Kinase-likeness and Kinase-Privileged Fragments: Toward Virtual Polypharmacology

Alex M. Aronov,* Brian McClain, Cameron Stuver Moody, and Mark A. Murcko

Vertex Pharmaceuticals, Inc., Cambridge, Massachusetts 02139-4242

Received August 16, 2007

Small molecule protein kinase inhibitors are widely employed as biological reagents and as leads in the design of drugs for a variety of diseases. We investigated the phenomenon of kinase-likeness, i.e., the propensity of ligands to inhibit protein kinases, in the context of kinase-specific substructural fragments. The frequency of occurrence of multiple structural fragments in kinase inhibitor libraries relative to nonkinase compounds has been analyzed. A combination of structural fragment counts, termed the “2-0” kinase-likeness rule, provides approximately 5-fold enrichment in kinase active compounds. This rule has been validated using in-house kinase counterscreening data and applied prospectively to uncover kinase activities in marketed drugs. In addition, the role of discriminating fragments in kinase recognition was interrogated using available structural data, providing an insight into their effect on inhibitor potency and selectivity. One of these fragments, bisarylaniline, has been characterized as a kinase-privileged fragment with specific binding preferences and a link to increased activity within kinases.

Introduction

Protein kinases have become the second most exploited group of drug targets after G-protein-coupled receptors (GPCRs), accounting for 20–30% of drug discovery projects at many pharmaceutical companies. Kinases play critical roles in cellular signaling networks, and many proteins in this class are established targets for pharmaceutical intervention. Small molecule kinase inhibitors have generated much interest, as both potential therapeutics and experimental tools for understanding the physiological roles of these enzymes. Various small molecule target-selective inhibitors of disease-relevant protein kinases are currently in different stages of clinical testing, and the first representatives of this class have already received Food and Drug Administration (FDA) approval.

As part of a larger paradigm shift in the industry from target “cherry picking” toward broader chemogenomics approaches to gene family drug discovery, pharmaceutical companies have started moving away from larger, general screening libraries of diverse compounds toward smaller, focused libraries often designed to target specific gene families. Some of these libraries have been assembled from the existing corporate compound collections; others have been synthesized through directed combinatorial synthesis; yet others have been purchased from commercial suppliers as prepackaged targeted databases. It is now generally accepted that a focused library approach to gene-family-centered drug discovery is of benefit to the outcome of library screening, in large part because of the overwhelming size of available unfocused chemical space.

The concept that particular molecules have a high propensity for binding to protein targets has been an active research area for nearly 20 years. According to the original definition by Evans, a “privileged substructure” is a “single molecular framework able to provide ligands for diverse receptors”. More recently, this term has taken on a broader meaning, coming to symbolize those substructures found to be promiscuous within a given target family. Privileged substructures have been identified for known drugs, protein binding, and GPCRs. The added implication that these “privileged structures” are thus specific for the given target family, is not always correct and requires a proper survey of both target family active and inactive compounds.

Structural requirements for the activity of a compound against a target family of interest can be derived from crystallographic information about the target gene family or from the analysis of known ligands, with an eye toward the presence in them of structural motifs that can be linked to increased activity for the target family. In addition, significant variation of molecular properties observed for drugs has been shown to be dependent upon the biological target class, lending itself to molecular property-based library focusing.

A number of studies have described library focusing in the context of the kinase gene family. Use of structural information in virtual ligand screening has led to enrichments of up to 17-fold but more typically in the range of 2–7-fold for crystal structures and homology models. While docking approaches are powerful when applied to specific biological targets, their utility in broader library focusing as directed toward gene families is less straightforward.

Machine learning has been applied to recognize compounds that act on kinases based on training that uses information about reported kinase inhibitors. Examples include neural networks trained using BCUT descriptors, a naïve Bayesian model using a combination of extended 2D topological fingerprints and basic molecular property descriptors, and a broad survey of machine-learning methods that used fragment-based Ghose–Crippen descriptors. Although enrichments were broadly similar in descriptor-based classification experiments relative to docking studies, machine learning approaches have demonstrated enrichment in compounds targeting specific kinases without the need to develop models for each individual kinase.

While the results of machine-based classification experiments have been encouraging, the approaches, such as neural nets, suffer from a number of drawbacks. The most significant of these is the “black box” character of the models, reflected in their inability to provide clearly interpretable rules that can be related directly to the chemical structure of classified com-
Similarly, the GVK BIO database of compounds from the drug-like molecular weight range between 250 and 600 a.e. kinase literature and patents, which were prefiltered to satisfy the by GVK BIO (Hyderabad, India) contained 18 817 compounds from
distributions in kinase compounds relative to general medicinal
example, the "0.001 µM" bin contains compounds that inhibit at
least one kinase target with subnanomolar $K_i$. The sets of kinase frequent hitters, and selective kinase inhibitors were used to probe the effect of bisarylanilines on kinase intrafamily selectivity. Additionally, a collection of 3647 in-house compounds synthesized between September 2003 and September 2004 and confirmed active ($K_i < 30 \mu M$) in at least one kinase assay was binned by the single highest activity level observed for each compound. For example, the "0.001 µM" bin contains compounds that inhibit at least one kinase target with subnanomolar $K_i$. The sets of kinase frequent hitters and selective kinase inhibitors were used to probe the effect of bisarylanilines on kinase intrafamily selectivity.

Because the vast majority of kinase inhibitors under investigation are ATP site inhibitors, special care was taken at the data set curation step to ensure the removal of functionalities most associated with other modes of kinase inhibition: peptidic structures are known to often act by binding to the peptide substrate groove; phosphate-

Table 1. Compound Databases Used in the Study

<table>
<thead>
<tr>
<th>compound setsa</th>
<th>number of compounds</th>
<th>number of frameworksb</th>
<th>similarity (standard deviation)c</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMC 5776</td>
<td>2893</td>
<td>0.22 (0.08)</td>
<td></td>
</tr>
<tr>
<td>JMC 238 858</td>
<td>51 521</td>
<td>NDd</td>
<td></td>
</tr>
<tr>
<td>kinase 18 817</td>
<td>5727</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>Vertex kinase</td>
<td>3647</td>
<td>0.33 (0.13)</td>
<td></td>
</tr>
<tr>
<td>PILLS 486</td>
<td>275</td>
<td>0.22 (0.09)</td>
<td></td>
</tr>
</tbody>
</table>

a A full listing of compound databases used in the study is provided in the Materials and Methods. b Molecular frameworks were derived as described by Bemis and Murcko. c Mean and standard deviation for internal pairwise Tanimoto similarity using Daylight fingerprints. d ND = not determined.

Results

Molecular Frameworks of Kinase Inhibitors. We have previously reported an analysis of kinase inhibitor frameworks for a set of published kinase inhibitors. Briefly, we used the topological decomposition rules from Bemis and Murcko to define the structures of kinase inhibitors in the context of their framework and side chain atoms. A drug framework is obtained by removing side chains from the original drug molecule. Each framework can be further partitioned into rings and linkers that comprise it. Rings are cycles within the graph representation of molecules and cycles sharing an edge. Linkers are defined as units consisting of atoms that are on the direct path connecting two ring systems. The analysis performed on 119 published kinase inhibitors revealed that structural diversity at the level of rings and linkers was unexpectedly low. Four rings and eight linkers were found to describe nearly 90% of all ring and linker occurrences in the fragmented data set.

In an effort to capture information about a broader set of kinase inhibitor building blocks, we applied the same deconvolution procedure to the set of 18 817 compounds from kinase literature and patents (the kinase set). As a control, we used the set of 238 858 compounds from Journal of Medicinal Chemistry publications (the JMC set) and the CMC database containing 5776 diverse compounds (the CMC set). Kinase inhibitors are underrepresented in both the JMC and CMC sets, because the explosion in kinase research activity dates back to mid or late 1990s. Hence, both the JMC and CMC collections can serve as a useful background comparison that reflects general topological tendencies in medicinal chemistry, more (CMC) or less (JMC) impacted by the considerations of metabolism, bioavailability, and likelihood of toxicity as pertains to specific moieties and fragments.

The top 10 rings most frequently encountered in kinase inhibitors are shown in Figure 1A. In agreement with previous reports for drug-like and kinase compounds, phenyl is the most widely encountered ring system, with similarly high representation in both kinase and general collections. An average kinase inhibitor contains 1.03 phenyls, while that number is somewhat lower for the JMC set (0.89 phenyls per structure) and the CMC set (0.67 phenyls per structure), an observation that can be likely attributed to the generally higher aromatic character of kinase compounds. A significantly larger discrepancy is observed for the occurrence of nitrogen-containing heteroaromatics, which includes both mono- and bicyclic ring systems. Heterobicyclics were previously proposed by Muegge and Enyedy11 as a structural driver for increased kinase activity, and our data confirms that these ring systems occur with higher frequency in the kinase set. For instance, kinase inhibitors contain an average of 0.13 quinazoline motifs per molecule, compared to 0.01 per molecule in the JMC set and none in the CMC set. Similarly, indolinone and pyrrolopyrimidine are present in the kinase set (0.09 and 0.06 per structure, respectively) and absent in the general compound collections. In fact, even monocyclic nitrogen-containing heteroaromatics, such as pyridine, pyrimidine, and pyrazole, are more often encountered in kinases: pyridine has an “abundance” of 0.21 per compound in kinase inhibitors and a mere 0.05 per compound in general collections.

A truly stark contrast between kinase and general compounds becomes apparent from the analysis of top 10 linkers found in kinase inhibitors (Figure 1B). The –NH– linker is by far the

and phosphate-containing compounds frequently bind to SH2 and SH3 domains.
most common linker found in kinase inhibitors, with 0.33 occurrences per ligand. At the same time, it occurs with a much lower frequency in nonkinase compounds, with 0.04 and 0.01 occurrences per molecule for JMC and CMC sets, respectively. In contrast, the –CONH– is the most common linker in medicinal chemistry, while it is only seventh on the list of most common kinase linkers. In general, it appears that kinase inhibitors are enriched in NH-containing linkers, with –CONH– (amide), –NH–, –NHCONH– (urea), and –CH₂NH– all being found with higher frequency in the kinase collection.

**Discriminating Kinase-Inhibitor Fragments.** The next step was to try and distill the framework fragmentation results (Figure 1) into a set of clearly interpretable rules for incorporation into the expert system. Because ring enrichments pointed to nitrogen-containing heteroaromatics as preferred kinase motifs, we analyzed the frequency of occurrence of heteroaromatic nitrogens (Naro), (ii) it contains one or more heteroaromatic NH groups (Naro); (iii) it contains one or more anilines (Ar\[NH\]); (iv) it contains one or more heteroaromatic nitrogens, accounting for 37% of the kinase set. Fewer kinase inhibitors contained the molecules with two heteroaromatic nitrogens, 37% for nonkinase compounds to 10% for kinase inhibitors. As shown in Figure 2A, more kinase than nonkinase compounds contain heteroaromatic nitrogens (Naro). Whereas 77 and 65% of CMC and JMC compounds, respectively, contain no heteroaromatic nitrogens, 77% of kinase inhibitors contained at least one and 59% of kinase inhibitors contained at least two such atoms. The largest single bin of kinase inhibitors contained the molecules with two heteroaromatic nitrogens, accounting for 37% of the kinase set. Fewer compounds contained heteroaromatic NH (NHaro); however, the compounds that did (Figure 2B) were more likely to be kinase inhibitors (25%) than general compounds from CMC (8%) or JMC (11%).

### Table 2. Mean Frequency of Occurrence of Discriminating Structural Fragments

<table>
<thead>
<tr>
<th>structural fragment</th>
<th>CMC</th>
<th>JMC</th>
<th>kinase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naro</td>
<td>0.53</td>
<td>0.84</td>
<td>1.88</td>
</tr>
<tr>
<td>NHaro</td>
<td>0.09</td>
<td>0.12</td>
<td>0.27</td>
</tr>
<tr>
<td>Ar[NH]</td>
<td>0.13</td>
<td>0.26</td>
<td>1.15</td>
</tr>
<tr>
<td>nitrile</td>
<td>0.02</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>Ar[NH]Ar[NH]</td>
<td>0.01</td>
<td>0.03</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Indeed, nitriles can be found in kinase inhibitors approximately 4–5-fold more frequently compared to nonkinase compounds (Table 2). Because of the ubiquity of the –NH– linker in kinase ligands (Figure 1B), we analyzed its relative rate of appearance as both aniline (Ar–NH) and bisarylaniline (Ar₁–NH–Ar₂). The results echoed our earlier dramatic observation. While anilines in general are 5–10-fold more likely to be found in kinase inhibitors, bisarylanilines occur 10-fold more frequently in the kinase set than in the JMC set and a stunning 31-fold more often than in the CMC set!

While the average frequency of occurrence of individual fragments in kinase inhibitors is informative, the more meaningful criterion is the extent to which these motifs are distributed among the compounds in the kinase collection relative to the nonkinase compounds. As shown in Figure 2A, more kinase than nonkinase compounds contain heteroaromatic nitrogens (Naro). Whereas 77 and 65% of CMC and JMC compounds, respectively, contain no heteroaromatic nitrogens, 77% of kinase inhibitors contained at least one and 59% of kinase inhibitors contained at least two such atoms. The largest single bin of kinase inhibitors contained the molecules with two heteroaromatic nitrogens, accounting for 37% of the kinase set. Fewer compounds contained heteroaromatic NH (NHaro); however, the compounds that did (Figure 2B) were more likely to be kinase inhibitors (25%) than general compounds from CMC (8%) or JMC (11%).

The **“2-0” Rule of Kinase-likeness.** By following the observations of structural fragment distribution across data sets, we propose the following general “rule of thumb” for discriminating kinase compounds from general compounds that do not possess kinase activity. A compound is likely to have kinase activity if (i) it contains two or more heteroaromatic nitrogens (Naro), (ii) it contains one or more heteroaromatic NH groups (NHaro), (iii) it contains one or more anilines (Ar–NH), and (iv) it contains one or more nitriles (R–CN). Alternatively, this “rule of thumb” can be formulated in numeric form, which we termed the “2-0” rule of kinase-likeness. A compound is kinase-like if either of the following requirements is fulfilled:

\[
\sum (N_{aro}) + \sum (NH_{aro}) > 2 \quad \text{or} \quad \sum (Ar-NH) + \sum (R-CN) > 0
\]

(1)

Indeed, this rule is able to discriminate the kinase set from the general collections (Figure 3). Nearly four of every five kinase compounds (78%) pass the “2-0” rule, while the number of
passing compounds drops to 16% for the CMC and 22% for the JMC compound collections. It is tempting to speculate that the difference between the results for the CMC and JMC sets reflects the higher probability of encountering kinase compounds in the JMC database, which included compounds up to 2003, relative to the CMC set.

As retrospective validation of the kinase-likeness rule, we applied it to a set of data from 10 randomly selected kinase HTS campaigns at Vertex (Figure 4). For the purposes of this paper, we chose not to disclose the identities of individual kinases, and they were labeled as kinases 1–10. The number of confirmed actives in these campaigns varied greatly, from 26 compounds for kinase 8 to 877 compounds for kinase 1. We analyzed the extent to which these confirmed actives abide by the “2-0” kinase-likeness rule. The percentage of actives passing the kinase-likeness criteria ranges from a high of 98% for kinase 4 to a low of 70% for kinase 6. On average, 89% of HTS actives show agreement with the kinase-likeness criteria.

The relative success of the HTS campaign in terms of the number of uncovered and confirmed hits does not seem to correlate with kinase-likeness of the hit set. Of the five lowest yielding HTS campaigns in terms of confirmed actives (kinases 3 and 5–8), kinase-likeness is low (<80%) for kinases 5 and 6 and 88–96% for kinases 3, 7, and 8.

Application of Kinase-likeness to Virtual Screening. We chose to illustrate the kinase-likeness rule by uncovering hidden kinase activities of marketed oral drugs. Of the 486 entries in the PILLS24 database, 240 were available for testing. Of the 240 compounds, 37 (8%) satisfied the kinase-likeness criteria, while 203 compounds did not pass. All of the compounds were then screened for activity against three kinases: SGK1, p70 S6K, and PAK1. Testing was performed at or near 30 µM ligand concentration in triplicate. Kinase activity for the purposes of this study was defined as an observation of 25% or greater inhibition of the target kinase that can be supported by either a titration dose response or dynamic structure–activity relationship (SAR) for a series of closely related compounds. A total of 3 of the 37 kinase-like drugs exhibited kinase activity (Figure 5). Triamterene, a diuretic that targets Na–K–Cl cotransporter 2, inhibited greater than 90% of activity for all three kinases.
Thioguanine, an antineoplastic thought to act by inhibiting HGPRT and IMPDH, showed weak activity in the PAK1 assay (31% inhibition at 31 µM), while anti-inflammatory drug sulfasalazine, known to block acetyl-CoA acetyltransferase, was active against SGK1 (70% inhibition at 33 µM). Indeed, upon assaying analogues of thioguanine, several purines were confirmed as micromolar inhibitors of PAK1 (Table S1 in the Supporting Information). Of the 204 nonkinase-like drugs, the sole activity was uncovered in the case of antineoplastic drug methoxsalen (64% inhibition of p70 S6K), indicating an overall 0.5% hit rate for this group of compounds. Overall, the application of a simple “2-0” rule resulted in a 5-fold enrichment in kinase-active compounds. This is generally comparable to previously reported results for classification (typically 5−10-fold enrichment, 79−95% accuracy) and docking to kinase homology models (1.2−7-fold enrichment), as well as our in-house model built using a Random Forest-based classifier and a set of topological torsions descriptors.

Some of the structures of newly uncovered kinase inhibitors have precedent in the literature. For instance, triamterene shares the diaminopteridine core with 1 (TG100-115, Figure 6), a phosphoinositide 3-kinase (PI3K) inhibitor currently undergoing clinical trials. It is thought to form three hydrogen bonds to the hinge region of PI3K. Thioguanine also has a potential to form three hydrogen bonds to the kinase hinge from C2NH2, N3, and N9H of adenine, similar to the previously reported structures of kinase-bound purines. Finally, flavonoids and wortmannin are examples of oxygen-rich aromatics lacking nitrogen reminiscent of methoxsalen.

Figure 3. Performance of the “2-0” kinase-likeness rule on external compound collections. Percent of compounds surviving the filter is shown for general and kinase-specific compound libraries.

Figure 4. Retrospective performance of the “2-0” kinase-likeness rule on confirmed actives (Ki < 30 µM) in Vertex kinase HTS campaigns. Results of 10 randomly selected HTS screens (4 tyrosine and 6 serine/threonine kinases) are shown. Red line indicates the number of confirmed actives in each screen.

The diaminopteridine core with 1 (TG100-115, Figure 6), a phosphoinositide 3-kinase (PI3K) inhibitor currently undergoing clinical trials. It is thought to form three hydrogen bonds to the hinge region of PI3K. A similar hydrogen-bonding pattern can be expected in the case of triamterene. Thioguanine also has a potential to form three hydrogen bonds to the kinase hinge from C2NH2, N3, and N9H of adenine, similar to the previously reported structures of kinase-bound purines. Finally, flavonoids and wortmannin are examples of oxygen-rich aromatics lacking nitrogen reminiscent of methoxsalen.

Role of Bisarylanilines in Kinase-Inhibitor Binding. As outlined in Table 2, bisarylanilines are found 10-fold less frequently in the JMC set and 31-fold less often in the CMC than they occur in the kinase set. Further analysis of bisarylanilines revealed that >95% of occurrences contains a heteroaromatic nitrogen ortho to the −NH linker. We examined the distribution pattern for this structural motif among the compounds in the kinase collection relative to the nonkinase
compounds. A quarter of all kinase compounds contain this motif (Figure 2E), including 2% of ligands that incorporate two bisarylanilines within the same molecule. Further, bisarylanilines as defined in Figure 2E appear to be a selective marker of kinase inhibition and not found in nonkinase compound sets (<1%).

To investigate the possible link between the presence of a bisarylaniline in a molecule and the ability of the molecule to potently inhibit kinases, we selected 3647 in-house compounds synthesized within a single year, between September 2003 and September 2004, which were confirmed active ($K_i > 30 \mu M$) in at least one kinase assay. The set was binned by the single highest activity level observed for each compound. For example, the “0.001 μM” bin contains compounds that inhibit at least one kinase target with subnanomolar $K_i$. The “>10 μM” bin contains molecules for which every measured kinase $K_i$ exceeded the 10 μM threshold. The average compound was tested in between 10 and 16 kinase assays (62% of the library). Additionally, the kinase frequent hitter and selective kinase inhibitor data sets described earlier by Aronov and Murcko are profiled. The number of compounds passing the kinase-likeness filter increases moderately with increasing ligand potency (empty bars). The number of compounds containing anilinic bisaryl−NH− linkers appears more sensitive to the ligand potency level (shaded bars). Promiscuous kinase inhibitors are more likely to contain bisarylanilines than selective ligands. The number of compounds in each bin is indicated (red).

As expected, the kinase-likeess rule showed a link to inhibitor potency (Figure 7). Only 51% of compounds in the 10−30 μM range were kinase-like as defined in eq 1; however, the degree of kinase-likeness is higher (87%) for compounds with at least one sub-10 μM activity and exceeds 90% for the more potent kinase inhibitors. Importantly, the bisarylaniline moiety appeared as a more sensitive marker of inhibitor potency. Whereas 77% of subnanomolar inhibitors contain the bisarylaniline signature, only 39% of sub-10 μM inhibitors do. Furthermore, the bisarylaniline was not found in weak (>10 μM) ligands.

Two additional data sets described previously were evaluated for the presence of kinase-like features and the bisarylaniline signature motif. The kinase frequent hitter set contained 43 ligands characterized as potent and promiscuous inhibitors of multiple kinase targets, and the corresponding “selective set” covered 209 compounds with one potent activity as defined using a set of five common kinases. Both of the sets were greater than 99% kinase-like, which likely stems from their respective potency on kinase targets, but the bisarylaniline content was 2-fold higher in the frequent hitters (Figure 7).

When the results from Figure 7 are taken together, they confirm a link between kinase-likeness and inhibitor potency and indicate that the bisarylaniline is an important kinase signature motif whose presence corresponds to higher affinity of a compound for a kinase target in a rather nonspecific manner, leading to broader inhibition of targets across the kinase family, the phenomenon previously termed “kinase frequent hitters”.

To assess the role of bisarylanilines from the structural perspective, we selected 164 in-house and public kinase-bound ligand structures that contain this motif. The 164 structures represented 139 unique ligands; they were aligned in a common frame of reference using heavy atoms of the kinase hinge region as described previously. The dihedral angle for the bisarylanilines as a measure of the coplanarity of the two linked aryl rings was found to vary from 0° to 90°, with a median of 23° and a mean of 31° (Figure 8A). Additionally, CNC fragments that represented bisarylaniline linkers were extracted from all 164 structures and converted to 18 clusters. 

When the results from Figure 7 are taken together, they confirm a link between kinase-likeness and inhibitor potency and indicate that the bisarylaniline is an important kinase signature motif whose presence corresponds to higher affinity of a compound for a kinase target in a rather nonspecific manner, leading to broader inhibition of targets across the kinase family, the phenomenon previously termed “kinase frequent hitters”. To assess the role of bisarylanilines from the structural perspective, we selected 164 in-house and public kinase-bound ligand structures that contain this motif. The 164 structures represented 139 unique ligands; they were aligned in a common frame of reference using heavy atoms of the kinase hinge region as described previously. The dihedral angle for the bisarylanilines as a measure of the coplanarity of the two linked aryl rings was found to vary from 0° to 90°, with a median of 23° and a mean of 31° (Figure 8A). Additionally, CNC fragments that represented bisarylaniline linkers were extracted from all 164 structures and converted to 18 clusters. 

To assess the role of bisarylanilines from the structural perspective, we selected 164 in-house and public kinase-bound ligand structures that contain this motif. The 164 structures represented 139 unique ligands; they were aligned in a common frame of reference using heavy atoms of the kinase hinge region as described previously. The dihedral angle for the bisarylanilines as a measure of the coplanarity of the two linked aryl rings was found to vary from 0° to 90°, with a median of 23° and a mean of 31° (Figure 8A). Additionally, CNC fragments that represented bisarylaniline linkers were extracted from all 164 structures and converted to 18 clusters. 

To assess the role of bisarylanilines from the structural perspective, we selected 164 in-house and public kinase-bound ligand structures that contain this motif. The 164 structures represented 139 unique ligands; they were aligned in a common frame of reference using heavy atoms of the kinase hinge region as described previously. The dihedral angle for the bisarylanilines as a measure of the coplanarity of the two linked aryl rings was found to vary from 0° to 90°, with a median of 23° and a mean of 31° (Figure 8A). Additionally, CNC fragments that represented bisarylaniline linkers were extracted from all 164 structures and converted to 18 clusters.
above. The aniline in the Scios TGFβ compound\(^3\) belongs to cluster C. Finally, the Vertex promiscuous kinase inhibitor disclosed previously\(^2\) contains two bisarylanilines that belong to clusters A and C.

**Role of Other Discriminating Fragments.** In a manner similar to bisarylanilines, we analyzed the role of other functional groups involved in the “2-0” rule in ligand binding to kinases. The 3D clustering of nitriles (Figure 9A) revealed a fan-like pattern. A higher proportion of nitriles are concentrated in proximity to the catalytic lysine. Formation of the $-\text{C}_\equiv\text{N} \cdots \text{H}_3\text{N}^+$ hydrogen bond satisfies the acceptor feature in the five-point frequent hitter pharmacophore.\(^2\) Of note is the absence in Figure 9A of nitriles fulfilling the hinge acceptor role; however, compounds of this type have been reported.\(^3\) Aromatic nitrogen acceptors (Figure 9B) can be found near and between the two acceptor features of the frequent hitter pharmacophore.\(^2\) Not surprisingly, the aromatic nitrogen donors are preferentially clustered around the two hinge donor features, and their primary role appears to be anchoring the scaffold to the kinase hinge.

**Discussion**

We investigated the phenomenon of kinase-likeness, i.e., the propensity of ligands to inhibit protein kinases, in the context

---

**Table 3. Kinase Inhibitors on the Market and in Phase-III Testing**

<table>
<thead>
<tr>
<th>name</th>
<th>“2-0” kinase-likeness features present?</th>
<th>bisarylaniline present?</th>
</tr>
</thead>
<tbody>
<tr>
<td>imatinib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>sorafenib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>gefitinib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>erlotinib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>sunitinib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>dasatinib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>lapatinib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>vandetanib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>cediranib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>vatalanib</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>pazopanib</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

---

**Figure 8.** Geometry of kinase recognition for bisaryl $-\text{NH}-$ linker fragments. (A) Distribution plot of observed dihedral angles for the bisarylaniline motif extracted from 164 (36 public and 128 proprietary) bisarylaniline-containing kinase inhibitor structures in the Vertex structural database. The median dihedral angle equals 23°. (B) Binding location of bisarylanilines in kinases in the context of the five-point kinase frequent hitter pharmacophore.\(^2\) CNC fragments (bisarylaniline linkers) from 164 ligands were converted to 18 clusters (1 Å rmsd), and cluster centroids were visualized in the context of the kinase active site. CNC fragments are colored as follows: carbons (yellow) and nitrogens (blue). The frequent hitter pharmacophore is shown as colored spheres: hydrogen bond acceptors (red), hydrogen bond donors (blue), and aromatic (green). (C) Examples of bisarylanilines in kinase inhibitors. Experimentally determined hinge hydrogen bond pivot is shown. The bisaryl $-\text{NH}-$ linker is shown in red.

**Figure 9.** Kinase recognition geometry in case of nitriles (A), heteroaromatic nitrogen acceptors Naro (B), and heteroaromatic nitrogen donors NHaro (C).
of kinase-specific substructural fragments. The frequency of occurrence of multiple structural fragments in kinase inhibitor libraries relative to nonkinase compounds has been analyzed. A combination of structural fragment counts, termed the "2-0" kinase-likeness rule, has been shown to accurately describe anywhere from approximately 80% to nearly 100% of kinase inhibitors. The performance of this rule is data set-dependent and is most applicable to potent kinase inhibitors. The rule has been internally validated using historic HTS data and has provided approximately 5-fold enrichment in kinase-active compounds in the prospective virtual screening experiment aimed at uncovering kinase activities in marketed drugs.

In addition, the aniline linker has been uncovered as a selective marker for kinase-active compounds, with a dramatically higher prevalence in kinase-targeted small molecules relative to other protein families. The presence of bisarylanilines appears to be a highly specific, albeit less general, signature of a propensity for inhibition across a wide variety of kinase family members, related to the likelihood of the compound to exhibit kinase frequent hitter characteristics. The frequency of occurrence of the bisarylaniline motif in kinase inhibitors contrasted with its relative scarcity in other compound classes, and the demonstrated link of this marker to increased activity within kinases points to bisarylanilines as a kinase-privileged fragment. Importantly, we have been able to link this kinase-privileged signature to its 3D structural role in facilitating ligand binding to kinases. Examination of the binding mode for bisarylaniline-containing ligands has uncovered three major binding locations, whereby the aniline serves as a linking point for aromatic ligand fragments that occupy any of the four main hydrophobic areas in the ATP site.

Ability to define an entire class of compounds as a combination of substructural fragments begs the following question: how much of it is truly intrinsic to kinases, and how much is due to conservative “me too” approaches in kinase chemistry? The components of the "2-0" rule, namely, heteroaromatic nitrogenos, heteroaromatic NH groups, anilines, and nitriles, are likely to be sufficiently elemental in nature to be largely reflective of the kinase affinity for lipophilic heteroaromatics anchored in the ATP site by a set of hydrogen bonds. The demonstrated link of bisarylaniline to higher potency and kinase promiscuity (Figure 7) coupled with a structural preference for specific regions of the ATP site (Figure 8B) argues for its kinase-privileged status. Whether or not the proliferation of bisarylanilines in kinase compounds has come because of the associated potency boost, is related to the ease of amine substitution chemistry on ubiquitous heteroaromatic rings, or stems from the lack of creativity in the kinase field remains an open question.

Better understanding of the kinase-likeness phenomenon has a number of potential applications in drug discovery. The kinase-likeness expert system can be used as a computationally inexpensive filter in the course of virtual screening, thus minimizing the number of false positives because of shortcomings in scoring functions. We have used it internally to prioritize compounds for screening against new kinase targets, as well as a way to rapidly identify kinase-like compounds in commercial libraries and catalogs. Another option is to remove kinase-like compounds upfront and prosecute a HTS campaign on nonkinase-like compounds to find the most novel motifs, albeit at the expense of the hit rate.

Our results suggest that there is potential value in subjecting drugs, drug candidates, or advanced leads to kinase counter-screening, which could be triggered when a ligand is character-ized as kinase-like. The observation of cross-reactivity for known drugs, whereby a chemical entity is able to bind to more than one target, is not entirely uncommon. Indeed, the “target-hopping” strategy, where chemical matter for one target can be considered as an attractive basis for the design of active agents against another target, has been an extremely fruitful approach in drug discovery. Kinase inhibitors have been previously described as exhibiting some of the greatest interfamily promiscuity, and our data further illustrates this. While a broader characterization of kinase activities of drugs is beyond the scope of this study, such an undertaking in the context of polypharmacology network elucidation could both increase our understanding of mechanisms of action for existing drugs and improve the chances of finding new chemical entities devoid of undesirable side effects.

Supporting Information Available: List of purine-containing PAK1 inhibitors similar to thioguanine. This material is available free of charge via the Internet at http://pubs.acs.org.

References


(23) Comprehensive Medicinal Chemistry (CMC-3D); MDL Information Systems, Inc.: San Leandro, CA, 1996.


(35) Chakravarty, S.; Dugar, S. Atk5 (TGFβ R1) inhibitors: In vitro and in vivo characterization of novel orally active pteridines. *Inhibitors of Protein Kinases; Polish Academy of Sciences (PAS): Warsaw, Poland, 2005.*


(40) Chakravarty, S.; Dugar, S. Atk5 (TGFβ R1) inhibitors: In vitro and in vivo characterization of novel orally active pteridines. *Inhibitors of Protein Kinases; Polish Academy of Sciences (PAS): Warsaw, Poland, 2005.*
