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[1] have become the second most exploited group of drug targets after G-protein-coupled receptors (GPCRs),[a] accounting for 20–30% of drug discovery projects at many pharmaceutical companies.[2] 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.[3,4] 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.[5]

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,[6] 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,[7] protein binding,[8] and GPCRs.[4,9] 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.[10]

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[3,11,12] or from the analysis of known ligands,[5] 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,[13,14] 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.[11,15] 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,[16] a naïve Bayesian model using a combination of extended 2D topological fingerprints and basic molecular property descriptors,[17] and a broad survey[18] of machine-learning methods that used fragment-based Ghose–Crippen[19] 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.[20] The most significant of these is the "black box" character of the models, reflected in their inability to provide clearly interpretable rules than can be related directly to the chemical structure of classified com-

* To whom correspondence should be addressed: Telephone: (617) 444-6804. E-mail: alex_aronov@vrtx.com.

[a] Abbreviations: ATP, adenosine triphosphate; GPCR, G-protein-coupled receptor; PCH, polar-charge hydrophobe; PKA, protein kinase A; rmsd, root-mean-square deviation.

10.1021/jm701021b CCC: $40.75
Published on Web 02/21/2008

Table 1. Compound Databases Used in the Study

compound sets[a]	number of compounds	number of frameworks[b]	similarity (standard deviation)[c]
CMC	5776	2893	0.22 (0.08)
JMC	238 858	51 521	ND[d]
kinase	18 817	5727	ND
Vertex kinase	3647	1300	0.33 (0.13)
PILLS	486	275	0.22 (0.09)

[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.[7] [c] Mean and standard deviation for internal pairwise Tanimoto similarity using Daylight fingerprints.[42] [d] ND = not determined.

pounds. We decided to develop an expert system that would be able to complement the successful machine learning approaches. This expert system would contain a clearly interpretable set of rules, which could be applied to compound selection for prioritized high-throughput screening (HTS) or external compound acquisition. The advantage of the expert system is its relative ease of use and interpretability at the level of individual chemical structures. Expert systems have been previously developed for the prediction of drug toxicity,[21,22] metabolism,[22] and drug-likeness.[20]

To this end, we investigated the phenomenon of kinase-likeness, i.e., the propensity of ligands to inhibit protein kinases, in the context of the presence or absence of kinase-specific substructural fragments. By analyzing the frequency of occurrence of multiple structural fragments in kinase inhibitor libraries relative to nonkinase compounds, we propose a rule for rapid identification of kinase-like structures. 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 in the context of the protein kinase family.

Materials and Methods

Data Sets. The data sets used for assessment of ring and linker distributions in kinase compounds relative to general medicinal chemistry libraries are listed in Table 1. The kinase set provided by GVK BIO (Hyderabad, India) contained 18 817 compounds from kinase literature and patents, which were prefiltered to satisfy the drug-like molecular weight range between 250 and 600 a.e. Similarly, the GVK BIO database of compounds from the *Journal of Medicinal Chemistry* publications between 1959 and 2003 encompassed 238 858 compounds prefiltered by molecular weight as mentioned above. The CMC[23] database contained 5776 diverse compounds, and the PILLS[24] data set contained 486 marketed oral drugs. All of the compound sets were sufficiently diverse, as exemplified by the large number of molecular frameworks and low pairwise Tanimoto similarity (Table 1).

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[25] and selective kinase inhibitors[25] 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- and phosphonate-containing compounds frequently bind to SH2 and SH3 domains.

Results

Molecular Frameworks of Kinase Inhibitors. We have previously reported an analysis of kinase inhibitor frameworks for a set of published kinase inhibitors.[12] Briefly, we used the topological decomposition rules from Bemis and Murcko[7] 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.[12]

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.[26] 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[7] and kinase[12] 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 Enyedy[11] 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

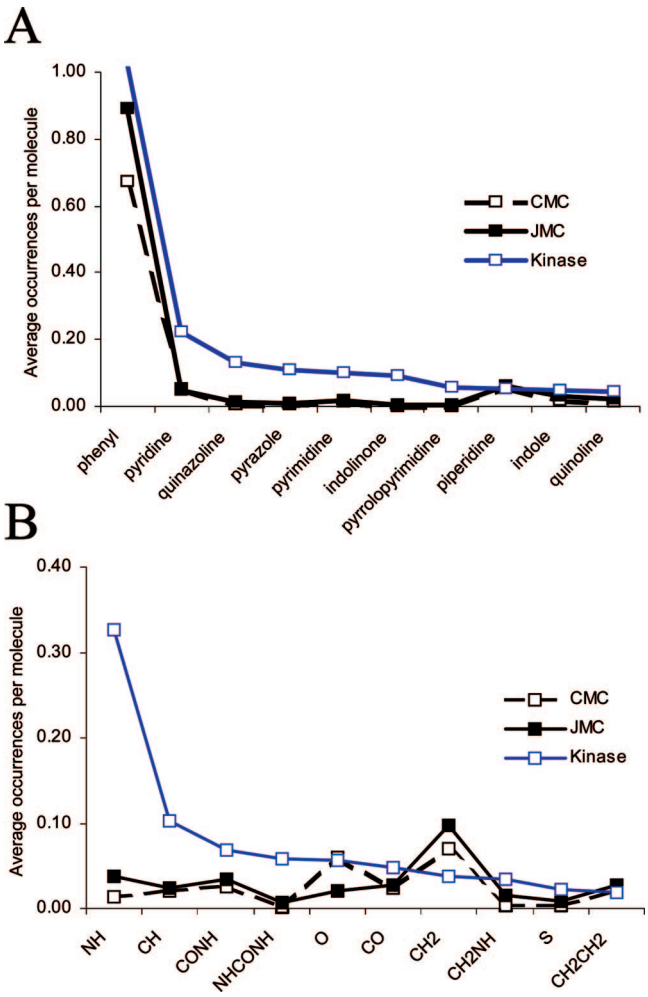


Figure 1. Frequences of occurrence for typical ring/linker building blocks in general and kinase-specific compound libraries. (A) Top 10 rings most frequently encountered in kinase inhibitors. The kinase library is particularly enriched with respect to nitrogen-containing heteroaromatics. (B) Top 10 linkers most frequently encountered in kinase inhibitors. The kinase library is highly enriched with NH-containing linkers, especially −NH− (>10-fold).

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 $-CH_2-$ 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), −NHCONH− (urea), and $-CH_2NH-$ 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 (N_{aro}, e.g., in pyridines or quinazolines) and heteroaromatic NH (NH_{aro}, e.g., in pyrroles or pyrazoles) in kinase and general compounds sets. In accordance with earlier observations, both motifs were found with approximately 2−3-fold higher frequency in the kinase set (Table 2). Evaluation of kinase inhibitor side chains pointed to nitriles as a functional group that makes a frequent appearance in kinase inhibitors (data not shown).

Table 2. Mean Frequency of Occurrence of Discriminating Structural Fragments

structural fragment[a]	CMC	JMC	kinase
N_{aro}	0.53	0.84	1.88
NH_{aro}	0.09	0.12	0.27
Ar[NH]	0.13	0.26	1.15
nitrile	0.02	0.03	0.11
$Ar_1[NH]Ar_2$	0.01	0.03	0.31

[a] Structural fragments are as described in the Results: N_{aro} is an heteroaromatic nitrogen (e.g., in pyridine); NH_{aro} is a heteroaromatic NH group (e.g., in pyrrole); Ar[NH] is an aryl-linked NH group; and $Ar_1[NH]Ar_2$ is a NH-linked bisaryl.

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 ($Ar_1-NH-Ar_2$). 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 (N_{aro}). 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 (NH_{aro}); 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 gap between kinase and nonkinase compounds increases for anilines, as seen previously (Table 2). Anilines of any kind could only be found in 10% of CMC and 19% of JMC compounds, whereas nearly two-thirds (64%) of the kinase compounds contain at least one and 19% employ two or more aniline motifs (Figure 2C). Relatively few compounds contain nitriles (Figure 2D); however, this number increases from 2 to 3% for nonkinase compounds to 10% for kinase inhibitors.

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 (N_{aro}), (ii) it contains one or more heteroaromatic NH groups (NH_{aro}), (iii) it contains one or more anilines (Ar−NH), and (iv) it contains one or more nitriles (R−C≡N).

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:

$$\sum(N_{aro}) + \sum(NH_{aro}) > 2 \quad \text{or} \quad \sum(\text{Ar-NH}) + \sum(\text{R-CN}) > 0 \tag{1}$$

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

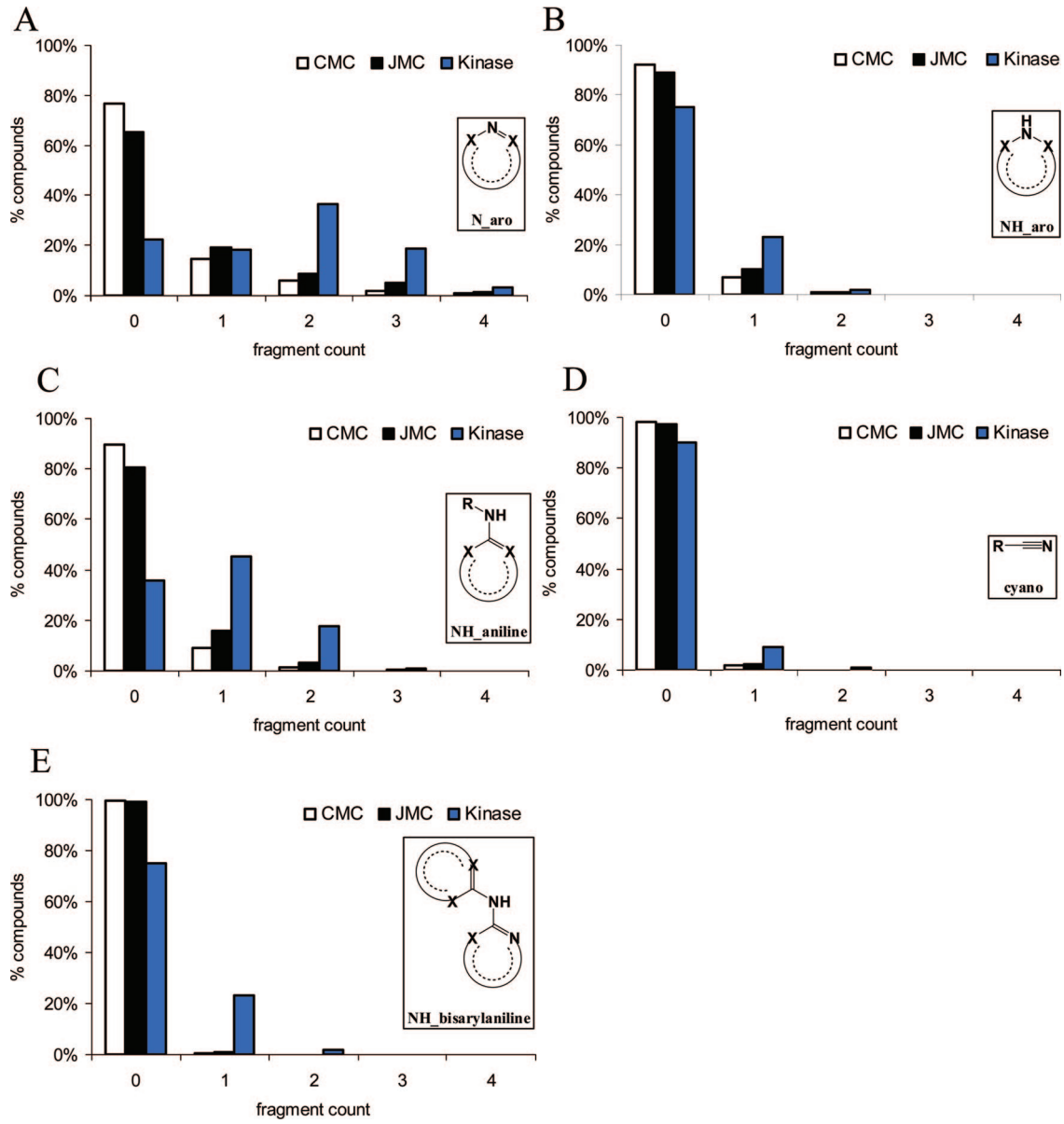
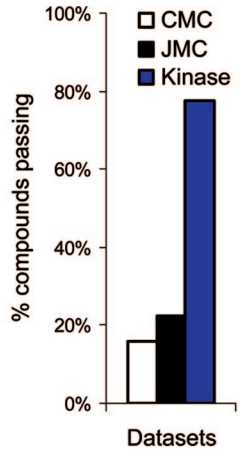


Figure 2. Distribution of select structural fragments in general and kinase-specific compound libraries. The fragment count reflects the number of occurrences per molecule of a select structural fragment.

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-likess 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-likess 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 PILLS[24] 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 cotrasporter 2,[27] inhibited greater than 90% of activity for all three kinases.



$$\Sigma(\text{N\_aro}) + \Sigma(\text{NH\_aro}) > 2$$
OR
$$\Sigma(\text{NH\_aniline}) + \Sigma(\text{CN}) > 0$$

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.

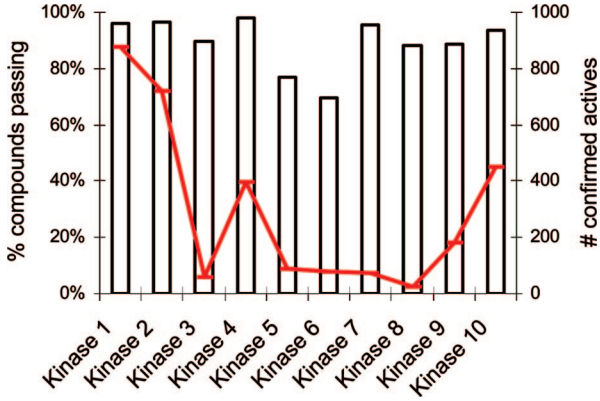


Figure 4. Retrospective performance of the "2-0" kinase-likeness rule on confirmed actives (K_I < 30 μM) in Vertex kinase HTS campaigns. Results of 10 randomly selected HTS screens (4 tyrosine and 6 serine/threonine kinases) are shown. Red line indicates the number of confirmed actives in each screen.

Thioguanine, an antineoplastic thought to act by inhibiting HGPRT and IMPDH,[27] showed weak activity in the PAK1 assay (31% inhibition at 31 μM), while anti-inflammatory drug sulfasalazine, known to block acetyl-CoA acetyltransferase,[27] 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[16–18] (typically 5−10-fold enrichment, 79–95% accuracy) and docking to kinase homology models[15] (1.2−7-fold enrichment), as well as our in-house model[28] built using a Random Forest-based[29] classifier and a set of topological torsions[30] descriptors.

Some of the structures of newly uncovered kinase inhibitors have precedent in the literature. For instance, triamterene shares

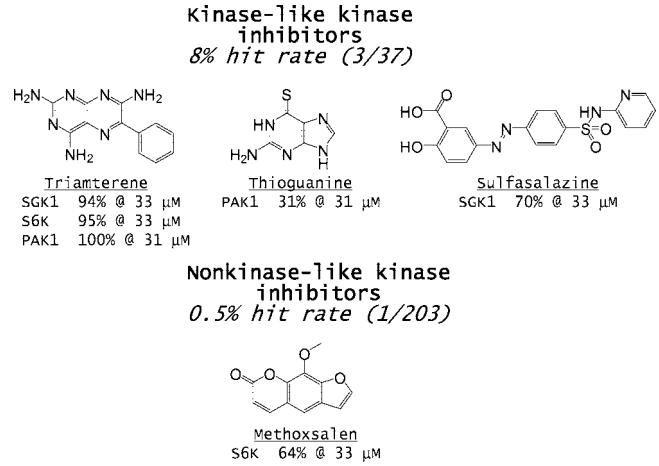


Figure 5. Application of the "2-0" kinase-likeness rule to virtual screening. Part of the PILLS[24] data set of 486 marketed oral drugs, 240 compounds were profiled against three kinases: SGK1, p70 S6K, and PAK1, expressed as the percentage of inhibition in triplicate at 30 μM. A total of 3 of 37 kinase-like drugs (8%) displayed activity (defined as greater than 30% inhibition at 30 μM), compared to 1 of 203 (0.5%) nonkinase-like oral drugs.

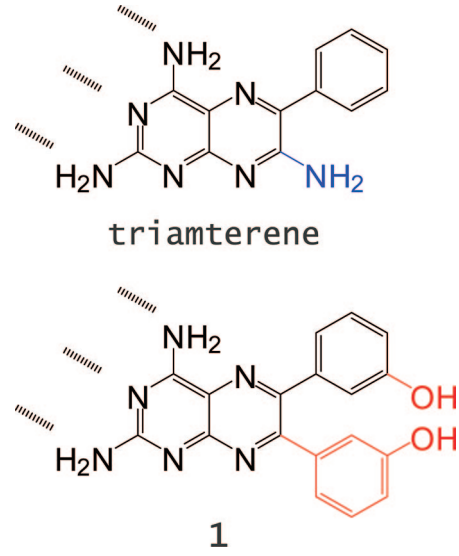


Figure 6. Structural similarity between triamterene and **1**. Likely hydrogen bonding complementarity to the kinase hinge region is shown based on the model by Doukas et al.[32]

the diaminopteridine core with **1** (TG100-115, Figure 6), a phosphoinositide 3-kinase (PI3K) inhibitor currently undergoing clinical trials.[31] It is thought to form three hydrogen bonds to the hinge region of PI3K.[32] 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 C^2NH_2, N^3, and N^9H of adenine, similar to the previously reported structures of kinase-bound purines.[33,34] 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

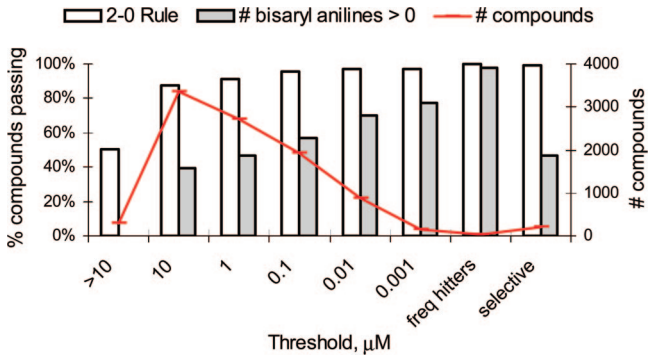


Figure 7. Performance of the "2-0" kinase-likeness rule on the Vertex kinase inhibitor library. A collection of 3647 in-house compounds synthesized and confirmed active (K_I > 30 μM) between September 2003 and September 2004 in at least one kinase assay was binned by the single highest activity level 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 "> 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[25] 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).

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 μM) in at least one kinase assay. The set was binned by the single highest activity level observed for each compound; e.g., the "0.001 μM" bin contained compounds that inhibit at least one kinase target with subnanomolar K_I. The "> 10 μM" bin contained compounds, for which their most potent activity fell in the 10−30 μM range.

As expected, the kinase-likess 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[25] 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".[25]

To assess the role of bisarylanilines from the structural perspective, we selected 164 in-house and public[35] 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.[12] 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 [1 Å root mean square deviation (rmsd)]. The nitrogen atom of CNC represents the −NH− linker, and the two carbons are the respective attachment points on the aromatic rings. Visualization in the context of the kinase active site (Figure 8B) indicates that bisarylanilines tend to coalesce in three regions of the ATP site. The largest concentration of biarylanilines (cluster A) is found within hydrogen bonding distance from the kinase hinge. The anilinic −NH− matches the hydrogen bond donor feature from the kinase frequent hitter pharmacophore (Figure 8B). One of the aryls can then occupy the adenine binding site, and the other uses hydrophobic contacts with the variable portion of the site that has been referred to at various times as either the specificity surface[36] or the secondary hydrophobic patch.[3] The second notable cluster of bisarylanilines is situated in the "north" portion of the ATP site, in proximity to the gatekeeper residue (cluster B). Aniline linkers enable the connection between the adenine binding site and the hydrophobic selectivity pocket that exists in kinases with smaller gatekeeper side chains, e.g., Abl, EGFR, or p38 MAP kinase. Alternatively, anilines can link the aromatic core to the fragments that occupy the phosphate-binding region of the site around the catalytic lysine side chain. In yet another role, aniline linkers provide a turn that ties together the phosphate-binding region with the variable hydrophobic patch (cluster C). Ultimately, it appears that bisarylanilines serve as important connector elements that enable the kinase ligands to span the main hydrophobic binding subsites and pick up additional interactions within the ATP cavity. The utility of bisarylanilines is enhanced by the hydrogen bonding ability of the aniline, an important factor in kinases that rely to a significant extent on hydrogen bonding for ligand recognition. It is also amplified by their relatively flat 3D characteristics that make the bisarylanilines compatible with the constraints of the active site.

Many of the kinase inhibitors undergoing clinical trials,[31] as well as most of the launched kinase inhibitor drugs, contain bisarylanilines. An investigation of chemical structures for advanced (Phase III clinical testing) or marketed kinase inhibitors (Table 3) reveals that 8 of 11 structures contain the bisarylaniline motif, with the 3 exceptions being sunitinib, sorafenib, and cediranib. Bisarylanilines found in kinase literature are known to bind in different locations of the ATP site. A number of examples included in Figure 8C illustrate this observation. Anilines in gefitinib, erlotinib, imatinib, and the Johnson & Johnson KDR inhibitor[37] fall in cluster B as defined

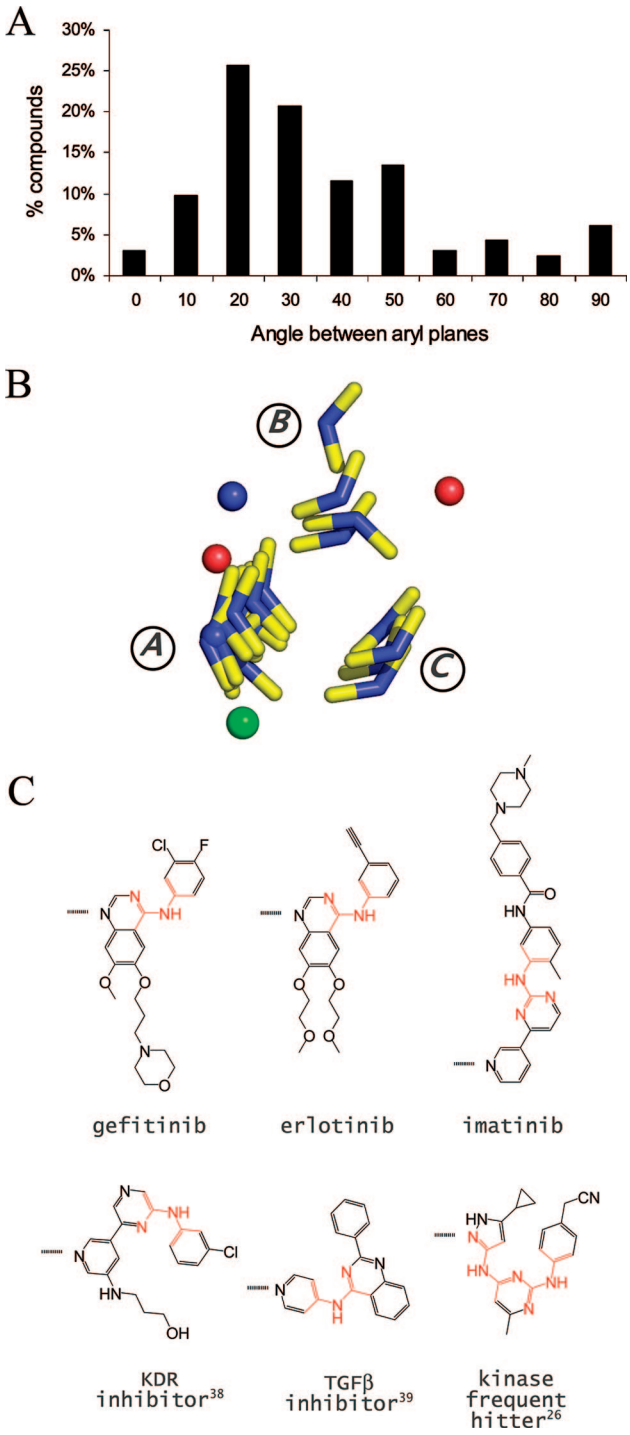


Figure 8. Geometry of kinase recognition for bisaryl −NH− linker fragments. (A) Distribution plot of observed dihedral angles for the bisarylaniline motif extracted from 164 (36 public and 128 proprietary) bisarylaniline-containing kinase inhibitor structures in the Vertex structural database. The median dihedral angle equals 23°. (B) Binding location of bisarylanilines in kinases in the context of the five-point kinase frequent hitter pharmacophore.[25] 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 −NH− linker is shown in red.

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

name	"2-0" kinase-likeness features present?	bisarylaniline present?
imatinib	×	×
sorafenib		
gefitinib	×	×
erlotinib	×	×
sunitinib		
dasatinib	×	×
lapatinib	×	×
vandetanib	×	×
cediranib	×	
vatalanib	×	×
pazopanib	×	×

above. The aniline in the Scios TGFβ compound[38] belongs to cluster C. Finally, the Vertex promiscuous kinase inhibitor disclosed previously[25] 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 $-C{\equiv}N\cdots H_3N^+-$ hydrogen bond satisfies the acceptor feature in the five-point frequent hitter pharmacophore.[25] Of note is the absence in Figure 9A of nitriles fulfilling the hinge acceptor role; however, compounds of this type have been reported.[39] Aromatic nitrogen acceptors (Figure 9B) can be found near and between the two acceptor features of the frequent hitter pharmacophore.[25] 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

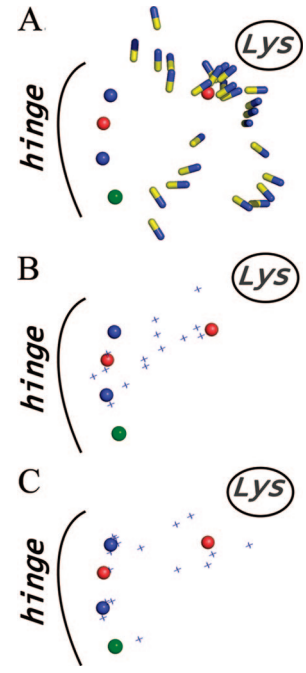


Figure 9. Kinase recognition geometry in case of nitriles (A), heteroaromatic nitrogen acceptors N_{aro} (B), and heteroaromatic nitrogen donors NH_{aro} (C).

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 nitrogens, 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.[5] 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 characterized 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.[14] 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.[14,40,41] Kinase inhibitors have been previously described as exhibiting some of the greatest interfamily promiscuity,[14] 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

(1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912–1934.

(2) Cohen, P. Protein kinases—The major drug targets of the twenty-first century. *Nat. Rev. Drug Discovery* **2002**, *1*, 309–315.

(3) Lowrie, J. F.; Delisle, R. K.; Hobbs, D. W.; Diller, D. J. The different strategies for designing GPCR and kinase targeted libraries. *Comb. Chem. High Throughput Screening* **2004**, *7*, 495–510.

(4) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**, *3*, 928–944.

(5) Brunne, R. M.; Hessler, G.; Muegge, I. Virtual compound libraries and molecular modeling. *Handbook of Combinatorial Chemistry. Drugs, Catalysts, Materials*; Wiley-VCH: Weinheim, Germany, 2002; pp 761–783.

(6) Evans, B. E.; Rittle, K. E.; Bock, M. G.; DiPardo, R. M.; Freidinger, R. M.; Whitter, W. L.; Lundell, G. F.; Veber, D. F.; Anderson, P. S.; Chang, R. S. Methods for drug discovery: Development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **1988**, *31*, 2235–2246.

(7) Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893.

(8) Hajduk, P. J.; Bures, M.; Praestgaard, J.; Fesik, S. W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* **2000**, *43*, 3443–3447.

(9) Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: Overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251–3264.

(10) Schnur, D. M.; Hermsmeier, M. A.; Tebben, A. J. Are target-family-privileged substructures truly privileged? *J. Med. Chem.* **2006**, *49*, 2000–2009.

(11) Muegge, I.; Enyedy, I. J. Virtual screening for kinase targets. *Curr. Med. Chem.* **2004**, *11*, 693–707.

(12) Aronov, A. M.; Bemis, G. W. A minimalist approach to fragment-based ligand design using common rings and linkers: Application to kinase inhibitors. *Proteins* **2004**, *57*, 36–50.

(13) Vieth, M.; Sutherland, J. J. Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* **2006**, *49*, 3451–3453.

(14) Paolini, G. V.; Shapland, R. H.; van Hoorn, W. P.; Mason, J. S.; Hopkins, A. L. Global mapping of pharmacological space. *Nat. Biotechnol.* **2006**, *24*, 805–815.

(15) Oshiro, C.; Bradley, E. K.; Eksterowicz, J.; Evensen, E.; Lamb, M. L.; Lanctot, J. K.; Putta, S.; Stanton, R.; Grootenhuis, P. D. Performance of 3D-database molecular docking studies into homology models. *J. Med. Chem.* **2004**, *47*, 764–767.

(16) Manallack, D. T.; Pitt, W. R.; Gancia, E.; Montana, J. G.; Livingstone, D. J.; Ford, M. G.; Whitley, D. C. Selecting screening candidates for kinase and G protein-coupled receptor targets using neural networks. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1256–1262.

(17) Xia, X.; Maliski, E. G.; Gallant, P.; Rogers, D. Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem.* **2004**, *47*, 4463–4470.

(18) Briem, H.; Gunther, J. Classifying "kinase inhibitor-likeness" by using machine-learning methods. *ChemBioChem* **2005**, *6*, 558–566.

(19) Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure−activity relationships. I. Partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565–577.
(20) Muegge, I.; Heald, S. L.; Brittelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* **2001**, *44*, 1841–1846.
(21) Dearden, J. C. In silico prediction of drug toxicity. *J Comput.-Aided Mol. Des.* **2003**, *17*, 119–127.
(22) Greene, N.; Judson, P. N.; Langowski, J. J.; Marchant, C. A. Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. *SAR QSAR Environ. Res.* **1999**, *10*, 299–314.
(23) *Comprehensive Medicinal Chemistry (CMC-3D)*; MDL Information Systems, Inc.: San Leandro, CA, 1996.
(24) Egan, W. J.; Walters, W. P.; Murcko, M. A. Guiding molecules towards drug-likeness. *Curr. Opin. Drug Discovery Dev.* **2002**, *5*, 540–549.
(25) Aronov, A. M.; Murcko, M. A. Toward a pharmacophore for kinase frequent hitters. *J. Med. Chem.* **2004**, *47*, 5616–5619.
(26) Cohen, P. The development and therapeutic potential of protein kinase inhibitors. *Curr. Opin. Chem. Biol.* **1999**, *3*, 459–465.
(27) Drugbank. http://redpoll.pharmacy.ualberta.ca/drugbank/.
(28) Feng, J. Vertex Pharmaceuticals, Inc., 2007, unpublished results.
(29) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958.
(30) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological torsions: A new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82–85.
(31) Investigational Drugs Database. Thomson Scientific Ltd., Stamford, CT, 2006.
(32) Doukas, J.; Wrasidlo, W.; Noronha, G.; Dneprovskaia, E.; Fine, R.; Weis, S.; Hood, J.; Demaria, A.; Soll, R.; Cheresh, D. Phosphoinositide 3-kinase γ/δ inhibition limits infarct size after myocardial ischemia/reperfusion injury. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19866–19871.
(33) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.* **2000**, *43*, 2797–2804.
(34) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E.; Tucker, J. A.; Wang, L.; Whitfield, H. J. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745–749.
(35) 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*, 235–242.
(36) McInnes, C.; Fischer, P. M. Strategies for the design of potent and selective kinase inhibitors. *Curr. Pharm. Des.* **2005**, *11*, 1845–1863.
(37) Kuo, G. H.; Wang, A.; Emanuel, S.; Deangelis, A.; Zhang, R.; Connolly, P. J.; Murray, W. V.; Gruninger, R. H.; Sechler, J.; Fuentes-Pesquera, A.; Johnson, D.; Middleton, S. A.; Jolliffe, L.; Chen, X. Synthesis and discovery of pyrazine-pyridine biheteroaryl as a novel series of potent vascular endothelial growth factor receptor-2 inhibitors. *J. Med. Chem.* **2005**, *48*, 1886–1900.
(38) Chakravarty, S.; Dugar, S. Alk5 (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.
(39) Angell, R. M.; Atkinson, F. L.; Brown, M. J.; Chuang, T. T.; Christopher, J. A.; Cichy-Knight, M.; Dunn, A. K.; Hightower, K. E.; Malkakorpi, S.; Musgrave, J. R.; Neu, M.; Rowland, P.; Shea, R. L.; Smith, J. L.; Somers, D. O.; Thomas, S. A.; Thompson, G.; Wang, R. *N*-(3-Cyano-4,5,6,7-tetrahydro-1-benzothien-2-yl)amides as potent, selective, inhibitors of JNK2 and JNK3. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1296–1301.
(40) Wermuth, C. G. Selective optimization of side activities: Another way for drug discovery. *J. Med. Chem.* **2004**, *47*, 1303–1314.
(41) Siegel, M. G.; Vieth, M. Drugs in other drugs: A new look at drugs as fragments. *Drug Discovery Today* **2007**, *12*, 71–79.
(42) Daylight Chemical Information Systems. http://www.daylight.com.

JM701021B