



2014

An Integrated Workflow for Proteome-Wide Off-Target Identification and Polypharmacology Drug Design

Thomas Evangelidis

Department of Pharmaceutical Chemistry, Faculty of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece.

Lei Xie

Department of Computer Science, Hunter College, The Graduate Center, The City University of New York, NY 10065, USA.

Follow this and additional works at: <https://tsinghuauniversitypress.researchcommons.org/tsinghua-science-and-technology>

 Part of the [Computer Sciences Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Thomas Evangelidis, Lei Xie. An Integrated Workflow for Proteome-Wide Off-Target Identification and Polypharmacology Drug Design. *Tsinghua Science and Technology* 2014, 19(03): 275-284.

This Research Article is brought to you for free and open access by Tsinghua University Press: Journals Publishing. It has been accepted for inclusion in *Tsinghua Science and Technology* by an authorized editor of Tsinghua University Press: Journals Publishing.

An Integrated Workflow for Proteome-Wide Off-Target Identification and Polypharmacology Drug Design

Thomas Evangelidis and Lei Xie*

Abstract: Polypharmacology, which focuses on designing drugs to target multiple receptors, has emerged as a new paradigm in drug discovery. To rationally design multi-target drugs, it is fundamental to understand protein-ligand interactions on a proteome scale. We have developed a Proteome-wide Off-target Pipeline (POP) that integrates ligand binding site analysis, protein-ligand docking, the statistical analysis of docking scores, and electrostatic potential calculations. The utility of POP is demonstrated by a case study, in which the molecular mechanism of anti-cancer effect of Nelfinavir is hypothesized. By combining structural proteome-wide off-target identification and systems biology, it is possible for us to correlate drug perturbations with clinical outcomes.

Key words: drug discovery; structural proteomics; polypharmacology; off-target; systems biology

1 Introduction

In the conventional drug discovery process, a reductionist model of “one-drug-one-gene” is widely adopted in the pharmaceutical industry. Drugs are screened or designed to modify the activity of a single disease-causing gene or its encoded protein, which is termed as drug target (or “on-target”). However, unexpected drug-target interactions often exist^[1]. The unexpected gene (or protein) involved in the drug action is categorized as “off-target”. The off-target may cause unwanted drug side effects. This is highlighted in a recent withdrawal of Torcetrapib from phase III clinical trial. Torcetrapib is intended to lower cholesterol, thus reduce the risk of heart disease. Unfortunately, it causes serious side effect of heart attack due to unexpected off-target effects. On the other hand, the off-target is

beneficial. One drug X designed for disease A can be used to treat another disease B, if the off-target of the drug X happens to be responsible for the disease B. Drug repurposing (or repositioning) is a practice to reuse existing drugs for new purposes^[2]. Moreover, in contrast to “one-drug-one-gene” paradigm, multi-target therapies (also called polypharmacology) are needed to treat systematic diseases such as cancer, given the observed robustness and resilience of complex biological networks to single drug treatments. In principal, multi-target therapeutics can exhibit greater efficacy and be less vulnerable to drug resistance by impacting multiple nodes at the system level. Indeed, a large number of the existing anti-infectious and anti-cancer therapeutics is either multi-target agents developed through drug combination or single multi-target drugs discovered serendipitously. While it is clear that a multi-target therapy enhances clinical efficacy, there is an increased possibility of serious side effects. Thus it remains a great challenge to rationally design effective multi-target therapies.

One of central topic in polypharmacology is to identify drug-target interactions on a proteome-scale. A large number of drug-target interactions is already known, and has been collected in the databases such Protein Data Bank (PDB)^[3], DrugBank^[4],

• Thomas Evangelidis is with the Department of Pharmaceutical Chemistry, Faculty of Pharmacy, National and Kapodistrian University of Athens, Athens, Greece. E-mail: tevang3@gmail.com.

• Lei Xie is with the Department of Computer Science, Hunter College, The Graduate Center, The City University of New York, NY 10065, USA. E-mail: lei.xie@hunter.cuny.edu.

* To whom correspondence should be addressed.

Manuscript received: 2014-04-18; accepted: 2014-04-28

PubChem^[5,6], and ChEMBL^[7]. Taking advantage of existing knowledge on the on-target, multi-target drug design can be reformulated as a proteome-wide off-target identification problem. Given a validated on-target, what are its off-targets within and across human and pathogen proteomes? A number of methodologies have been developed to answer this question. Relevant proteins can be identified through their ligand chemistry^[8,9], biological profiles (e.g., binding activities^[5], expression profiles^[10], or side effects^[11]), and promiscuities based on their sequence, structure, and ligand binding site similarity^[12]. Moreover, advances in genome sequencing and systems biology have made it possible to reconstruct, model, and simulate metabolic, signal transduction, and regulation pathways of human and a large number of pathogens. A review of related computational techniques can be found in Ref. [1]. Recently, we have successfully applied this combination of off-target identification and systems biology to reuse safe pharmaceuticals to treat multi-drug and extremely drug resistance tuberculosis^[13] and to reduce bacterial infections^[14], identify drug leads to target multiple receptors in *T. brucei*^[15], predict side effect profiles for several major pharmaceuticals^[16,17], propose molecular mechanisms of anti-cancer effect of anti-AIDS drug Nelfinavir^[18], understand the genetic disposition of drug responses^[19], and reconstruct structural proteome scale drug-target network of *M. tuberculosis* genome Drugome/TB^[20]. These works have demonstrated the power of integrating ligand binding comparison, protein-ligand docking, surface electrostatic potential analysis, Molecular Dynamics (MD) simulation, and various functional genomics techniques in the polypharmacology.

In spite of the success of aforementioned integrative strategy, each component in the computational pipeline was carried out sequentially and manually in the previous studies. A user friendly software that is able to predict off-targets through automatically linking multiple algorithms will be extremely useful. In this paper, we describe a Proteome-wide Off-target Pipeline (POP) that will facilitate drug off-target identification and polypharmacology drug design. POP implements the proven successful algorithms for ligand binding site characterization and comparison, protein-ligand docking, statistical analysis of docking scores, and electrostatic potential calculations. The details of individual algorithms and their performance evaluations

can be found in Refs. [13-23]. In addition, we present a case study that applies POP to the identification of human targets of anti-AIDS drug Nelfinavir. We find new evidences to support that the anti-cancer effect of Nelfinavir comes from its interactions with multiple human protein kinases, as suggested in our previous studies^[18]. By combining POP with biological network reconstruction and systems biology simulation, we are able to correlate molecular interactions to clinical outcomes. POP is free for academic users, available from <https://sites.google.com/site/offtargetpipeline/>.

2 Methods

2.1 Overview

The workflow of POP is outlined in Fig. 1. One of the core components in POP is software *SMAP* for ligand binding site characterization and analysis, which is available from <http://funsite.sdsc.edu>^[21,22]. The premise is that similar ligand binding sites between two proteins will bind to similar ligands. First, the location and the boundary of the ligand binding site in the given on-target is determined from the co-crystal

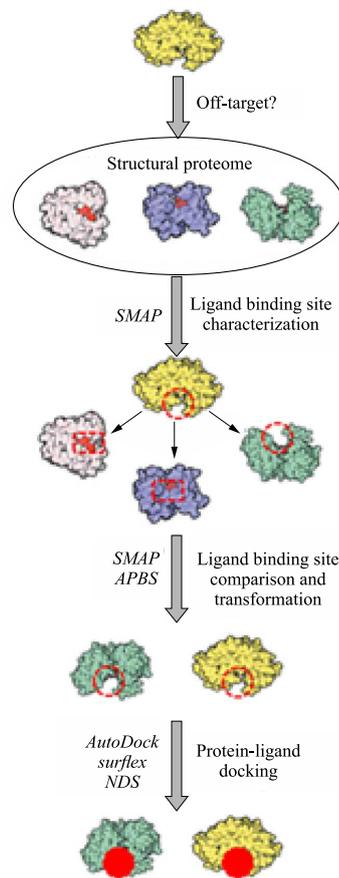


Fig. 1 A proteome-wide off-target pipeline.

structure or predicted by the pocket detection algorithm using *SMAP*. Subsequently, *SMAP* characterizes the properties of the binding site by the Geometric Potential (GP)^[21] and evolutionary profile^[22]. The electrostatic potential of the binding site is calculated using software APBS^[24]. The delineation of the binding site significantly reduces the search space for the structure comparison made during the next step. Second, *SMAP* searches the delineated binding site against all 3-D structures in a given proteome, either experimental determined or homology modeled, using a Sequence-Order Independent Profile-Profile Alignment (SOIPPA) algorithm^[22], which is implemented in *SMAP*. SOIPPA is able to detect similar ligand binding sites between two structurally un-related proteins. The statistical significance of the matched ligand binding site similarity is estimated by an Extreme Value Distribution mode^[23]. Thus off-targets with similar ligand binding sites to the on-target are identified. Third, *SMAP* superimposes the ligand binding site of the on-target and its bound ligand onto those top-ranked off-targets, thus predicts an initial binding pose for the ligand bound to each off-target. Finally, starting from the superimposed binding pose and the predicted ligand binding site for the off-targets, protein-ligand docking software Surflex^[25], AutoDock 4^[26], and AutoDock Vina^[27], all of which are free for academic users, are applied to further optimize the binding pose. The predicted binding affinity is normalized by the size of ligands and the nature of receptors using a novel statistical model for the protein-ligand docking score (details below). In addition, the electrostatic potential binding energy and similarity between the binding sites can be calculated based on the binding pose of the ligand.

2.2 Parallelization of *SMAP*

SMAP consists of three major steps as follows. (1) The protein structure is characterized by a geometric potential, a shape descriptor using the reduced C α only structure representation^[22]. It has been shown that both the location and the boundary of the ligand binding site can be accurately predicted by GP^[21]. Moreover, the reduced representation of the protein structure makes the algorithm tolerant to protein flexibility and experimental uncertainty, thus *SMAP* can be applied to low resolution structures and homology models. (2) Two structures are aligned using an SOIPPA algorithm^[22]. SOIPPA finds the most similar local

patch in the spirit of local sequence alignment but independent of the sequence order. This feature makes *SMAP* not only appropriate for practical problems since typically the boundary of the ligand binding site is not clearly defined and depends on the bound ligand, but also able to detect novel functional relationships across fold space. (3) The aligned surface patches are ranked by a scoring function that combines evolutionary, geometric, and physical information. Moreover, we have developed a new Extreme Value Distribution (EVD) statistical model that can rapidly estimate the statistical significance of the match between two ligand binding sites^[23]. In a benchmark evaluation, *SMAP* outperformed existing algorithms for both ligand binding site alignment quality and database search sensitivity and specificity^[22]. The ability of *SMAP* to establish cross-fold ligand binding relationships and its applications in drug discovery has been validated experimentally by several recent studies, for example, reuse of protein kinase inhibitors to target bacterial carboxyl-transferases^[22,28]. In another study, the safe pharmaceutical Entacapone, which targets human COMT for Parkinsons disease, has been predicted and validated to directly inhibit *M. Tuberculosis* InhA, the most used drug target to treat tuberculosis^[13]. Most recently, we have applied *SMAP* to identify target-drug interaction networks of pathogen *P. aeruginosa*. We have predicted that the estrogen receptor ligand binding site is significantly similar to that of two virulence factor genes responsible for the infection of *P. aeruginosa*. Subsequent experiments show that Selective Estrogen Receptor Modulators (SERMs) are able to increase the survival rate of *C. elegans* that are infected by *P. aeruginosa* by reducing the production of virulence factors^[14]. Although *SMAP* is proven to be efficient and robust, parallelization can further improve its performance for genome-wide ligand binding site comparisons. Moreover the software requires either the 3-D structure per se, or a pre-built *SMAP* object file. To avoid excessive disk space occupation, these 3-D coordinate and metadata files often need to be downloaded from Protein Data Bank web server (<http://www.rcsb.org>) before *SMAP* initiates the actual calculations. POP offers a further boost in performance by downloading PDB files needed for forthcoming calculations, while constantly running a preset number of jobs (the thread-pool size). Upon completion, the structural proteome is prioritized according to the predicted *SMAP* *p*-value of the ligand binding site

match, and an arbitrary number of highly-scored proteins is retained to undergo explicit docking with the molecule of interest. *SMAP* also superimposes the predicted off-target binding site to that in the on-target, thus provides an initial binding pose of ligand in the off-target.

2.3 Protein-ligand docking and statistical model for docking score normalization

Currently, three docking programs are supported, Surflex^[25], AutoDock 4^[26], and AutoDock Vina^[27]. These programs implement different scoring functions and have varying computation times. Following completion of docking, the set of proteins can be sorted according to docking score. Although docking is able to generate reliable binding poses, most of the time, efficient, and accurate estimation of the binding affinity is a challenging task^[29]. The docking scoring functions are additive in nature, leading to favorable scoring of high molecular weight molecules^[30]. Therefore, a docking score normalization based on the size of molecule is required to reduce the bias toward large compounds. Moreover, in the reverse docking problem where one ligand is docked to multiple receptors, the docking scores need to be normalized across the receptors so that the relative binding affinity is comparable. For this purpose, a statistical model for the docking score normalization is developed to cancel out systematic errors in the docking scoring functions^[17]. The Normalized Docking Score (NDS) is a relative docking score that is the derivation of the raw docking score from the background distribution of a large set of random molecules. The background distribution depends on both the size of the ligand and the nature of the receptor. A more negative value of the NDS indicates a higher chance of real binding. The normalization starts by assembling a library of drug-like molecules of varying number of carbon atoms, which are downloaded from ZINC (<http://zinc.docking.org>)^[31]. These molecules are docked to the protein receptors. We found that the docking score is linearly dependent on the number of carbon atoms. To save the computational time, only hundreds of decoy molecules with 5, 10, 15, 20, and 25 carbon atoms for each are selected. The correlation of the docking score to the number of carbon atoms is derived from linear regressions. From the linear fitting curve, the average docking score for molecules with a certain number of carbon atoms can

be computed. Based on the fitted average docking score, a normalized docking score NDS is computed as a z-score:

$$\text{NDS} = (S_i - \mu_i) / \sigma,$$

where S_i is the raw docking score for the molecule with i carbon atoms, μ_i is the fitted average docking score for the molecules with i carbon atoms, σ is the standard deviation, which is not dependent on the size of molecules and is approximately 1.0 for all of cases. With the Graphical User Interface (GUI) application normalization is conducted in high through-put mode and the standardized docking score is estimated with *R* software package embedded in the POP.

2.4 Electrostatics potential analysis

As docking scores alone may be not sufficient to prove efficient binding of the drug, a further electro-chemical analysis is carried out to establish whether a significantly similar off-target ligand site is present. Physiochemical features like electrostatic binding energy of the predicted docking pose, and comparison of electrostatic potential distribution between the on- and off-target ligand binding sites, can be measured with POP. Both the binding energy and the Electrostatic Potential (EP) are calculated with APBS. The EP similarity between two binding pockets is measured by the Pearson correlation between their EP distributions, as calculated with APBS. The EP distribution is based on the electrostatic potential grids in the binding pocket, which contain those grid points overlapping with atoms of the ligand docked in the binding pocket. The radius of atoms is fixed as 0.5 nm.

2.5 Integration of biological knowledge

At this point the putative off-targets have reached a number which permits further investigations over the biological function. POP facilitates this step as well, by connecting to PDB through the SOAP interface and retrieving basic information about biological function, which now includes protein names and UniProt accession numbers^[32]. The functional relationships among the on-target and the putative off-target can be inferred by the co-occurrence of the protein names in the literature using iHOP literature network^[33] (<http://www.ihop-net.org/UniPub/iHOP/>), or Gene Ontology (GO) semantic similarity by supplying FunSimMat^[34] (<http://funsimmat.bioinf.mpi-inf.mpg.de/index.php>) with the UniProt accession

number. If the putative off-target is involved in, for example, the same physiological/biological process as the on-target, it will give a further support to the prediction. Both of the automatic literature mining and the GO similarity search have contributed to the identification of side effect profiles of CETP inhibitors in our recent study^[17]. Following this round of calculations, the methodology imposes review of all available clinical and literature information regarding the top scored receptors. Hopefully, some of these receptors will be implicated in signaling pathways affected by the drug. These proteins have both favorable biological function and binding pockets which exhibit high affinity to the drug.

3 Results and Discussion

We applied POP to repositioning Human Immunodeficiency Virus (HIV) protease inhibitor Nelfinavir to target human kinome for the rational design of multiple protein kinase inhibitors. Nelfinavir is an FDA approved antiretroviral medication which belongs to the class of drugs known as protease inhibitors. Its primary use is to treat Acquired Immunodeficiency Syndromes (AIDS). However, according to recent studies it is also associated with cancer therapy^[35,36]. *In vivo* Nelfinavir causes caspase-dependent apoptosis and non-apoptotic death, as well as endoplasmic reticulum stress and autophagy^[35]. In addition, it suppresses tumor cell migration by blocking angiogenesis^[37] and inducing tumor oxygenation of the tumor^[38]. Given its wide spectrum of activity, oral bioavailability, and fine ADME properties, Nelfinavir is potentially an ideal candidate for repositioning towards cancer therapy.

However, there is an uncertainty around the signaling pathways affected by Nelfinavir. The majority of published data lead to the assumption that drug may suppress the Akt signaling pathway^[39]. In human, there are three genes in the Akt family: Akt1, Akt2, and Akt3. These genes code for enzymes that are members of the serine/threonine-specific protein kinase family. Akt1 is involved in cellular survival and protein synthesis pathways as well as glucose metabolism. Not surprisingly, Akt1 has been implicated as a major factor in many types of cancer, including breast, colorectal, and ovarian cancer^[40-42]. Akt2 plays a key role in glucose transport inductions and is considered as an important component of Insulin signaling pathway. It is

also found that some Akt2 alterations lead to ovarian carcinoma pathogenesis^[43], whereas down-regulation can block extravasation of circulating breast cancer cells^[44]. This gene is also over-expressed in a subset of human ductal pancreatic cancers^[45,46]. Akt3 is known to be stimulated by Platelet-Derived Growth Factor (PDGF), insulin, and Insulin-like Growth Factor 1 (IGF1)^[47]. Thus inhibition of Akt pathway may cause insulin resistance and diabetes, phenomena observed as side effects of HIV proteases. However, there is no experimental proof to support that Nelfinavir can bind directly to Akt. It is likely that the drug acts upstream in the signaling pathway^[48].

To identify the human off-targets of Nelfinavir, the Nelfinavir binding pocket of the HIV protease dimer structure (PDB id: 1ohr) was searched against 5985 human protein structures and models using SMAP. Top ranked SMAP hits (p -value $< 10^{-4}$) are enriched by multiple members of protein kinase superfamily (14/18). The SMAP alignments between the PKs and the Nelfinavir binding sites reveal that co-crystallized ATP and its competitive inhibitors bind in the vicinity of the predicted binding sites, as shown as an example in Fig. 2 for the case of the insulin-like growth factor 1 receptor kinase domain (PDB id: 2OJ9). If amino acid residues with atomic distances less than 0.5 nm to the inhibitors are considered as the binding site, around two-third of these residues overlap between known and predicted sites for 2OJ9.

Using MD simulation, we found that Nelfinavir might

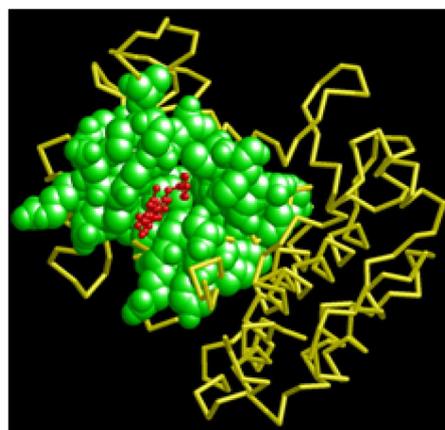


Fig. 2 The predicted Nelfinavir binding sites in insulin-like growth factor 1 receptor kinase domain (PDB id: 2OJ9) with its co-crystal protein kinase inhibitor 3-[5-(1H-imidazol-1-yl)-7-methyl-1H-benzimidazol-2-yl]-4-[(pyridin-2-ylmethyl)amino]pyridin-2(1H)-one (BMI). The predicted binding site residues are represented by green (big) spheres and BMI by red (small) balls and sticks, respectively.

weakly interact with multiple members of protein-kinase like superfamily^[18]. In addition, the most of predicted kinase off-targets are on the upstream of Akt1 and Akt2, as shown in the reconstructed AKT pathways (Fig. 3). Our predictions provide a mechanistic explain of both diabetes side effect and anti-cancer activity of Nelfinavir. In our initial work reported in Ref. [18], the ligand binding site analysis and the protein-ligand docking were carried out manually. POP makes this process automatically, thus may facilitate the drug off-target identification on a large scale.

Moreover, POP offers a new functionality for the comparison of EP between two binding pockets. EP similarity provides new evidences to support our hypothesis as protein kinases are the off-targets of Nelfinavir. EP similarities between the HIV protease and protein kinase binding pockets are quantitatively determined across the human kinome using EP similarity score. Among 105 members of protein

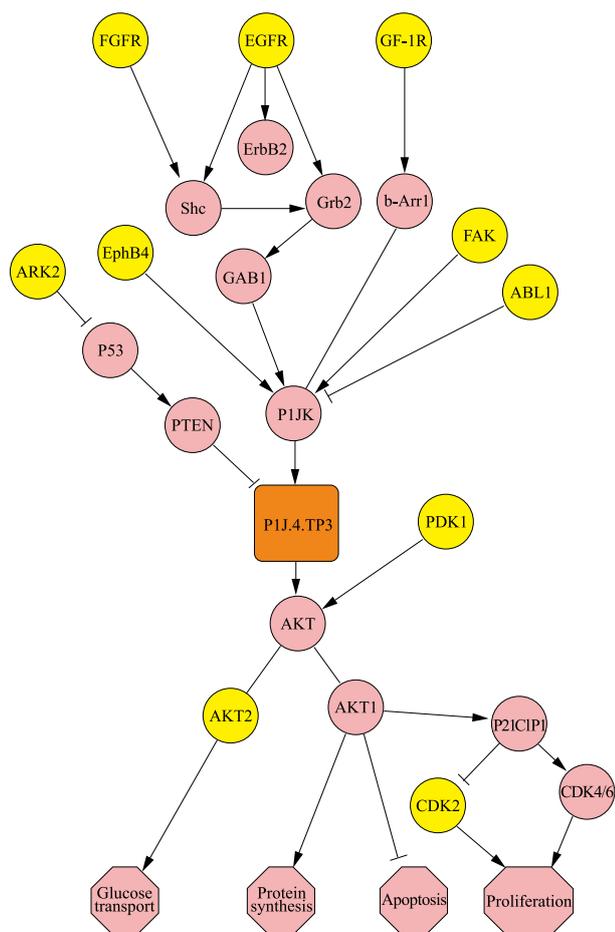


Fig. 3 Reconstructed PI3K/AKT pathway. Predicted off-targets are shown as yellow circles. Downstream activation and inhibition is represented by arrow-headed and bar-headed line, respectively.

kinase superfamily, 29% and 8% of them have EP similarity score larger than 0.50 and 0.80 (p -value = 0), respectively, as shown in Fig. 4. Among top 13 protein kinase off-targets that are predicted by SMAP, the percentage of them that bear the similar EP to the HIV protease in the binding pocket is higher than that of the whole kinome, with 50% and 15% EP similarity score larger than 0.50 and 0.80, respectively. Thus, the high EP similarities between these predicted off-targets and the HIV protease binding pocket through the whole human kinome is consistent with the high enrichment of protein kinases in the SMAP off-target search, as reported in Ref. [18]. As shown in Fig. 5, Nelfinavir molecule can be fragmented into five moieties: the 2-methyl-3-hydroxy-benzamide portion A, the S-phenyl group B, the tert-butyl carboxamide moiety C, the lipophilic dodecahydroisoquinoline ring D, and the central hydroxyl group E. Figure 6 shows the EP surfaces of the HIV protease and five top ranked protein kinase off-targets by SMAP. The EP similarity score is 0.85, 0.81, 0.67, 0.51, and 0.12 for ABL1, Akt2, IGF1R, EGFR, and EphA2, respectively. In general, the fragment B of Nelfinavir contacts more positive charged (blue in the figure) and deeper cavity patches than the fragment A in the HIV protease binding. The EP of the predicted protein kinase off-targets shows the similar pattern to that of the HIV protease. Combining with the binding site shape similarity and the docking pose comparisons, the detailed analysis of electrostatic potential variations in the PK binding pocket will provide useful clues on optimizing Nelfinavir to target the multiple PK families.

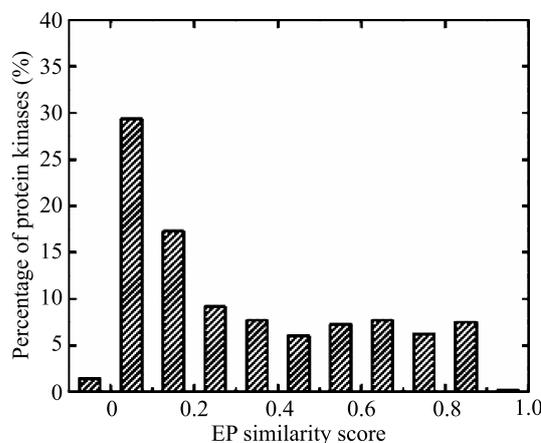


Fig. 4 Distribution of electrostatic potential similarity scores between Nelfinavir binding pockets of HIV protease and ATP binding pockets of protein kinases.

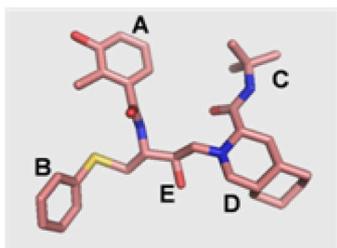


Fig. 5 Structure of Nelfinavir. A, 2-methyl-3-hydroxybenzamide moiety; B, S-phenyl group; C, tert-butyl carboxamide moiety; D, lipophilic dodecahydroisoquinoline ring; E, central hydroxyl group.

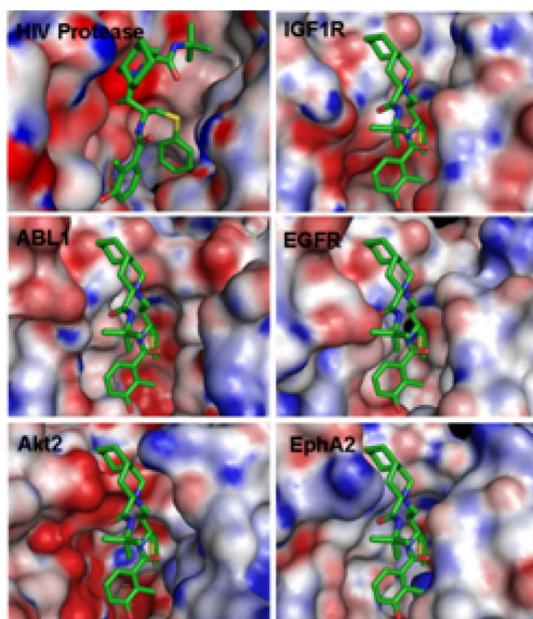


Fig. 6 The electrostatic potential surface of the ligand binding site in HIV protease and protein kinases (IGF1R, ABL1, EGFR, Akt2, and EphA2). The referenced Nelfinavir in the protein kinases is represented as a stick model. Carbon atoms are colored green; oxygen red; nitrogen blue; sulphur orange. The surface is colored according to the EP calculated from APBS. The color scale is set from -77.56 mV to 77.56 mV using a linear scale.

4 Conclusions

All the procedures described in this paper can be addressed with minimal effort by specially designed routines. Moreover, POP facilitates rapid screening of proteins groups with similar function or structural features, by querying PDB with keywords, assembling the dataset, and analyzing it automatically. Although POP is in its early stage, it demonstrates an automatic computational pipeline could facilitate proteome-scale off-target identifications and polypharmacology. The re-implementation of POP using scientific workflow

system may make it a valuable tool in computational systems biology and drug discovery.

5 Future Work

One of major drawbacks of the existing implementation of POP is that it is not flexible enough for users to incorporate new components. In addition, POP does not support distributed computing environment that is required by time-consuming computations such as MD simulation. We plan to re-implement POP using scientific workflow system WINGS (<http://www.wings-workflows.org>). Two unique features of WINGS make it a particularly suitable choice for the proteome-scale modeling of drug-target interactions. First, WINGS incorporates semantic constraints about datasets and workflow components. It will facilitate the integration of heterogeneous data set of chemicals, sequences, structures, and functions. Second, WINGS is built on Pegasus, a scientific workflow management system that allows applications to execute in different environments including desktops, computer clusters, grids, and clouds. Under the framework of WINGS, other computational components, which include all-against-all structure comparison and clustering of putative off-targets, MD simulations of drug-target bindings, and structure and network visualization tools, can be more easily incorporated into POP.

Acknowledgements

This work was supported by the National Institutes of Health GM078596, and High Performance Computing Center at The City University of New York. We thank the reviewers for their comments.

References

- [1] L. Xie, L. Xie, S. L. Kinnings, and P. E. Bourne, Novel computational approaches to polypharmacology as a means to define responses to individual drugs, *Annu. Rev. Pharmacol. Toxicol.*, vol. 52, pp. 361-379, 2012.
- [2] E. L. Tobinick, The value of drug repositioning in the current pharmaceutical market, *Drug News Perspect.*, vol. 22, no. 2, pp. 119-125, Mar. 2009.
- [3] 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.
- [4] D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, and M. Hassanali, DrugBank: A knowledgebase for drugs, drug actions and

- drug targets, *Nucleic Acids Res.*, vol. 36, no. Database issue, pp. D901-906, Jan. 2008.
- [5] B. Chen, D. Wild, and R. Guha, PubChem as a source of polypharmacology, *J. Chem. Inf. Model.*, vol. 49, no. 9, pp. 2044-2055, Sep. 2009.
- [6] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, and S. H. Bryant, PubChem: A public information system for analyzing bioactivities of small molecules, *Nucleic Acids Res.*, vol. 37, no. Web Server issue, pp. W623-633, Jul. 2009.
- [7] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, and J. P. Overington, ChEMBL: A large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D1100-1107, Jan. 2012.
- [8] 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.
- [9] M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijjer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. H. Thomas, D. D. Edwards, B. K. Shoichet, and B. L. Roth, Predicting new molecular targets for known drugs, *Nature*, vol. 462, no. 7270, pp. 175-181, Nov. 2009.
- [10] F. Iorio, R. Tagliaferri, and D. di Bernardo, Identifying network of drug mode of action by gene expression profiling, *J. Comput. Biol.*, vol. 16, no. 2, pp. 241-251, Feb. 2009.
- [11] 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, Jul. 2008.
- [12] L. Xie, L. Xie, and P. E. Bourne, Structure-based systems biology for analyzing off-target binding, *Curr. Opin. Struct. Biol.*, vol. 21, no. 2, pp. 189-199, Apr. 2011.
- [13] S. L. Kinnings, N. Liu, N. Buchmeier, P. J. Tonge, L. Xie, and P. E. Bourne, Drug discovery using chemical systems biology: Repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis, *PLoS Comput. Biol.*, vol. 5, no. 7, p. e1000423, Jul. 2009.
- [14] S. J. Ho Sui, R. Lo, A. R. Fernandes, M. D. G. Caulfield, J. A. Lerman, L. Xie, P. E. Bourne, D. L. Baillie, and F. S. L. Brinkman, Raloxifene attenuates *Pseudomonas aeruginosa* pyocyanin production and virulence, *Int. J. Antimicrob. Agents*, vol. 40, no. 3, pp. 246-251, Sep. 2012.
- [15] J. D. Durrant, R. E. Amaro, L. Xie, M. D. Urbaniak, M. A. J. Ferguson, A. Haapalainen, Z. Chen, A. M. Di Guilmi, F. Wunder, P. E. Bourne, and J. A. McCammon, A multidimensional strategy to detect polypharmacological targets in the absence of structural and sequence homology, *PLoS Comput. Biol.*, vol. 6, no. 1, p. e1000648, Jan. 2010.
- [16] L. Xie, J. Wang, and P. E. Bourne, In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators, *PLoS Comput. Biol.*, vol. 3, no. 11, p. e217, Nov. 2007.
- [17] L. Xie, J. Li, L. Xie, and P. E. Bourne, Drug discovery using chemical systems biology: Identification of the protein-ligand binding network to explain the side effects of CETP inhibitors, *PLoS Comput. Biol.*, vol. 5, no. 5, p. e1000387, May 2009.
- [18] L. Xie, T. Evangelidis, L. Xie, and P. E. Bourne, Drug discovery using chemical systems biology: Weak inhibition of multiple kinases may contribute to the anti-cancer effect of nelfinavir, *PLoS Comput. Biol.*, vol. 7, no. 4, p. e1002037, Apr. 2011.
- [19] R. L. Chang, L. Xie, L. Xie, P. E. Bourne, and B. ø. Palsson, Drug off-target effects predicted using structural analysis in the context of a metabolic network model, *PLoS Comput. Biol.*, vol. 6, no. 9, p. e1000938, 2010.
- [20] S. L. Kinnings, L. Xie, K. H. Fung, R. M. Jackson, L. Xie, and P. E. Bourne, The Mycobacterium tuberculosis drugome and its polypharmacological implications, *PLoS Comput. Biol.*, vol. 6, no. 11, p. e1000976, 2010.
- [21] L. Xie and P. E. Bourne, A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites, *BMC Bioinformatics*, vol. 8, no. Suppl 4, p. S9, 2007.
- [22] L. Xie and P. E. Bourne, Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments, *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 14, pp. 5441-5446, Apr. 2008.
- [23] L. Xie, L. Xie, and P. E. Bourne, A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery, *Bioinforma. Oxf. Engl.*, vol. 25, no. 12, pp. i305-312, Jun. 2009.
- [24] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. USA*, vol. 98, no. 18, pp. 10037-10041, Aug. 2001.
- [25] A. N. Jain, Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search, *J. Comput. Aided Mol. Des.*, vol. 21, no. 5, pp. 281-306, May 2007.
- [26] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.*, vol. 19, no. 14, pp. 1639-1662, Nov. 1998.
- [27] O. Trott and A. J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.*, vol. 31, no. 2, pp. 455-461, 2010.

- [28] J. R. Miller, S. Dunham, I. Mochalkin, C. Banotai, M. Bowman, S. Buist, B. Dunkle, D. Hanna, H. J. Harwood, M. D. Huband, A. Karnovsky, M. Kuhn, C. Limberakis, J. Y. Liu, S. Mehrens, W. T. Mueller, L. Narasimhan, A. Ogden, J. Ohren, J. V. N. V. Prasad, J. A. Shelly, L. Skerlos, M. Sulavik, V. H. Thomas, S. VanderRoest, L. Wang, Z. Wang, A. Whitton, T. Zhu, and C. K. Stover, A class of selective antibacterials derived from a protein kinase inhibitor pharmacophore, *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 6, pp. 1737-1742, Feb. 2009.
- [29] O. Guvench and A. D. MacKerell Jr, Computational evaluation of protein-small molecule binding, *Curr. Opin. Struct. Biol.*, vol. 19, no. 1, pp. 56-61, Feb. 2009.
- [30] G. Carta, A. J. S. Knox, and D. G. Lloyd, Unbiasing scoring functions: A new normalization and rescoring strategy, *J. Chem. Inf. Model.*, vol. 47, no. 4, pp. 1564-1571, Aug. 2007.
- [31] 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, Feb. 2005.
- [32] A. Bairoch, R. Apweiler, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. ODonovan, N. Redaschi, and L.-S. L. Yeh, The universal protein resource (UniProt), *Nucleic Acids Res.*, vol. 33, no. Database issue, pp. D154-159, Jan. 2005.
- [33] J. M. Fernandez, R. Hoffmann, and A. Valencia, iHOP web services, *Nucleic Acids Res.*, vol. 35, no. Web Server issue, pp. W21-26, Jul. 2007.
- [34] A. Schlicker and M. Albrecht, FunSimMat: A comprehensive functional similarity database, *Nucleic Acids Res.*, vol. 36, no. Database issue, pp. D434-439, Jan. 2008.
- [35] J. J. Gills, J. Lopiccio, J. Tsurutani, R. H. Shoemaker, C. J. M. Best, M. S. Abu-Asab, J. Borojerdi, N. A. Warfel, E. R. Gardner, M. Danish, M. C. Hollander, S. Kawabata, M. Tsokos, W. D. Figg, P. S. Steeg, and P. A. Dennis, Nelfinavir, a lead HIV protease inhibitor, is a broad-spectrum, anticancer agent that induces endoplasmic reticulum stress, autophagy, and apoptosis *in vitro* and *in vivo*, *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.*, vol. 13, no. 17, pp. 5183-5194, Sep. 2007.
- [36] J. J. Gills, J. Lopiccio, and P. A. Dennis, Nelfinavir, a new anti-cancer drug with pleiotropic effects and many paths to autophagy, *Autophagy*, vol. 4, no. 1, pp. 107-109, Jan. 2008.
- [37] N. Pore, A. K. Gupta, G. J. Cerniglia, and A. Maity, HIV protease inhibitors decrease VEGF/HIF-1 α expression and angiogenesis in glioblastoma cells, *Neoplasia N. Y. N.*, vol. 8, no. 11, pp. 889-895, Nov. 2006.
- [38] K. C. Cuneo, T. Tu, L. Geng, A. Fu, D. E. Hallahan, and C. D. Willey, HIV protease inhibitors enhance the efficacy of irradiation, *Cancer Res.*, vol. 67, no. 10, pp. 4886-4893, May 2007.
- [39] J. P. Plastaras, N. Vapiwala, M. S. Ahmed, D. Gudonis, G. J. Cerniglia, M. D. Feldman, I. Frank, and A. K. Gupta, Validation and toxicity of PI3K/Akt pathway inhibition by HIV protease inhibitors in humans, *Cancer Biol. Ther.*, vol. 7, no. 5, pp. 628-635, May 2008.
- [40] J. D. Carpten, A. L. Faber, C. Horn, G. P. Donoho, S. L. Briggs, C. M. Robbins, G. Hostetter, S. Boguslawski, T. Y. Moses, S. Savage, M. Uhlik, A. Lin, J. Du, Y.-W. Qian, D. J. Zeckner, G. Tucker-Kellogg, J. Touchman, K. Patel, S. Mousses, M. Bittner, R. Schevitz, M.-H. T. Lai, K. L. Blanchard, and J. E. Thomas, A transforming mutation in the pleckstrin homology domain of AKT1 in cancer, *Nature*, vol. 448, no. 7152, pp. 439-444, Jul. 2007.
- [41] Q. Meng, C. Xia, J. Fang, Y. Rojanasakul, and B.-H. Jiang, Role of PI3K and AKT specific isoforms in ovarian cancer cell migration, invasion and proliferation through the p70S6K1 pathway, *Cell. Signal.*, vol. 18, no. 12, pp. 2262-2271, Dec. 2006.
- [42] X. Ju, S. Katiyar, C. Wang, M. Liu, X. Jiao, S. Li, J. Zhou, J. Turner, M. P. Lisanti, R. G. Russell, S. C. Mueller, J. Ojeifo, W. S. Chen, N. Hay, and R. G. Pestell, Akt1 governs breast cancer progression *in vivo*, *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 18, pp. 7438-7443, May 2007.
- [43] Z. Q. Yuan, M. Sun, R. I. Feldman, G. Wang, X. Ma, C. Jiang, D. Coppola, S. V. Nicosia, and J. Q. Cheng, Frequent activation of AKT2 and induction of apoptosis by inhibition of phosphoinositide-3-OH kinase/Akt pathway in human ovarian cancer, *Oncogene*, vol. 19, no. 19, pp. 2324-2330, May 2000.
- [44] A. Faridmoayer, M. A. Fentabil, M. F. Haurat, W. Yi, R. Woodward, P. G. Wang, and M. F. Feldman, Extreme substrate promiscuity of the Neisseria oligosaccharyl transferase involved in protein O-glycosylation, *J. Biol. Chem.*, vol. 283, no. 50, pp. 34596-34604, Dec. 2008.
- [45] S. Tanno, S. Tanno, Y. Mitsuuchi, D. A. Altomare, G. H. Xiao, and J. R. Testa, AKT activation up-regulates insulin-like growth factor I receptor expression and promotes invasiveness of human pancreatic cancer cells, *Cancer Res.*, vol. 61, no. 2, pp. 589-593, Jan. 2001.
- [46] X. H. Shi, Z. Y. Liang, X. Y. Ren, and T. H. Liu, Combined silencing of K-ras and Akt2 oncogenes achieves synergistic effects in inhibiting pancreatic cancer cell growth *in vitro* and *in vivo*, *Cancer Gene Ther.*, vol. 16, no. 3, pp. 227-236, Mar. 2009.
- [47] L. Castagnoli, A. Costantini, C. Dall'Armi, S. Gonfloni, L. Montecchi-Palazzi, S. Panni, S. Paoluzi, E. Santonico, and G. Cesareni, Selectivity and promiscuity in the interaction network mediated by protein recognition modules, *FEBS Lett.*, vol. 567, no. 1, pp. 74-79, Jun. 2004.
- [48] A. K. Gupta, G. J. Cerniglia, R. Mick, W. G. McKenna, and R. J. Muschel, HIV protease inhibitors block Akt signaling and radiosensitize tumor cells both *in vitro* and *in vivo*, *Cancer Res.*, vol. 65, no. 18, pp. 8256-8265, Sep. 2005.



Thomas Evangelidis is currently a PhD candidate in the Dept. of Pharmaceutical Chemistry, University of Athens, Greece. He received his BS in molecular biology & genetics (2007) from Democritus University of Thrace, Greece, his MRes in computational biology (2008) from University of York, UK, and his

MPhil in pharmaceutical sciences (2011) from University of Manchester, UK. His research interests include molecular modeling, computer-aided drug design, molecular dynamics simulations, and protein NMR.



Lei Xie is an associate professor in computer science at Hunter College, and PhD program in computer science, biology and biochemistry in the City University of New York. He has worked in the area of bioinformatics, molecular modeling, systems biology, and drug discovery for more than 10 years. His current research

focuses on multi-scale modeling of causal genotype-phenotype associations, structural systems pharmacology of existing and preclinical drugs, and polypharmacological drug design for complex diseases. His researches have been supported by National Institute of Health, National Science Foundation, and pharmaceutical/biotechnology companies such as Johnson & Johnson. He currently serves in the editorial board of *Journal of Computational Biology and Drug Design*. He is the winner of Young Investigator of Genome Technology in 2010 and 2012 due to his contributions to systems pharmacology. From 2004 to 2010, he was a principle scientist at San Diego Supercomputer Center (SDSC). Prior to his stint at SDSC, he worked in Roche and Eidogen to develop enterprise platform for drug discovery. He was trained in Computational Biology as a postdoctoral fellow at Prof. Barry Honig's group in Columbia University and Howard Hughes Medical Institute from 2000-2002. He obtained his PhD in organic chemistry and MS in computer science from Rutgers University–New Brunswick, both in 1999, and BS in polymer physics from University of Science and Technology of China in 1990.