A NOVEL MULTI-MODAL DRUG REPURPOSING APPROACH FOR IDENTIFICATION OF POTENT ACK1 INHIBITORS‡

SHARANGDHAR S. PHATAK
Integrated Molecular Discovery Laboratory (iMDL), The University Texas M.D. Anderson Cancer Center
School of Biomedical Informatics, The Univ. Texas Health Science Center
7000 Fannin St. Ste 600, Houston, Texas, 77030, USA
Email: sharangdhar@gmail.com

SHUXING ZHANG*
Integrated Molecular Discovery Laboratory (iMDL), The University Texas M.D. Anderson Cancer Center
1901 East Road, Unit 1950, Houston, TX, 77030, USA
Email: shuzhang@mdanderson.org

Exploiting drug polypharmacology to identify novel modes of actions for drug repurposing has gained significant attentions in the current era of weak drug pipelines. From a serendipitous to systematic or rational ways, a variety of unimodal computational approaches have been developed but the complexity of the problem clearly needs multi-modal approaches for better solutions. In this study, we propose an integrative computational framework based on classical structure-based drug design and chemical-genomic similarity methods, combined with molecular graph theories for this task. Briefly, a pharmacophore modeling method was employed to guide the selection of docked poses resulting from our high-throughput virtual screening. We then evaluated if complementary results (hits missed by docking) can be obtained by using a novel chemo-genomic similarity approach based on chemical/sequence information. Finally, we developed a bipartite-graph based on the extensive data curation of DrugBank, PDB, and UniProt. This drug-target bipartite graph was used to assess similarity of different inhibitors based on their connections to other compounds and targets. The approaches were applied to the repurposing of existing drugs against ACK1, a novel cancer target significantly overexpressed in breast and prostate cancers during their progression. Upon screening of ~1,447 marketed drugs, a final set of 10 hits were selected for experimental testing. Among them, four drugs were identified as potent ACK1 inhibitors. Especially the inhibition of ACK1 by Dasatinib was as strong as IC_{50}=1nM. We anticipate that our novel, integrative strategy can be easily extended to other biological targets with a more comprehensive coverage of known bio-chemical space for repurposing studies.

1. Introduction

The continual decline of the number of new small molecular entities from the pharmaceutical industry pipelines has been well documented[1]. The stop-gap measures such as mergers and outsourcing associated with the modern drug discovery process are unlikely to improve the drug discovery success rates in the long run[2]. Of several approaches under consideration to improve the

* To whom correspondence should be addressed.
‡ This work was partially supported by DOD BC085871 and Perot Foundation.

pipeline output, drug repositioning is the one that aims to increase the applicability of already discovered therapeutics to hitherto unknown clinical conditions. This approach may save time and costs associated with the discovery phase[2]. Drug repurposing certainly comes with some distinct advantages and the efforts have been driven by several important factors including: the access to increasing amounts of experimental data (e.g. kinase profiling[3]), better understanding of compound polypharmacology[4], biological data mining (BioCreative III)[5], and regulatory impetus from FDA and NIH[2]. Current successful examples are mostly from serendipitous discoveries such as the repurposing of buproprion from depression to smoking cessation as Zyban[6] and Duloxetine[7] from depression to stress urinary incontinence. Without doubt, there is an unmet need to develop novel, comprehensive methods for systematic drug repositioning to improve the efficiency.

In silico methods, either receptor-based or ligand-based, have been applied to drug repurposing projects. Keiser et al. predicted and validated 23 novel drug-target associations using two-dimensional chemical similarity approach (SEA)[8]. Recently the approach was employed for a large-scale prediction and testing of drug activity on side-effect targets[9]. Ligand-based quantitative structure-activity relationship (QSAR) models have been used by Yang et al. to predict indications for 145 diseases using the side effects as features[10]. With structure-based techniques, inverse docking was also used for drug repositioning[11, 12]. Likewise by mining drug phenotypic side effect similarities, Campillos et al. identified novel drug-target interactions[13]; Oprea et al. incorporated semantic method-based text mining for predicting novel drug actions[2]. With bipartite graph-based methods, Yildirim et al. linked FDA approved drugs to targets using binary associations[14], and Yamanishi predicted drug-target interactions using a combination of graph and chem-genomic approaches[15]. Our group recently conducted a comprehensive review of using molecular networks for drug discovery and development[16]. By developing models with other publicly available data, Dudley et al. repositioned Topiramate, an anti-convulsant drug to potential usage as an inflammatory bowel disease drug[17]. However, these unimodal approaches are likely to be limited by their respective shortcomings, e.g. inverse docking by scoring limitations[18]. Thus we propose that multimodal approaches may offer better solutions by offsetting the weakness of individual methods. In this study, we describe an integrative computational framework based on structure-based drug design and chemical-genomic similarity methods, combined with molecular network theories for drug repurposing. The approaches were applied to identification of existing drugs to target ACK1 for cancer treatment.

ACK1 (activated CDC42 kinase 1) is a ubiquitously expressed atypical non-receptor tyrosine kinase that integrates and delivers signals from multiple ligand-activated receptor tyrosine kinases such as EGFR, HER2 and PDGFR[19]. It also regulates several downstream proteins (e.g. AR, AKT and Wwox) implicated in cell survival roles[19, 20]. The activated ACK1 phosphorylates androgen receptor at Tyr-267 that leads to increased transcription of androgen receptors involved in the development of advanced metastatic prostate cancer or androgen independent prostate cancer[21, 22]. The knockdown of ACK1 increases cell apoptosis in prostate cancer cell lines, suggesting its importance as an anti-oncogenic drug target[22, 23]. Unlike the limited efficacies of conventional targeted therapeutics against RTKs, it has been hinted that ACK1 inhibitors may have higher efficacy for cancer treatment as it integrates signals from multiple RTKs and thus restraining the compensatory mechanisms of RTK signaling[20]. Although inhibitors targeting ACK1 have been

developed, publicly available data on them are still limited and few late stage clinical trials are being conducted to date. Therefore, it is an attractive cancer target for drug repurposing.

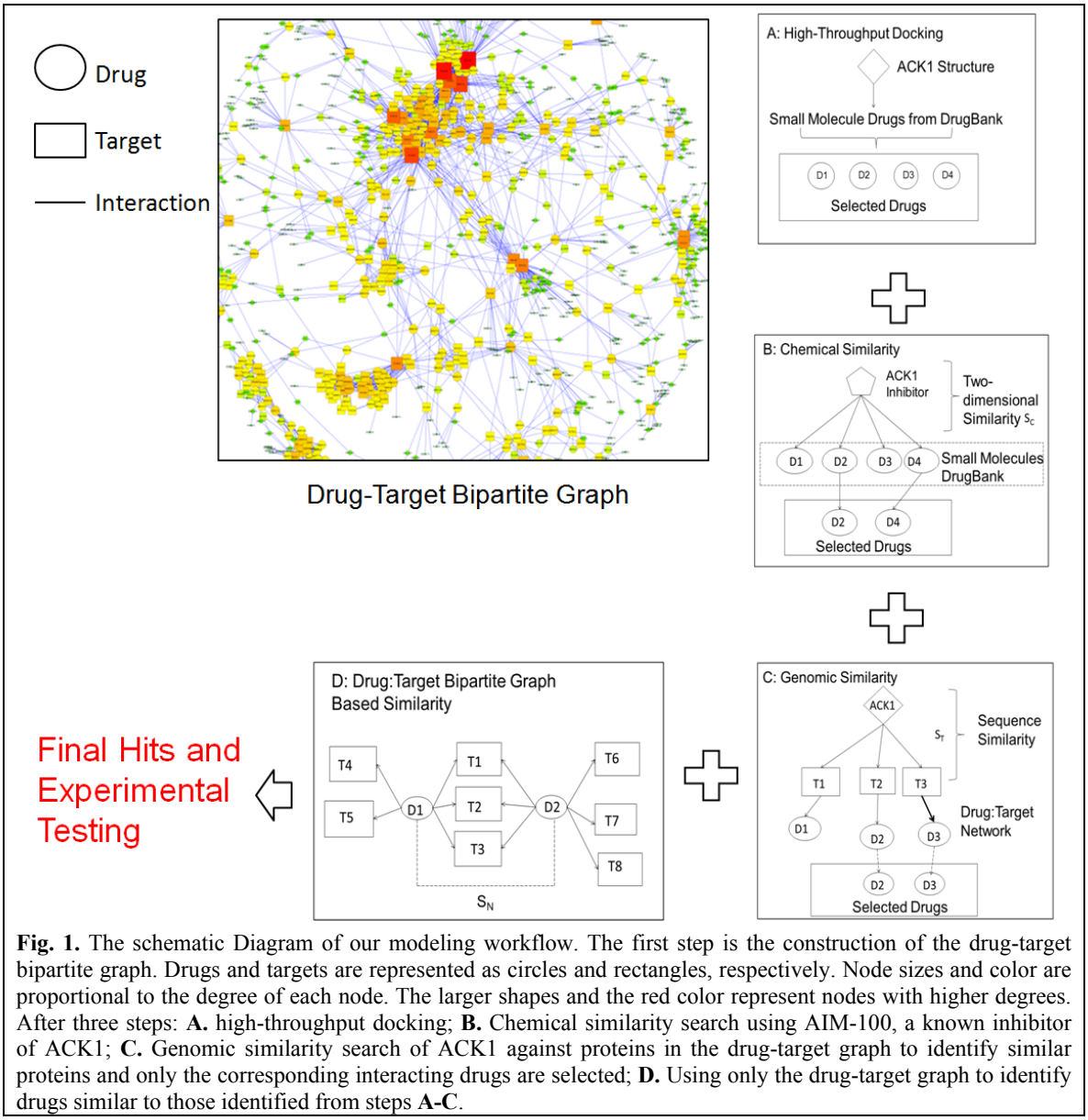

Fig. 1. The schematic Diagram of our modeling workflow. The first step is the construction of the drug-target bipartite graph. Drugs and targets are represented as circles and rectangles, respectively. Node sizes and color are proportional to the degree of each node. The larger shapes and the red color represent nodes with higher degrees. After three steps: **A.** high-throughput docking; **B.** Chemical similarity search using AIM-100, a known inhibitor of ACK1; **C.** Genomic similarity search of ACK1 against proteins in the drug-target graph to identify similar proteins and only the corresponding interacting drugs are selected; **D.** Using only the drug-target graph to identify drugs similar to those identified from steps **A-C**.

With our integrative approach consisting of classical structure-based drug design and chem-genomic similarity analysis approaches in tandem with the bipartite drug-target graph method, we identified 10 drugs for experimental testing. Four of them (Dasatinib, Sunitinib, Flavopiridol and Gefitinib) were confirmed active with IC_{50}<20uM. In particular, the IC_{50} of Dasatinib is as low as 1nM. Our results showed that integrative analysis of chemical-genomic features and molecular networks of drug-targeted interactions, combined with structure-based high-throughput docking could be successfully applied to drug repurposing for potent inhibitor discovery.

2. Methods and Materials

2.1. *Overall Approach*

Our drug repositioning workflow is illustrated by **Fig. 1** using an integrated three-level approach consisting of virtual screening, chemical genomic similarity, and bipartite-graph methods. The bipartite-graphs were developed based on the extensive data curation of DrugBank[23], Protein Data Bank (PDB)[24], and Protein Knowledge Base (UniProt)[25] using in-house developed python scripts. In brief, we employed high-throughput virtual screening followed by a pharmacophore-guided method to select a set of drugs as potential ACK1 inhibitors. Next, we evaluated if complementary results (hits missed by docking) can be obtained by using a novel chemo-genomic similarity approach based on chemical/sequence information. Finally, employing only the drug-target bipartite graph-based similarity, we identified a third set of drugs as potential ACK1 inhibitors. These three sets were further evaluated and merged into our final set consisting of 10 drugs which were evaluated using a qPCR-based kinase assays[26]. Four hits showed strong inhibition of ACK1 (1nM~20μM) and they can be potentially used for prostate cancer treatment.

2.2. *Virtual Screening*

Several structures of the ACK1 kinase domain are available in PDB. For virtual screening we chose two of them (3EQR and 1U4D) which are co-crystallized with very different ligands (T74 and DBQ, respectively). This strategy would implicitly accommodate for receptor flexibility and also possibly help us identify diverse chemotypes. Analysis of these two crystal structures revealed the importance of residues Ala208, Thr205, Glu206, Ala208 and Asp270 because they form hydrogen-bonding interactions with ligands. Particularly in 3EQR, the amine moiety on the 2,6-dimethylphenyl group of T74 interacts with the hydroxyl group on the conserved Thr205 residue. This hydrogen bond was found to significantly enhance the ACK1 inhibition in both biochemical and auto-phosphorylation assays as compared to its parent compound (N-aryl pyrimidine-5-carboxamide series)[27]. It suggests the importance of using this interaction as a pharmacophoric feature for subsequent hit selection. The high-throughput docking was conducted with the Glide software (www.schrodinger.com). Default parameters were used unless otherwise stated. The grid box with size 10Å X 10Å X 10Å was centered on the centroid of ligands (T74 or DBQ), and the active site flexibility was addressed with the induced-fit protocol. Only the approved/experimental drugs from DrugBank were selected for screening, and they were prepared with Epik, including their protonation and tautomer states at pH 7.0. The standard-precision (SP) mode was used for docking and scoring. To validate our protocol, both T74 and DBQ were re-docked into their respective co-crystallized crystal structure. In both cases, the ligands were docked within 1Å of their crystal structure binding poses. The Glide docking scores for T74 and DBQ were -10.4 and -9.26, respectively. Therefore, screened compounds with Glide scores above –9.26 were retained during hit selection via pharmacophore-based visual inspection. The pharmacophores were derived using MOE based on the analysis of the crystal structures and known ACK1 inhibitors (e.g., AIM-100)[28]. To be selected, the hits have to mimic at least three pharmacophoric features: **1).** a hydrophobic moiety in the nucleotide binding pocket surrounded by residues Ile190, Met203, and Leu207; **2).** hydrogen bonds with either Ala208, Thr205, Glu206 or Asp270; and **3).** a polar

solvent exposed group in the phosphate binding region of ACK1 surrounded by Asp215 and Arg216. With this strategy, the aim was to reduce the false positives by eliminating the dependence on docking scores as the only parameter because frequently many high ranked compounds could have completely wrong poses due to inaccuracies in scoring functions.

2.3. *Chem-genomic Similarity*

To compensate for the limitations of docking methods (e.g. inaccurate scoring functions), we implemented a novel approach by combining chemical and genomic similarity metrics. This was to identify those missing ACK1 inhibitors from virtual screening. The underlying assumption of our chemical similarity metric is that similar chemistry may result in similar biological activity. To this end, the MACCS fingerprints were employed as they represent chemical substructures within compounds as a *bitstring* using pre-defined substructures and are suitable for such applications. The similarity was expressed with Tanimoto coefficient defined as

$$Tc(d_{ij}) = |Ai \cap Bj| / |Ai \cup Bj| \qquad \textbf{Eq. 1.}$$

Where: Tc(d_{ij}) = Tanimoto coefficient between drugs *i* and *j*. A_i = number of *on* bits (1 is for *on* and 0 is for *off*) in drug *i*, B_j = number of *on* bits in drug *j*. This cheminformatics approach was implemented using the Openbabel toolkit (www.openbabel.org). Briefly, a known ACK1 inhibitor AIM-100[29] was used as the query compound and compared with all of the small molecule drugs in our curation. In order to determine the cutoff Tanimoto coefficient, AIM-100 was compared with Dasatinib (Tc = 0.61) as it was shown to be active against ACK1 in our virtual screening study. Therefore, only those drugs that were similar to AIM-100 with ± 5% of Tc = 0.61 were selected, and their affiliated targets in our curated data were obtained.

Genomic-based approaches in such studies were reported to be complementary to their cheminformatics counterparts[30]. Hence to enable rational selection of hits for experimental testing, all protein sequences from PDB were compared with the ACK1 kinase domain. For those sequences/targets with a meaningful genomic similarity with ACK1 (defined as sequence identity>40%), their corresponding drugs, if available in our data curation, were selected for experimental testing. For this step, the Needleman-Wunsch algorithm was employed to identify proteins from PDB similar to ACK1 and the proteins must be represented in our bipartite drug-target graph (described below). We considered the drugs connected to these proteins in the bipartite-graph as likely inhibitor candidates against ACK1.

2.4. *The Unweighted Drug-Target Bipartite Graph*

To use drug-target networks[14] in this study, we extensively curated data (e.g., structures, annotations, etc.) from multiple databases including DrugBank, PDB and UniProt, and developed an unweighted drug-target bipartite graph[16, 23]. Once the proteins were identified (e.g. based on genomic similarity), the respective PDB codes would be obtained from PDB and their corresponding co-crystallized drugs would also be derived. However we only selected those drugs that were present in the drug-target bipartite graph but not identified either from virtual screening or from chem-genomic similarity search. To this end, the DrugBank database was downloaded from the website (www.drugbank.ca). The initial database containing 6,711 drug entries included

6,580 small molecule drugs. For this study, entries containing biotech/nutraceuticals, withdrawn, illicit and other non-small molecule like (as defined by the chemical filter developed for this study) were excluded. This eventually resulted in 1,447 approved drugs in our curation. At the time of this work, the drugcard information did not contain The PDB codes were mapped to their respective UniProt codes using a a Biopython (www.biopython.org) based protocol to rationally reduce the complexity of the drug-target bipartite graph by eliminating redundant degrees as one UniProt code can effectively represent multiple pdb codes. Denoting the drug set as D = $\{d_1, d_2, \ldots, d_n\}$ and the target UniProt set as U =$\{u_1, u_2, \ldots, u_n\}$, the drug-target bipartite graph was developed as G(D,U,E) where E= $\{e_{ij}: d_i \in D, u_j \in U\}$. A link ($e_{ij}$ in E) is established between d_i and u_j only when there is an explicit association in the respective drug record.

2.5. *Graph-based Similarity*

The unweighted and undirected bipartite graph of drugs from DrugBank is shown in **Fig. 1**. Here, drugs are represented as vertices and their corresponding proteins as edges. Since this graph follows the power-law probability distribution[31], it is feasible to calculate the similarity between two vertices (drugs) based on the shared edges (proteins). Once the similarity of two drugs is established, their affiliated edges (proteins), even unshared ones, may be established as a likely target for the drugs respectively. In our study, we attempted to identify those drugs that shared graph-based similarity with any hit identified from docking and chem-genomic approaches. For the similarity metric we utilized the *Salton's* cosine measure as it normalizes the similarity measures and does not penalize/favor vertices that may have larger number of edges. This graph could easily be represented as an $n \times m$ adjacent matrix $\{a_{ij}\}$ where $a_{ij} = 1$ if d_i and u_j (drug and UniProt, respectively) were connected, or 0 if not. In an undirected network as in our case, the number n_{ij} of common neighbors of vertices *i* and *j* is given by:

$$n_{ij} = \sum_k A_{ik} A_{jk} \qquad \textbf{Eq. 2.}$$

Where *A* is the matrix. Thus, as proposed by Salton, the cosine similarity can be represented as:

$$\sigma_{ij} = \cos\theta = (\sum_k A_{ik} A_{jk}) / (\sqrt{\sum_k A_{ik}^2} \sqrt{\sum_k A_{jk}^2}) \qquad \textbf{Eq. 3.}$$

As our drug-protein network is an unweighted graph, the elements of the adjacency matrix take only the values of 0 and 1, so that $A_{ij}^2 = A_{ij}$ for all *i, j*. Then $\sum_k A_{ik}^2 = \sum_k A_{ik} = k_i$ where k_i is the degree (number of connections) of vertex *i.* Thus:

$$\sigma_{ij} = \frac{\sum_k A_{ik} A_{kj}}{\sqrt{k_i k_j}} = \frac{n_{ij}}{\sqrt{k_i k_j}} \qquad \textbf{Eq. 4.}$$

In simple terms, the cosine similarity of *i* and *j* is therefore the number of common neighbors (in our case, proteins represented by UniProt IDs) between two vertices (represented as drugs) divided by the geometric mean of their degrees. Therefore in this approach, only graph-based geometric similarity is considered without including any chemical/biological information.

2.6. *Experimental Testing*

To validate our predictions, the selected drugs were experimentally tested using the proprietary screening platform with a quantitative qPCR-based assay[26]. This approach measures the amount of DNA-tagged kinase that is unable to bind to an immobilized ligand attached to a fixed support. The kinase assays were developed as kinase-tagged T7 phage strains that are grown in parallel in 24-well blocks in an *E. coli* host derived from the BL21 strain and tagged with DNA for qPCR detection. Streptavidin-coated magnetic beads treated with biotinylated small molecule ligands for 30 minutes at room temperature were used to measure binding affinities for kinase assays. All hits were prepared as 40x stocks in 100% DMSO and directly diluted in the assays. All reactions were performed in polypropylene 384-well plates in final volume of 0.04 ml. The assay plates were incubated at room temperature with shaking for 1 hour, and the affinity beads was washed with buffer (1 X PBS, 0.05% Tween 20). The beads were re-suspended in elution buffer (1 X PBS, 0.05% Tween 20 0.5µM non-biotinylated affinity ligand). The kinase concentration in the eluates was measured by qPCR. The compounds were screened at 0.1µM and 10µM. In addition to ACK1, five other kinases of our interest and implicated in important cancer signaling pathways were used to evaluate selectivity of these inhibitors. The results for primary screen binding interactions were reported as %Ctrl where lower numbers indicate stronger hits:

$$\%\text{Ctrl calculation} = \frac{\text{test compound signal} - \text{positive control signal}}{\text{negative control signal} - \text{positive control signal}} * 100 \qquad \textbf{Eq. 5}.$$

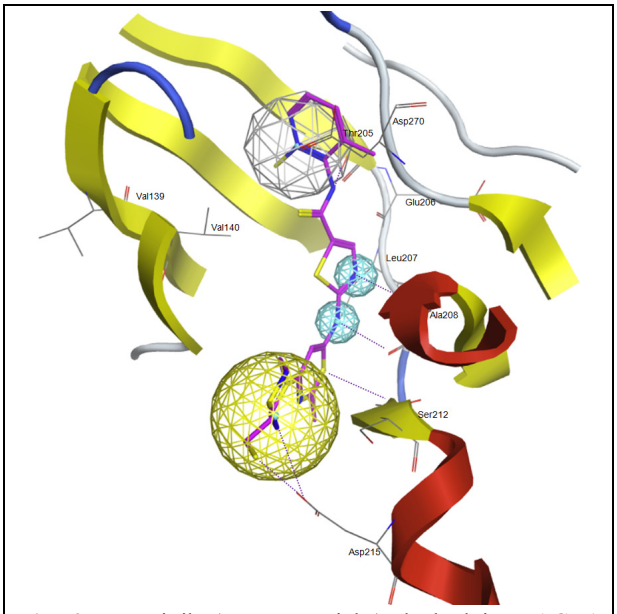

Fig. 2. Dasatinib (magenta sticks) docked into ACK1 (ribbon display). It was ranked top and has reasonable interactions with ACK1. The gray lines are critical residues in the active site. Hydrogen bonds are in magenta dashed lines. The spheres are pharmacophores: gray for hydrophobic, cyan for hydrogen bonds, and yellow for solvent exposed groups.

3. Results

3.1. *High-throughput virtual screening*

As described in the **Methods** section, small molecule drugs were docked and scored against two ACK1 crystal structures. Drugs scored above -9.26 were selected, also based on specific pharmacophoric features characterizing the binding poses. We particularly were interested in those hits with a hydrophobic moiety in the nucleotide binding pocket and forming hydrogen bonds with the Thr205 pocket. For example, Indinavir, a HIV protease inhibitor, was discarded despite being the best ranked hit (data not shown). On the other hand, although Dasatinib only ranked the 8th, it was selected because the drug demonstrated consistent binding pose with ACK1 (**Fig. 2**). Similarly, Amodiaquine, Flavoxate, Imatinib and Lapatinib were also selected based on our

docking studies with 3EQR and Mebendazole with 1U4D crystal structures. These hits also exhibited similar shape properties to the co-crystallized ligands of the respective crystal structure. We found that hits from 3EQR had high average molecular weight of 461Da (T74 MWT is 514Da). Screening with 1U4D which has the smaller co-crysallized ligand (DBQ, MWT =254Da) resulted in smaller hit (e.g., Mebendazole MWT=295Da). This was in-line with our hypothesis that diverse chemotypes might be obtained when different crystal structures are used.

3.2. *Chem-genomics based inhibitor identification*

The fundamental principle behind this approach is: **a).** compounds with similar chemistry are likely to possess similar biological profiles, and **b).** if there is meaningful genomic similarity (e.g., high sequence identity) between two proteins (thus also similar tertiary profile), compounds binding to one protein may interact with the other protein as well. We employed AIM-100 inhibitor for chemical similarity search. To determine the Tanimoto coefficient (Tc) threshold, AIM-100 was compared with Dasatinib (a promising binder based on docking) and we obtained Tc=0.61. Hence, all similar drugs within ±5% of Tc were kept. The small range of Tc is to ensure that the hits would maintain a certain degree of both chemical similarity and diversity. Based on drug-target bipartite graph, the corresponding targets of these selected drugs were also identified.

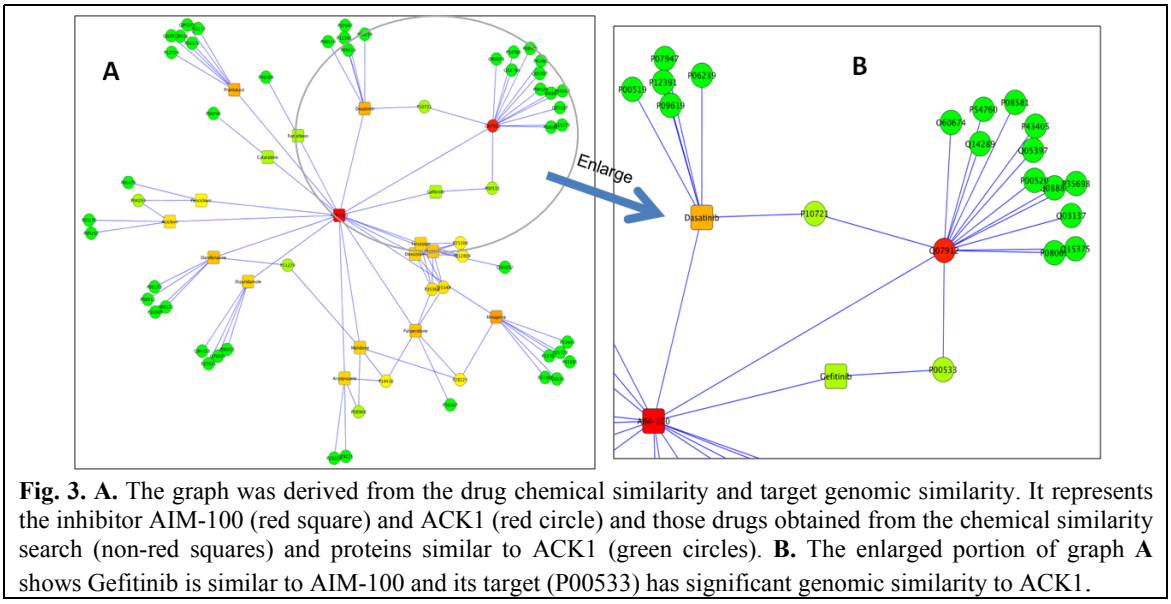

Fig. 3. A. The graph was derived from the drug chemical similarity and target genomic similarity. It represents the inhibitor AIM-100 (red square) and ACK1 (red circle) and those drugs obtained from the chemical similarity search (non-red squares) and proteins similar to ACK1 (green circles). **B.** The enlarged portion of graph **A** shows Gefitinib is similar to AIM-100 and its target (P00533) has significant genomic similarity to ACK1.

On the other hand, all proteins in our dataset were identified based on their genomic similarity to ACK1 (sequence identity>40%), and then their corresponding bound drugs were also obtained. These two sets of selected drug-target pairs were merged if two pairs shared the same target or the same drug. This resulted in a graph as demonstrated in **Fig 3**. Based on this combined chem-genomic similarity approach, Gefitinib, Sorafenib and Sunitinib were identified after excluding those (e.g., Imatinib) already identified by molecular docking. These observations were consistent with our postulation that combining *in silico* approaches, e.g. classical structure-based methods with molecular networks, might help identify unique and complementary sets of inhibitors.

3.3. *Graph-based similarity*

In this step, we attempted to identify potential ACK1 inhibitors based on their similarity to those already identified in the previous steps. However, the strategy was not based on chemical structure or genomic sequence similarities. Instead, the similarity was defined purely with our drug-target graph-based geometry (e.g., vertices and edges) without considering other chemical/biological information. We tried to investigate if this could provide us any extra hits. Using *Salton's* cosine index we calculated a similarity matrix based on the bipartite graph with the shared edges (proteins). A snapshot of the entire matrix is shown in **Fig. 4**. The hypothesis was that any small molecule drugs that showed some similarity to the previously identified inhibitors from the docking and chem-genomic similarity steps might be an inhibitor as well. As expected, we were able to identify the majority of the common hits such as Dasatinib and Imatinib (identified by both docking and similarity search methods). But we also identified new hits such as Flavopiridol as one of the ACK1 inhibitors, based on its graph similarity to Lapatinib. Though several other drugs were also identified, only Flavopiridol, along with another 9 drugs, was purchased for experimental testing due to the constraints of their commercial availability and our budgets.

3.4. *Experimental Results*

The *Kinomescan's* proprietary platform based on several thousands of profiled kinase inhibitors allowed the estimation of binding affinities of any compound based on their primary screening. The specific assay details of this approach are described elsewhere[26]. In addition to ACK1, we screened our selected compounds against several other kinases including EGFR, MEK1, PDPK1, PIK3CA and ABL2, because these targets are suggested to play important and diverse roles in various cancer pathways. EGFR, PDPK1 and PIK3CA are located in the signal transduction pathways that aid tumor growth and reduce apoptosis. MEK1 is located in the MAPK cell signaling that might affect the prognosis of the androgen-independent prostate cancer. We also tested ABL2 as it is the reported target of several drugs (e.g., Imatinib and Dasatinib).

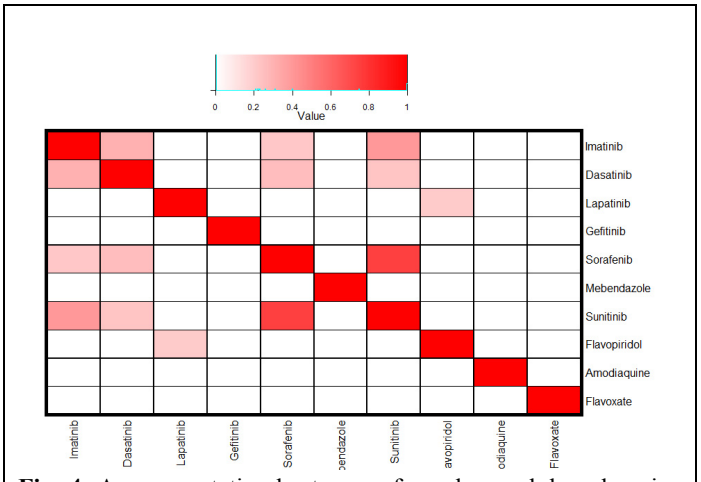

Fig. 4. A representative heat-map of purely graph-based cosine similarities of Flavopiridol against drugs identified from docking and chem-genomic similarity. The higher values (darker red) means higher graph-based similarity.

At the end, 10 hits were purchased and tested. Among them, four drugs including Dasatinib, Sunitinib, Flavopiridol and Gefitinib, showed significant inhibition of ACK1 with estimated IC_{50}<25μM. The activities of these compounds are illustrated in **Table 1**. These true ACK1 inhibitors were originally designed for different kinases, demonstrating the well-known polypharmacological properties of kinase inhibitors. In particular, Dasatinib was originally designed as a multi-BCR/ABL and Src family

tyrosine kinase inhibitor approved for chronic myelogenous leukemia (CML). Here we demonstrated that it also strongly inhibited ACK1 (further experiments showed IC_{50}=1nM) which is implicated in advanced prostate cancer patients. This provided a strong mechanistic support of using Dasatinib to treat prostate cancer. Interestingly, just after our experimental testing of these ACK1 inhibitors, Dr. Whang's group from UNC Chapel Hill published their evaluation of Dasatinib on inhibiting ACK1-related prostate cancer progression *in vitro* and *in vivo*[32]. Their discovery highly conformed to our *in silico* predictions. Currently we are teaming up to further explore repurposing of our identified drugs to treat advanced prostate cancer by targeting ACK1.

Table1. Experimental screening results of in silico drug hits against six kinases.

Target / Compounds	ACK1		PIK3CA		PDPK1		ABL2		EGFR		MEK	
	0.1μM	10μM	0.1μM	10μM	0.1μM	10μM	0.1μM	10μM	0.1μM	10μM	0.1μM	10μM
Amodiaquine	89	100	97	93	100	93	97	100	100	91	95	96
Gefitinib	100	77	100	100	100	1000	100	83	2.2	0	100	83
Lapatinib	100	93	98	92	100	100	99	100	0.25	0.05	95	87
Imatinib	100	82	100	92	100	100	43	3.2	100	71	95	94
Dasatinib	4	0	95	98	100	100	0.15	0	59	1.2	89	3.2
Sorafenib	100	89	100	99	100	85	97	33	100	96	100	99
Mebendazole	100	98	100	100	100	100	100	33	100	83	100	38
Flavoxate	100	94	100	100	100	100	100	94	100	88	94	95
Sunitinib	93	33	100	83	100	39	92	51	92	72	51	0.1
Flavopiridol	100	74	100	97	100	100	100	80	100	53	92	92

Sunitinib, Flavopiridol and Gefitinib were originally developed as PDGFR-Beta, CDK-2, and EGFR inhibitors, respectively, but also inhibited ACK1 based on our results. Imatinib and Sorafenib only showed moderate inhibition of ACK1. Flavoxate and Mebendazole were initially considered interesting as they are not kinase inhibitors but were predicted to inhibit ACK1. Unfortunately experimental results indicated that they were either false positives or weak ACK1 inhibitors. Therefore no further work is being performed on them but our efforts of identifying new chemotypes (non-kinase inhibitors) as ACK1 inhibitors are still undergoing.

3.5. *Comparison of Different Methods*

Our multi-modal approach clearly differs from other unimodal methods developed for drug repurposing such as SEA[30] and AERS-based method[33]. Cheng et al. recently evaluated multiple schemes and they found that their network-based interference (NBI) approach obtained better results in their cases[34]. In this study, we focused on a combined strategy but also investigated in details how each method is different from the others in their ability to identify ACK1 inhibitors. **Table 2** demonstrates docking-based virtual screening could reveal more diverse chemotypes including both kinase and non-kinase inhibitors. As expected, drugs uncovered with chemical structure and protein sequence based similarity analysis are all kinase inhibitors. Gefitinib and Sunitinib were shown to have low micromolar affinity to ACK1. Lastly, the graph-based similarity method, which does not include any chemical, biological, or sequence/structure information,

identified a different chemotype drug -- Flavopiridol. It exhibits ~25uM inhibition of ACK1. The common hits (blue in **Table 2**) by these methods are Imatinib and Dasatinib, and in particular, the later demonstrates a nanomolar IC_{50}. Clearly this multi-modal approach shows improved performance over each individual methods in the present study.

Table 2. Drugs identified by different methods.

High-throughput Docking	Chem-genomic Similarity Analysis	Graph-based Similarity Analysis
Imatinib	Imatinib	Imatinib
Dasatinib	Sunitinib	Dasatinib
Lapatinib	Gefitinib	Flavopiridol
Mebendazole	Sorafenib	
Amodiaquine		
Flavoxate		

However, certain limitations still exist. For structure-based docking methods, the target 3D structures are usually required. To reduce false predictions, we incorporated as much known expert knowledge as possible such as using multiple ACK1-inhibitor complex structures to partially compensate for target flexibility[18]. We also filtered the top-ranked hits with protein pharmacophores[27]. The chemical similarity based methods are generally reliable, but combination with shape-based techniques may give better results[35]. For the graph-based analysis we were limited to the publicly available drug-protein interaction information. As the data increases, we expect our predictions will be continuously improved.

4. Conclusions

Understanding the drug polypharmacology may hold a great promise in our next generation of drug discovery and development. Along the line, drug repurposing applications are getting more and more attention as it may provide an efficient and effective way to fuel the current drug discovery engines. Both FDA and NIH have recently put a significant amount of funding and effort to promote drug repurposing. From *in silico* point of view, more multi-modal approaches and data integration are needed to increase our opportunity of success. To this end, our present study is to integrate the classical structure-based methods with chem-genomic similarity approaches, along with molecular graph theories to develop new strategies for drug repurposing. Our approach was applied to identification of existing drugs as ACK1 inhibitors for prostate cancer treatment, and multiple potent inhibitors have been discovered.

Our three-pronged approach consisted of curating currently available drug-target information into high-quality bio-chemical databases. Next, by combining the high-throughput molecular docking, chem-genomic similarity search and our in-house drug-target bipartite graphs, we identified 10 promising hits. Subsequent experimental profiling of these selected drugs against six kinases indicated that four of them, including Dasatinib, Sunitinib, Flavopiridol, and Gefitinib, could significantly inhibit ACK1. In particular, the IC_{50} of Dasatinib was as low as 1nM. Therefore we have demonstrated that, extensive analysis of chemical-genomic features, characterization of drug-target relations with graph-based approaches, and classical high-throughput docking are complementary to each other. The combination use of these methods can efficiently and accurately reveal strong inhibitors, corroborating our hypothesis of the need for an integrative approach for drug repurposing. In principle, this approach can be easily extended to other biological targets and chemical databases as a general tools for drug repurposing.

5. Acknowledgements

We specially thank Dr. Young Whang from UNC Chapel Hill School of Medicine for discussing ACK1 inhibitors for treating prostate cancer. We also thank Lu Chen for helping generate pharmacophores. This work was also supported by the MDACC IRG program and the Prostate SPORE Career Award (P50CA140388 SPORE). SSP was supported by UT-Health Innovation for Cancer Prevention Research Pre-Doctoral Fellowship from CPRIT Grant # RP101503.

References

1. B. H. Munos and W. W. Chin, *Sci Transl Med* **3**, 89cm16 (2011)
2. T. I. Oprea, S. K. Nielsen, O. Ursu et al., *Mol Inform* **30**, 100-111 (2012)
3. J. T. Metz, E. F. Johnson, N. B. Soni et al., *Nat Chem Biol* **7**, 200-2 (2011)
4. G. V. Paolini, R. H. Shapland, W. P. van Hoorn et al., *Nat Biotechnol* **24**, 805-15 (2006)
5. C. N. Arighi, P. M. Roberts, S. Agarwal et al., *BMC Bioinformatics* **12 Suppl 8**, S4 (2011)
6. M. S. Boguski, K. D. Mandl and V. P. Sukhatme, *Science* **324**, 1394-5 (2009)
7. T. T. Ashburn and K. B. Thor, *Nat Rev Drug Discov* **3**, 673-83 (2004)
8. M. J. Keiser, V. Setola, J. J. Irwin et al., *Nature* **462**, 175-81 (2009)
9. E. Lounkine, M. J. Keiser, S. Whitebread et al., *Nature* **486**, 361-7 (2012)
10. L. Yang and P. Agarwal, *PLoS One* **6**, e28025 (2011)
11. Y. Z. Chen and D. G. Zhi, *Proteins* **43**, 217-26 (2001)
12. Y. Y. Li, J. An and S. J. Jones, *Genome Inform* **17**, 239-47 (2006)
13. M. Campillos, M. Kuhn, A. C. Gavin et al., *Science* **321**, 263-6 (2008)
14. M. A. Yildirim, K. I. Goh, M. E. Cusick et al., *Nat Biotechnol* **25**, 1119-26 (2007)
15. Y. Yamanishi, M. Araki, A. Gutteridge et al., *Bioinformatics* **24**, i232-40 (2008)
16. J. K. Morrow, L. Tian and S. Zhang, *Crit Rev Biomed Eng* **38**, 143-56 (2010)
17. J. T. Dudley, M. Sirota, M. Shenoy et al., *Sci Transl Med* **3**, 96ra76 (2011)
18. L. Chen, J. K. Morrow, H. T. Tran et al., *Curr Pharm Des* **18**, 1217-39 (2012)
19. K. Mahajan and N. P. Mahajan, *J Cell Physiol* **224**, 327-33 (2010)
20. N. P. Mahajan, Y. E. Whang, J. L. Mohler et al., *Cancer Res* **65**, 10514-23 (2005)
21. M. E. Grossmann, H. Huang and D. J. Tindall, *J Natl Cancer Inst* **93**, 1687-97 (2001)
22. C. D. Chen, D. S. Welsbie, C. Tran et al., *Nat Med* **10**, 33-9 (2004)
23. L. Tian and S. Zhang, *Conf Proc IEEE Eng Med Biol Soc* **2009**, 2336-9 (2009)
24. H. M. Berman, K. Henrick, H. Nakamura et al., *Nat Biotechnol* **25**, 845-6 (2007)
25. R. Apweiler, A. Bairoch, C. H. Wu et al., *Nucleic Acids Res* **32**, D115-9 (2004)
26. M. A. Fabian, W. H. Biggs, 3rd, D. K. Treiber et al., *Nat Biotechnol* **23**, 329-36 (2005)
27. D. J. Kopecky, X. Hao, Y. Chen et al., *Bioorg Med Chem Lett* **18**, 6352-6 (2008)
28. L. Du-Cuny, L. Chen and S. Zhang, *J Chem Inf Model* **51**, 2948-60 (2011)
29. K. Mahajan, D. Coppola, Y. A. Chen et al., *Am J Pathol* **180**, 1386-93 (2012)
30. M. J. Keiser, B. L. Roth, B. N. Armbruster et al., *Nat Biotechnol* **25**, 197-206 (2007)
31. P. Sheridan, T. Kamimura and H. Shimodaira, *PLoS One* **5**, e13580 (2010)
32. Y. Liu, M. Karaca, Z. Zhang et al., *Oncogene* **29**, 3208-16 (2010)
33. M. Takarabe, M. Kotera, Y. Nishimura et al., *Bioinformatics* **28**, i611-i618 (2012)
34. F. Cheng, C. Liu, J. Jiang et al., *PLoS Comput Biol* **8**, e1002503 (2012)
35. S. R. Vasudevan, J. B. Moore, Y. Schymura et al., *J Med Chem* **55**, 7054-60 (2012)