31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Mapping Drug-Target Interaction Networks

Longzhang Tian, and Shuxing Zhang*

Abstract—**Molecular polypharmacological studies have gained more and more attention as they are important in predicting drug off-target properties and potential toxicity/side effect. The explosive growth of biomedical data provides us an opportunity to develop novel strategies to conduct such studies by analyzing molecular interaction networks. In this paper, we present an integrated web application that is implemented based on more than 5,000 drugs and 56,000 biological macromolecule structures. With efficient search of drug information (biological targets, pharmacology, side effect, etc.) and chemical similarity, molecular maps can be constructed to demonstrate the relationships among multiple drugs and receptors. In addition, receptor information can also be employed to map the interaction network. The 3D structures of available drug-receptor complexes can be visualized via our web server, and the query results will be used to identify similar structures for any given drugs as well as their cross interactions with other biological targets. Our implementation provides an efficient way to evaluate the safety and polypharmacological properties of chemical compounds.**

I. Introduction

BIOLOGICAL research has produced vast amounts of data (e.g., from HTS experiments) and most of them are publicly available. Based on these data, molecular interaction networks have been recently studied using systems biology approaches and various databases have been constructed [1-6]. These databases are used to probe the relationships of all elements (using top-down or bottom-up methods) rather than approaching them separately. The perturbing effect of a molecule on the biological systems can be predicted by identifying its role across signaling pathways. This is certainly of great importance as it helps us to understand the systems and to address the problems occurred to the systems such as cancer, which is a particularly challenging realm as emerging evidence continues to corroborate the notion that cancers involves complex signaling networks. For instance, we have been applying the Ingenuity's pathway analysis software (http://www.ingenuity.com) and its database to the identification of all possible targets which can be perturbed by our newly designed Akt (a.k.a. protein kinase B) pleckstrin homology (PH) domain inhibitors [7;8]. Our study demonstrated that some of these inhibitors have strong inhibitions on a particular pathway (such as PI3K/Akt pathway) while some others affect complex signaling networks (unpublished results).

On the other hand, chemical database techniques have been becoming powerful tools in drug discovery. As a result of the NIH Roadmap Molecular Libraries Initiative (http://mli.nih.gov/mli/), various chemical databases have been constructed, including PubChem [9], PDB [10], MOAD [11], ZINC [12], etc. In particular, PubChem [9] has provided a large collection of chemical compounds which are annotated with ample bioassays performed by various research groups or high-throughput screening centers. In addition, protein-ligand databases, such as MOAD and PDB, enabled queries of structure and functions of tens of thousand targets as well as their complexes. Obviously these databases are very useful by providing tools for virtual screening or relating ligand to their binding receptors. However there has been little effort of building ligand networks to probe the biological and pharmacological molecular networks until recently. Bork and colleagues [13] used phenotypic side-effect similarities to infer whether two drugs share a target. They applied to 746 marketed drugs with a network of 1018 side effect followed up with experimental validations, and found that 11 out of 13 implied drug-target interactions reveal inhibition constant equal or less than 10μM. This demonstrated the feasibility of using annotated information to infer molecular interactions and possible new use or off-target activities of the marketed drugs [13]. In addition, Shoichet group employed ligand similarity to study the protein functions as well as chemical compound polypharmacology and they found several compounds indeed had previously unknown off-target properties [14]. Another notable work is the database of STITCH [15] used to search interaction networks of chemicals and proteins. STITCH integrates information about interactions from crystal structures, binding experiments and drug–target relationships. These work demonstrated the importance of using cheminformatics approaches to relate ligands with their receptors, which are complementary to the bioinformatics methods.

During the last several years, we have been engaged in the study of protein-ligand interactions using cheminformatics approaches [16;17]. These approaches have been employed in virtual screening for the identification of potent inhibitors as well as in off-target property studies [18-20]. In particular, we developed a fast chemometric approach, termed CoLiBRI [17], where each studied compound in the database was screened against

Manuscript received April 7, 2009. This work was supported in part by the M.D. Anderson Cancer Center Faculty Startup fund.

S. Zhang is with the M.D. Anderson Cancer Center, Houston, TX 77030 USA (corresponding author to provide phone: 713-745-2958; fax: 713-792-5577; e-mail: shuzhang@mdanderson.org).

L. Tian is with the M.D. Anderson Cancer Center, Houston, TX 77030 USA (e-mail: ltian1@mdanderson.org).

proteins in the database to identify possible cross interactions. CoLiBRI is based on the representation of both receptor binding sites and their respective ligands in a space of universal chemical descriptors. The representation of both ligands and active sites using chemical descriptors allows the application of well known chemometric techniques in order to correlate chemical similarities between active sites and their respective ligands. We have established a protocol to map patterns of nearest neighbor active site vectors in a multidimensional descriptor space onto those of their complementary ligands, and vice versa. This protocol affords straightforward and efficient identification of the complementary ligands in large databases of chemical compounds for given active sites. Conversely, starting from the ligand chemical structure, one may identify possible complementary receptor cavities as well. Therefore, we can map the potential interactions between a compound and a receptor, and thus build molecular networks among chemical compounds and biological targets. Herein we present a completely different approach by using text mining and chemical similarity search techniques map molecular interaction networks.

II. Procedures

Our method was built upon Java web application technology. Tomcat 6 was employed as the back-end Java application server while Apache 2 was used as the front-end web server. MySQL databases stored the drug and protein data for Java application. The drug and protein databases and other pre-generate files for the web applications were generated using server side python scripts.

Drug data of both structures and text information were collected from various literature and resources (e.g., DrugBank) [21;22]. The protein data were obtained from PDB [10] including their text information and structure files. Python scripts were developed to process these data, and MySQL databases were constructed for both drugs and receptors. The JChem software [23] were used for the structure similarity search which attempted to find similar chemicals based on their likeness to a specified structure. Herein similarity search was performed on structures using structure fingerprints generated by JChem [23]. 2D structure images for all the drugs were generated using Pybel package [24]. Maps were created to evaluate the relationships between drugs and receptors. We collected the targets for every drug based on existing literatures and databases with drug names and functions. We examined all information for each drug and built maps for them with proteins as part of the graph nodes. Similarly, we examined all target information and built maps for each target with drugs as their nodes. Drug-target-drug maps were further created for every drug to help building possible links between drugs with the underlying principle that a drug can binds to multiple receptors and a receptor can bind to multiple drugs; that multiple targets are involved in the same signaling pathways; and that drug-drug interactions may affect their efficacy and toxicities. Our implementation is demonstrated by Figure 1.

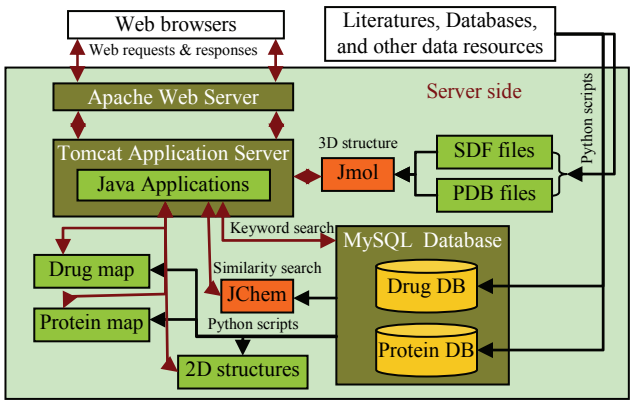

Figure 1. Workflow of our web applications

III. Results

Our implementation has integrated chemical databases, biological macromolecule databases, and many computational functions (e.g., similarity search) into one unified web framework. Figure 2 demonstrated our web interface of this development. Each drug was offered a unique ID and its related descriptive information along with some quantitative properties (e.g., experimental logP, plasm protein binding percentage, etc.) is also included. Through our website, users can perform similarity search, text mining, data curation, and so on.

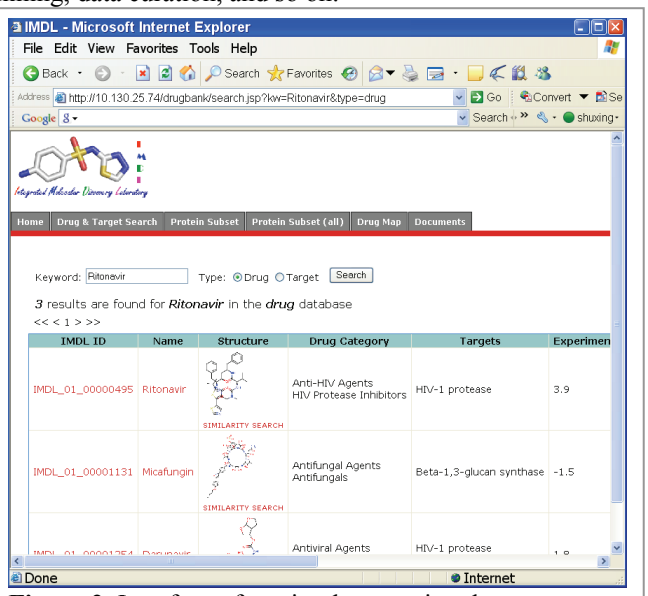

Figure 2. Interface of our implementation demonstrates that our tools can be used to perform chemical similarity search, text mining, molecular interaction network construction, and many other tasks.

As we mentioned above, molecular interaction networks can be studied based on their relationships of drugs and receptors, drugs and drugs, as well as receptors and

receptors. The interaction maps can be really complex, but our implementation allows users to focus on particular molecules of their interest. For instance, Figure 3 demonstrated how to map the molecular interaction network of Gleevec (Imatinib). We found that the c-Abl tyrosine kinase with available 3D structure information in PDB (1OPL) was targeted by Gleevec, but simultaneously was also the target of two other molecules, desatinib and ATP. When the mouse was put on the icon of a molecule, a pop-up window will display the related information (structure, literature, etc.) of the molecule. If 1OPL was used as query, similar result was obtained. However, if ATP was used for searching (Figure 2e), the network was much more complex, which is consistent to the fact that ATP can bind to many molecules (in particular kinases) and numerous crystal structures complexed with ATP have been solved in PDB.

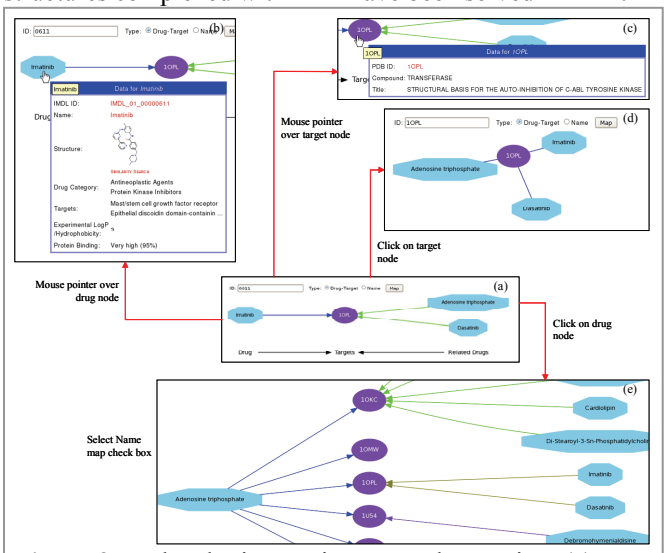

Figure 3. Molecular interaction network mapping: (a) drug-target-drug maps with drug Gleevec as an example; (b) pop-up window for drug details (Gleevec); (c) pop-up window for target information; (d) drug-target-drug maps built with the receptor as the query; (e) molecular interaction network built using ATP as the query.

In order to give an overall estimation on the molecular interaction network, we also performed some statistical studies. It was found that 1170 out of 5058 drugs have no indentified targets with available 3D structures. In comparison, 3522 drugs were found to have 1~10 targets while 80 drugs have more than 10 targets with 3D structures in PDB. This distribution analysis showed that many drugs do not have identified biological targets with available 3D structures, indicating the fact that they were originally developed using traditional medicinal synthesis without known targets or structures. On the other hand, the results also demonstrated that the majority of the existing drugs have clear targets and their structures have been solved. However, many drugs obviously could target multiple receptors and may lead to off-target properties.

As demonstrated in Figure 4, fluorouracil is usually used as an anticancer drug in chemotherapy. Based on our molecular interaction network analysis, it can also interact with multiple biological targets such as thymidylate synthase (1JUJ), dihydropyrimidine dehydrogenase (1GTE), uracil phosphoribosyltransferase (1JLR), and fatty acid binding protein (1TOU). This means that fluorouracil, in addition to its expected function as a thymidylate synthase inhibitor, potentially also interferes with those biological pathways involved in the above proteins, and thus leading to possible side effect and off-target properties. For instance, it is known that fluorouracil can cause both acute CNS damage and progressively worsening delayed degeneration of the CNS in mice. This may be due to the binding of fluorouracil to fatty acid binding protein in brain.

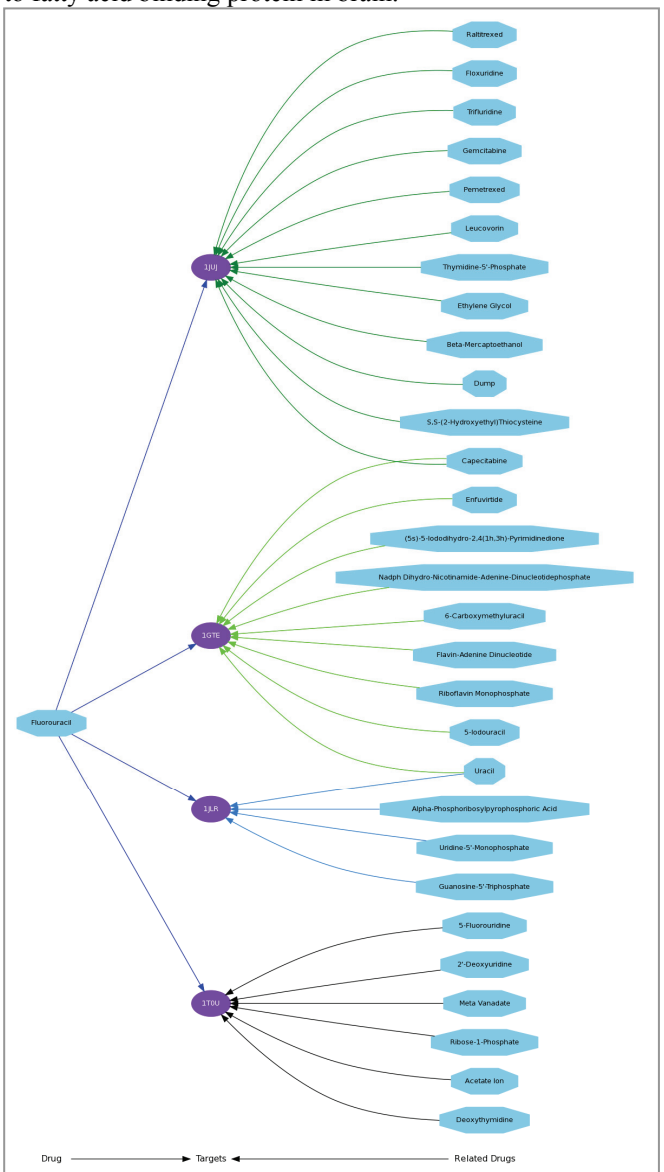

Figure 4. Anticancer agent fluorourocil could target to multiple receptors such as 1JUJ, 1GTE, 1JLR, and 1TOU. Eclipses represent receptors and polygons are for ligands. On the right side of the receptors, ligands are clustered based on their targets (connection arrows are in the same color if ligands bind to the same receptor).

IV. Conclusion

Modern drug discovery is focused on the development of effective and safe therapeutics with insignificant adverse drug reactions and side effect. Methods for *in silico* off-target property prediction and polypharmcology studies in early discovery stage is critical for effective drug development and safety evaluation [25-27]. To this end, we herein presented a novel strategy to analyze molecular interaction networks. As demonstrated, we can efficiently map the drug-target interaction network and predict the potential off-target property for any give drugs. In the future, we will also integrate various predictive ligand-based (e.g., QSAR) and structure-based (e.g., molecular docking) approaches [16;19;28-30] to generated more novel testable hypotheses for experimental validation.

Acknowledgment

We thank Saif Abayazid for his help on the development of chemical similarity search module.

References

[1] E. L. Berg, E. J. Kunkel, and E. Hytopoulos, "Biological complexity and drug discovery: a practical systems biology approach," *Syst. Biol. (Stevenage.)*, vol. 152, no. 4, pp. 201-206, Dec.2005.

[2] J. Boyle, C. Cavnor, S. Killcoyne, and I. Shmulevich, "Systems biology driven software design for the research enterprise," *BMC. Bioinformatics*, vol. 9, p. 295, 2008.

[3] A. L. Mayburd, A. Martlinez, D. Sackett, H. Liu, J. Shih, J. Tauler, I. Avis, and J. L. Mulshine, "Ingenuity network-assisted transcription profiling: Identification of a new pharmacologic mechanism for MK886," *Clin. Cancer Res.*, vol. 12, no. 6, pp. 1820-1827, Mar.2006.

[4] A. Splendiani, "RDFScape: Semantic Web meets systems biology," *BMC. Bioinformatics*, vol. 9 Suppl 4, p. S6, 2008.

[5] D. Steinhauser, B. Usadel, A. Luedemann, O. Thimm, and J. Kopka, "CSB.DB: a comprehensive systems-biology database," *Bioinformatics*, vol. 20, no. 18, pp. 3647-3651, Dec.2004.

[6] C. Wierling, R. Herwig, and H. Lehrach, "Resources, standards and tools for systems biology," *Brief. Funct. Genomic. Proteomic.*, vol. 6, no. 3, pp. 240-251, Sept.2007.

[7] D. Mahadevan, G. Powis, E. A. Mash, B. George, V. M. Gokhale, S. Zhang, K. Shakalya, L. Du-Cuny, M. Berggren, M. A. Ali, U. Jana, N. Ihle, S. Moses, C. Franklin, S. Narayan, N. Shirahatti, and E. J. Meuillet, "Discovery of a novel class of AKT pleckstrin homology domain inhibitors," *Mol. Cancer Ther.*, vol. 7, no. 9, pp. 2621-2632, Sept.2008.

[8] S. A. Moses, M. A. Ali, S. Zuohe, L. Du-Cuny, L. L. Zhou, R. Lemos, N. Ihle, A. G. Skillman, S. Zhang, E. A. Mash, G. Powis, and E. J. Meuillet, "In vitro and In vivo Activity of Novel Small-Molecule Inhibitors Targeting the Pleckstrin Homology Domain of Protein Kinase B/AKT," *Cancer Res.*, June2009.

[9] http://pubchem.ncbi.nlm.nih.gov/.

[10] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235-242, Jan.2000.

[11] L. Hu, M. L. Benson, R. D. Smith, M. G. Lerner, and H. A. Carlson, "Binding MOAD (Mother Of All Databases)," *Proteins*, vol. 60, no. 3, pp. 333-340, Aug.2005.

[12] J. J. Irwin and B. K. Shoichet, "ZINC--a free database of commercially available compounds for virtual screening," *J. Chem. Inf. Model.*, vol. 45, no. 1, pp. 177-182, Jan.2005.

[13] M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen, and P. Bork, "Drug target identification using side-effect similarity," *Science*, vol. 321, no. 5886, pp. 263-266, July2008.

[14] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet, "Relating protein pharmacology by ligand chemistry," *Nat. Biotechnol.*, vol. 25, no. 2, pp. 197-206, Feb.2007.

[15] M. Kuhn, M. C. von, M. Campillos, L. J. Jensen, and P. Bork, "STITCH: interaction networks of chemicals and proteins," *Nucleic Acids Res.*, vol. 36, no. Database issue, p. D684-D688, Jan.2008.

[16] S. Zhang, A. Golbraikh, and A. Tropsha, "Development of quantitative structure-binding affinity relationship models based on novel geometrical chemical descriptors of the protein-ligand interfaces.," *J. Med. Chem.*, vol. 49, no. 9, pp. 2713-2724, May2006.

[17] S. Oloff, S. Zhang, N. Sukumar, C. Breneman, and A. Tropsha, "Chemometric analysis of ligand receptor complementarity: identifying Complementary Ligands Based on Receptor Information (CoLiBRI).," *J. Chem. Inf. Model.*, vol. 46, no. 2, pp. 844-851, Mar.2006.

[18] S. Oloff, R. B. Mailman, and A. Tropsha, "Application of validated QSAR models of D1 dopaminergic antagonists for database mining," *J. Med. Chem.*, vol. 48, no. 23, pp. 7322-7332, Nov.2005.

[19] S. Zhang, A. Golbraikh, S. Oloff, H. Kohn, and A. Tropsha, "A Novel Automated Lazy Learning QSAR (ALL-QSAR) Approach: Method Development, Applications, and Virtual Screening of Chemical Databases Using Validated ALL-QSAR Models," *J Chem Inf. Model.*, vol. 46, no. 5, pp. 1984-1995, Sept.2006.

[20] W. F. Zheng and A. Tropsha, "Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 1, pp. 185-194, Jan.2000.

[21] A. Heinrichs, "PubMed Central," *Trends Mol. Med.*, vol. 7, no. 3, p. 102, Mar.2001.

[22] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, and J. Woolsey, "DrugBank: a comprehensive resource for in silico drug discovery and exploration," *Nucleic Acids Res.*, vol. 34, no. Database issue, p. D668-D672, Jan.2006.

[23] F. Csizmadia, "JChem: Java applets and modules supporting chemical database handling from web browsers," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 2, pp. 323-324, Mar.2000.

[24] N. M. O'Boyle, C. Morley, and G. R. Hutchison, "Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit," *Chem. Cent. J.*, vol. 2, p. 5, 2008.

[25] D. E. Clark and P. D. Grootenhuis, "Progress in computational methods for the prediction of ADMET properties," *Curr. Opin. Drug Discov. Devel.*, vol. 5, no. 3, pp. 382-390, May2002.

[26] A. M. Davis and R. J. Riley, "Predictive ADMET studies, the challenges and the opportunities," *Curr. Opin. Chem. Biol.*, vol. 8, no. 4, pp. 378-386, Aug.2004.

[27] J. C. Dearden, "In silico prediction of ADMET properties: how far have we come?," *Expert. Opin. Drug Metab Toxicol.*, vol. 3, no. 5, pp. 635-639, Oct.2007.

[28] S. Zhang, K. Kumar, X. Jiang, A. Wallqvist, and J. Reifman, "DOVIS: an implementation for high-throughput virtual screening using AutoDock," *BMC. Bioinformatics*, vol. 9, no. 1, p. 126, Feb.2008.

[29] S. Zhang and L. Du-Cuny, "Development and evaluation of a new statistical model for structure-based high-throughput virtual screening," *Int. J. Bioinform. Res. Appl.*, vol. 5, no. 3, pp. 269-279, 2009.

[30] S. Zhang, L. Wei, K. Bastow, W. Zheng, A. Brossi, K. H. Lee, and A. Tropsha, "Antitumor agents 252. Application of validated QSAR models to database mining: discovery of novel tylophorine derivatives as potential anticancer agents," *J. Comput. Aided Mol. Des*, vol. 21, no. 1-3, pp. 97-112, Jan.2007.