

# Novel Kinase Inhibitors by Reshuffling Ligand Functionalities Across the Human Kinome

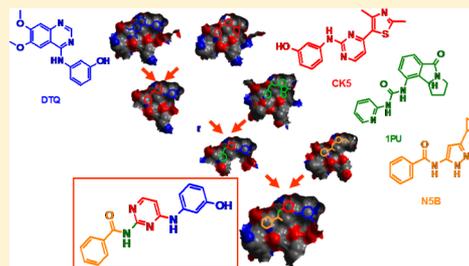
Dušica Vidović,<sup>†</sup> Steven M. Muskal,<sup>‡</sup> and Stephan C. Schürer<sup>\*,†,§</sup>

<sup>†</sup>Center for Computational Science and <sup>§</sup>Department of Pharmacology, Miller School of Medicine, University of Miami, Miami, Florida 33136, United States

<sup>‡</sup>Eidogen-Sertanty, Inc., Oceanside, California 92056, United States

**S** Supporting Information

**ABSTRACT:** Protein kinases remain among the most versatile and prospective therapeutic drug targets with currently 15 distinct compounds approved for use in humans and numerous clinical development programs. The vast majority of kinase inhibitors bind at the ATP site. Here we present an integrated workflow to amplify the rapidly increasing space of structurally resolved small molecule kinase ligands to generate novel inhibitors. Our approach considers both receptor-based similarity constraints in cocomplexes and ligand-based filtering/refinement methods to generate novel, drug-like matter. After building a comprehensive database of the structural kinome and identifying ATP-competitive ligands, we leverage local site similarities and site alignments to shuffle ligand fragments across the kinome. After extensive curation and standardization, our automated protocol starting from 936 cocrystal ATP-competitive binding sites generated about 150 000 new ligand structures among them over 26 000 lead-/drug-like compounds; the majority of those are novel based on structural similarity and scaffolds. In a retrospective analysis we demonstrate that our protocol produced known potent kinase inhibitors and we show how docking can be applied to prioritize the most likely efficacious compounds. Our workflow emulates a common strategy in medicinal chemistry to identify and swap corresponding moieties from known inhibitors to generate novel and potent leads. Here, we systematize and automate this approach leveraging available knowledge covering the entire human Kinome.



## ■ INTRODUCTION

During the past decade, protein kinases have emerged as one of the most versatile and prospective family of drug targets.<sup>1</sup> More than 500 protein kinases, typically referred to as the human Kinome,<sup>2</sup> are involved in literally every signal transduction cascade. Selective inhibition of protein kinases has therefore been considered an attractive therapeutic strategy for a wide variety of disorders, including cancer, immunological, neurological, metabolic, and infectious diseases.<sup>3,4</sup> However, due to the ubiquity of kinases in cellular functions, the long-term use of kinase drugs is often related to the side-effects and toxicity.<sup>5,6</sup> This has been a justifiable trade-off for otherwise untreatable and severe diseases. To date, 15 small molecule kinase inhibitors have received US Food and Drug Administration approval for use in humans to treat various forms of cancer (compiled from several sources) while more than 300 kinase inhibitors are in clinical development (ChEMBL); most of them also for cancer.

The vast majority of kinase inhibitors target the ATP-binding site of the kinase catalytic domain, which is highly conserved across the human kinome. This has important consequences for the development of novel compounds. Achieving selectivity across the more than 500 kinases can be a major challenge, although several selectivity strategies exist including exploiting gatekeeper residue differences<sup>7</sup> and targeting the kinase inactive vs active conformation.<sup>8</sup> Available kinase profiling platforms

enable the assessment of kinase inhibitor selectivity and various studies suggest that the development of selective ATP-competitive compounds is possible.<sup>9–12</sup>

On the other hand, conservation of the kinase ATP binding site can be exploited to optimize existing lead compounds toward alternative, originally unintended, kinase targets. Perhaps most importantly, similarities in the kinase ATP sites provide an opportunity for the development of poly pharmacology inhibitors.<sup>9,13</sup> In principle, kinases with the most similar ATP sites may be targetable with the same (optimized) ligand. We have previously shown that protein binding site similarities are correlated with corresponding (experimentally determined) small molecule binding affinities among protein tyrosine phosphatases<sup>14</sup> and protein tyrosine kinases (unpublished results).

It is for these reasons that the ATP site remains an attractive target for kinase drug discovery. The large number of kinase inhibitor drug and probe discovery projects in the pharmaceutical industry and in academia has led to crowded space, in particular with respect to intellectual property, and consequently to efforts of developing novel kinase inhibitor scaffolds.<sup>15</sup> However, there is also a vast increase in publically available kinase ligand and structural knowledge, and thus, there

Received: August 16, 2012

exists an opportunity to systematically amplify the current space.<sup>16</sup> For example the Protein Data Bank (PDB)<sup>17</sup> now covers over 100 distinct human kinase domains and over 1000 distinct kinase ATP cocrystal sites (vide infra).

Fragment-based discovery of lead compounds has evolved into an established discipline during the last ten years based on the rationale of the reductionist approach to sample chemical space by low complexity “fragment” ligands while still exploring drug-like chemical space by combining privileged fragments.<sup>18,19</sup> The importance of privileged chemical fragments to understand and model small molecule kinase activity has been demonstrated.<sup>20</sup> Leveraging the well-known Breed algorithm,<sup>21</sup> we developed a fully automated workflow to amplify the current space of kinase cocrystal ligands by systematically reshuffling ligand fragments across the entire human Kinome based on alignments of the local ATP-binding sites. Besides well-defined characterization of the ATP-binding sites, the abundance of the experimental structural information of the ATP binding site ligands qualify them as great candidates for the fragment-based approach presented here (in contrast to allosteric sites for example that lack such extensive structural information).

Because the input ligands are kinase ATP-site binders and the algorithm shuffles fragments across aligned binding sites, the resulting ligand-cross products are likely enriched with potential kinase inhibitors. We describe the workflow components and their integration and characterize the generated kinase ligand space. In a retrospective analysis, we demonstrate that our protocol produced novel and potent kinase inhibitors. We make available all generated structures, scaffolds, and the workflow protocol.

## MATERIALS AND METHODS

### Database of Kinase Structures and Cocrystal Ligands.

To systematically explore the corpus of available protein kinase structure and their ligands, we created a focused kinase TIP (Target Informatics Platform, Eidogen-Sertanty)<sup>22</sup> database. We searched the PDB<sup>17</sup> for entries which contain keyword “kinase” and the experimental method “X-ray”. The search resulted in 2610 records (as of Jan 2, 2009). For each record we retrieved the PDB entry code, PDB sequence, resolution, *R*-value, *R*-free, gene, species, Swiss-Prot accession, and the PDB ligand code. To identify the kinase domain PDB chains, we split the PDB records by chains and then compared each PDB chain sequence to the human kinase domain sequences retrieved from the Sugen/Salk Kinbase database (<http://kinase.com/kinbase/>; June 2007). Before we used them as a reference, we curated the kinase domain sequences to remove duplicate kinase domains, pseudogenes entries, and atypical kinase domains. We also removed the domains that did not match the corresponding Uniprot/Swiss entries after comparing their sequences as an exact string match and by BLAST score.<sup>23</sup> This curation procedure resulted in 488 human kinase domain sequences that were further used for the identification of kinase PDB chains among selected PDB entries.

BLAST searches were performed for each PDB chain to identify the ones that correspond to human kinase domains. 1117 PDB chains (one chain per PDB entry) were kept with an expectation value of less than 10; it should be noted here that the presence of nonkinase domain structures would not negatively influence the modeling in TIP. From the corresponding PDB records, we created a TIP database.

Genes corresponding to the PDB chains were assigned based on the consensus of PDB SwissProt annotations and sequence comparison against the Sugen kinase domains and Uniprot. Where no consensus was obtained, annotations were done manually considering protein symbol synonyms and domain sequence descriptions from Uniprot.

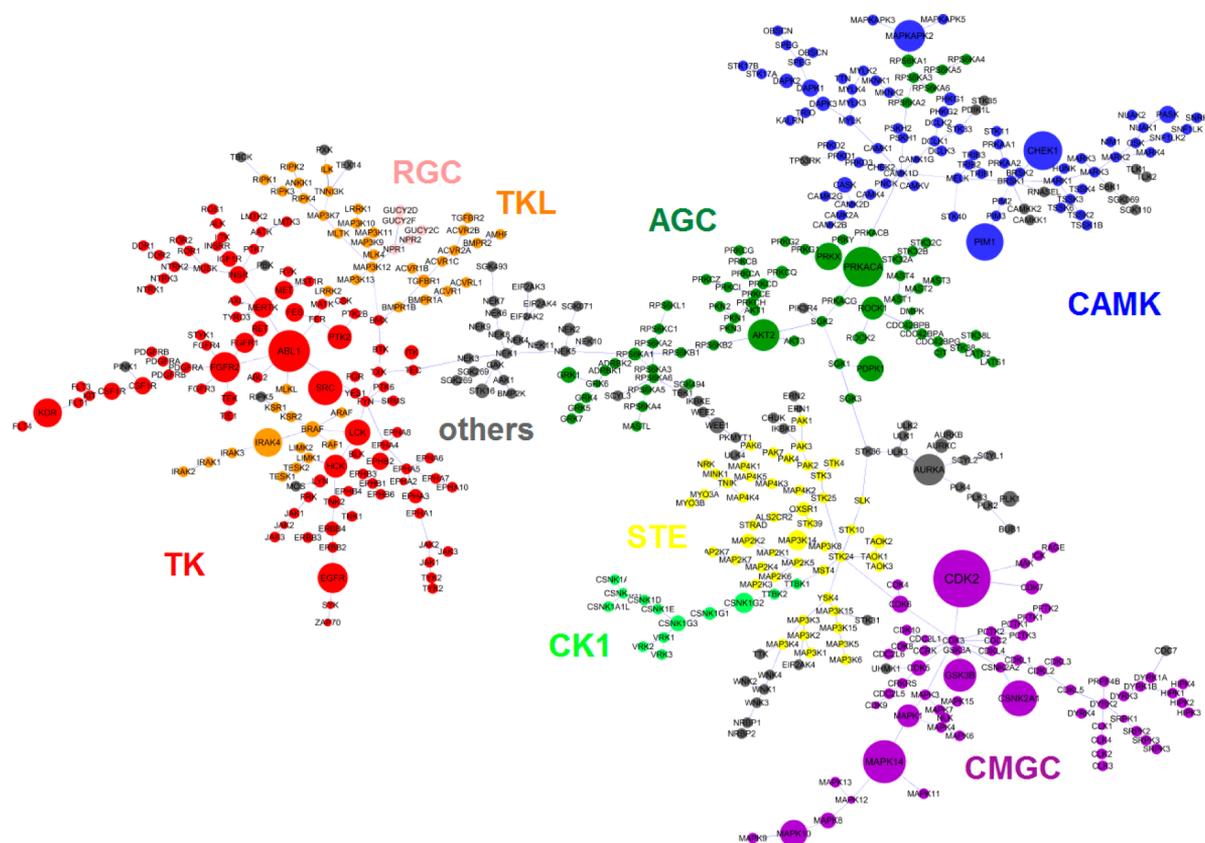
**Identification of ATP Binding Sites.** All cocrystal ligand structures and cocrystal sites were extracted from the TIP kinase database (3168 unique sites). The PDB 2D ligand structures from the PDB ligand database (downloaded on Jun 9, 2009) were merged with the cocrystal sites in the TIP database by PDB ligand code and chain identifier resulting in 2280 unique sites.

Not all of those annotated sites are catalytic sites. The ATP binding sites were identified based on the chemical structure of the cocrystal ligands. Salt, solvent, and addends used for crystallization were removed and the remaining ligands were reviewed manually. In cases where the cocrystallized ligand was not a heterocyclic compounds with typical ATP-binding characteristics,<sup>24</sup> the cocrystal complex was manually inspected to verify the ATP binding site. This process resulted in 1018 ATP cocrystal binding sites corresponding to binding sites annotated in the TIP database.

**Kinase ATP Site Similarities.** The SiteSorter algorithm implemented in the TIP software system computes pairwise 3D similarities between sites based on their graph representations and the optimal overlay of the two considered sites determined by a clique detection algorithm similar to Klebe’s approach.<sup>25</sup> This optimal overlay of the two sites is further used to derive a score on the basis of chemical group similarity incorporating side chain and backbone atoms. Using this site similarity score, any reference site in TIP can be queried for similar sites. Here we further normalized the site similarity values based on site size as follows:  $S_{AB\text{normalized}} = S_{AB}/(S_{AA} + S_{BB} - S_{AB})$ , where  $S_{AB}$  is the calculated score for the overlaid sites A and B and where  $S_{AA}$  and  $S_{BB}$  are self-site similarities for site A and site B, respectively. The rank-order of sites the most similar to a query reference was rescored according to normalized site similarity. In the process of computing pairwise site similarities, sites are aligned to the query; this provides the reference coordinates to which the cocrystal ligands that correspond to the ATP sites are aligned. A TIP project that contains the query results with the site alignment information can be generated, downloaded, and further processed using the EVE (Eidogen Visualization environment) client application.<sup>26</sup>

### ATP Ligand Functionality Shuffling and Parameters.

An algorithm to systematically shuffle ligand functionality (ligand cross) is implemented in EVE. EVE aligns the project structures based on the selected reference site; this superimposes the corresponding cocrystal ligands as they bind to the ATP site. On the basis of this overlay, EVE identifies matching ligand bonds. Bonds were considered matching if the distance between the corresponding atoms of the two overlapping bonds is less than 1 Å and the angle between the bonds is less than 15° (when one bond is translated so the corresponding atoms overlap). For each of the identified matching bonds of each pair of superimposed ligands, the new ligands are generated by swapping the fragments of the ligands on the opposite sides of the matching bond. This process can be repeated over several iterations. We performed three iterations; the results of the first iteration were used as the input to the second iteration and the results of the second iteration were the input ligands to the third iteration. This ligand cross algorithm



**Figure 1.** Phylogenetic diversity and coverage of the structurally resolved Kinome with corresponding ATP-binding cocrystal ligands (node size corresponds to number of records).

is similar to Breed.<sup>21</sup> In addition to the GUI interface, EVE is command line enabled, which facilitated scripting and automation.

**Workflow Integration of Site Similarity Query and Ligand Functionality Shuffling Across the Kinome.** An integrated workflow was developed and implemented to query the TIP database by reference sites, extract the most similar sites with their corresponding (PDB) ligands, generate an EVE project, execute the ligand cross ligand functionality shuffling algorithm, and extract the newly generated ligands. The protocol was implemented in Pipeline Pilot 7 (Accelrys). Pipeline Pilot Web service components were built to authenticate to the TIP server, query TIP, and extract the results. EVE was run on a local client using the Pipeline Pilot command line execution component. All required components and the entire ligand cross protocol are available from Eidogen-Sertanty upon request (<http://www.eidogen-sertanty.com/>).

This protocol was executed for all 1018 ATP sites with a valid cocrystal ligand as identified above. For each reference site up to 50 most similar ATP binding sites including the site itself were kept based on the normalized site similarity score. The input 3D ligands were extracted and written to a SDF file, an EVE project was generated, and ligand cross functionality shuffling was executed in EVE using the parameters above in three iterations. The resulting generated 3D molecule structures were written to an SDF file. A total of 1 287 206 molecules were generated.

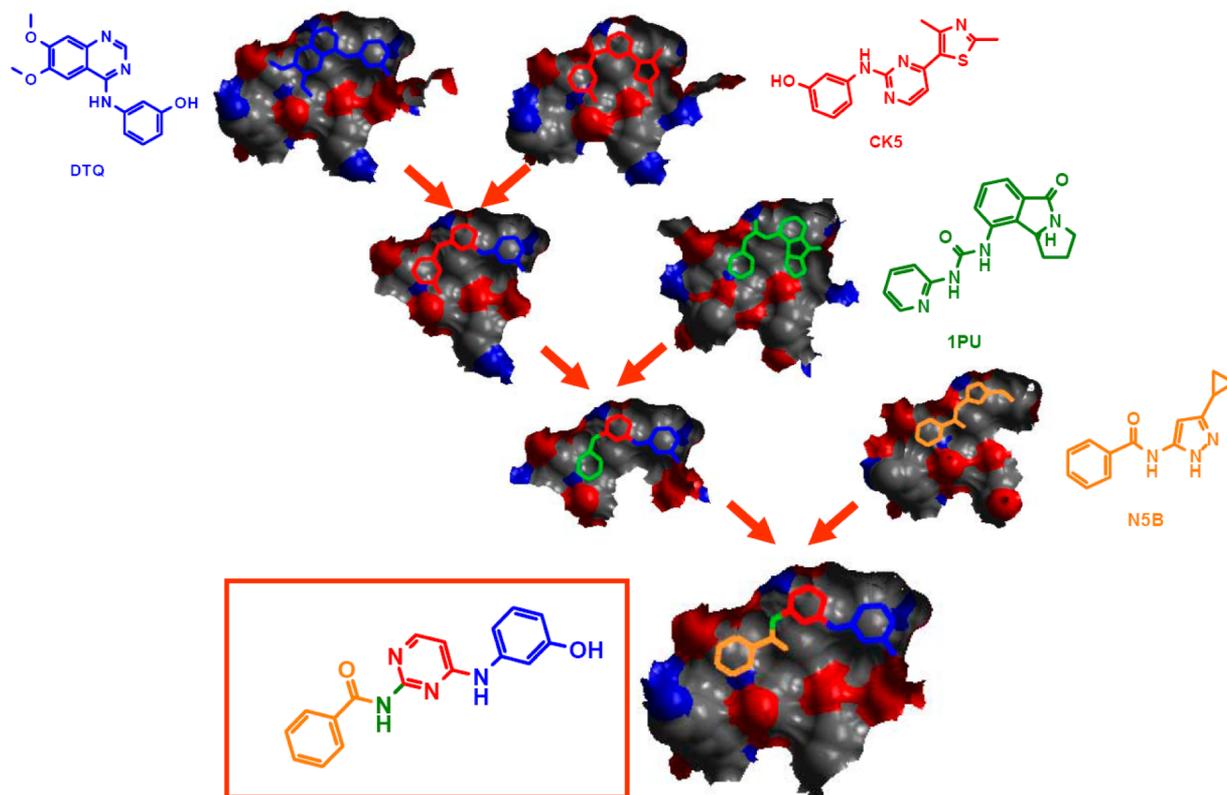
**Ligand-Cross Results Corrections.** Upon reviewing the results, we observed structural inconsistencies in some of the generated cross-products. Several structures with overlapping atoms were generated. In order to eliminate such structures

from the further analysis, we used the van der Waals (VDW) distance between the atoms as a criteria as was implemented in the Accelrys Pipeline Pilot<sup>27</sup> component “Bump Check Filter” (with the value for VDWRadiusScaleFactor of 0.8). We also found structures with incorrect bond order, mainly in ring systems; this is due to missing bond order definitions on the 3D PDB representations. To correct bond orders, we applied a bond order assignment algorithm implemented in Pipeline Pilot (Accelrys) as well as a set of transformations implemented as intramolecular reactions using the ChemAxon<sup>28</sup> Pipeline Pilot components.

**Lead- and Drug-Likeness.** In order to identify lead- and drug-like molecules among generated structures, we used drug-like and lead-like filters as implemented in Filter from OpenEye.<sup>29</sup>

**Novelty of the Generated Kinase Inhibitors.** In order to estimate novelty of the generated ligand-cross products we compared the Murcko assemblies of the generated molecules to the starting material, i.e. to the PDB ligands. Murcko assemblies are contiguous ring systems including the chains that link two or more rings.<sup>30</sup> Here we used the extended-connectivity topological fingerprints (ECFP4)<sup>31</sup> to represent the structures and the Tanimoto similarity metric, which is the most widely used feature-based similarity measure.<sup>32,33</sup>

**Identification of Known Kinase Inhibitors among the Generated Molecules.** The corrected and filtered (drug- and/or lead-like) chemical structures were processed by a Pipeline Pilot protocol to generate a canonical structure representation (including a canonical tautomer). Known kinase inhibitors and their corresponding experimental data were extracted from the Kinase Knowledge Base (KKB, Q4 2009,



**Figure 2.** Illustration of the ligand cross algorithm. Similar sites are selected and overlaid. The corresponding fragments of two ligands are swapped along an aligned single bond given distance and angle constraints (here, the bonds are considered a match if the angle between the bonds is less than  $15^\circ$  and the distance between the corresponding atoms is less than 1 Å). The ligand crossing using four input PDB structures (shown as blue, red, green, and yellow) generate several new ligands that fit into the aligned binding sites.

<http://www.eidogen-ertanty.com/kinasekb.php>) from Eidogen-Sertanty<sup>34</sup> and were standardized using the same procedure after stripping salt and addends. Ligand cross results and known kinase inhibitors were compared by unique canonical SMILES. Among other known inhibitors, 92 CDK2 kinase inhibitors with the  $pIC_{50}$  values greater than 6 were identified and further analyzed.

**Kinase Inhibitors Docking.** To illustrate how ligand-cross generated molecules can be further prioritized with respect to their potential as kinase inhibitors, we docked the 92 generated CDK2 inhibitors and correlated their experimental activities with the docking scores. The ligands were prepared by LigPrep (Schrödinger LLC)<sup>35</sup> to generate tautomers and ionization forms ( $pH = 7 \pm 2$ ). Standard precision (SP) flexible docking was carried out using Glide (Schrödinger LLC).<sup>36</sup>

## RESULTS AND DISCUSSION

**Structural Kinome and Cocrystal Ligands.** To systematically explore the available structural knowledge of the human Kinome, we created TIP database derived of PDB records with similarity to human kinase domains (see Materials and Methods). We identified those cocrystal ligands that bind to the ATP site. We considered 1018 ATP binding sites and their corresponding ligands. We inspected the ligands as candidate input structures of (our implementation of) the ligand-cross algorithm (see below). Ligand bond orders often cannot be directly inferred from the available experimental PDB records or, in some cases, were found to be erroneous or entirely omitted. EVE assigns bond orders based on atom distance and angular constraints; however, due to nonideal ligand geometry,

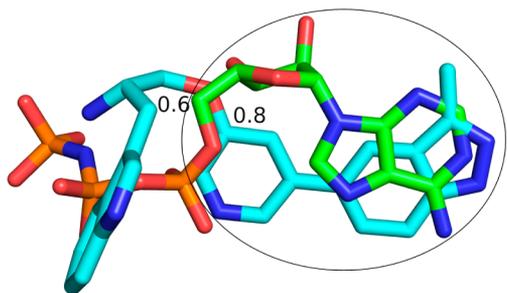
in several cases bond lengths and angles can deviate from the predefined ranges and lead to bond order errors. In order to remove such structures, we downloaded curated ligand structures from Ligand Expo.<sup>37</sup> Ligand Expo ligands of interest were identified by assigned codes of the corresponding PDB ligands. For each Ligand Expo, structure we generated ionization and tautomerization states to compare them to the EVE 3D representation of the corresponding PDB record. Here, 82 ligands among 1018 structures did not match any of the enumerated protonation/ionization and tautomerization states and were therefore eliminated as input structures. Figure 1 illustrates phylogenetic coverage of the structurally resolved Kinome, and the corresponding inhibitor ligands considered here. The 936 binding sites correspond to 108 kinase family members spanning the entire family. They include 176 unique small molecule ligands.

Although allosteric binding sites represent an attractive approach to selectively target kinases, they are not nearly as well-characterized as the ATP binding site and their identification across the entire protein kinase family would be very challenging.

**Ligand-Cross Implementation, Automation, and Results.** Our goal was to generate new likely kinase inhibitors by systematically swapping fragments of known ligands across the structurally resolved Kinome. To identify the best compatible input structures we explored similarities of the corresponding binding sites (see Materials and Methods); this was independent from the ligand structures, i.e. no similarity of the corresponding kinase ligands was used. For each reference ATP site, we identified 50 most similar ones (including the

reference site itself) as input for the ligand cross algorithm. TIP computes all pairwise site alignments and stores them in the database. The 50 most similar ATP binding sites of the (filtered) PDB input records were overlaid in EVE relative to the reference site, thus superimposing the corresponding ligands. EVE identifies matching single bonds based on distance and angle constraints (see Materials and Methods). For each matching bond, the algorithm swaps the corresponding aligned fragments of the ligands and saves the resulting structures. Three iterations of such ligand fragment crossing were executed using products from the previous generation as input for the next and in each step producing novel ligands that fit into the overlaid sites (Figure 2). The newly created structures were named by combining names of the original ligands and numbers to enumerate all possible fragment crossings for each site pair. The entire process to generate new ligands from cocrystal ligands of similar ATP sites was implemented and automated in Pipeline Pilot using several Web Services, Pipeline Pilot internal components, and local installation of EVE as described in Materials and Methods. From all input records after three generations of ligand crossing of the 50 most similar sites for each reference structure, a total of 1 287 206 ligands were generated and combined into one SDF file for further curation and analysis.

**Correction and Filtering of Cross Product Molecular Structures.** Initial analysis showed several generated ligand structures with overlapping atoms. Such atom overlapping occurs when the swapped fragments of two overlaid ligands do not expand along (approximately) opposite directions of the matching bond vector, but the two substituents occupy the same space at one side of the aligned bond (Figure 3).

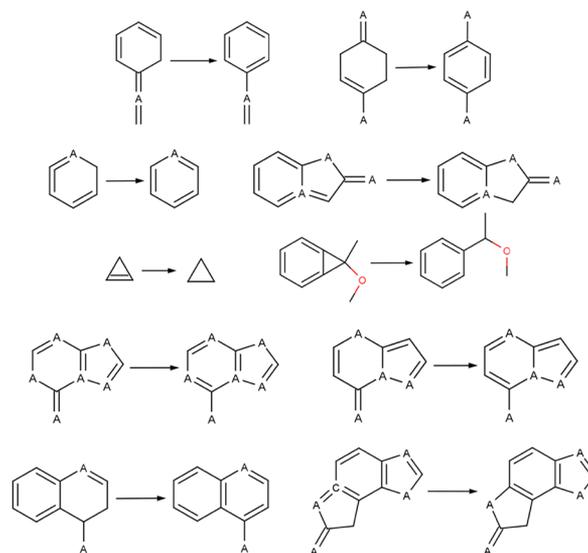


**Figure 3.** Overlapping atoms in compounds generated by swapping fragments of PDB ligands ANP and L20 (in green and blue, respectively) by the ligand-cross algorithm. Although the aligned bonds obey the matching criteria (atom distance 0.6 and 0.8 Å, and the angle between the bonds less than 15°) swapping from the circled fragments generates a new structure with overlapping moieties.

In order to eliminate such resulting structures, we used the “Bump Check Filter” component implemented in Pipeline Pilot with the van der Waals radius scale factor of 0.8. This factor is used to multiply the VDW radius for each heavy atom and allows for partial overlap, anticipating some degree of conformational relaxation of the new ligand will reduce the atom overlap within the ATP-binding site. After applying this filter, the total number of structures decreased to 845 148.

Further, we found that some of the newly generated ligands have an incorrect bond order. Although we already filtered the PDB input ligands to remove incorrectly perceived bonds, the same problem can persist with newly generated structures. Tolerance in the bond matching criteria (up to 1 Å atom

distance and 15° bond angle) to account for ligand flexibility in some cases resulted in slightly distorted geometry of the ligand-cross product structure (dihedral angles and lengths of the newly generated bond fall outside the typical ranges) and resulted in incorrectly assigned bond orders. However, we did not find this issue to be extensive. We observed most of the irregularities in ring systems, and as a consequence in many cases, the bound atoms have an incorrect valence. We identified 17 329 compounds (about 2% of all generated structures) with bad valences. In order to correct those structure, we assigned bond orders based on the 3D structure coordinates obtained from EVE using algorithm implemented in Pipeline Pilot. The remaining 275 structures that still had incorrect bond orders were corrected by applying 10 generic intramolecular transformations using ChemAxon Reactor (Figure 4).



**Figure 4.** Set of 10 intramolecular transformations to correct bond orders in ring systems (A represents any atom except hydrogen).

We also removed from further analysis 133 ruthenium complexes (originating from the ligands JM1 and HB1) leaving a total of 845 015 structures.

Following filtering and structure corrections, we standardized all structural representations by a series of chemical manipulator functions in Pipeline Pilot including standardization of the formal charges of functional groups, defining stereochemical configuration from the 3D coordinates, kekulization, ionization at physiological pH, and generating a canonical tautomer. Canonical SMILES were generated from the standardized structures and duplicates removed. This process resulted in 149 538 unique structures.

**Lead- and Drug-Likeness.** We applied OpenEye Filter to identify lead- and drug-like molecules. They resulted in 12 110 and 26 321 structures, respectively. The combined set amount to 26 673 unique compounds and is referred to as the lead-/drug-like subset. We provide their chemical structures as Supporting Information. It should be noted that only 14 and 16 of the original unique 176 PDB ligands correspond to lead- and drug-like, respectively, using the same criteria.

**Known Kinase Inhibitors.** Among the lead-/drug-like subset of generated structures, we identified 299 (unique) previously reported kinase inhibitors. Kinase inhibitory activity and the chemical structures were obtained from the Kinase

Knowledge Base (KKB, Eidogen-Sertanty, <http://www.eidogen-ertanty.com/kinasekb.php>). For several of these 299 kinase inhibitors, activities against multiple kinases were reported, resulting in 1869 experimental data points shown in the heat map in Figure S1 in the Supporting Information. Compounds with submicromolar  $IC_{50}$  (i.e., with the  $pIC_{50} \geq 6$ ) were considered active with the majority of the experimental data for CDK2. We identified 92 CDK2 inhibitors reported  $pIC_{50}$  of greater than 6.

#### Prioritization of New Ligands by Ensemble Docking.

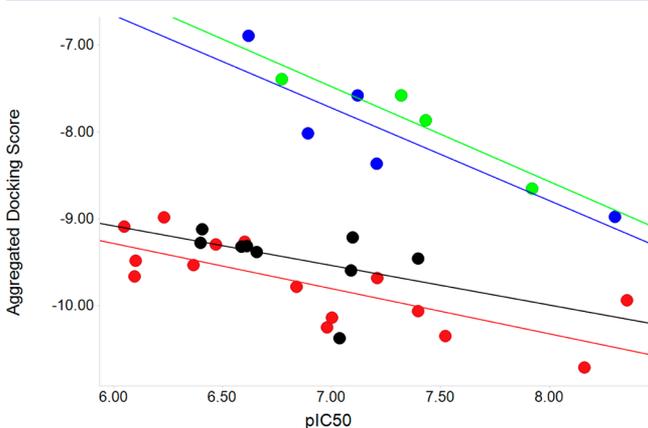
We performed ensemble docking of the 92 generated, but previously known CDK2 kinase inhibitors to demonstrate that virtual screening can be successfully applied to prioritize active compounds among the new compounds generated by the ligand cross protocol.

In order to select an ensemble of CDK2 crystal structures, we identified 175 CDK2 PDB ligands and explored their diversity. We selected crystal structures of five most diverse CDK2 ligands (diversity based on functional-class fingerprints). They include PDB entries 1e1x, 1ke9, 1p5e, 2c5x, and 2r3f. The 92 ligands were prepared and flexibly docked to each of the five CDK2 structures as described in Materials and Methods.

Ligand-specific scores for CDK2 were derived from the ensemble docking protocol. The best scoring pose of each specific combination of ligand representation for each CDK2 structure was selected. These scores were then averaged across the different ligand ionization and tautomerization states for each protein structure. The so-derived ligand-averaged docking scores were then congregated across the different CDK2 structures by selecting for each ligand the best (averaged) score from the five different CDK2 protein structures. The rationale for that was the known flexibility of kinase structures and ligand diversity.<sup>38</sup>

We found, in general, a good trend between the docking scores and the experimental  $pIC_{50}$  values within the structural clusters (series) of the ligands. We used maximal common substructure to cluster the CDK2 inhibitors. Figure 5 illustrates for several chemotypes the correlation of aggregate docking scores and experimental inhibitory activity. We did not include any series with less than 4 members into the analysis.

It should be emphasized that, rather than validating the docking algorithm to predict experimental values, our primary



**Figure 5.** Correlation of aggregate docking scores and reported kinase activity for four chemical series of CDK2 inhibitors. Data points are colored by cluster membership/chemotype. Clustering was performed based on the maximal common substructure.

aim was to demonstrate on the example of reported kinase inhibitors, how the automated ligand-cross assembly can be combined with other tools to identify and prioritize useful novel compounds.

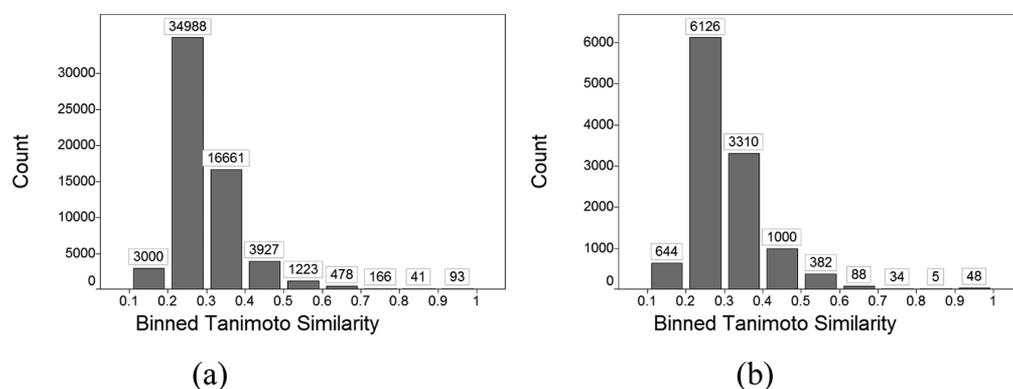
**Selectivity of Generated Ligands.** Our retrospective analysis suggests that the ligand-cross protocol can generate reasonably selective ligands and that selectivity is to some extent related to the kinases (on the kinase group level) used as input to the protocol. For example, the most potent CDK2 inhibitor among the previously known ligands ( $pIC_{50} = 8.36$ , the most right-side red data point in the Figure 5) shows selectivity toward members of different kinase groups and some selectivity toward members of the same group. Available data extracted from the KKB (Q2 2012) is summarized in Supporting Information Table S1. The compound is also active against CDK9 ( $IC_{50} = 1$  nM), but less active against several other members of the same kinase group, CMGC, including various CDKs, GSKs, and CSNK and inactive against CAMK2 and MAPKs. The CDK inhibitor is also selective against TK group kinases although it shows activity (decreasing for KDR, FLT, LCK, PDGFR, ABL). It is also less active against Aurora kinases (“other” group), and it is inactive for AGC kinases (AKT, PKA, PKC). Activity against TK kinases can be rationalized from the input ligands: three originate from CDK2 and one from FADK2 (TK group). Although not perfect, there is significant group selectivity of this CDK inhibitor. We observe higher activity toward more similar kinases and less activity against dissimilar ones. This is not surprising, but our results suggest that selectivity of ligand-cross products may be influenced to a certain extent by the input kinase structures/binding sites (with their corresponding ligands).

**Diversity Analysis of Generated Ligands.** In order to explore the novelty and diversity of the newly created structures, we performed a structure similarity analysis based on the Murcko assemblies of the input (original PDB ligands) and the structures generated by the ligand functionality shuffling protocol (ligand-cross products). We also generated Murcko assemblies for the lead-/drug-like subset and compared them to the PDB ligands by their Tanimoto similarity based on extended-connectivity fingerprints. The distribution of the maximum similarity of the newly generated ligand-cross products against the PDB ligand input structures based on their corresponding Murcko assemblies is shown in Figure 6a; Figure 6b illustrates the distribution for the lead-/drug-like subset.

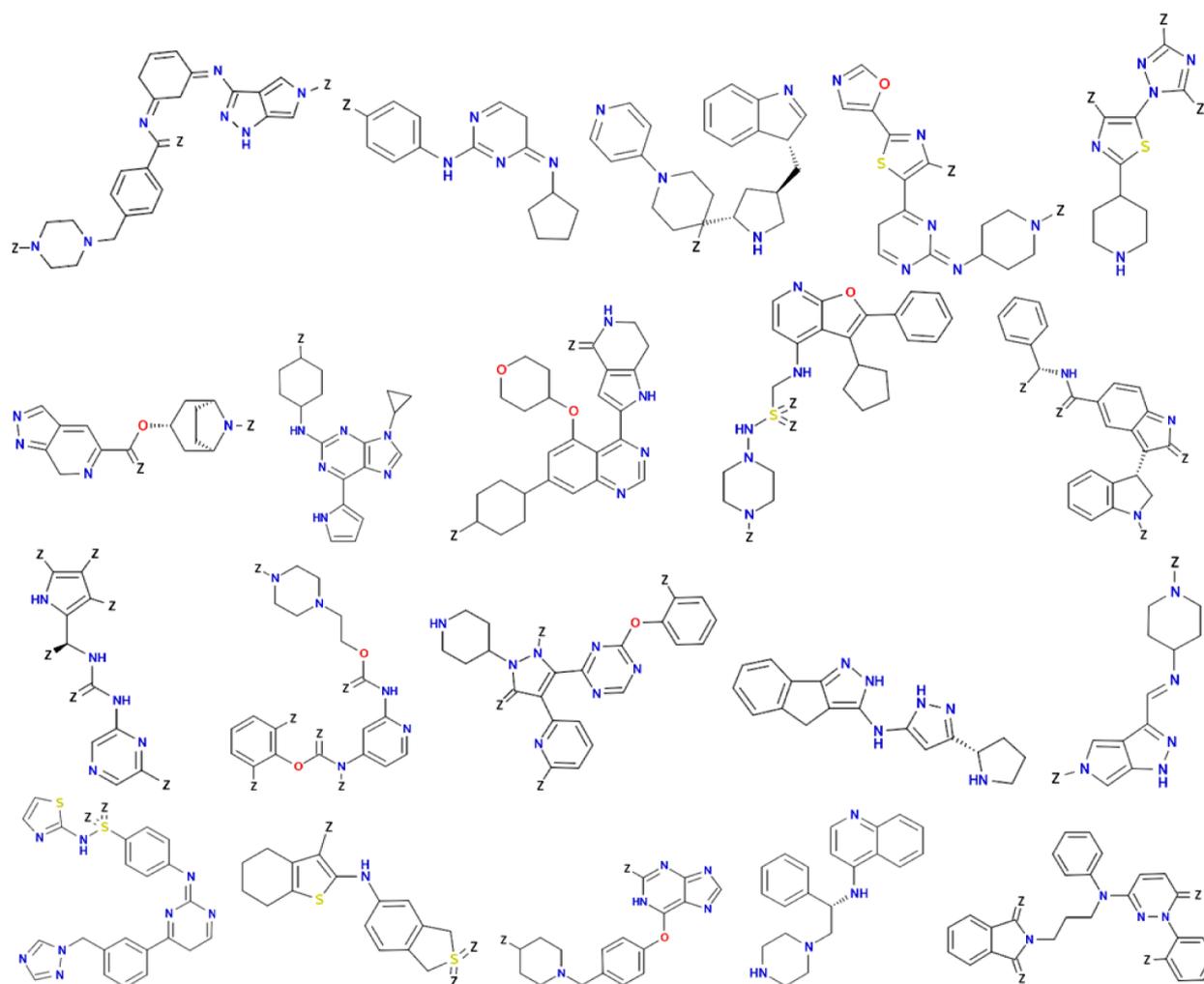
The low Murcko assembly base similarity of the majority of the ligand-cross products compared to their input structures (Figure 6) demonstrated that the generated compounds are novel. These results also suggest that novel kinase inhibitor scaffolds (Murcko assemblies) are generated in the ligand functionality shuffling protocol. We provide all generated Murcko assemblies as Supporting Information.

To illustrate diversity among generated structures among the lead- and drug-like subset, we first identified the 100 most diverse Murcko assemblies and from those selected the 20 most novel scaffolds based on their lowest closest similarity to the Murcko assemblies of the PDB input ligands. These are shown in Figure 7. The entire set of 100 most diverse Murcko fragments is provided in the Supporting Information (Figure S2).

In the same way we compared the known kinase inhibitors identified among the generated cross-products to the original



**Figure 6.** (a) Distribution of maximum Tanimoto similarity of the ligand-cross products against the PDB input ligands based on the extended-connectivity fingerprints of the corresponding Murcko assemblies: (a) all generated structures and (b) the lead-/drug-like subset of these compounds.



**Figure 7.** Twenty of the most novel among 100 most diverse Murcko assemblies corresponding to the lead- and drug-like subset of structures generated by the ligand functionality shuffling protocol; Z represents attachment points of side chains or atoms to the Murcko fragments also indicating the valence.

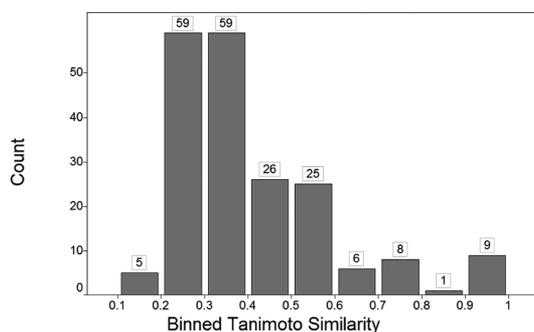
input PDB ligands. Murcko assemblies corresponding to the structures were compared by Tanimoto similarity based on extended-connectivity fingerprints. Figure 8 shows their maximum similarity distribution.

The results illustrated in Figure 8 demonstrate that novel (with respect to the input ligands) and diverse kinase inhibitors

can be obtained by the exhaustive ligand functionality shuffling protocol (in this case, three generations of ligand crossing).

## CONCLUSION

Identifying and swapping corresponding moieties from known small molecules of interest is a strategy frequently applied in



**Figure 8.** Distribution of maximum Tanimoto similarity of known kinase inhibitor ligand-cross products against the PDB input ligands based on the extended-connectivity fingerprints of the corresponding Murcko assemblies.

medicinal chemistry to generate novel and active lead compounds. We developed and implemented a protocol to systematize and automate this approach across the entire human Kinome. Our protocol integrates different technologies including binding site detection, alignment, and similarity computation implemented in TIP and the ligand-cross algorithm in EVE. Several postprocessing steps were performed to ensure high quality of the generated structures. From about 1000 input structures, roughly 150 000 unique compounds were generated. After applying strict lead-/drug-like filters, we were left with close to 27 000 structures. The vast majority of these compounds showed very low similarity to the input ligands and can be considered different scaffolds (Mucko assemblies). Among the structures generated by the ligand functionality shuffling protocol, we identified several known kinase inhibitors. Although not unexpected, it demonstrated the utility of the algorithm to generate potent kinase inhibitors. Because the protocol swaps corresponding ligand features based on overlaid protein ligand binding sites, it is highly likely that many more (unknown) active kinase inhibitors are among the ligand-cross products. Some of those are likely in the hands of pharmaceutical companies and can be compared based on the provided structures. We demonstrated that docking is an effective method to prioritize among the generated compounds to identify likely and potent inhibitors. This is not unexpected considering the method of swapping ligand features within their overlaid binding sites. Beyond the specific compounds, the generated scaffolds may be of particular interest to develop novel ATP competitive kinase inhibitors. We are making the generated structures and scaffolds available as Supporting Information.

As the body of available cocrystal structures steadily increases, an automated and efficient protocol to mine this information and generate novel starting points for lead optimization is likely to gain even more significance.

As our protocol overlays ligands by reference of their (local) binding sites and not their corresponding protein structures, it can be applied across different protein families, as long as a site similarity can be detected. In that context, the ATP binding site is of particular relevance, because it is present in many families beyond protein kinases.

Another potential application of the presented method is to strategically select protein kinases based on a desired poly pharmacological profile and align those sites to swap ligand features to achieve activity against a set of kinase targets. Such an approach would focus on only a few kinases with most

similar sites rather than a systematic ligand crossing applied in this work. The generated ligands could then be further prioritized using docking as we have demonstrated here for CDK2.

In summary, our in-silico workflow generates novel, drug-like matter within the structurally resolved Kinome by utilizing a combination of receptor-based similarity, ligand feature shuffling, and ligand-based filtering/refinement methods. The generated compounds and scaffolds are likely of interest to researchers developing novel kinase inhibitors.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Figure S1: heat map depicting 1869 experimental data points for 299 known kinase inhibitors (also generated by ligand-cross method) against multiple kinases. Figure S2: set of 100 most diverse Murcko assemblies derived from the newly generated structures. File ci3003842\_si\_003.txt contains SMILES for all newly generated drug- and lead-like structures. File ci3003842\_si\_004.txt contains SMILES for all Murcko assemblies of newly generated drug and lead-like structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Fax: (+1) 305-243-9304. E-mail: [sschurer@med.miami.edu](mailto:sschurer@med.miami.edu).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank OpenEye (Santa Fe, NM, USA) for providing licenses for their software. We also acknowledge resources by the Center for Computational Science at the University of Miami (Miami, FL, USA).

## ■ ABBREVIATIONS

ATP, adenosine triphosphate; CDK2, cyclin-dependent kinase 2; ECFP, extended-connectivity fingerprints; EVE, Eidogen visualization environment; KKB, kinase knowledge base; PDB, protein data bank; TIP, target informatics platform; VDW, van der Waals

## ■ REFERENCES

- (1) Cohen, P. Protein kinases-the major drug targets of the twenty-first century? *Nat. Rev. Drug. Discov.* **2002**, *1*, 309–315.
- (2) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912–1934.
- (3) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28–39.
- (4) Fedorov, O.; Muller, S.; Knapp, S. The (un)targeted cancer kinome. *Nat. Chem. Biol.* **2010**, *6*, 166–169.
- (5) Lenihan, D. J. Tyrosine kinase inhibitors: can promising new therapy associated with cardiac toxicity strengthen the concept of teamwork? *J. Clinical Oncol.* **2008**, *26*, 5154–5.
- (6) Joensuu, H.; Trent, J. C.; Reichardt, P. Practical management of tyrosine kinase inhibitor-associated side effects in GIST. *Cancer Treatment Reviews* **2011**, *37*, 75–88.
- (7) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. Through the “gatekeeper door”: exploiting the active kinase conformation. *J. Med. Chem.* **2010**, *53*, 2681–2694.
- (8) Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.

- (9) Anastassiadis, T.; Deacon, S. W.; Devarajan, K.; Ma, H.; Peterson, J. R. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011**, *29*, 1039–1045.
- (10) Davis, M. I.; Hunt, J. P.; Herrgard, S.; Cicceri, P.; Wodicka, L. M.; Pallares, G.; Hocker, M.; Treiber, D. K.; Zarrinkar, P. P. Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2011**, *29*, 1046–1051.
- (11) Goldstein, D. M.; Gray, N. S.; Zarrinkar, P. P. High-throughput kinase profiling as a platform for drug discovery. *Nat. Rev. Drug Discov.* **2008**, *7*, 391–397.
- (12) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Cicceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **2008**, *26*, 127–132.
- (13) Knight, Z. A.; Lin, H.; Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nat. Rev. Cancer* **2010**, *10*, 130–137.
- (14) Vidović, D.; Schürer, S. C. Knowledge-based characterization of similarity relationships in the human protein-tyrosine phosphatase family for rational inhibitor design. *J. Med. Chem.* **2009**, *52*, 6649–59.
- (15) Akritopoulou-Zanze, I.; Hajduk, P. J. Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors. *Drug Discov. Today* **2009**, *14*, 291–297.
- (16) Ghose, A. K.; Herberich, T.; Pippin, D. A.; Salvino, J. M.; Mallamo, J. P. Knowledge Based Prediction of Ligand Binding Modes and Rational Inhibitor Design for Kinase Drug Discovery. *J. Med. Chem.* **2008**, *51*, 5149–5171.
- (17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–42.
- (18) Leach, A. R.; Hann, M. M. Molecular complexity and fragment-based drug discovery: ten years on. *Current Opin. Chem. Biol.* **2011**, *15*, 489–96.
- (19) Hoffer, L.; Renaud, J.-P.; Horvath, D. Fragment-Based Drug Design: Computational and Experimental State of the Art. *Comb. Chem. High Throughput Screening* **2011**, *14*, 500–20.
- (20) Vieth, M.; Erickson, J.; Wang, J.; Webster, Y.; Mader, M.; Higgs, R.; Watson, I. Kinase inhibitor data modeling and de novo inhibitor design with fragment approaches. *J. Med. Chem.* **2009**, *52*, 6456–6466.
- (21) Pierce, A. C.; Rao, G.; Bemis, G. W. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. *J. Med. Chem.* **2004**, *47*, 2768–75.
- (22) *Target Informatics Platform (TIP)*; Eidogen-Sertanty, Inc. <http://www.eidogen-sertanty.com/tip.php>.
- (23) Altschup, S. F.; Gish, W.; Pennsylvania, T.; Park, U. Basic Local Alignment Search Tool 2. Department of Computer Science. *J. Mol. Biol.* **1990**, *215*, 403–410.
- (24) Aronov, A. M.; McClain, B.; Moody, C. S.; Murcko, M. A. Kinase-likeness and kinase-privileged fragments: toward virtual polypharmacology. *J. Med. Chem.* **2008**, *51*, 1214–1222.
- (25) Schmitt, S.; Kuhn, D.; Klebe, G. A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *J. Mol. Biol.* **2002**, *323*, 387–406.
- (26) *Eidogen Visualization Environment (EVE)*; Eidogen-Sertanty, Inc. <http://www.eidogen-sertanty.com/eve.php>.
- (27) *Accelrys Pipeline Pilot 8.0*; Accelrys, San Diego, CA; 2010.
- (28) *ChemAxon JChem Software Suite*; ChemAxon, Budapest, HU; 2011.
- (29) *Open Eye Filter*; OpenEye Scientific Software, Inc., Santa Fe, NM, 2011; <http://www.eyesopen.com/filter>.
- (30) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.
- (31) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
- (32) Flower, D. On the properties of bit string-based measures of chemical similarity. *J. Chem. Inf. Comp. Sci.* **1998**, *38*, 379–386.
- (33) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical similarity searching. *J. Chem. Inf. Comp. Sci.* **1998**, *38*, 983–996.
- (34) *Kinase Knowledge Base (Q4 2009)*; Eidogen-Sertanty Inc., Oceanside, CA, USA, 2009.
- (35) *LigPrep*; Schrodinger LLC, Portland, OR, USA, 2011.
- (36) *Glide*; Schrodinger LLC, Portland, OR, USA, 2011.
- (37) Ligand Expo 2011. <http://ligand-expo.rcsb.org/>.
- (38) Tuccinardi, T.; Botta, M.; Giordano, A.; Martinelli, A. Protein Kinases: Docking and Homology Modeling Reliability. *J. Chem. Inf. Model.* **2010**, *50*, 1432–1441.