54 J. Med. Chem. 2011, 54, 54–66
DOI: 10.1021/jm101195a

Journal of Medicinal Chemistry Article

Trends in Kinase Selectivity: Insights for Target Class-Focused Library Screening

Shana L. Posy,[†] Mark A. Hermsmeier,[‡] Wayne Vaccaro,[§] Karl-Heinz Ott,[∥] Gordon Todderud,[⊥] Jonathan S. Lippy,[⊥] George L. Trainor,[§] Deborah A. Loughney,[†] and Stephen R. Johnson*,[†]

[†]*Computer-Assisted Drug Design, Applied Biotechnology, Bristol-Myers Squibb Research and Development, P.O. Box 4000, Princeton, New Jersey 08543, United States,* [‡]*Chemistry Informatics, Bristol-Myers Squibb Research and Development, P.O. Box 4000, Princeton, New Jersey 08543, United States,* [§]*Discovery Chemistry, Bristol-Myers Squibb Research and Development, P.O. Box 4000, Princeton, New Jersey 08543, United States,* [∥]*Bioinformatics, Applied Biotechnology, Bristol-Myers Squibb Research and Development, P.O. Box 4000, Princeton, New Jersey 08543, United States, and* [⊥]*Lead Evaluation, Applied Biotechnology, Bristol-Myers Squibb Research and Development, P.O. Box 4000, Princeton, New Jersey 08543, United States*

Received March 26, 2010

A kinome-wide selectivity screen of > 20000 compounds with a rich representation of many structural classes has been completed. Analysis of the selectivity patterns for each class shows that a broad spectrum of structural scaffolds can achieve specificity for many kinase families. Kinase selectivity and potency are inversely correlated, a trend that is also found in a large set of kinase functional data. Although selective and nonselective compounds are mostly similar in their physicochemical characteristics, we identify specific features that are present more frequently in compounds that bind to many kinases. Our results support a scaffold-oriented approach for building compound collections to screen kinase targets.

Introduction

The human genome contains more than 500 protein kinases, which transfer the γ-phosphate of ATP to the hydroxyl group of substrate tyrosine, serine, or threonine residues.[1] Exogenous kinase inhibitors compete with ATP to occupy the ATP binding site, a cleft between the two lobes of the kinase catalytic domain that is highly conserved across the kinome.[1,2] Kinase inhibitors have been developed as therapeutic agents that selectively target one or more kinases, and marketed kinase drugs exhibit a wide range of selectivity. More selective inhibitors include gefitinib[3,4] and fasudil. Examples of less selective inhibitors are sorafenib, which targets stem cell factor receptor (KIT),[a] platelet-derived growth factor receptor (PDGFR), B-RAF, and vascular endothelial growth factor receptor-2 (VEGFR2);[5] and dasatinib, which inhibits Src family kinases, ABL1, KIT, PDGFR-β, focal adhesion kinase (FAK), and EphA2.[6−9]

Lead compounds in a kinase drug discovery project are screened against multiple kinases to monitor their selectivity for the target (or targets) of interest. Selectivity screening typically involves close homologues of the intended target, which have ATP binding sites highly similar to those of the target and thus are likely to bind similar compounds. Recent large-scale kinase selectivity screens have confirmed that selective inhibitors tend to bind kinases that are grouped together on the kinome phylogenetic tree, and less commonly to distantly related kinases.[10−12] While these studies have characterized the targets of select inhibitors, they have not systematically explored the molecular characteristics that influence whether compounds are selective or nonselective in kinome-wide screens. Moreover, the previous screening sets maximized structural diversity and therefore included few representatives of each class of chemical structure. Fully characterizing the selectivity patterns of specific structural classes would require many examples of each class.

Here, we report the results of a nearly kinome-wide selectivity screen of > 20000 compounds with a rich representation of many structural classes. We describe the selectivity patterns

*To whom correspondence should be addressed. E-mail: stephen.johnson@bms.com. Telephone: (609) 252-3003. Fax: (609) 252-6030.

[a] Abbreviations: AAK1, AP2-associated kinase 1; Abl, Abelson murine leukemia viral oncogene; ABL1, Abelson murine leukemia viral oncogene homologue 1; Ack, activated Cdc42-associated kinase; Alk, anaplastic lymphoma kinase; ARG, Abelson murine leukemia viral oncogene homologue 2; ASK2, apoptosis signal-regulating kinase 2; Axl, AXL receptor tyrosine kinase; BLK, B lymphoid tyrosine kinase; B-RAF, v-raf murine sarcoma viral oncogene homologue B1; CDF, cumulative distribution function; CDK2, cyclin-dependent kinase 2; CK1D/CSNK1D, casein kinase I isoform δ; CK1E/CSNK1E, kinase I isoform ε; Csk, c-src tyrosine kinase; CTK, Csk-type protein tyrosine kinase; DDR, discoidin domain receptor tyrosine kinase; EGFR, epidermal growth factor receptor; Eph, ephrin receptor; FAK, focal adhesion kinase; Fer, (fps/fes-related) tyrosine kinase; FGFR, fibroblast growth factor receptor; FGR, Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homologue; FLT3, fms-related tyrosine kinase 3; GPCR, G-protein-coupled receptor; GSK, glycogen synthase kinase; HCK, hemopoietic cell kinase; HTS, high-throughput screen; IKKβ, inhibitor of the κ light polypeptide gene enhancer in B cells, kinase β; InsR, insulin receptor; JakA, Janus kinase; Jnk3, c-Jun N-terminal kinase 3; KIT, Hardy-Zuckerman 4 feline sarcoma viral oncogene homologue; LCK, lymphocyte-specific protein tyrosine kinase; LKB1, serine/threonine kinase 11; MAPK, mitogen-activated protein kinase; MAP3K4, mitogen-activated protein kinase kinase kinase 4; MAP3K6, mitogen-activated protein kinase kinase kinase 6; MEKK, MAP/ERK kinase kinase; Met, met proto-oncogene (hepatocyte growth factor receptor); MST3, mammalian STE20-like protein kinase 3; Musk, muscle, skeletal, receptor tyrosine kinase; MW, molecular weight; NAK, NF-κB-activating kinase; PDGFR, platelet-derived growth factor receptor; Ret, ret proto-oncogene; Sev, v-ros UR2 sarcoma virus oncogene homologue; SMARTS, Smiles (simplified molecular input line entry) ARbitrary Target Specification; Src, v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homologue; Syk, spleen tyrosine kinase; Tc, Tanimoto coefficient; Tek, TEK tyrosine kinase, endothelial; Tie, tyrosine kinase with immunoglobulin-like and EGF-like domains; TK, tyrosine kinase; Trk, neurotrophic tyrosine kinase; VEGFR2, vascular endothelial growth factor receptor-2; YSK1, yeast Sps1/Ste20-related kinase 1.

 Published on Web 12/03/2010

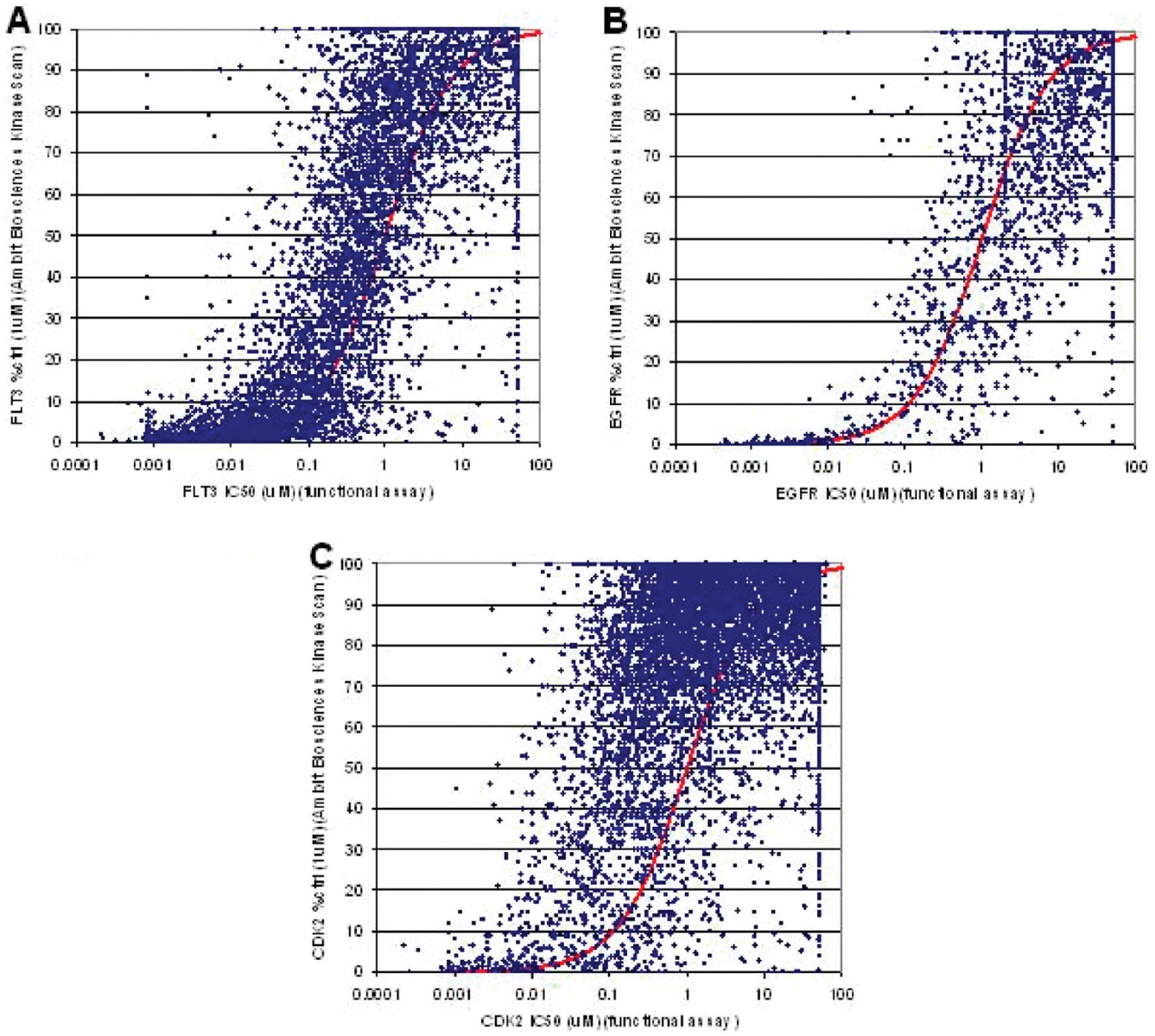


Figure 1. Relationship between the Ambit BioSciences binding assays and functional IC_{50} determinations performed in house for three representative assays: (A) FLT3, (B) EGFR, and (C) CDK2. Functional IC_{50} values are plotted on a log scale. Plots were generated with Excel 2003.

for each class and show that a broad spectrum of structural scaffolds can achieve specificity for most kinases. We also map a trade-off between kinase potency and selectivity that is consistent across multiple assay platforms. Although we find that selective and nonselective compounds are mostly similar in their physicochemical characteristics, we identify specific features that are present more frequently in compounds that bind to many kinases. Finally, we discuss the implications of our results for building compound collections to screen kinase targets.

Results

Summary of the Data Set. A total of 21851 compounds were screened at Ambit BioSciences at 1 μM in an assay panel of 317−402 kinases; 12960 compounds were screened against the full 402-kinase panel. Compounds were selected for screening by two mechanisms. Approximately 13000 compounds were culled from preexisting sets of kinase inhibitors that had exhibited activity in one or more kinase functional assays. The remainder were recently synthesized for ongoing kinase projects and were submitted to Ambit BioSciences for screening regardless of their activity in in-house assays. Because of the two-track selection process, the collection includes both a diverse set of compounds and deep sampling of select chemotypes.

Activity Cutoff Selections. The binding assays performed at Ambit BioSciences were compared to functional data generated internally for a smaller set of compounds to determine reasonable activity thresholds. The two formats were compared for 40 targets for which substantial data were available. Figure 1 shows the comparison between the IC_{50} values determined by functional assays and the Ambit BioSciences binding data (% control) for CDK2 (19678 compounds), FLT3 (6880 compounds), and EGFR (3155 compounds). A theoretical IC_{50} can be computed from a single-concentration binding result using the following

Table 1. Measure of Correspondence between the Ambit BioSciences Binding Assays and Functional Assay IC_{50} Determinations

	log(estimated IC_{50})[a] vs functional log(IC_{50})			activity threshold correspondence		
target	slope	intercept	R^2	5%/50 nM[b]	13%/150 nM[c]	33%/500 nM[d]
CDK2	0.42	1.18	0.17	64	69	76
FLT3	0.88	0.34	0.67	84	88	93
EGFR	0.81	0.24	0.47	83	76	69

[a] Log(estimated IC_{50}) is calculated from the estimated IC_{50} for the Ambit BioSciences binding assay (eq 1). [b] Percentage of all compounds with Ambit BioSciences binding values that are ≤5% control that are also more potent than 50 nM in the functional assay. [c] Percentage of all compounds with Ambit BioSciences binding values that are ≤13% control that are also more potent than 150 nM in the functional assay. [d] Percentage of all compounds with Ambit BioSciences binding values that are ≤33% control that are also more potent than 500 nM in the functional assay.

equation (see the Experimental Section for its derivation):

$$\text{estimated } IC_{50} = \left(\frac{0 - \%\ \text{control}}{\%\ \text{control} - 100}\right) \quad (1)$$

The red curves in Figure 1 indicate the theoretical IC_{50} values that correspond to the single-point binding values on the *y*-axes as calculated with eq 1. If the single-point binding data perfectly corresponded to the functional data, all of the data points would lie along the calculated curves. On the basis of visual inspection of the relationship between the binding data and functional data, it was determined that for most targets a clear correspondence was apparent. The three assays shown are representative of the range of comparisons observed, with FLT3 being among the best, CDK2 being among the worst, and EGFR being typical of the majority of cases. The relatively poor correspondence for CDK2 is likely due to the presence of the cofactor cyclin E in the functional assay and not in the binding assay (see the Experimental Section).

Given the reasonable relationship between the red lines calculated from the single-concentration binding data in Figure 1 and the IC_{50} values from functional assays, we used eq 1 to select cutoffs of 5% control (very potent activity), 13% control (potent activity), and 33% control (moderate activity), corresponding to estimated IC_{50} values of roughly 50, 150, and 500 nM, respectively. Table 1 illustrates the agreement between the binding assay cutoffs and the IC_{50} values from the functional assays for the three representative kinases. For example, 3155 compounds were tested in both the EGFR Ambit BioSciences binding assay and the in-house EGFR functional assay; 83% of the compounds with binding activity below the cutoff for very potent activity (≤5% control) had IC_{50} values in the functional assay of ≤50 nM, and 76% of the compounds with binding activity below the cutoff for potent activity (≤13% control) had IC_{50} values of ≤150 nM. For CDK2, 64% of the compounds with binding activity that was ≤5% control had IC_{50} values in the CDK2 functional assay of ≤50 nM and 69% of the compounds with binding activity that was ≤13% control had IC_{50} values of ≤150 nM.

Of the 21851 compounds screened, 4187 (19.1%) compounds bound none of the assayed kinases at even a moderate level (≤33% control). Of the active compounds, 7790 demonstrated both selectivity and extreme potency [the "SP set" (see the Experimental Section)]. It is important to note that throughout this work, selectivity is defined by the number of kinase targets bound below a certain cutoff. Thus, selective compounds bind few targets, and nonselective compounds bind many targets. The selectivity index SI(*x*) quantifies selectivity for a given activity threshold *x*. For example, an SI(33) value of 0.2 indicates that a compound bound 20% of the targets assayed at ≤33% control. This definition of selectivity does not take into account the relative affinities of the compound for its multiple targets. A compound that was assayed in 100 binding assays and bound two targets is considered selective, regardless of whether it bound the two targets with equal affinity or whether it bound target A 10 times more strongly than target B.

Figure 2A shows a heat map of the Ambit BioSciences activities for the SP set. Each row represents a single kinase and each column a single compound. Kinases are grouped by family/subfamily classification, and compounds are clustered by structural similarity. Bright regions of the heat map indicate clusters of similar compounds that bound to particular families of kinases. Figure 2B presents a detailed view of one heat map segment. The two boxed regions highlight two clusters of compounds: compounds in the first cluster bound Tec family kinases as well as select Src family kinases (primarily BLK and HCK); compounds in the second cluster did not bind Tec kinases and bound to almost all Src family kinases.

Kinase Hit Rates. Each kinase was potently bound by at least one compound, but the hit rates for the kinases varied from 0.03 to 18% (<13% control). Hit rates for all targets for all 21851 compounds, and for only the compounds in the SP set, are provided in the Supporting Information. Of all 402 kinases, the kinases with the lowest hit rates were the tyrosine kinase CTK (4 hits) and two serine/threonine kinases that function in the MAPK pathway, MAP3K6/ASK2 (6 hits) and MAP3K4 (13 hits). Their low hit rates may indicate that it would be difficult to develop inhibitors to target these kinases.

Because most kinases targeted for drug discovery efforts belong to the tyrosine kinase (TK) branch of the kinome, we examined the hit rates for TKs in more detail. Table 2 lists for each TK family the average hit rates for all compounds and for the SP set only. The family of type III receptor tyrosine kinases (family PDGFR in Table 2) had the highest average hit rate for all compounds (13%). This family includes KIT and Flt3, which are known to bind many structurally diverse compounds.[13−17] When only the SP set was considered, the average hit rate for the PDGFR family (10.4%) was roughly the same as for all compounds, showing that both selective and nonselective compounds bound to PDGFR family kinases. The RET and DDR families had similar average hit rates for all compounds of 8 and 7.6%, respectively. However, they had very different average hit rates when only the SP set was considered. For DDR kinases, the SP set average hit rate was comparable to the hit rate for all compounds (7.3%), while for RET kinases, the average hit rate dropped to 2.7%. The reduction in the hit rate for the SP set indicates that most of the compounds that bound to RET kinases are nonselective.

Activity Homology and Sequence Identity. The compound activity profiles that were generated for each kinase can be

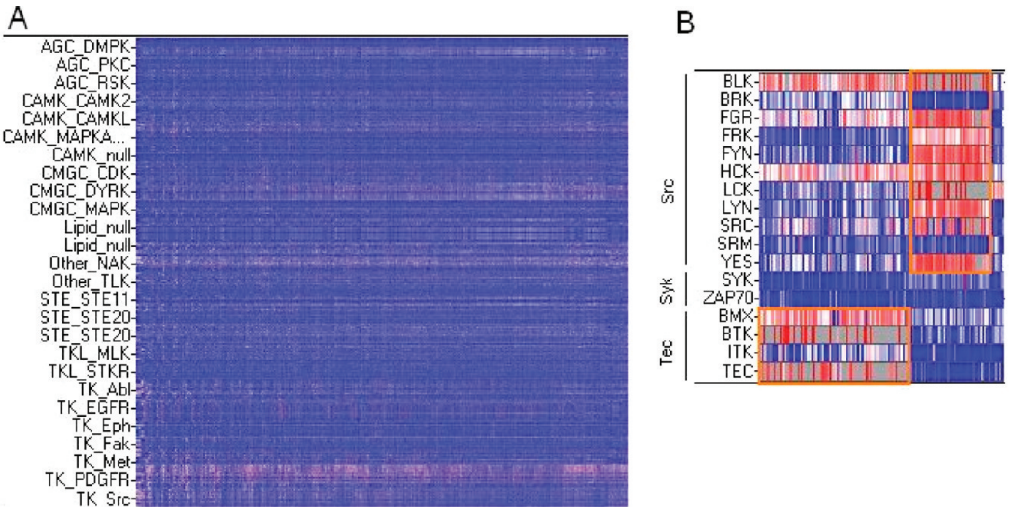


Figure 2. Heat map of binding data for potent, selective compounds. Each row represents a single kinase and each column a single compound. Kinases are grouped by family and compounds by structural similarity. Heat map values range from 0% (red, maximum binding activity) to 100% (blue, minimum binding) of control. (A) Binding data for all potent, selective compounds (the SP set). (B) Detailed view of binding to Tec, Syk, and Src family kinases for two clusters of structurally similar compounds. The heat map was generated with Spotfire DecisionSite version 9.1.

Table 2. Average Hit Rates for Members of Each Tyrosine Kinase Family

family	hit rate (all)	hit rate (SP set)
PDGFR	13.09	10.44
Ret	7.97	2.73
DDR	7.61	7.29
Src	7.41	4.51
InsR	5.95	4.58
JakA	5.91	6.01
Trk	5.5	2.6
Ack	4.94	1.25
Abl	4.76	2.76
Tec	4.42	3.01
VEGFR	4.25	3.15
Axl	4.19	1.45
Tie	3.98	1.98
EGFR	3.93	4.72
Fak	3.73	0.93
Sev	3.7	1.54
Eph	3.66	1.58
Alk	3.59	0.96
Musk	3.46	1.47
Csk	2.65	2.09
Met	2.44	2.06
FGFR	2.26	0.74
Fer	1.09	0.28
Syk	0.89	0.42

used to calculate activity-based measures of kinase homology. We define activity homology as the prior probability that a compound will be active for kinase B given that it is active for kinase A (see the Experimental Section). Figure 3 shows the cumulative distribution function of activity homology (CDF_H) for pairs of kinases which were binned by the sequence identity of their kinase catalytic domains. In general, kinase pairs with higher sequence identity had greater activity homology; half the kinase pairs with 80−90% sequence identity had >60% activity homology, while only 20% of kinase pairs with 50−60% sequence identity had activity homologies in the same range.

Exceptions to the general trend of sequence homology correlating with activity homology exist. For example, the

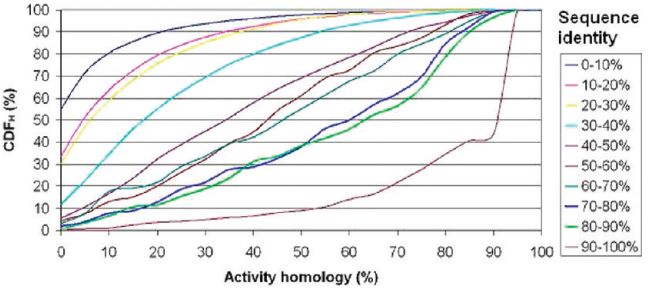


Figure 3. CDF of activity homology. Pairs of kinases were binned by sequence identity. For the kinase pairs in each sequence identity bin, the cumulative distribution function of activity homology is plotted as a separate function. Activity homology for kinases A and B is the fraction of compounds active for kinase A that are also active for kinase B.

kinase domains of YSK1 and MST3 share 88% sequence identity, with only one residue difference (Ile-78 vs Val-94) in their active sites. We would therefore predict that most compounds that bind YSK1 would also bind MST3 (and vice versa) and that these two kinases would have similar hit rates. Both kinases are very selective, with hit rates of 1.1 and 1.2%, respectively (corresponding to 244 and 262 hits, respectively, out of 21851 compounds tested). Surprisingly, however, only 53.3% of compounds that were active in YSK1 were active in MST3, and only 49.6% of MST3 actives bound YSK1. Despite their high sequence similarity, activity in the YSK1 assay is not necessarily predictive of activity in the MST3 assay, and vice versa. The observed lack of overlap between compounds active for YSK1 and MST3 may be caused in part by the low sensitivity of the YSK1 and MST3 binding assays and the limitations of data obtained by screening at a single compound concentration. Nonetheless, the activity homology of these two kinases is unusually low as compared to those of kinase pairs with similar sequence homology (Figure 3).

Cases of low sequence identity and high activity homology can also be found. Bamborough et al.[10] found that LKB1 and AAK1, which are distantly related in sequence, had

unexpectedly similar profiles of inhibitor activity. Our data confirm this activity homology: 87% of compounds that bound LKB1 also bound AAK1. Similarly, the kinase domains of the EphA and EphB kinases share low overall sequence homology with PDGFRB (19−26% identity). When only the 26 residues of the ATP binding site are considered, EphA and EphB kinases are more similar to PDGFRB (~50% identity). Despite their dissimilarity in sequence, some EphA and EphB kinases had high activity homology with PDGFRB. For example, EphB1 has low sequence similarity to PDGFRB but high activity homology; 75% of EphB1 actives were also active in PDGFRB (Table 3). EphA6, in contrast, has the same level of sequence similarity to PDGFRB as EphB1 but much lower activity homology: only 46% of EphA6 actives showed activity in PDGFRB.

Table 3. Sequence Similarity and Activity Homology of EphB1 and EphA6 to PDGFR

kinase	PDGFRB sequence identity (%)	PDGFRB binding site sequence identity (%)	PDGFRB activity homology (%)	gatekeeper
PDGFR	100	100	100	T
EphB1	23	50	75	T
EphA6	20	50	46	V

One explanation for this disparity is the identity of the so-called gatekeeper, a residue in the active site that can affect kinase selectivity.[9] Both EphB1 and PDGFRB have a threonine in the gatekeeper position, while EphA6 has a valine (Table 3). This single position may account for the greater activity homology of EphB1 to PDGFRB than EphA6.

Activity homology can be used to identify kinases that may present selectivity issues when targeting a specific kinase. For example, 20 kinases had >90% activity homology with EphB2 (i.e., they bound to 90% or more of the 372 compounds that bound EphB2). These 20 activity homologues include three EphA and three EphB kinases, which as close sequence homologues would be expected to have high activity similarity to EphB2, but they also include more distantly related tyrosine kinases such as LCK (98%), ARG (95%), and FGR (91%).

Potency and Selectivity. Figure 4A plots maximum binding activity versus selectivity [SI(13) (see the Experimental Section for the definition)] for very potent compounds. Because binding is measured as percent control, a compound that is more potent has a lower value of maximum activity. As seen in the figure, compounds that are slightly less potent bind very potently to fewer kinases. Compounds with maximum activity of 4−5% control all have SI(13) values of <0.04 (i.e., they bind potently to 16 or fewer kinases) and a mean SI(13)

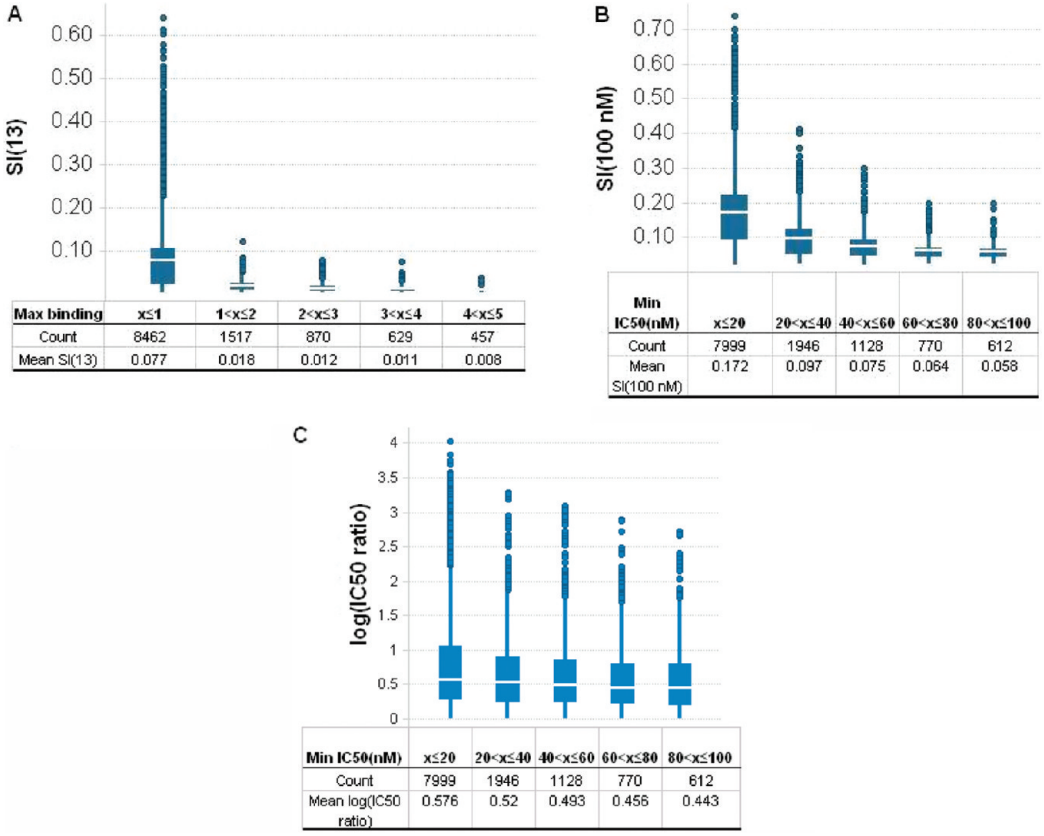
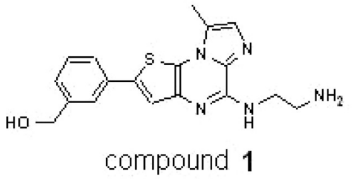
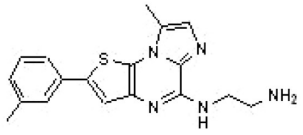


Figure 4. Plots of potency vs selectivity. (A) Maximum binding activity vs selectivity index [SI(13)] for potent compounds. More potent compounds have lower values of maximum binding activity, and more selective compounds have lower selectivity indices. The maximum binding values were binned using a bin width of 1.0, and the SI(13) values were plotted as a box plot. For each bin, the top of the corresponding box indicates the 75th percentile of SI(13), the bottom of the box indicates the 25th percentile, and the white line inside the box indicates the mean value. The number of compounds in each bin and the mean SI(13) value are shown in the table below the plot. (B) Minimum IC_{50} vs selectivity index [SI(100 nM)] for compounds tested in at least 10 in-house kinase functional assays. The minimum IC_{50} values were binned using a bin width of 20 nM, and the SI(100 nM) values for each bin were plotted as a box plot as in panel A. (C) Minimum IC_{50} vs log(IC_{50} ratio) for compounds tested in at least 10 in-house kinase functional assays. See the Experimental Section for calculation of the log(IC_{50} ratio) values. Plots were generated with TIBCO Spotfire version 2.2. The scatter plots corresponding to panels A and B are included in the Supporting Information.



compound **1**

Target	Activity (% control)
IKKB	0.2
IKKA	4.6
CDK7	0.5
CDC2L2_ISO1	6.3
CDC2L1_ISO4	2.1



compound **2**

Target	Activity (% control)
IKKB	4.8
IKKA	15
CDK7	58
CDC2L2_ISO1	68
CDC2L1_ISO4	42

Figure 5. Structures and kinome binding profiles of a representative pair of compounds. The structures of compounds **1** and **2** are shown at the left and their binding activities at the right. Binding data are shown for all targets that were bound by compounds **1** and **2** (≤13% control).

value of 0.008 (i.e., on average they bind potently to three kinases). In contrast, compounds that are slightly more potent can bind with high affinity to more kinases. As potency increases in the plot (maximum activity decreases), the range of SI(13) values increases as well. Compounds with maximum activity of 0−1% control have a mean SI(13) value of 0.077, indicating potent binding to 31 kinases on average, and 25% of compounds in this bin have SI(13) values of ≥0.10. Compounds that are more potent generally acquire activity for multiple kinases,[18] probably because of the conservation of ATP binding site features among kinases.

Figure 4A plots the trend of increased potency correlating with decreased selectivity for the data set as a whole, for which the more potent and less potent compounds being compared include structurally diverse molecules. The same trend holds true for pairs of structurally similar compounds that vary in potency; 20003 pairs of structurally similar compounds in which both compounds are very potent but one compound is more potent than the other were assembled. In 14298 cases (71.5%), the more potent compound is also less selective. Compounds **1** and **2** are one such pair: both are imidazo[1,2-*a*]thieno[3,2-*e*]pyrazines and were designed to inhibit IKKβ.[19] They differ only in the substitution pattern of the pendant phenyl (Figure 5). Compound **1** shows potent activity in five Ambit BioSciences assays and is most potent in the IKKβ assay, with an activity of 0.2% control. Compound **2**, although structurally similar to **1**, is both more selective and less potent and bound only IKKβ at 4.8% control.

Importantly, the selectivity−potency trend outlined above is also observed with a panel of kinase enzyme assays. Figure 4B shows the correlation of greater potency with decreased selectivity [SI(100 nM)] for this data set. Both panels A and B of Figure 4 highlight the difficulty of achieving potency and selectivity for a single target. It is critical to note that in these plots, selectivity is calculated as the fraction of assays for which a compound is active below a given threshold. The relative activity of the compound for different targets that meet the activity threshold is ignored. This trend is particularly relevant in the hit identification stage of drug discovery programs.

An alternative measure of kinase selectivity commonly used with functional data is a compound's fold selectivity for its primary target over other, secondary targets. Even if a compound potently inhibits a number of kinases, it may be considered selective if its activity for the target of interest is 1 order of magnitude greater than its activity for any other kinase. To quantify this relative selectivity, we compute the log of the ratio of the compound's second-most potent IC_{50} (for kinase 2) to its most potent IC_{50} (for kinase 1). A higher log(IC_{50} ratio) indicates greater selectivity; a compound with a log(IC_{50} ratio) value of 1 exhibits 10-fold selectivity for its primary target (kinase 1). When the log(IC_{50} ratio) is plotted versus potency, only a weak correlation is seen between selectivity and potency (Figure 4C), with the average log(IC_{50} ratio) exhibiting a slight decrease as potency decreases. This confirms that within a background of activity against a number of kinases, compounds can nonetheless be made to be extremely potent and relatively selective for a single target during the course of lead optimization.

Structural Features of Nonselective and Selective Compounds. The data set includes compounds that span the selectivity spectrum from 0.002 ≤ SI(33) ≤ 0.76. A set of 286 compounds are particularly nonselective, with moderate or potent activity for 40% or more of assayed kinases. To elucidate what structural features may confer such nonselective kinase activity, the physicochemical properties of nonselective compounds were compared to the properties of the SP set. Both the SP set and the nonselective compounds were filtered by structural similarity to avoid any bias introduced by including multiple similar structures. Table 4 lists the structural descriptors used and the coefficient for each descriptor, which indicates how well-separated the descriptor distributions are for the selective and nonselective classes.

As shown in Table 4, no descriptor completely separates the two classes; for each structural feature, there is significant overlap between the distributions for selective and nonselective compounds. Nonetheless, the results do highlight features that are more prominent in nonselective compounds. One of the descriptors that most significantly discriminates between the two classes is the number of rings: only 9.6% of selective compounds have at least 6 rings, while 33% of nonselective compounds do (Figure 6A). Another feature that tends to differ between the two classes is a shape descriptor that corresponds to the "width" of the molecules. Steric quadroples describe the extents of a molecule, with the *X*, *Y*, and *Z* quadrupoles mapping to the length, width, and height, respectively. Nonselective compounds have larger values of the *Y* steric quadrupole, indicating that they may span the width of the ATP binding site more fully than selective compounds and thereby have the potential to interact with more residues. In addition, nonselective compounds are

Table 4. Physicochemical Descriptors and Their Separation Coefficients for Selective and Nonselective Compounds[a]

descriptor	coefficient	descriptor	coefficient
no. of polar hydrogens	0.39	no. of heavy atoms	0.04
no. of rings	0.36	monopole	0.04
quadrupole *Y*	0.35	calculated LogP (VolSurf)	0.04
no. of NH and OH groups	0.31	quadrupole *Z*	0.03
no. of N and O atoms	0.28	molecular volume (VolSurf)	0.03
no. of amides	0.10	molecular surface area (VolSurf)	0.02
no. of halogens	0.08	MW	0.02
quadrupole *X*	0.05	no. of rotors	0.00

[a] Additional descriptors and coefficients are given in the Supporting Information.

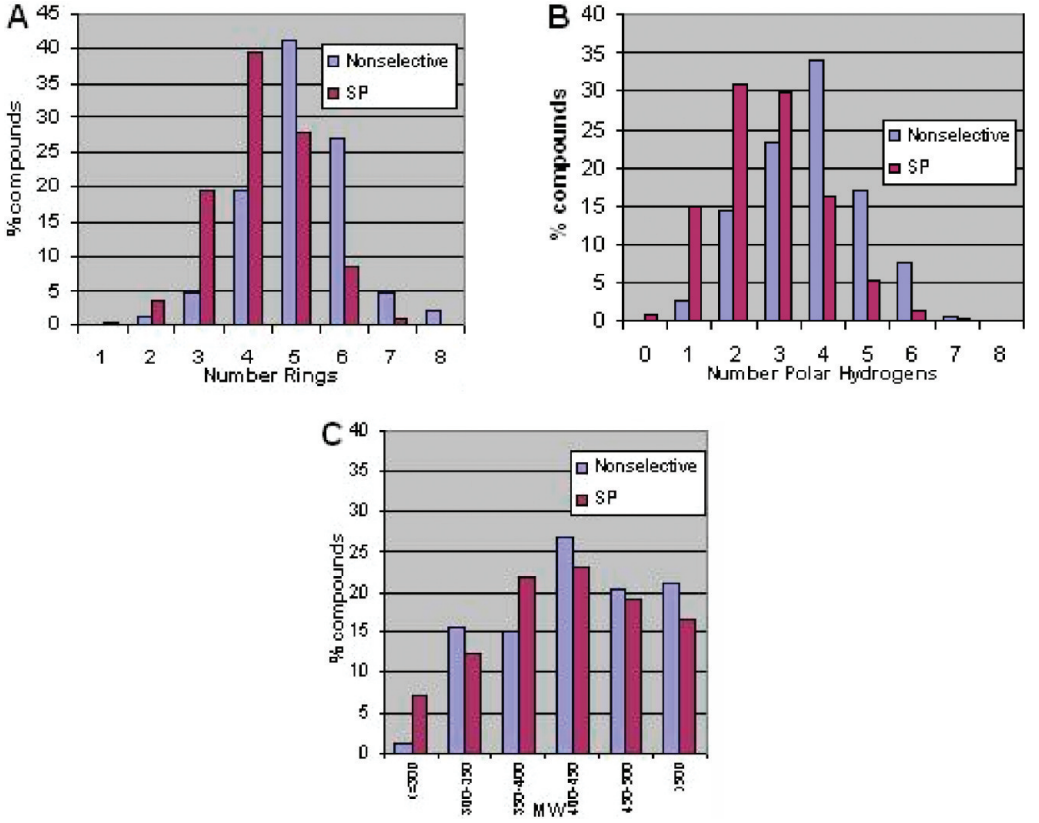


Figure 6. Structural characteristics of selective and nonselective compounds: (A) number of rings, (B) number of polar hydrogens, and (C) molecular weight. The distributions are shown for the filtered nonselective and SP sets. Plots were prepared with Excel 2003.

more likely to have many polar hydrogens: 23% of selective compounds have at least four polar hydrogens, while 59% of nonselective compounds have at least four polar hydrogens (Figure 6B). One possible explanation is that the greater number of hydrogen bond donors in nonselective compounds increases the likelihood of forming hydrogen bonds with the residues lining the ATP binding site in a greater variety of protein conformations. Many, though not all, of the additional polar hydrogens in nonselective compounds are due to the more frequent presence of primary amines, which have two polar hydrogens each. Positively charged amines can potentially interact with the catalytic aspartate and with the negatively charged residues that coordinate Mg^{2+} ions when ATP is bound in the active site.

Interestingly, no significant difference is seen in the molecular weight (MW) distributions of the two compound sets. It might be expected that larger compounds would tend to be more specific, because functional groups are typically added to a molecular core to achieve greater specificity. In this data set, however, no such trend is observed. The MW distributions for both selective and nonselective compounds are skewed toward higher molecular weights, with a slightly increased skew for nonselective compounds. 68% of nonselective compounds have molecular weights of >400, as do 59% of selective compounds. As one can see in Figure 6C, the MW distribution of the nonselective compounds does not differ substantially from that of the selective compounds.

Chemotypes and Specificity. Using the chemotype assignment scheme described in the Experimental Section, 34 chemotypes are represented in the SP set by at least 50 examples. Figure 7 plots the frequency with which each chemotype is active ($\leq$13% control) against each kinase family/subfamily. While there are gaps in the family coverage, many chemotypes have at least a few examples that are active against every family. An equivalent plot with less selective compounds included has even fewer gaps (data not shown). For many of the chemotypes, the defining substructure is a heteroatomic ring system that can in principle bind to the hinge region of the ATP binding site. For these chemotypes, the common structural motif can be

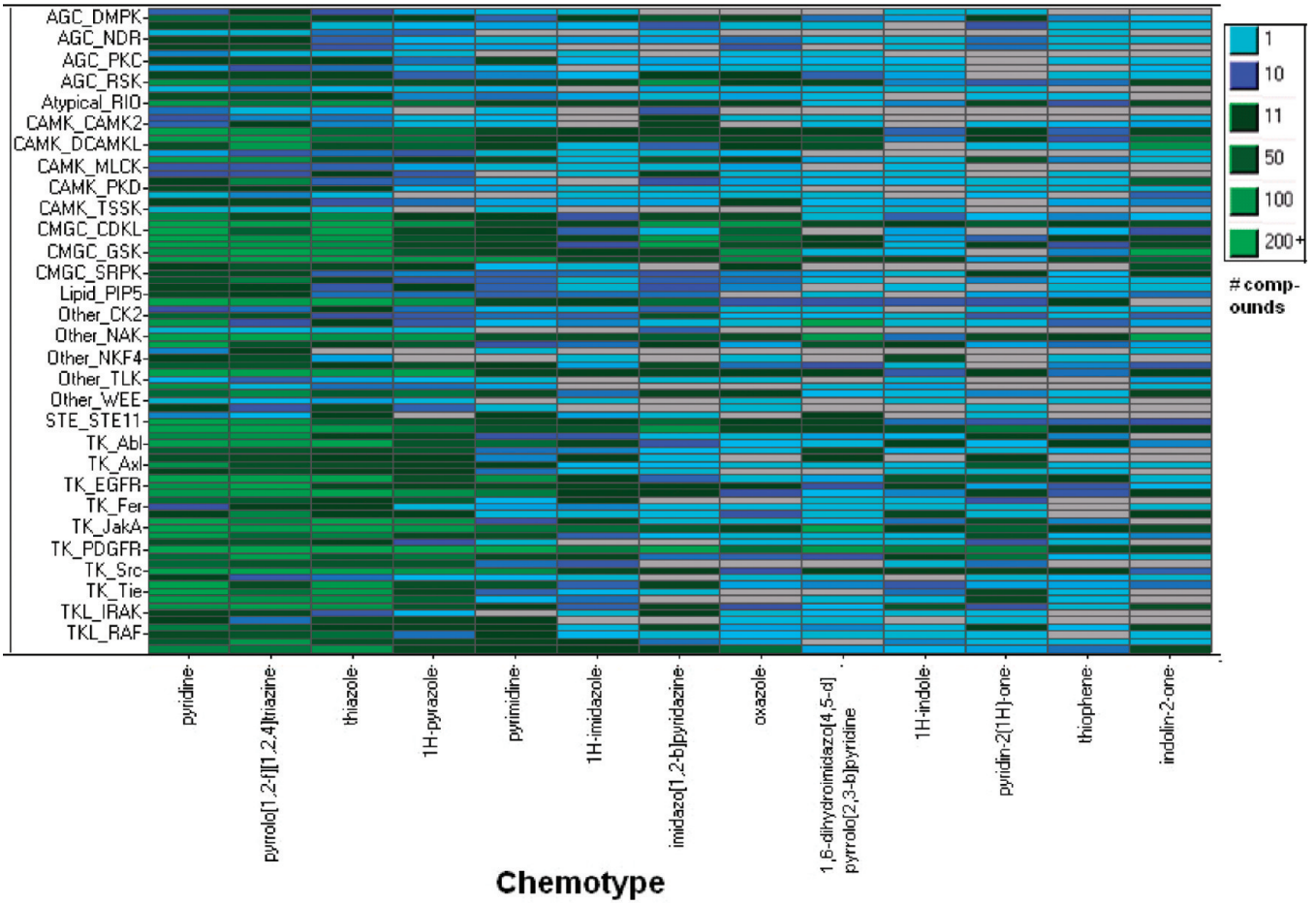


Figure 7. Kinase families and chemotype specificity. Thirteen chemotypes with ≥50 representatives in the SP set are represented along the *X*-axis, and kinase families are listed on the *Y*-axis. The color of each rectangle indicates the number of compounds in the given chemotype that potently inhibit one or more kinases in the specified family. Plots were prepared with TIBCO Spotfire version 2.2.

viewed as a core scaffold, and the members of the chemotype differ in the identity of the substituents appended to the core. Because most chemotypes have representatives that can bind to most kinase families, the results suggest that for a specific compound, the identity of the substituents may drive its activity pattern as much or more than the identity of the core. Some chemotypes, such as the pyridines and 1*H*-pyrazoles, have relatively evenly distributed activities across many different families, while other chemotypes are skewed toward a particular family. For instance, 74% of the indolin-2-ones are active against NAK kinases (283 of 318), and 83% (318 of 383) exhibit GSK activity. This does not necessarily reflect an intrinsic affinity of this particular core for the NAK and GSK families, because the screening collection contains many structurally similar analogues and does not constitute an unbiased set. Despite its skewed target distribution, the indolin-2-one core is capable of binding many different kinase families (Figure 7). Our data suggest that many different core scaffolds can potentially be modified to produce selective compounds for most kinase targets.

Structural Similarity and Profile Similarity. Chemically similar compounds would be expected to inhibit the same targets and thus to have similar activity profiles. Bamborough et al.[10] observed that structural similarity correlated with activity profile similarity in their data set, and our larger compound set follows the same trend. Profile−profile similarities were calculated as a continuous Tanimoto score as

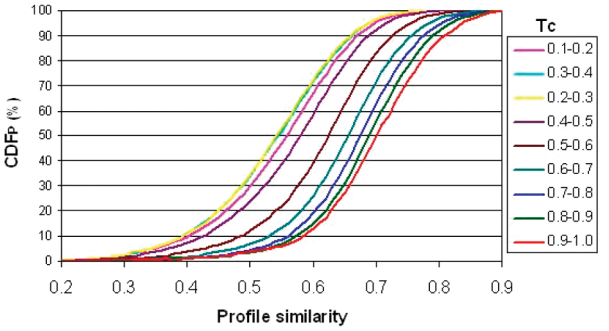


Figure 8. CDF of profile similarity (CDF_P) by structural similarity bin. Pairs of compounds were binned by structural similarity (Tc). For the compound pairs in each structural similarity bin, the cumulative distribution function of profile similarity is plotted as a separate function.

described in the Experimental Section. A higher Tanimoto coefficient indicates greater profile−profile similarity. Figure 8 plots the CDF of the activity profile similarity (CDF_P) for pairs of compounds that were binned by their structural similarity [Tc (see the Experimental Section)]. As shown in the figure, pairs of compounds that have more similar structures tend to have more similar activity profiles, and therefore, the CDFs are shifted to the right as structural similarity increases. For example, approximately half (52%) of compound

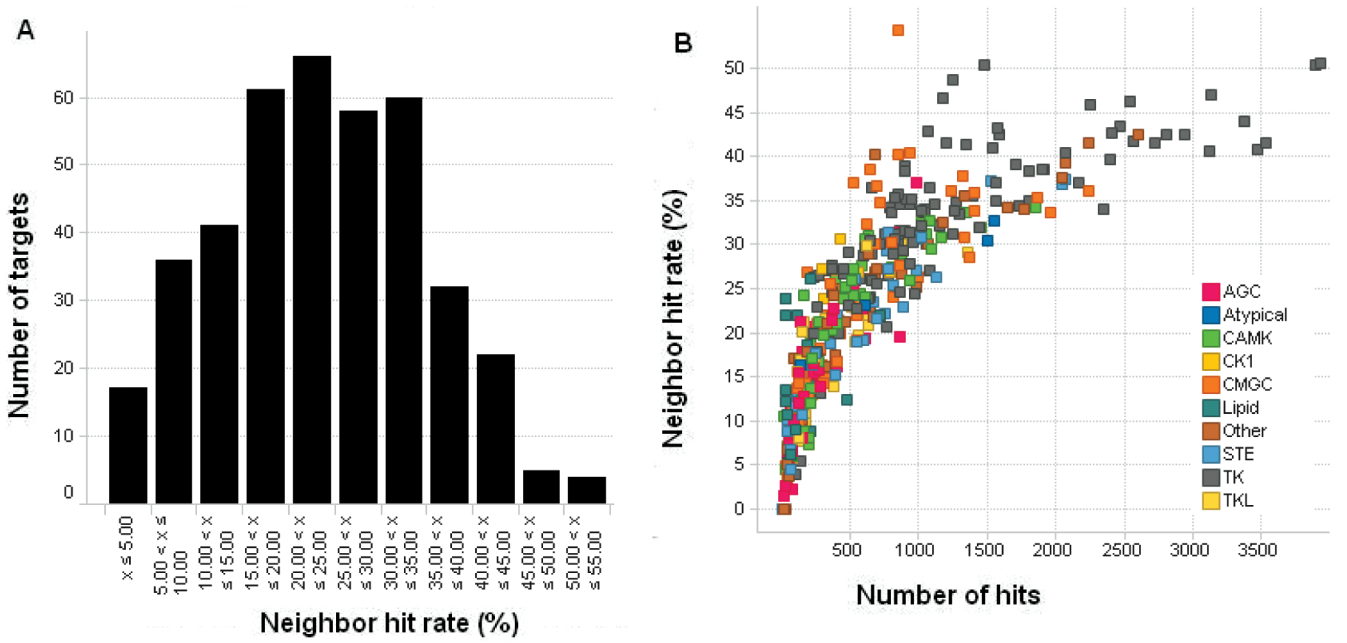


Figure 9. Neighbor hit rates. (A) Distribution of neighbor hit rates over the set of kinase targets. The neighbor hit rate is defined as the percent of structural analogues of a target's actives that are themselves active for the same target. (B) For each target, the number of hits is plotted vs the neighbor hit rate. The targets are colored by kinase class.

pairs with moderate structural similarity (0.5−0.6) have an activity profile similarity of >0.73. In contrast, 83% of highly similar compound pairs (structural similarity of >0.9) have activity profile similarities in that range.

Although structurally similar compounds tend to have similar activity profiles across the entire panel (Figure 8), the likelihood that structural neighbors will have similar activities for a particular kinase differs from target to target. For each target, we identified compounds that are structurally similar to compounds that bind the target and computed the fraction of these structural neighbors that themselves bind the target. The distribution of the neighbor hit rates by target, shown in Figure 9A, is centered around 22%. On average, ~22% of structural analogues of a known active compound can be expected to be active. As the number of active compounds for a target increases, the neighbor hit rate also increases (Figure 9B); if a target is inhibited by relatively few compounds, a smaller fraction of structural analogues of the active compounds are themselves active.

Discussion

Kinome Profiling and Target "Druggability". Recent studies have described Ambit KinomeScan screening results for small sets of 200−500 structurally diverse compounds that were tested against panels of 119−317 kinases.[10−12] Our compound set is much larger and, while it includes a structurally diverse set, was selected to include multiple representatives of many structural clusters. While the distribution of hit rates found here largely agrees with that found by Bamborough et al., there are significant differences in hit rates for a number of kinases that are likely due in large part to the differences in the compound sets. GAK, JNK3, CK1D (CSNK1D), CK1E (CSNK1E), and JNK3 had exceptionally high hit rates in the earlier study, while in our data set, GAK had a moderately high hit rate (8.1% of all compounds) and JNK3, CK1D, and CK1E all had low hit rates (1.4, 2.6, and 3.6%, respectively). In a more typical case in which our hit rates confirm previous findings, kinases of the MEKK/MAP3K family had very few hits in our collection, in agreement with the earlier study's low hit rates for MAP3K5 (ASK1) and MAP3K4. The lack of hits in two such different compound sets indicates that these kinases may be difficult targets for drug discovery efforts.

In contrast, the kinases of the PDGFR family had high hit rates in both screens. This led us to examine whether PDGFR kinases have any unusual binding site features that may contribute to their binding versatility. Alignment of the PDGFR kinases reveals a conserved Cys-x-Tyr-Gly-Asp motif in the extended hinge region at the outer edge of the binding site. While the Gly and Asp of the motif are commonly found in other kinase families, the Cys-x-Tyr sequence is unique to the PDGFR family. The large tyrosine is in a site commonly occupied by Gly. It is not obvious why this unusual combination of residues would allow PDGFR kinases to bind to structurally diverse ligands. One possibility is that the Cys-x-Tyr motif may allow sampling of alternative conformations of the extended hinge in PDGFR kinases as compared to other kinases.

The neighbor hit rate, coupled with the hit rate for all compounds, provides a measure of a kinase's potential tractability as a drug target. The correlation between the two rates suggests that the more hits that are found in an initial kinase screen, the more likely it is that analogues of the first hits will themselves be active. Interestingly, our finding that on average 22% of analogues of active compounds tend to be actives themselves agrees with an earlier study showing that 30% of close analogues (Tanimoto coefficient of ≥ 0.85) of hits in high-throughput assays are themselves active.[20] However, the range we find is 0−54%, suggesting that this metric should be computed for each target.

Implications for Kinase-Focused Screening Collections. Many chemical substructures have been identified as kinase-specific privileged structures; i.e., compounds containing these structure fragments are enriched for kinase activity as

compared to other target classes.[21−24] These fragments have flat aromatic ring systems containing hydrogen bond acceptors and donors that can interact with the hinge residues of the ATP binding site. The chemotypes in our compound set represent many such substructures, including well-known kinase-binding scaffolds such as pyrrolo[2,1-*f*][1,2,4]triazines[25] and 1*H*-pyrazoles.[26] While these core substructures may indeed bind preferentially to kinases as compared to other target classes such as GPCRs or ion channels, within the kinome the concept of privileged substructures may be less applicable. Our data show that, in general, most chemotypes can bind selectively to a broad range of kinases; we do not find many chemotypes that are capable of binding to only a few specific kinase families. This indicates that kinase family specificity depends less on the identity of the core scaffold and more on the specific substituents appended to the core.

Various procedures for selecting compound sets for high-throughput screening (HTS) of kinase targets have been described.[27−33] By screening a small, diverse collection of compounds that are enriched for compounds likely to bind kinases, one can find leads for novel kinase targets without conducting a more extensive HTS campaign of more than 1 million compounds. To assemble such kinase-focused screening libraries, candidate compounds are assigned a score reflecting their predicted kinase binding potential. Such scores may incorporate various similarity metrics based on comparisons to known kinase inhibitors,[32] docking scores or interaction fingerprints for one or more kinase structures,[30−32] descriptor-based activity predictions,[27,28] and/or activities from previous kinase screens. High-scoring compounds are typically identified and then passed through a diversity filter. This step ensures that multiple diverse scaffolds are included and that no single scaffold is overrepresented in the focused library.

Our results support a potentially complementary paradigm. In this approach, a small number of known kinase scaffolds would be selected on the basis of their synthetic accessibility and potential for conferring intellectual property. The library would comprise as many analogues containing these scaffolds as possible. If a diversity pick were required to select a subset from many analogues, it would be based only on comparison of the structural fragments appended to the core scaffold. Such an approach would take advantage of the proven kinase binding potential of known cores and allow exploration of more diverse fragments in the noncore regions than may be found in standard kinase-focused collections. Scaffold-oriented kinase libraries in line with this paradigm have been synthesized and successfully screened for active kinase inhibitors.[34−36] Of particular note is a recent effort to systematically identify and build libraries around novel scaffolds with potential hinge-binding elements that were not represented in the kinase literature.[36]

A scaffold-oriented screening strategy may be especially appropriate for kinases with low neighbor hit rates. Harper et al.[37] have shown that a target's neighbor hit rate has a large impact on the number of compounds that need to be screened to find an expected number of leads. For kinases with high neighbor hit rates, hits will likely be found by screening a compound collection of many diverse chemotypes (where each is represented by a small number of compounds). Because analogues of active compounds have a high probability of being active, active chemotypes can be identified even if they are sparsely represented in the screening collection. For a kinase with a low neighbor hit rate, however, even if a particular scaffold can bind the target of interest, many analogues containing the scaffold may need to be sampled to identify a lead compound. Because most chemotypes have the potential to bind most kinase families, screening a large number of analogues of a single chemotype may be more productive for more difficult kinase targets than screening fewer examples of more chemotypes.

Molecular Determinants of Selectivity. In general, we find that selective and nonselective compounds have overlapping profiles of physicochemical features. We did not see a significant difference in the molecular weights of nonselective compounds as compared to selective compounds, in contrast to an earlier work that found that molecular weight greatly affected selectivity.[10] We did find that extremely nonselective compounds tend to have more rings and more polar hydrogens than selective compounds, suggesting that these features should be avoided when feasible in designing new kinase inhibitors. Initial attempts to select feature combinations that would predict compound selectivity have shown some promise (data not shown). Future refinement of the models is expected to improve their predictive capabilities.

Conclusions

The kinome profiling data described in this work can be used to evaluate the potential druggability of kinase targets and to identify potential off-target liabilities. Because of the observed variation in target tractability and our finding that most chemotypes can bind to most kinase families, we propose a scaffold-oriented screening strategy that complements other methods of designing kinase-focused libraries. The observed variability in hit rates and neighbor hit rates for the highly conserved kinases suggests that targets in less conserved classes may have similar screening variability. The considerations that we outline for kinase screening strategies may therefore apply to target classes such as GPCRs and ion channels as well.

Experimental Section

Binding Assays. The binding assays were conducted by Ambit BioSciences as described previously.[11,12] Compounds were screened at a single concentration of 1 μM. The assays measure a compound's ability to inhibit binding of a bait ligand, with 0% control corresponding to full inhibition and 100% control to no inhibition. Three different activity cutoffs were used in this work to define moderate, potent, and extremely potent binding. On the basis of comparisons between Ambit BioSciences binding data and in-house functional assay data (see the following section), moderate activity was defined as $\leq$33% control, potent activity as $\leq$13% control, and extremely potent activity as $\leq$5% control. The hit rates listed in Table 1 were calculated with the potent activity threshold ($\leq$13% control). For each compound, a selectivity index at an activity threshold x, SI(x), is calculated as the number of assays for which the compound was active $\leq x$ % control divided by the total number of kinases for which the compound was assayed. The range of potential SI values is 0−1; an SI of 0 indicates that no kinases were bound at the given threshold, and an SI of 1 means that all kinases were bound.

Functional CDK2, FLT3, and EGFR Assays. The kinase functional assays were performed in 384-well plates using a 30 μL reaction volume. The reactions were run for 1 h in reaction buffer [100 mM HEPES (pH 7.4), 10 mM $MgCl_2$, 0.015% Brij35, and 4 mM DTT]. The reaction was initiated by combination of kinase CDK2/cyclin E (Proqinase), FLT3 (Invitrogen), or HER1/EGFR (Invitrogen) and 1.5 μM peptide substrate fluorescein-QSPKKG-NH_2 (for CDK2/cyclin E) or fluorescein-EAIYAAPFAKKK-NH_2

(for FLT3 and HER1/EGFR) in the presence of ATP in wells containing the compound to be tested. The ATP concentrations used in these assays were 30 μM for CDK2, 200 μM for FLT3, and 6 μM for HER1/EGFR. The final concentration of DMSO in these reaction mixtures was 1.6%. The reaction mixture was analyzed on the Caliper LabChip 3000 (Caliper, Hopkinton, MA) by the electrophoretic mobility shift of the fluorescent substrate and phosphorylated product. Inhibition data were calculated by comparison to no enzyme control reactions for 100% inhibition and DMSO-only reactions for 0% inhibition. Dose−response curves were generated to determine the concentration required to inhibit 50% of kinase activity (IC_{50}). Compounds were dissolved at 10 mM in dimethyl sulfoxide (DMSO) and evaluated at 11 concentrations, each in duplicate. IC_{50} values were derived by nonlinear regression analysis.

Selection of Activity Cutoffs. The Logit transformation is related mathematically to the Hill equation with the assumption that the Hill slope is 1:

$$\text{effect} = \frac{\text{dose}}{K_i + \text{dose}}$$

which can be rearranged to

$$K_i = \text{dose} \times \frac{1 - \text{effect}}{\text{effect}}$$

which is mathematically equivalent to

$$K_i = \text{dose} \times \frac{\text{Maxeffect} - \text{effect}}{\text{effect} - \text{Mineffect}}$$

where Maxeffect and Mineffect are the maximum and minimum effects possible in the assay, respectively. In the case of the Ambit BioSciences binding assays, the maximum effect is 0% control, the minimum effect is 100% control, and the dose is 1 μM. The following equation was utilized to calculate the estimated IC_{50} values (red lines shown in Figure 1).

$$\text{estimated } IC_{50} = \left(\frac{0 - \%\ \text{control}}{\%\ \text{control} - 100}\right) \quad (2)$$

Structural Similarity. To compare compounds by their chemical structures, we retrieved the connection table for each compound from the corporate database. An atom-pair fingerprint[38] was generated for each structure. The Tanimoto coefficient (Tc) was used as a measure of structural similarity.[39] Two compounds are considered structurally similar if their fingerprints have a Tc of $\geq$0.7. The all-against-all matrix of compound pairwise similarities was calculated.

Neighbor Hit Rates. Structural neighbors of a compound make up the set of compounds with Tc values of $\geq$0.7 to the query structure. A compound is not considered its own neighbor even though its Tanimoto similarity to itself is equal to 1. For each kinase, all structural neighbors of all potently active compounds ($\leq$13% control) were identified, and the combined neighbor list was filtered to remove duplicates. The neighbor hit rate is defined as the number of structural neighbors that were active for the target of interest ($\leq$13% control) divided by the total number of structural neighbors.

SP Set. A total of 7790 compounds were identified on the basis of their selectivity and extreme potency. These compounds are collectively labeled the SP set. Selectivity denotes moderate or potent activity for at most 10% of the panel [i.e., SI(33) $\leq$ 0.1], and extreme potency is defined as activity for at least one kinase of $\leq$5% control and $>$0% control. Values of 0% control have been shown to have a higher false positive rate upon retesting than values greater than zero, especially in cases where a compound has 0% control activity for a single kinase and no non-zero activity for any other kinase (data not shown). Compounds were therefore required to have activity greater than zero and less than or equal to 5% control to be considered very potent and included in this set.

Potency and Selectivity. To compare potency and selectivity, we found 11935 compounds that met the criterion for extreme potency ($>$0 and $\leq$5% control for at least one kinase), regardless of their selectivity. For these compounds, the SI(13) was calculated to measure selectivity using a potent activity threshold; 20003 pairs of structurally similar compounds were identified as described above (both compounds in each pair were extremely potent).

Selectivity in Functional Assays. A total of 12455 compounds that were tested in at least 10 in-house assays of kinase catalytic activity and had an IC_{50} of $\leq$100 nM for at least one assay were identified. For each compound, a selectivity index, SI(100 nM), was calculated as the number of assays that gave IC_{50} values of $\leq$100 nM divided by the total number of kinases for which the compound was assayed. Selectivity was also calculated for each compound as the log of the ratio of the second-most potent IC_{50} to the most potent IC_{50}. This measure describes the relative selectivity of the compound for its highest-affinity target and does not involve an activity cutoff.

Kinase Sequences and Alignments. Ambit BioSciences targets were assigned to groups and families on the basis of the Manning classification scheme.[1] The DNA insert sequence of each target was obtained from Ambit BioSciences and translated into its corresponding protein sequence. Each pair of protein sequences was aligned with T-coffee version 6,[40] and the sequence identity was calculated as the number of identical aligned residues divided by the number of aligned positions. To compare the ATP binding sites of kinases in the PDGFR, EphA, and EphB families, we identified 26 binding site residues by visual inspection of a representative kinase crystal structure and mapped them to the corresponding positions in the pairwise sequence alignments.

Chemotype Assignment. ClassPharmer version 4.6[41] was used to systematically identify the common ring systems in a large database of known kinase inhibitors (data not shown). Two hundred ninety-seven ring systems that contain the hydrogen bonding characteristics of a kinase inhibitor were manually selected and converted to SMARTS patterns. The SMARTS generated from this procedure were used for substructure searching in Pipeline Pilot version 7.5.2[42] to assign compounds to "chemotype" classes. Compounds were annotated with all matching substructures and thus may belong to multiple chemotypes. The substructure motifs are not true chemotypes in the sense that they are not "cores", and many matching structural fragments may be noncore substituents. However, most of the substructures used to define the chemotype classes are known to function as kinase-binding cores and can form the critical hydrogen bonds to the hinge residues of the ATP binding site.

Thirty-four chemotypes contained at least 50 representatives in the SP set. For each of these 34 chemotypes, for each kinase family, the number of compounds in the SP set that bound potently ($\leq$13% control) to at least one family member was tallied. These raw counts were converted to frequencies for each chemotype by dividing the count by the total number of SP compounds in each chemotype.

Activity Homology. The activity homology for each pair of kinases A and B was calculated as the prior probability of activity for kinase B given activity for kinase A, i.e., the percent of active compounds ($\leq$13% control) for kinase B that were also active for kinase A. An activity homology of 65% means that 65% of the compounds that were active for kinase B also had activity for kinase A.

Kinase pairs were binned by pairwise sequence identity (as described above), and the cumulative distribution function of activity homology (CDF_H) was calculated separately for the kinase pairs in each sequence identity bin. $CDF_{Hi}(x)$ is the percent of kinase pairs in sequence bin i that have activity homology of $\leq x$.

Physicochemical Features of Selective and Nonselective Compounds. Two hundred eighty-six compounds were identified

with an SI(33) of ≥0.4. For each of the 286 nonselective compounds and the 7790 compounds in the SP set, a single three-dimensional conformation was generated from the connection table using Omega version 2.3.[43] Simple derived properties such as molecular weight, the number of polar hydrogens, etc., were calculated with molProps, an in-house tool based on the OpenEye OEChem toolkit.[43] Additional physicochemical features were calculated with VolSurf version 4.1.4.[44,45] For each of the 90 total descriptors, the separation of the nonselective and selective classes was calculated using the mean-differences test. The separation coefficient for descriptor i, S_i, equals $(\mathrm{avg}_{s,i} - \mathrm{avg}_{n,i})^2/(\mathrm{stddev}_{s,i}^2 + \mathrm{stddev}_{n,i}^2)$, where $\mathrm{avg}_{s,i}$ and $\mathrm{stddev}_{s,i}$ are the average and standard deviation of the values of descriptor i for selective compounds, respectively, and $\mathrm{avg}_{n,i}$ and $\mathrm{stddev}_{n,i}$ are the average and standard deviation of the values of descriptor i for nonselective compounds, respectively.

The analysis was repeated with the selective and nonselective compounds clustered by structure to reduce the bias introduced by having large clusters of very similar compounds in the compound sets. Using the structural similarity matrix described above, group-average linkage clustering was applied to both the SP set and the set of nonselective compounds. Two similarity cutoffs were tried. When a tight similarity cutoff of 0.7 was employed, 3945 selective clusters (including singletons) and 152 nonselective clusters were found. A looser similarity cutoff of 0.5 yielded 1302 selective clusters and 61 nonselective clusters. For both cutoffs, a single representative was chosen randomly for each cluster and the separation coefficients were recalculated. The calculations were also conducted with all nonselective compounds included and only the SP set clustered by structure. While the magnitude of the separation coefficients differed depending on how compounds were clustered, the observed trends are the same in all cases and the same features emerge as important. The values shown were derived from the analysis with the tighter cutoff and both selective and nonselective compounds clustered.

Activity Profile Similarity. To compare the profile similarity of two compounds, we developed an algorithm based on the Tanimoto coefficient. The normal Tanimoto algorithm uses a fingerprint in which each bit is a binary response. For a fingerprint based on activity, an arbitrary activity cutoff must be set to separate the actives from the inactives. This introduces a large discontinuity when activities fall close to the cutoff value and can make the similarity results rather dependent on that value. Additionally, it was found that the large amount of negative activity data (i.e., lack of activity) in the results for these assays often overwhelms the positive data upon calculation of the Tanimoto coefficient, especially for more selective compounds. The Continuous Tanimoto algorithm was designed to address both of these issues.

The Tanimoto algorithm was modified to remove this discontinuity at the cutoff threshold by making the contribution of one bit a linear function of the activities of one assay for the two molecules.

An activity fingerprint is comprised of the values for each assay (% control). A continuous version of the Tanimoto index that uses the raw activity values rather than binning values into discrete active and inactive categories was developed. The Tanimoto profile similarity for compounds **1** and **2** is

$$S_{\mathbf{1},\mathbf{2}} = [2Na + 2\sum_{i=1,n}(A_{\mathbf{2},\mathbf{1}}b/100 - A_{\mathbf{1},i}a/100)]/\{2Na + \sum_{i=1,n}[A_{\mathbf{1},i}(b-a)/100] + \sum_{i=1,n}[A_{\mathbf{2},i}(b-a)/100]\}$$

where N is the total number of assays, $A_{\mathbf{1},i}$ is the activity of compound **1** in assay i, and $A_{\mathbf{2},i}$ is the activity of compound **2** in assay i. The constants a and b were empirically set to 10 and 1, respectively. The contribution for each assay is 10 if both compounds are 0% control, 1 if they are both 100% control, and 0 if one is 100% and the other 0%. For intermediate values, the contribution is a linear extrapolation. There is a normalization factor to put the results between 0 and 100% similarity.

The ratio of a to b is the weight of the contribution given to active results versus inactive. A ratio of 1 generates similarity results that are determined primarily by the inactive data. This makes all highly selective compounds very similar, regardless of the few targets against which they are potent. A very high ratio would consider only the active data that are more appropriate for selective compounds. A low ratio works better for nonselective compounds. A ratio of 10 produced similarity results for most of our compounds that make intuitive sense upon comparison of histograms of the raw profile data.

Compound pairs were binned by pairwise structural similarity (Tc), and the cumulative distribution function of activity profile similarity ($\mathrm{CDF_P}$) was calculated separately for the compound pairs in each structural similarity bin. $\mathrm{CDF_{P}}_i(x)$ is the percent of compound pairs in structural similarity bin i that have activity profile similarity of $\leq x$.

Acknowledgment. We thank Lawrence Williams for his critical reading of the manuscript. John Tokarski and David Langley contributed to the selection of compounds for the screening set. Tai Wong was instrumental in facilitating the collaboration with Ambit BioSciences. Songmei Xu, Randi Brown, Liang Schweizer, and Litao Zhang contributed to the collection of the functional data plotted in Figure 1.

Supporting Information Available: Scatter plots of the data shown in panels A and B of Figure 4, the list of Ambit BioSciences targets and their group/family classification, hit rates for all compounds and for the SP set for all targets, separation coefficients for all physicochemical properties, neighbor hit rates for all targets, correspondence between Ambit BioSciences binding values and functional data for 37 additional assays, and pairwise activity homologies. 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) Sheinerman, F. B.; Giraud, E.; Laoui, A. High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J. Mol. Biol.* **2005**, *352*, 1134–1156.

(3) Barker, A. J.; Gibson, K. H.; Grundy, W.; Godfrey, A. A.; Barlow, J. J.; Healy, M. P.; Woodburn, J. R.; Ashton, S. E.; Curry, B. J.; Scarlett, L.; Henthorn, L.; Richards, L. Studies leading to the identification of ZD1839 (IRESSA): An orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1911–1914.

(4) Wakeling, A. E.; Guy, S. P.; Woodburn, J. R.; Ashton, S. E.; Curry, B. J.; Barker, A. J.; Gibson, K. H. ZD1839 (Iressa): An orally active inhibitor of epidermal growth factor signaling with potential for cancer therapy. *Cancer Res.* **2002**, *62*, 5749–5754.

(5) Wilhelm, S. M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; Cao, Y.; Shujath, J.; Gawlak, S.; Eveleigh, D.; Rowley, B.; Liu, L.; Adnane, L.; Lynch, M.; Auclair, D.; Taylor, I.; Gedrich, R.; Voznesensky, A.; Riedl, B.; Post, L. E.; Bollag, G.; Trail, P. A. BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res.* **2004**, *64*, 7099–7109.

(6) Lombardo, L. J.; Lee, F. Y.; Chen, P.; Norris, D.; Barrish, J. C.; Behnia, K.; Castaneda, S.; Cornelius, L. A.; Das, J.; Doweyko, A. M.; Fairchild, C.; Hunt, J. T.; Inigo, I.; Johnston, K.; Kamath, A.; Kan, D.; Klei, H.; Marathe, P.; Pang, S.; Peterson, R.; Pitt, S.; Schieven, G. L.; Schmidt, R. J.; Tokarski, J.; Wen, M. L.; Wityak, J.; Borzilleri, R. M. Discovery of N-(2-chloro-6-methyl- phenyl)-2-(6-(4-(2-hydroxyethyl)- piperazin-1-yl)-2-methylpyrimidin-4-ylamino)thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. *J. Med. Chem.* **2004**, *47*, 6658–6661.

(7) Bantscheff, M.; Eberhard, D.; Abraham, Y.; Bastuck, S.; Boesche, M.; Hobson, S.; Mathieson, T.; Perrin, J.; Raida, M.; Rau, C.; Reader, V.; Sweetman, G.; Bauer, A.; Bouwmeester, T.; Hopf, C.; Kruse, U.; Neubauer, G.; Ramsden, N.; Rick, J.; Kuster, B.; Drewes, G. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* **2007**, *25*, 1035–1044.
(8) Huang, F.; Reeves, K.; Han, X.; Fairchild, C.; Platero, S.; Wong, T. W.; Lee, F.; Shaw, P.; Clark, E. Identification of candidate molecular markers predicting sensitivity in solid tumors to dasatinib: Rationale for patient selection. *Cancer Res.* **2007**, *67*, 2226–2238.
(9) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting cancer with small molecule kinase inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28–39.
(10) Bamborough, P.; Drewry, D.; Harper, G.; Smith, G. K.; Schneider, K. Assessment of chemical coverage of kinome space and its implications for kinase drug discovery. *J. Med. Chem.* **2008**, *51*, 7898–7914.
(11) Fabian, M. A.; Biggs, W. H., III; Treiber, D. K.; Atteridge, C. E.; Azimioara, M. D.; Benedetti, M. G.; Carter, T. A.; Ciceri, P.; Edeen, P. T.; Floyd, M.; Ford, J. M.; Galvin, M.; Gerlach, J. L.; Grotzfeld, R. M.; Herrgard, S.; Insko, D. E.; Insko, M. A.; Lai, A. G.; Lelias, J. M.; Mehta, S. A.; Milanov, Z. V.; Velasco, A. M.; Wodicka, L. M.; Patel, H. K.; Zarrinkar, P. P.; Lockhart, D. J. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* **2005**, *23*, 329–336.
(12) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, 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) Schmidt-Arras, D.; Schwable, J.; Bohmer, F. D.; Serve, H. Flt3 receptor tyrosine kinase as a drug target in leukemia. *Curr. Pharm. Des.* **2004**, *10*, 1867–1883.
(14) Sawyers, C. L. Finding the next Gleevec: FLT3 targeted kinase inhibitor therapy for acute myeloid leukemia. *Cancer Cell* **2002**, *1*, 413–415.
(15) Lefevre, G.; Glotin, A. L.; Calipel, A.; Mouriaux, F.; Tran, T.; Kherrouche, Z.; Maurage, C. A.; Auclair, C.; Mascarelli, F. Roles of stem cell factor/c-Kit and effects of Glivec/STI571 in human uveal melanoma cell tumorigenesis. *J. Biol. Chem.* **2004**, *279*, 31769–31779.
(16) Dai, Y.; Guo, Y.; Frey, R. R.; Ji, Z.; Curtin, M. L.; Ahmed, A. A.; Albert, D. H.; Arnold, L.; Arries, S. S.; Barlozzari, T.; Bauch, J. L.; Bouska, J. J.; Bousquet, P. F.; Cunha, G. A.; Glaser, K. B.; Guo, J.; Li, J.; Marcotte, P. A.; Marsh, K. C.; Moskey, M. D.; Pease, L. J.; Stewart, K. D.; Stoll, V. S.; Tapang, P.; Wishart, N.; Davidsen, S. K.; Michaelides, M. R. Thienopyrimidine ureas as novel and potent multitargeted receptor tyrosine kinase inhibitors. *J. Med. Chem.* **2005**, *48*, 6066–6083.
(17) Manley, P. W.; Drueckes, P.; Fendrich, G.; Furet, P.; Liebetanz, J.; Martiny-Baron, G.; Mestan, J.; Trappe, J.; Wartmann, M.; Fabbro, D. Extended kinase profile and properties of the protein kinase inhibitor nilotinib. *Biochim. Biophys. Acta* **2010**, *1804*, 445–453.
(18) Goldstein, D. M.; Gray, N. S.; Zarrinkar, P. P. High-throughput kinase profiling as a platform for drug discovery. *Nat. Rev. Drug Discovery* **2008**, *7*, 391–397.
(19) Belema, M.; Bunker, A.; Nguyen, V. N.; Beaulieu, F.; Ouellet, C.; Qiu, Y.; Zhang, Y.; Martel, A.; Burke, J. R.; McIntyre, K. W.; Pattoli, M. A.; Daloisio, C.; Gillooly, K. M.; Clarke, W. J.; Brassil, P. J.; Zusi, F. C.; Vyas, D. M. Synthesis and structure-activity relationship of imidazo(1,2-a)thieno(3,2-e)pyrazines as IKK-β inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4284–4289.
(20) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350–4358.
(21) 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.
(22) Briem, H.; Gunther, J. Classifying "kinase inhibitor-likeness" by using machine-learning methods. *ChemBioChem* **2005**, *6*, 558–566.
(23) Garcia-Echeverria, C.; Traxler, P.; Evans, D. B. ATP site-directed competitive and irreversible inhibitors of protein kinases. *Med. Res. Rev.* **2000**, *20*, 28–57.
(24) Dumas, J. Protein kinase inhibitors: Emerging pharmacophores 1997−2000. *Expert Opin. Ther. Pat.* **2001**, *11*, 405–429.
(25) Hunt, J. T.; Mitt, T.; Borzilleri, R.; Gullo-Brown, J.; Fargnoli, J.; Fink, B.; Han, W. C.; Mortillo, S.; Vite, G.; Wautlet, B.; Wong, T.; Yu, C.; Zheng, X.; Bhide, R. Discovery of the pyrrolo[2,1-f]-[1,2,4]triazine nucleus as a new kinase inhibitor template. *J. Med. Chem.* **2004**, *47*, 4054–4059.
(26) Saxty, G.; Woodhead, S. J.; Berdini, V.; Davies, T. G.; Verdonk, M. L.; Wyatt, P. G.; Boyle, R. G.; Barford, D.; Downham, R.; Garrett, M. D.; Carr, R. A. Identification of inhibitors of protein kinase B using fragment-based lead discovery. *J. Med. Chem.* **2007**, *50*, 2293–2296.
(27) Deanda, F.; Stewart, E. L.; Reno, M. J.; Drewry, D. H. Kinase-targeted library design through the application of the PharmPrint methodology. *J. Chem. Inf. Model.* **2008**, *48*, 2395–2403.
(28) 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.
(29) Akritopoulou-Zanze, I.; Hajduk, P. J. Kinase-targeted libraries: The design and synthesis of novel, potent, and selective kinase inhibitors. *Drug Discovery Today* **2009**, *14*, 291–297.
(30) Martin, E. J.; Sullivan, D. C. Surrogate AutoShim: Predocking into a universal ensemble kinase receptor for three dimensional activity prediction, very quickly, without a crystal structure. *J. Chem. Inf. Model.* **2008**, *48*, 873–881.
(31) Gozalbes, R.; Simon, L.; Froloff, N.; Sartori, E.; Monteils, C.; Baudelle, R. Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. *J. Med. Chem.* **2008**, *51*, 3124–3132.
(32) Dongyu, S.; Claudio, C.; Zhan, D.; Scott, B.; Donovan, C.; Juswinder, S.; Patrick, C.; Gretchen, H.; Wen-Cherng, L.; Jason, D.; Jessica, F.; Serene, J. A kinase-focused compound collection: Compilation and screening strategy. *Chem. Biol. Drug Des.* **2006**, *67*, 385–394.
(33) 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.
(34) Akritopoulou-Zanze, I.; Darczak, D.; Sarris, K.; Phelan, K. M.; Huth, J. R.; Song, D.; Johnson, E. F.; Jia, Y.; Djuric, S. W. Scaffold oriented synthesis. Part 1: Design, preparation, and biological evaluation of thienopyrazoles as kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 96–99.
(35) Gracias, V.; Ji, Z.; Akritopoulou-Zanze, I.; Abad-Zapatero, C.; Huth, J. R.; Song, D.; Hajduk, P. J.; Johnson, E. F.; Glaser, K. B.; Marcotte, P. A.; Pease, L.; Soni, N. B.; Stewart, K. D.; Davidsen, S. K.; Michaelides, M. R.; Djuric, S. W. Scaffold oriented synthesis. Part 2: Design, synthesis and biological evaluation of pyrimidodiazepines as receptor tyrosine kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2691–2695.
(36) Kettle, J. G.; Ward, R. A. Toward the comprehensive systematic enumeration and synthesis of novel kinase inhibitors based on a 4-anilinoquinazoline binding mode. *J. Chem. Inf. Model.* **2010**, *50*, 525–533.
(37) Harper, G.; Pickett, S. D.; Green, D. V. Design of a compound screening collection for use in high throughput screening. *Comb. Chem. High Throughput Screening* **2004**, *7*, 63–70.
(38) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: Definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.
(39) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11*, 1046–1053.
(40) Notredame, C.; Higgins, D. G.; Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.
(41) *ClassPharmer*, version 4.5; simulationsplus, Inc.: Lancaster, CA, 2008.
(42) *Pipeline Pilot*, version 7.5; Accelrys: San Diego, 2008.
(43) *OEChem*, version 1.7.0; OpenEye Scientific Software, Inc.: Santa Fe, NM, 2008.
(44) Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204–2216.
(45) Cruciani, G.; Pastor, M.; Guba, W. VolSurf: A new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *11*, S29–S39.