–20.

Jacob, L., and Vert, J.P. (2008). Protein-ligand interaction prediction: an improved chemogenomics approach. Bioinformatics *24*, 2149–2156.

Jacoby, E. (2011). Computational chemogenomics. Comput. Mol. Sci. *1*, 57–67.

Kadoyama, K., Kuwahara, A., Yamamori, M., Brown, J.B., Sakaeda, T., and Okuno, Y. (2011). Hypersensitivity reactions to anticancer agents: data mining of the public version of the FDA adverse event reporting system, AERS. J. Exp. Clin. Cancer Res. *30*, 93.

Keiser, M.J., Setola, V., Irwin, J.J., Laggner, C., Abbas, A.I., Hufeisen, S.J., Jensen, N.H., Kuijer, M.B., Matos, R.C., Tran, T.B., et al. (2009). Predicting new molecular targets for known drugs. Nature *462*, 175–181.

MacDonald, M.L., Lamerdin, J., Owens, S., Keon, B.H., Bilter, G.K., Shang, Z., Huang, Z., Yu, H., Dias, J., Minami, T., et al. (2006). Identifying off-target effects and hidden phenotypes of drugs in human cells. Nat. Chem. Biol. *2*, 329–337.

Maggiora, G.M. (2011). The reductionist paradox: are the laws of chemistry and physics sufficient for the discovery of new drugs? J. Comput. Aided Mol. Des. *25*, 699–708.

Okuno, Y., Tamon, A., Yabuuchi, H., Niijima, S., Minowa, Y., Tonomura, K., Kunimoto, R., and Feng, C. (2008). GLIDA: GPCR—ligand database for chemical genomics drug discovery—database and tools update. Nucleic Acids Res. *36* (Database issue), D907–D912.

Paolini, G.V., Shapland, R.H., van Hoorn, W.P., Mason, J.S., and Hopkins, A.L. (2006). Global mapping of pharmacological space. Nat. Biotechnol. *24*, 805–815.

Schneider, G., and Fechner, U. (2005). Computer-based de novo design of drug-like molecules. Nat. Rev. Drug Discov. *4*, 649–663.

Schreiber, S.L. (2005). Small molecules: the missing link in the central dogma. Nat. Chem. Biol. *1*, 64–66.

Wassermann, A.M., Geppert, H., and Bajorath, J. (2009). Ligand prediction for orphan targets using support vector machines and various target-ligand kernels is dominated by nearest neighbor effects. J. Chem. Inf. Model. *49*, 2155–2167.

Yabuuchi, H., Niijima, S., Takematsu, H., Ida, T., Hirokawa, T., Hara, T., Ogawa, T., Minowa, Y., Tsujimoto, G., and Okuno, Y. (2011). Analysis of multiple compound-protein interactions reveals novel bioactive molecules. Mol. Syst. Biol. *7*, 472.

Yamamoto, K.N., Hirota, K., Kono, K., Takeda, S., Sakamuru, S., Xia, M., Huang, R., Austin, C.P., Witt, K.L., and Tice, R.R. (2011). Characterization of environmental chemicals with potential for DNA damage using isogenic DNA repair-deficient chicken DT40 cell lines. Environ. Mol. Mutagen. *52*, 547–561.