Assessment of Chemical Coverage of Kinome Space and Its Implications for Kinase Drug Discovery

Paul Bamborough,*,† David Drewry,‡ Gavin Harper,† Gary K. Smith,‡ and Klaus Schneider§

Molecular Discovery Research, GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, U.K., Five Moor Drive, Research Triangle Park, North Carolina 27709, U.S.A., and New Frontiers Science Park, Harlow, Essex CM19 5AW, U.K.

Received September 4, 2008

More than 500 compounds chosen to represent kinase inhibitor space have been screened against a panel of over 200 protein kinases. Significant results include the identification of hits against new kinases including PIM1 and MPSK1, and the expansion of the inhibition profiles of several literature compounds. A detailed analysis of the data through the use of affinity fingerprints has produced findings with implications for biological target selection, the choice of tool compounds for target validation, and lead discovery and optimization. In a detailed examination of the tyrosine kinases, interesting relationships have been found between targets and compounds. Taken together, these results show how broad cross-profiling can provide important insights to assist kinase drug discovery.

Introduction

The search for inhibitors of protein kinases is an area of intense activity. A keyword search for "kinase inhibitors" returned 1281 patents filed in 2007 alone. The approval of imatinib mesylate (4-[(4-methyl-1-piperazinyl)methyl]-*N*-[4-methyl-3-[[4-(3-pyridinyl)-2-pyrimidinyl]amino]phenyl]benzamide methanesulfonate)[1] for the treatment of chronic myeloid leukemia (CML) in 2001 has been followed by several more for oncology indications and increased activity for other therapeutic areas.[2,3] One lesson from the success of imatinib mesylate, which inhibits KIT[a] and other kinases in addition to its primary target c-Abl (Abelson murine leukemia viral oncogene homologue 1), is that a kinase inhibitor does not require absolute selectivity for a single target for an acceptable safety profile. Indeed, off-target effects may have beneficial effects for CML and other indications.[4] However, understanding the inhibition profile of a compound is highly desirable if not essential.

The human genome contains an estimated 518 protein kinases.[5,6] Because of the conservation of the ATP-binding site, compounds bind to multiple kinases with greater frequency than for any other large biological system with the possible exception of aminergic G-protein-coupled receptors.[7] Even compounds that are initially reported to be selective rarely turn out to be completely selective when screened against a large enough panel. Several recent publications have reported additional kinase activities of clinical compounds and biological tools to those that were originally known.[8−11] As inhibition of some kinases becomes more firmly linked to adverse effects, it will become increasingly important to avoid inhibiting these.

Here, results are reported from screening a set of 577 diverse compounds across a panel of 203 protein kinases. The compounds were intended to represent kinase inhibitor chemical space as known to the authors at the time this experiment was carried out. This differs from other recent kinase profiling studies, which made use of clinical compounds and structures taken from the literature. Examples are given of some of the many ways in which the data generated might be useful.

For instance, we describe the discovery that a known CHK1 inhibitor is also a potent PIM1 inhibitor. Another use is to determine the inhibition profile of compounds. The profiles of several literature compounds and the implications for their selectivity are discussed.

Compound inhibition profiles have also been used to compare kinases to one another, with special attention to the tyrosine kinases, the most common targets of kinase cancer drugs. A more general analysis relates kinase similarity calculated using the activity profile to that from sequence similarity, which highlights interesting differences between kinome branches. This is relevant to several problems in drug discovery, including the tractability of individual kinase targets, their selectivity, the use of compounds as probes for kinase cellular function, and the estimation of a threshold of sequence similarity at which cross-

* To whom correspondence should be addressed. Telephone: +44 1438 745745. Fax: +44 1438 3352. E-mail: Paul.A.Bamborough@gsk.com.
† GlaxoSmithKline R&D, Stevenage, U.K.
‡ GlaxoSmithKline R&D, Triangle Park, U.S.A.
§ GlaxoSmithKline R&D, Harlow, U.K.

[a] Abbreviations: AAK1, AP2-associated kinase 1; ABL, Abelson murine leukemia viral oncogene homologue 1; AKT, v-akt murine thymoma viral oncogene homologue; ARG, Abelson murine leukemia viral oncogene homologue 1; ASK1, Apoptosis signal-regulating kinase 1; ATP, adenosine triphosphate; BTK, Bruton's tyrosine kinase; CAMK, calcium/calmodulin-dependent protein kinase; CDK, cyclin-dependent kinase; CHK1, checkpoint kinase 1; CK1, casein kinase 1; CLK, CDC-like kinase 1; CMGC, CDK MAPK GSK3 and CLK family; CSK, C-src tyrosine kinase; EPH, ephrin receptor; ERB, erythroblastic leukemia viral oncogene homologue; ERK, extracellular signal-regulated kinase; FER, Fps/fes related tyrosine kinase; FES, feline sarcoma oncogene kinase; FLT, Fms-like tyrosine kinase; FMS, colony stimulating factor 1 receptor; FRK, Fyn-related kinase; GAK, cyclin G-associated kinase; GSK3, glycogen synthase kinase; IKK, IκB kinase; IR, insulin receptor kinase; JAK, Janus tyrosine kinase; JNK, c-Jun N-terminal kinase; KHS1, kinase homologous to SPS1/STE20; KIT, Hardy−Zuckerman 4 feline sarcoma viral oncogene homologue; LCK, lymphocyte cell-specific protein−tyrosine kinase; LKB1, serine/threonine kinase 11; LOK, lymphocyte-oriented kinase; MAPK, mitogen-activated protein kinase; MELK, maternal embryonic leucine zipper kinase; MPSK1, myristoylated and palmitoylated serine/threonine-protein kinase; PAK, p21/Cdc42/Rac1-activated kinase; PDGFR, platelet-derived growth factor receptor; PIM1, proviral integration site in Moloney murine leukemia virus; RIPK2, receptor-interacting ser/thr kinase 2; SAR, structure−activity relationships; SLK1, STE20-like kinase; STE, yeast sterile kinase branch; SYK, spleen tyrosine kinase; TGFβR1, transforming growth-factor β receptor; TK, tyrosine kinase family; TNK1, tyrosine kinase nonreceptor 1; ZAP70, ζ-chain (TCR) associated protein kinase 70 kDa.

10.1021/jm8011036 CCC: $40.75
Published on Web 11/26/2008

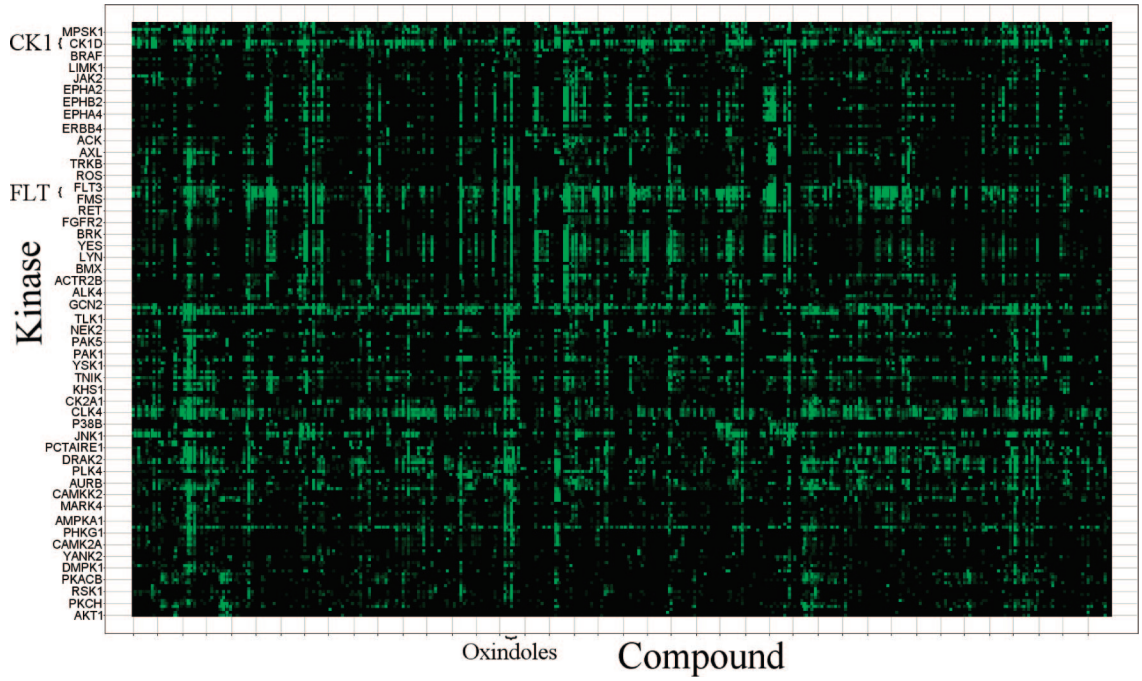

Figure 1. Heat plot showing 203 kinases (*y*-axis, ranked by position in Sugen alignment) against 577 compounds (*x*-axis, grouped by KCS class or chemotype). Bright spots indicate potent binding (low %control). Unselective compounds can be seen as bright vertical bands, for example, in the oxindole class marked. Horizontal bright bands indicate kinome sub-branches that are frequently inhibited by compounds in many chemical series, for example, FLT and CK1.

kinase activity of compounds becomes more likely, as well as how this varies over different kinome branches.

Results and Discussion

The 577 compounds intended to represent the diversity of a kinase compound collection were screened at 10 μM in a binding assay format against a panel of 203 protein kinases. The design of the compound set, the assay methodology, and the interpretation of the single-concentration data will be discussed in greater detail (see Experimental Section).

Overview of Results. Results from the 10 μM single-concentration profiling are displayed graphically in Figure 1. Compounds are grouped horizontally by their structural class (see Experimental Section). Kinases are ranked vertically, according to their order in the Sugen kinome alignment,[5] so that similar kinases are grouped together. Bright colors indicate high affinity. Bright horizontal bands show branches of similar kinases with similar activity profiles. Among these are the FLT tyrosine kinase branch, including KIT, FLT3, and PDGFRα/β, and the casein kinase 1 branch. Many compounds bind to these kinases, but few show selectivity. Similarly, bright vertical bands are also apparent. These show series of compounds that are able to bind to many different protein kinases, for example, the oxindoles. Conversely, dark bands indicate selective compounds or infrequently bound kinases. One key observation is that individual compounds tend to bind to kinases in groups colocated on the same branches of the kinome. Such kinases tend to track the activity patterns of one another. However, most compounds bind to several such small scattered groups.

Distribution of Results by Protein Kinase. For every one of the 203 kinases in the panel, at least one compound was identified that showed affinity (%control < 10; see Experimental Section for a definition). This significant achievement shows that even by use of a relatively small number of carefully chosen compounds, it is possible to find small molecule inhibitors of previously unknown kinases.

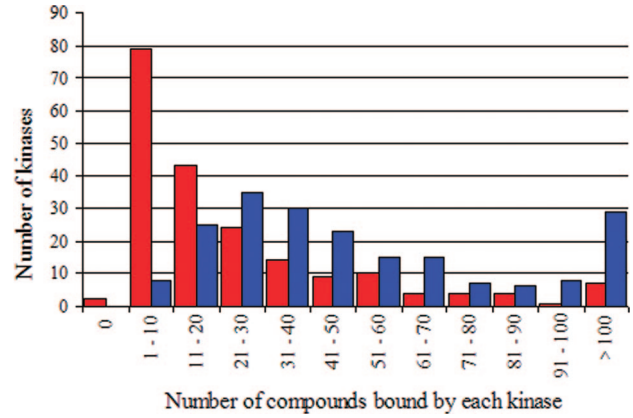

Figure 2. Number of compounds that bind to each kinase. The bar-chart shows the distribution of hit-rates per kinase, using cutoffs of 1 (red = higher affinity, %control < 1) and 10 (blue = moderate affinity, %control < 10).

Figure 2 shows the distribution of the number of compounds bound by each kinase. For example, nearly 30 kinases were able to bind over 100 compounds using a moderate binding threshold definition (%control < 10). By use of a more potent threshold (%control < 1), fewer than 10 kinases bound over 100 compounds. 60% of the kinases in the panel were bound potently by 20 compounds or fewer, but a significant number were bound by many more. In general, those bound by many compounds were also bound by many different chemotypes.

The number of compounds that bind to a given kinase (see Supporting Information, Table S1) may give an indication of its tractability as a target. Some kinases were bound potently (%control < 1) by many more compounds than others. Casein kinases CK1δ and CK1ε had high affinity for 148 and 170 compounds, respectively. GAK and AAK1 (AP2-associated kinase 1) were also bound frequently (181 and 89 compounds), as were PDGFRα/β, KIT, and FLT3 (95, 121, 117, and 80

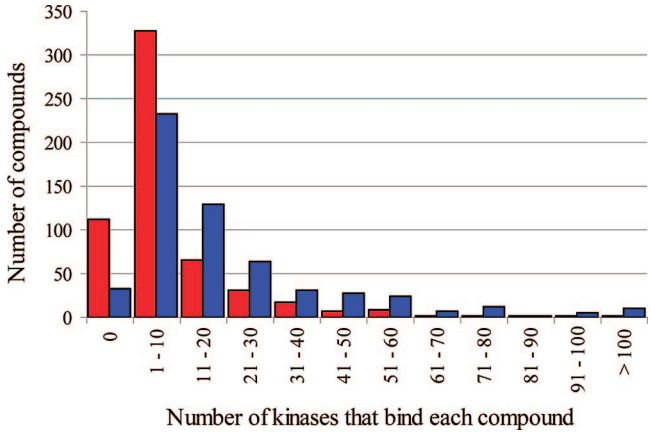

Figure 3. Number of kinases that bind each compound. The bar-chart shows the distribution of hit-rates per compound, using cutoffs of 1 (red = higher affinity, %control < 1) and 10 (blue = moderate affinity, %control < 10).

compounds). The least frequently bound kinases at the same threshold were ASK1 (*Arabidopsis* signal-regulating kinase 1) and ZAP70 (ζ-chain TCR associated protein kinase 70 kDa), which no compounds bound with %control < 1.

It would be interesting to speculate on the reasons for these widely different hit-rates. Some kinases are more forgiving, while some have more restrictive binding requirements. Caution is needed in interpreting these results, since some assays are more sensitive than others (see Experimental Section and Figure S5a). In addition, the assays used were binding assays, not activity assays, and were not conducted in the presence of ATP. The kinases, screened as phage display fusion proteins, are also not necessarily screened in their physiologically relevant state. The influence of activation state, partner proteins, and associated domains may affect the results greatly.

A recent report used thermal melting to measure binding of 156 compounds to 60 Ser/Thr kinases.[10] Noting that no potent inhibitors of the ERK family were found, the authors speculated that low hit-rates were associated with kinases present in their inactive states. In our study, ERK1 and ERK2 showed a more typical hit-rate and were bound with %control < 1 by 14 and 22 compounds, respectively, 11 of which were common to both. In the previous study, SLK and ASK1 were among the most frequently bound kinases. In the study reported here, SLK was bound by 61 compounds with %control < 1, a high hit-rate in accord with the previous results. In contrast, ASK1 had a low hit-rate and was not bound by any compounds with %control < 1. The differences between the two studies could be due to different activation states of ASK1 and the ERKs in the two assays, to the sensitivities of the different binding assays used, or to the use of fusion proteins but is perhaps more likely due to differences in the compound sets.

Despite the diversity of the compound set, the results will be influenced by the kinase targets on which GlaxoSmithKline has worked historically. However, these results still give an indication of which targets are likely to be more tractable, with a reasonable likelihood of finding hits within a diverse kinase compound collection.

Distribution of Results by Compounds. Figure 3 shows the distribution of the number of kinases bound by each compound. The number of kinases to which some inhibitors bind may be surprising. Hesperadin, the most extreme example, had high affinity (%control < 1) to 126 out of the 203 kinases tested. Sixteen compounds showed potent binding to at least 50 kinases. It is possible that even these compounds might show good windows of selectivity for certain targets if screened at lower concentration.

Table 1. Effect of Molecular Weight on Selectivity[a]

MW	N_{cpds}	N_{unsel}	% (N_{unsel}/N_{cpds})
≤300	165	10	6
>300−350	171	17	10
>350−400	117	16	14
>400−450	73	14	19
>450−500	35	7	20
>500	16	9	56

[a] Compounds were binned by molecular weight. The number of componds in each MW band, the number of unselective compounds (defined as the number of compounds binding ≥20 kinases with %control < 1), and the percentage of unselective compounds are shown.

At the other extreme, it is possible to find compounds with very good selectivity. A total of 230 compounds bound to ≤10 kinases in this panel with %control < 10, and 33 compounds did not bind to any of the kinases in this study. These are of particular interest, as they had previously shown activity against other kinases not included in this panel. Clearly it is possible to find highly selective inhibitors of at least some kinases, even using such a large kinase panel.

Influence of Molecular Weight. Molecular weight is one factor that greatly affects the number of targets to which a given compound binds. As shown in Table 1, the fraction of compounds showing low selectivity (defined as those binding ≥20 kinases with %control < 1) increases with molecular weight. One interpretation is that potent low-MW compounds are more likely to be selective than equipotent high-MW compounds. However, the higher hit-rate of larger compounds might simply result from the use of a fixed cutoff in a single compound concentration screen. Potency against a molecular target also correlates with molecular weight, so larger compounds may still maintain a window of selectivity. Nevertheless, this result shows the importance of molecular weight in interpretation of selectivity, particularly when screening is performed at a single compound concentration.

Compounds identified from cross-screening have multiple potential uses. Selective compounds may be valuable tools for target validation experiments. Even unselective inhibitors may have utility, for example, as versatile components of screening technology able to be applied to multiple targets or in X-ray crystallography to help to obtain a useful crystal system. Most importantly, hits are potential lead compounds for drug discovery. Examples of some of these will now be given.

Discovery of Hits against New Kinase Targets. One of the main benefits of this cross-screening exercise has been the discovery of new activities for existing inhibitors. Results for PIM1 will now be discussed as an example. PIM1 is a possible therapeutic target in several types of cancer.[12] The PIM family is located on the CAMK branch of the kinome tree. Twenty compounds showed %control < 1 at 10 μM. Encouragingly, these included analogues of the literature imidazopyridazine PIM1 inhibitor **1** (Figure 4, data not shown).[13] PIM1 K_d values were determined for 14 compounds from nine different chemical series. Five compounds from four series had K_d values below 100 nM. Among these was **2** (Figure 4), a member of a family of CHK1 inhibitors,[14] which bound to PIM1 with a K_d of 35 nM.

PIM1 is unusual because of the presence of a proline (Pro123) in the hinge region of the ATP site.[15] Because of this, it is unable to make the commonly conserved H-bonding interactions to inhibitors. One interesting outcome of this work was the discovery that despite the atypical binding site, many inhibitors of other protein kinases are able to bind to PIM1.

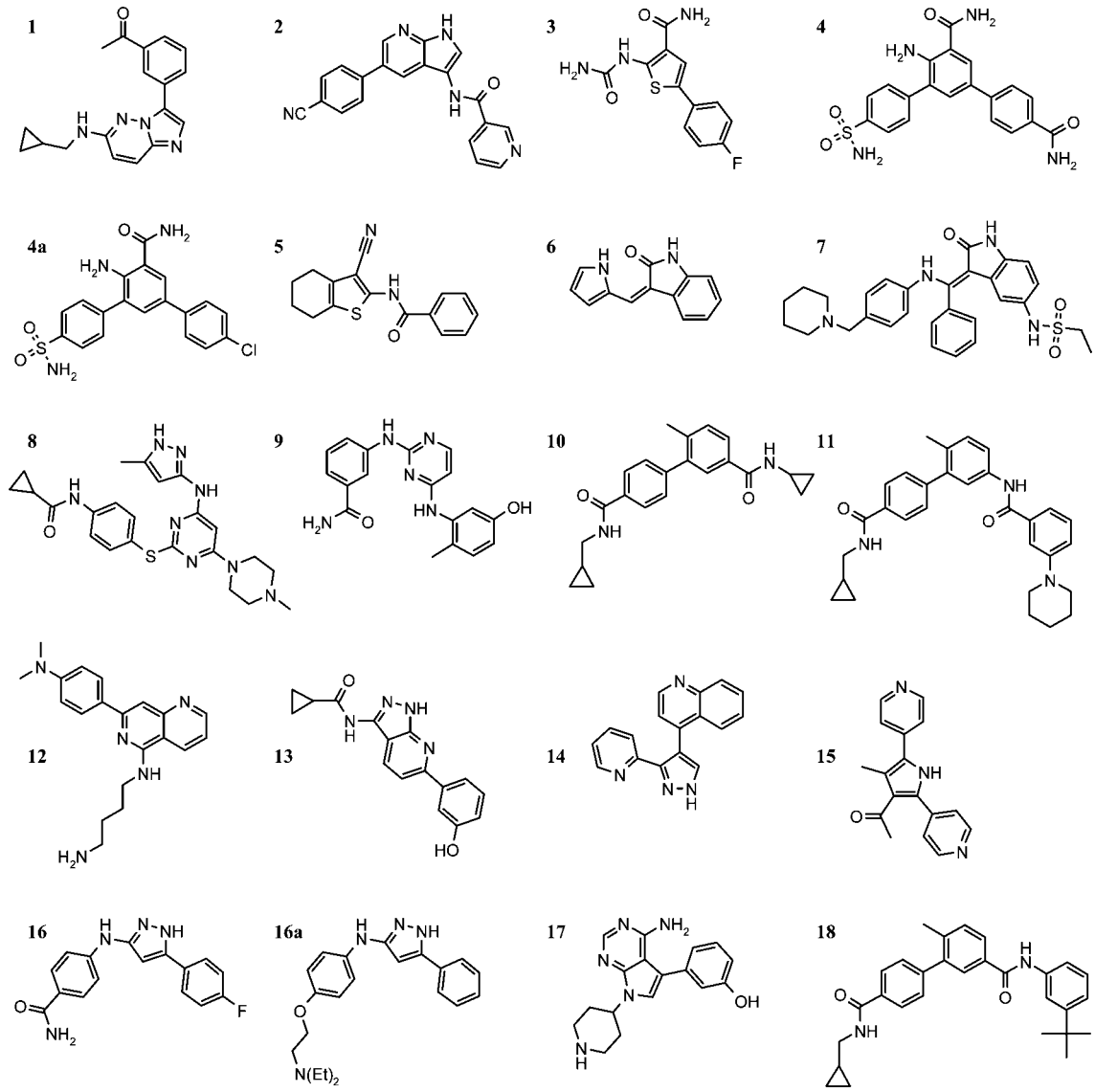

Figure 4. Structures of literature compounds **1**−**18**. Published targets include **1** PIM1,[13] **2** CHK1,[14] **3** IKK-β,[19] **4a** IKK-β,[21] **5** JNK3,[22] **6** VEGFR,[23] **7** and **8** Aurora B,[24,25] **9** LCK,[27] **10** and **11** p38α,[28,29] **12** SYK,[30] **13** GSK3β,[31] **14** ALK5,[32] **15** cytokine antagonist,[35] **16a** FLT3,[56] **17** SRC,[57] **18** p38α.[29]

Because its unusual ATP binding site gives a greater potential for compounds to adopt unusual binding modes, X-ray crystallographic binding data were especially valuable for PIM1. It was possible to solve the X-ray structures of three of the active compounds in complex with PIM1 rapidly, which was helpful in optimizing the potency and selectivity of the molecules.

Hits were also discovered against less well-studied kinases. MPSK1 (myristoylated and palmitoylated ser/thr kinase 1, also known as STK16 and PKL12) belongs to the NAK family on the STE branch of the kinome. It has been linked to TGFβ signaling and to extracellular matrix-cell adhesion.[16,17] It has been shown to phosphorylate developmentally regulated GTP binding protein 1, and a crystal structure of staurosporine bound to its kinase domain has been solved.[18] A total of 29 compounds bound to MPSK1 with high affinity (%control < 1). Six compounds from six series had K_d values below 100 nM, including the azaindole **2** (24 nM).

Expanded Kinase Profiles of Literature Compounds. Another outcome of this study is a wider knowledge of the off-target activities of published kinase inhibitors. This is important in understanding the results of target validation experiments that use these compounds. Results for 13 selected literature compounds (Figure 4) will now be summarized. Full results of screening at 10 μM are shown in Table S1 (Supporting Information), and K_d values are given in Table 2.

3 is TPCA-1, a thiophene carboxamide reported to be a selective inhibitor of IKK-β with IC_{50} of 19.5 nM.[19] Against this assay panel, which does not include any of the four IκB kinases, **3** was found to bind to 12 kinases with %control < 1. These included the second kinase domain of JAK2, KIT, cFMS, and EphB1 as well as others (Table S1). The compound bound to 41 other kinases with %control < 10. One K_d value was determined, against JAK2 (22 nM). In an assay measuring the inhibition of JAK2 catalytic activity, the K_i was determined to be 5 nM (unpublished results). This compound has been described as a selective IKK-β inhibitor and used to study the role of this kinase in cells[20] but would perhaps be better described as a dual IKK-β/JAK inhibitor with potential other kinase activities.

The phenylcarboxamides are a series structurally related to the thiophene carboxamides. The IKK-β IC_{50} of compound **4** is 650 nM, and that of the related compound **4a** is 100 nM.[21] The selectivity profile of **4** in this panel is cleaner than that of **3**, with only six kinases binding with %control < 10 (Table

Table 2. K_d Values (nM) Measured for Literature Compounds **2**−**18**[a]

compd	kinase	K_d	compd	kinase	K_d
2	PIM1	35	**9**	TYRO3	264
2	LKB1	380	**9**	EPHA3	0.1
2	TNIK	160	**9**	EPHA5	3
2	MST2	75	**9**	EPHA4	5
2	NEK2	483	**9**	EPHA6	20
2	AAK1	53	**9**	EPHB2	7
2	AXL	596	**9**	EPHB3	81
2	MPSK1	24	**9**	EPHA2	8
3	JAK2	22	**9**	EPHA1	5
4	P38G	79	**9**	ADCK4	129
5	AKT3	>40000	**12**	SLK	395
6	DRAK2	353	**12**	TXK	623
6	SRPK1	361	**13**	MARK1	>40000
7	MSK1	31	**13**	PIM2	207
7	MSK2	94	**13**	SKMLCK	16
7	AMPKA1	19	**13**	MPSK1	150
7	BRSK2	92	**14**	ARG	1870
7	KHS2	5	**15**	CLK2	10
7	PAK2	14000	**15**	CK1a2	225
7	FES	149	**15**	CK1d	58
7	LTK	52	**15**	CK1g1	59
7	ROS	77	**16**	MNK2	1230
7	FAK	33	**16**	MARK1	11000
7	PYK2	10	**16**	DRAK1	96
7	TYK2	75	**16**	NEK7	>40000
8	KHS1	53	**16**	GAK	367
8	INSR	506	**16**	KIT	111
8	IRR	276	**16**	PDGFRB	42
8	TNK1	85	**16**	FLT3	30
9	ACTR2B	1	**16**	LIMK2	>40000
9	BMX	36	**17**	TXK	80
9	TXK	10	**17**	FGR	7
9	BLK	1	**17**	FRK	2
9	YES	1	**17**	BRK	10
9	FYN	0.4	**17**	EPHA1	6
9	FGR	0.3	**18**	FRK	46
9	BRK	2	**18**	EPHA3	122
9	CSK	28			

[a] Structures are shown in Figure 4.

S1). Of these, binding to CSNK1ε and GAK (cyclin G-associated kinase) was also seen with compound **3**. One K_d was determined, against p38γ (79 nM).

5 is an example of a cyanothiophene series of JNK3 inhibitors with selectivity over other JNK isoforms.[22] **5** itself has a reported JNK3 IC_{50} of 1.5 μM. The X-ray structure of this compound shows a possibly unique binding mode in which the nitrile forms a hydrogen-bond to the hinge region. The only kinases in this panel showing %control < 30 were JNK3 and AKT3 (Table S1; the AKT3 result was a false positive, as the K_d was found to be >40 μM; see Experimental Section for a discussion of the frequency of these results). No binding was detected against JNK1 and only a marginal effect was seen on JNK2, confirming the report of the selectivity of this compound in a different assay format.

Originally reported as a 390 nM VEGFR inhibitor,[23] **6** is a member of the oxindole series. It bound to 23 targets with %control < 1, including VEGFR2 (Table S1), and a further 25 kinases with %control < 10, confirming its low selectivity. K_d values were obtained for two targets, DRAK2, also known as STK17B (350 nM), and SRPK1 (360 nM). Other workers have pointed out the broad inhibition profile of related oxindoles.[10]

7 is Hesperadin, another oxindole and a literature inhibitor of cell proliferation and of Aurora B among other kinases.[24] Defects in mitosis caused by this compound were attributed to AurB inhibition, a conclusion supported by RNA*i* studies. It is possible that this compound may still show a window of selectivity for AurB, but in this assay panel it was the least selective compound in terms of the number of kinases inhibited at 10 μM (126 with %control < 1, Table S1). These included AurA, AurB, and AurC. K_d values were obtained against 12 targets, one of which, PAK2, was inactive (a false positive in the single-concentration screening; see Experimental Section for a discussion of the frequency of these). The rest had K_d around or below 100 nM, as shown in Table 2.

VX680, compound **8**, is an Aurora kinase inhibitor with K_i(app) of 0.6 nM (AurA), 18 nM (AurB), and 4.6 nM (AurC).[25] It was reportedly 100-fold selective over 55 kinases, with one exception being FLT3 (30 nM). The compound binds to many other kinases in this panel (including the Aurora kinases, there were 32 kinases with %control < 1 and 37 more with %control < 10). Four K_d values were determined (Table 2) including KHS1, also known as MAP4K5 (53 nM), and TNK1 (85 nM). These values compare well with those reported independently by Ambit Biosciences (83 nM for both).[9,11] Another group has studied this compound and report that it inhibits MELK and SRC with 10-fold lower affinity than Aurora.[8] A clinical trial of the compound was recently suspended after QTc prolongation was found in one patient;[26] the extent to which off-target kinase activities may have caused this is unclear.

9 is a bis-anilinopyrimidine inhibitor of tyrosine kinases including LCK, BTK, LYN, SYK, and TXK.[27] In this panel it inhibited 48 kinases with %control < 1 (33 of them tyrosine kinases, almost half of the 71 tyrosine kinases in the panel). A further 27 kinases were bound with %control < 10. K_d values for 16 kinases were determined and found to be below 100 nM (Table 2). These included TXK (10 nM).

Biphenylamides such as **10** and **11** are inhibitors of p38α/β. While **10** binds to p38 in its apo-like DFG-in conformation, **11** utilizes the DFG-out binding mode.[28,29] **10** inhibited no kinases in the panel apart from p38α and p38β. While still quite selective, **11** bound to PDGFRα and PDGFRβ with %control < 1 and to KIT, LOK, RAF1, and p38γ with %control < 10. This is a counterexample to the view that DFG-out inhibitors are more selective than DFG-in compounds.

12, a [1,6]naphthyridine, is a SYK inhibitor with IC_{50} of 34 nM.[30] To our knowledge no selectivity data have been published for this series. In this panel, the compound appears relatively selective. It bound to SYK with %control of 0.3, as well as to 13 other kinases with %control < 1 (Table S1). A further 27 kinases had %control < 10. The K_d for SLK and TXK were determined to be 365 and 623 nM.

The pyrazolo[3,4-*b*]pyridine **13** is an 8 nM inhibitor of GSK3β.[31] While GSK3β was not a member of the panel, the compound bound to GSK3α with %control of 0 (indicating potent binding). Otherwise, it was relatively selective, binding with high affinity (%control < 1) to seven other kinases including CDK2 (which was identified in the original publication) and PCTAIRE1. Twenty-one additional kinases were bound with %control < 10 (Table S1). Three potent K_d values were measured: skeletal muscle MLCK (16 nM), MPSK1 (150 nM), and PIM2 (207 nM) (Table 2).

14 is a 27 nM inhibitor of TGFβ receptor kinase 1, also known as ALK5.[32] It appears to be highly selective. As well as TGFβR1, it bound to only three kinases with %control < 1 (GAK, CK1δ, and CK1ε). A further five kinases bound **14** with %control < 10 (Table S1). The K_d value for one of these (ARG, also known as Abl2) was determined and found to be quite weak, 1.9 μM. The high selectivity of this compound suggests that TGFβR1 may indeed be the molecular target responsible for effects seen in mouse rheumatoid arthritis[33] and tumor metastasis[34] models.

15 is included in patent claims for interleukin and TNF antagonists, with members of the series active in cellular assays.[35] The activity of **15** itself was not disclosed. The compound bound six kinases in the panel with %control < 1 (CLK1/2 and CK1 isoforms α2, δ, ε, and γ1). K_d values were obtained against four kinases: CLK2 (10 nM), CK1α2 (225 nM), CK1γ1 (60 nM), and CK1δ (70 nM). Only 10 other kinases were inhibited with %control < 10 (Table S1). This result is of potential interest, as further confirmation of any relationship between TNFα modulation and these targets could be sought using unrelated inhibitors with overlapping kinase activities.

In conclusion, the detailed profiles of these literature compounds illustrate the wide-ranging selectivities of these molecules. Some are indeed selective, although screening against this broad panel has revealed many additional activities. Such detailed profiles give useful information about the utility of compounds as tools and can suggest potential causes of any observed off-target effects.

Comparing Compounds Using Their Kinase Activity Fingerprints. Molecular similarity between pairs of compounds is a familiar concept in cheminformatics.[36] This is normally calculated by comparing bit-strings of molecular descriptors such as the presence of functional groups or patterns of connected atoms. Another way to quantify the similarity between compounds is by comparing their biological profiles, sometimes called affinity fingerprints. This has been done previously with diverse sets of compounds and with targets spanning multiple target classes[7,37,38] but not with a focus on kinase inhibitors. The "compound SAR similarities" between all pairs of compounds (a 577 × 577 matrix) were calculated from their kinase activity fingerprints using two thresholds for activity, Tan1 and Tan10 (see Experimental Section for definitions). The aim was to compare these matrices to one obtained using chemical similarity. The two metrics are plotted against one another in Figure S1 (Supporting Information).

Since one of the objectives when choosing the set was to maximize the coverage of chemical diversity, there are few pairs of compounds with high structural similarity. Over 99.5% of pairs have structural similarity of <0.6. The few compounds with high structural similarity are less alike than their similarity score implies. For example, the few pairs with similarity of >0.8 are typically enantiomers, or regioisomers, identical apart from the positions of large substituents. These would project into different parts of the ATP site, so it is not surprising that such compounds have different activity profiles. This is the case for compounds plotted in the lower right-hand corner of Figure S1.

Even with allowance for the low number of truly structurally similar compounds, the data show a trend for these to have similar kinase profiles. Because of the large number of points in Figure S1, the trend is more obvious when plotted in bar-chart form (Figure 5). Almost 10% of compounds with high (>0.6) structural similarity have high (>0.4) Tan10 compound SAR similarity. Compound pairs with lower structural similarity have a much lower chance of having a similar compound SAR profile. Some structurally unrelated inhibitors have high compound SAR similarity, but these are very uncommon. Only 0.7% of pairs with structural similarity of <0.4 show compound SAR similarity of >0.4. Expanding the data set to include more structurally related analogues would make this result more obvious.

These results provide some quantification of a trend that would be expected to hold true for other diverse compound sets and that has an important consequence. Two structurally dissimilar kinase inhibitors that have the same primary target will probably inhibit a different spectrum of off-target kinases. This observation confirms that it is important not to rely on a single compound to elucidate the role of a kinase in cells.[39,40] The complete selectivity profile of tool compounds is rarely known, and so it is difficult to say with confidence that a cellular effect is due to a single kinase. The more tool compounds sharing a common target that show the same effect, the more likely that this is due to that target, provided that the compounds are structurally dissimilar and not too unselective.

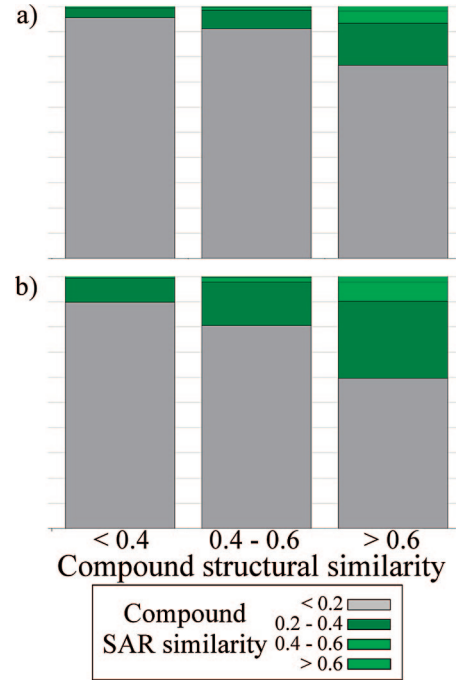

Figure 5. Normalized bar-charts showing the increasing compound SAR similarity with compound structural similarity. Two different SAR similarity thresholds are shown: (a) %control = 1; (b) %control = 10.

This concludes our chemocentric discussion of the data. We now turn to an analysis of the biological targets, first a general overview and then a closer inspection of the tyrosine kinases. This is an important kinome branch for drug discovery, containing as it does the targets for many of the marketed kinase oncology drugs.

Comparing Kinases Using Their Compound Activity Fingerprints. The matrix of kinase/compound affinities can be used to compare pairs of kinases as well as to compare pairs of compounds. This concept has been termed SAR homology (SARAH)[41] and has been investigated using kinase-focused sets[42,43] and with diverse sets of compounds and target classes.[7,37,38] The conclusions of these experiments can now be tested using a much larger set of compounds and kinases and a complete data matrix.

Kinase SAR similarities between all pairs of kinases were computed from their chemical fingerprints (see Experimental Section). The resulting 203 × 203 kinase SAR similarity matrix is shown in Figure 6. As before, kinases are ranked by their position in the Sugen kinome alignment.[5] Higher similarity between pairs of kinases is indicated with brighter green spots. Figure 6a shows the sequence similarity using the percentage sequence identity over the amino acids of the kinase domain. Squares close to the diagonal represent pairs of kinases that are more closely related to one another by sequence and for which similar compound profiles would be expected. Blocks of closely related kinases correspond to distinct branches of the kinome

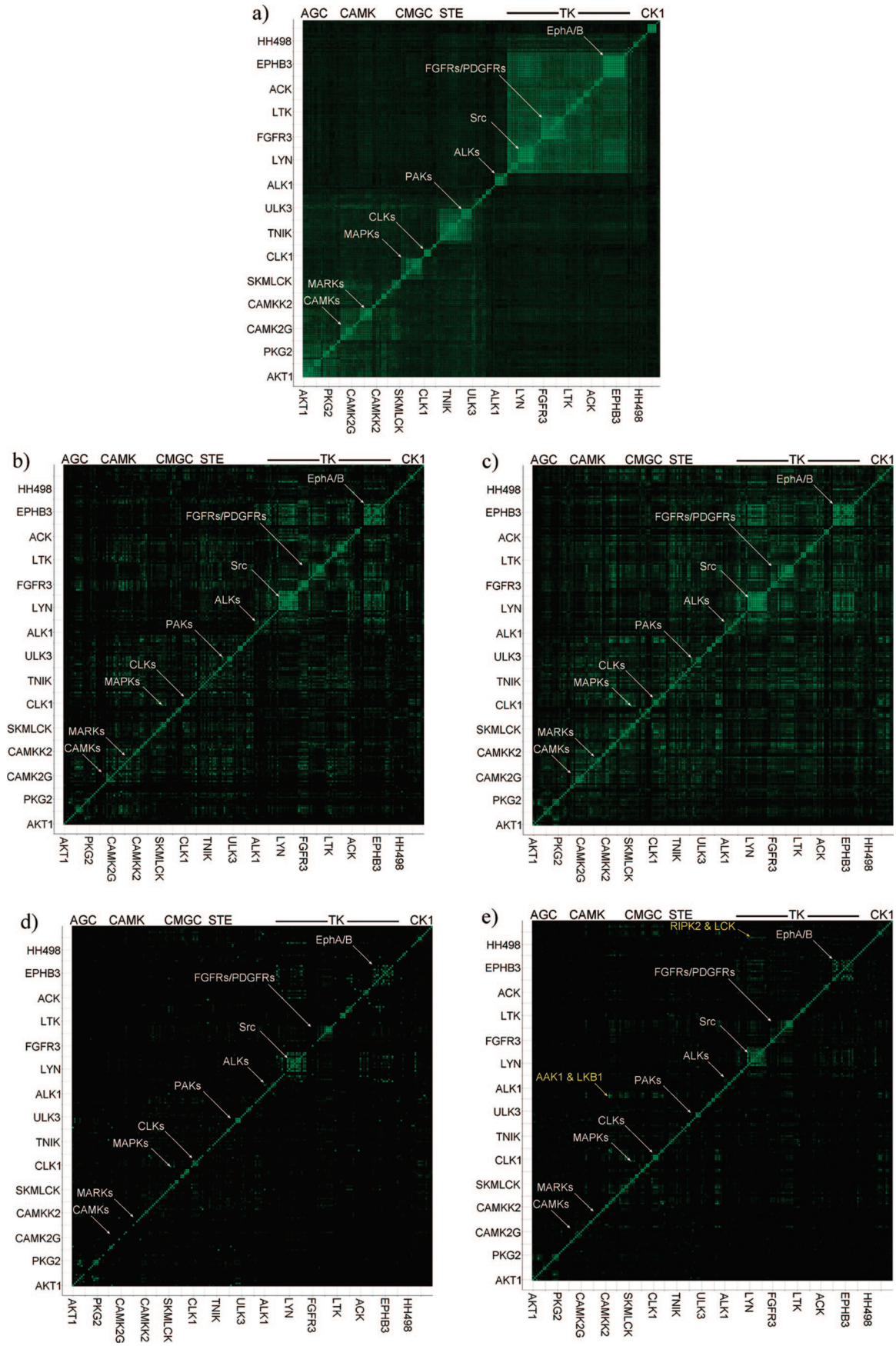

Figure 6. Similarity between pairs of kinases, ranked by their position in the Sugen tree. The color represents similarity (black = 0, green = 1). Major kinome branches are marked above the plots. Significant sub-branches are indicated with white labels: (a) amino acid sequence identity over the kinase domain; (b) kinase SAR similarity, Tan1 metric (see Experimental Section); (c) kinase SAR similarity, Tan10 metric; (d) kinase SAR similarity, Tan1 metric, including only relatively selective compounds inhibiting less than 20 kinases; (e) kinase SAR similarity, Tan10 metric, including only compounds inhibiting less than 20 kinases.

Table 3. Kinase SAR Similarities between Pairs of Kinases Referred to in the Text

						kinase 1[e]		kinase 2[e]	
kinase 1	kinase 2	%ID[a]	Tan1[b]	Tan10[c]	Tan10-u[d]	$n < 1$	$n < 10$	$n < 1$	$n < 10$
LKB1	AAK1	25	0.45	0.58	0.49	70	127	85	186
RIPK2	LCK	26	0.46	0.45	0.38	34	80	45	103
RIPK2	P38A	24	0.17	0.18	0.12	34	80	15	34
CDK2	GSK3A	37	0.38	0.40	0.40	40	71	44	86
P38A	P38B	75	0.40	0.63	0.63	15	34	27	41
P38A	P38G	67	0.00	0.14	0.10	15	34	1	22
P38B	P38G	65	0.00	0.17	0.13	27	41	1	22
CSK	BTK	45	0.15	0.38	0.25	26	37	4	17
CSK	ABL1	47	0.28	0.31	0.07	26	37	11	44
FLT3	KIT	54	0.46	0.57	0.47	79	200	115	180
FLT3	FLT1	49	0.39	0.42	0.28	79	200	50	110
FLT3	CSF1R	55	0.36	0.42	0.30	79	200	56	122
FLT3	PDGFRA	54	0.37	0.50	0.41	79	200	93	148
FLT3	PDGFRB	49	0.40	0.60	0.51	79	200	119	186
EPHA1	FRK	42	0.70	0.48	0.30	15	56	19	49
EPHA1	SYK	38	0.05	0.21	0.20	15	56	6	41
average (203 × 203)		28	0.08	0.13	0.05	25	57	25	57

[a] The %identity of the two sequences over the kinase domain. [b] The kinase SAR similarity scores using Tan1 metrics (see Experimental Section). [c] The kinase SAR similarity scores using Tan10 metrics (see Experimental Section). [d] The Tan10 metric omitting unselective compounds. [e] The number of compounds binding to each kinase using %control thresholds of 1 and 10. Average values for all pairs of kinases are shown. A full matrix of 203 × 203 kinases is in Supporting Information.

(which are labeled above the matrix plots). Some of the especially highly conserved sub-branches are indicated with white labels. This view can be compared to the corresponding matrices colored by kinase SAR similarity scores derived from their affinity profiles. Two different activity cut-offs are shown.

The most important observation is that many of the blocks of high-similarity kinases in Figure 6a are still visible in Figure 6b. These features are also present when the similarities are calculated with a weaker activity threshold as shown in Figure 6c. These plots show that despite the underlying complexity of the data, when it is viewed in the format of Figure 1, there is a statistical tendency for kinases with higher sequence similarity to bind to similar compounds. These trends are especially clear for sub-branches such as the SRC and EphA/B families of tyrosine kinases.

Figure 6b and Figure 6c contain some bright spots far from the diagonal. These indicate targets from different branches of the kinome that have high kinase SAR similarity (pairs of kinases sharing a similar inhibition profile that is unexpected from sequence homology). These are potentially interesting results. However, it is difficult to judge the significance of most of these matches without K_d data to confirm the results, so they should be regarded as tentative.

It is to be expected that the results of this exercise will be compound-dependent. A few unselective compounds might account for the kinase pairs with low sequence similarity that show high SAR similarity. To investigate this, the 73 least selective compounds (defined as those inhibiting ≥20 kinases with %control < 1) were removed and the analysis was repeated. Many of the remote-homology high kinase SAR similarity pairs disappeared (Figure 6d,e).

Individual compounds often inhibit kinases that are far apart in the kinome tree. However, this is much more common for unselective pan-kinase inhibitors. When these are neglected, the kinase SAR similarity follows patterns similar to those that would be predicted from a simple sequence comparison of the kinase domains, where compounds bind to closely related kinases more often than to unrelated ones.

Kinases Sharing Unexpectedly High SAR Similarity. Not all off-diagonal relationships disappear when unselective compounds are removed. One such relationship is between LKB1 and AAK1 (Figure 6e). The Tan10 kinase SAR similarity between these two assays is 0.58 (Table 3). The LKB1 tumor suppressor,[44] also known as STK11, is associated with Peutz−Jeghers syndrome and is located on the CAMK branch of the kinome close to CHK1. AAK1 (AP2-associated kinase 1)[45] is located away from the main kinome branches and has not been strongly linked to disease. Both kinases bound many compounds with high potency: LKB1 bound 70 compounds with %control < 1 and AAK1 bound 85. Forty-eight compounds were common to both of which 19 were fairly selective (binding less than 20 kinases with %control < 1). Five out of six compounds whose K_d values were measured against both AAK1 and LKB1 showed $K_d < 1$ μM against both kinases. For example, compound **2** (Figure 4) had K_d of 380 nM against LKB1 and 53 nM against AAK1 (Table 2).

The significance of this unexpected similarity in inhibition profile is unclear. AAK1 and LKB1 share 25% of residues over their kinase domains. Within the 30 residues lining the ATP-binding site, 42% of residues are identical, but kinase ATP-site conservation is usually higher than domain conservation, and this is is not an unusually high value. It is possible that these compounds depend strongly on individual residues for their interactions with the ATP sites of LKB1 and AAK1 which are lost in the averaging of sequence similarity calculations, although those that are conserved do not appear to be in any way unusual. Indeed, LKB1 appears to have quite a typical kinase ATP site, while AAK1 is unusual in lacking some of the conserved glycine residues in the glycine-loop rich region of the ATP-site (Ala53, Ala58: Figure S2), so this result is difficult to explain. However, our data suggest that any effort directed at targeting AAK1 or LKB1 should monitor selectivity against the other kinase.

Another unexpected relationship was seen between RIPK2 and a cluster of homologous Src-family kinases related to LCK. RIPK2, receptor-interacting ser/thr kinase 2, is also known as RICK or CARDIAK.[46−48] It has been associated with innate and acquired immunity,[49] although the requirement for kinase activity for these functions is unclear. Despite sharing only 26% kinase-domain sequence identity, RIPK2 and LCK have Tan10 kinase SAR similarity of 0.45 (Table 3, Figure 6e). Thirty-four compounds bound to RIPK2 with %control < 1, and 45 compounds bound to LCK. Twenty-five of these compounds are common to both targets, and 8 are fairly selective (binding

less than 20 kinases with %control < 1). Of the 25 compounds active against both kinases, three were submitted for RIPK2 K_d determination and binding was confirmed with K_d < 200 nM. Two of the three had previously shown IC_{50} < 200 nM against LCK in a different binding assay format (data not shown).

The p38 kinase inhibitor SB-203580 has been reported to inhibit RIPK2.[11,50,51] This compound was not included in this study, but a number of other p38α inhibitors were. The Tan10 kinase SAR similarity between p38α and RIPK2 was low, 0.18 (Table 3). This implies that in spite of the result for SB-203580, p38α kinase inhibitors are no more likely to inhibit RIPK2 than any other kinase inhibitor.

It is possible that pairs of kinases with unexpectedly high SAR similarities share unusual conformational features that are not obvious from their sequences that are responsible for the similar binding profiles. Such conformational differences might be intrinsic to the sequence, perhaps due to N- or C-terminal extensions to the kinase domain. They might be induced by particular phosphorylation states of the two kinases, or non-native due to the use of a T7 phage fusion protein. Alternatively, when 203 × 203 kinase pairs are compared, some of which have high hit-rates, a certain number of similar patterns would be expected by chance alone. Results such as these deserve further investigation, perhaps by measuring catalytic activity of the full-length kinase sequence against full-length substrate.

Comparing Kinase SAR Similarity to Phylogenetic Similarity. Another way to represent the data in Figure 6 is to plot the kinase SAR similarity derived from activity fingerprints against the amino acid sequence identity over the kinase domain (Figure 7, Figure S3).

Whichever kinase SAR-similarity metric is used, the conclusion is the same. There is no linear relationship between kinase sequence similarity and SAR similarity. However, there is a difference between pairs of kinases with below 40−50% sequence identity and those with above that threshold. When two kinases share over 40−50% sequence identity in their kinase domains, it is likely that they will show reasonable SAR similarity and that compounds are likely to behave similarly against them. More distantly related kinases with <40% sequence identity are much less likely to have significant SAR similarity. Rarely, some distantly related kinases show high SAR similarity and some closely related kinases show low SAR similarity. These uncommon results become less frequent when unselective compounds are removed from the analysis (Figure 7b). These results accord well with an analysis of sparse IC_{50} literature data over a smaller number of kinases and compounds.[42] This investigation used a different SAR-similarity metric but gave rise to a similar diagram that appeared to show the same change in behavior at 40−50% sequence identity. This study considered more complex sequence metrics, including the use of an amino acid similarity matrix over the kinase domain and over the residues of the ATP binding site, but concluded that these correlate closely with the simple %identity over the kinase domain and so would not alter the outcome.

A consequence of these results is that it is relatively likely that compounds that inhibit one kinase will show activity against others from the same branch, provided that these kinases are related by over 40−50% sequence identity. Outside these tight kinome sub-branches, less closely related kinases might also be inhibited by the same compound but with lower probability. A given compound can inhibit distantly related targets across diverse parts of the kinome, as can be seen by examining Figure 1. However, this occurs less frequently, and it is not readily predictable which targets those will be.

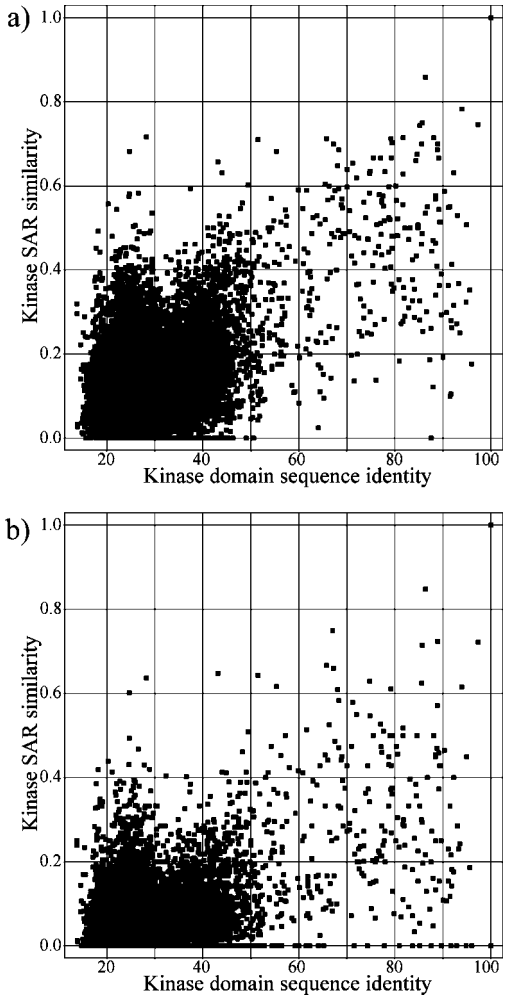

Figure 7. Kinase SAR similarity plotted against kinase sequence similarity (amino acid %identity over kinase domain): (a) Tan10 metric (see Experimental Section); (b) Tan10 metric, with unselective compounds removed (those inhibiting 20 or more kinases with %control < 1). Similar plots are shown for the Tan1 metric in Figure S3 (Supporting Information).

Differences within Kinome Branches. A closer investigation reveals that this 40−50% rule does not apply equally to all branches of the kinome. For example, in Figure 6b−e the tyrosine kinase sub-branches are clearly visible as areas within which kinases have high SAR-similarity. In contrast, in the MAPK region of the CMGC branch, kinase SAR similarity is less apparent than the structural similarity would suggest (note the presence of a bright box in this region in Figure 6a and its absence in Figure 6b−e). To investigate this interesting result, pairs of kinases within the same branch of the kinome were considered separately for five branches (AGC, CAMK, CMGC, STE, and TK) as shown in Figure 8. For each pair of kinases in a branch, the kinase domain sequence identity and SAR similarity were partitioned into bands (<30%, 30−40%, etc.). The pie charts in Figure 8a illustrate, for pairs of kinases tested that lie within each branch of the kinome, the way that the kinase SAR similarity increases with increasing domain sequence identity. While some parts of this plot are more sparsely populated than others, it appears that different branches behave differently. For example, pairs of kinases on the AGC and CMGC branches that share 40−50% sequence identity (red boxes) have lower SAR similarity than pairs with the same sequence identity on other branches.

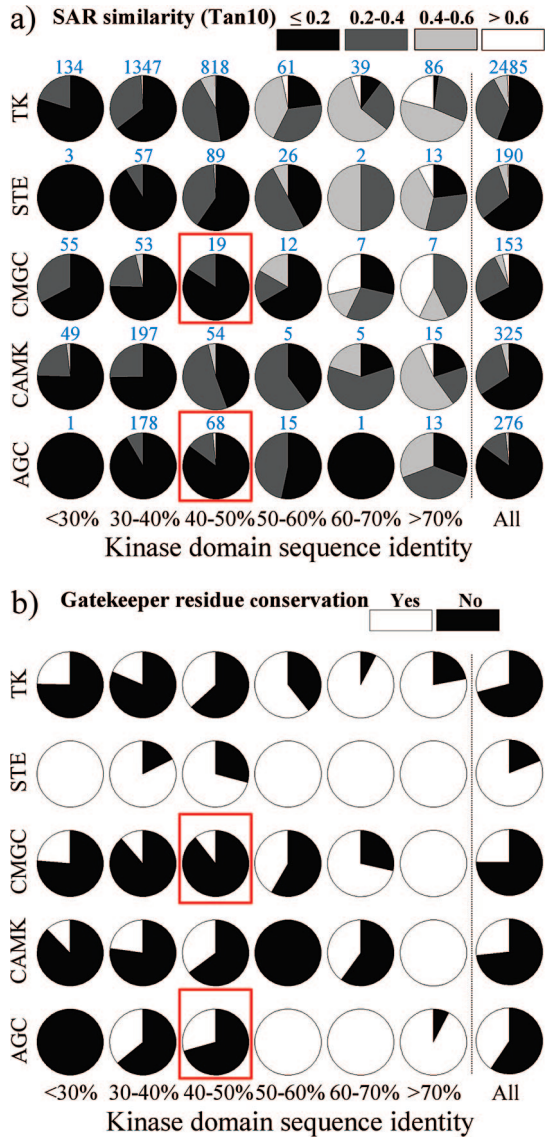

Figure 8. Dependence of kinase SAR similarity on sequence identity for different kinome branches. Each pie-chart represents pairs of protein kinases on each branch of the kinome within given ranges of kinase domain sequence identity. (a) The pie segments represent the proportion of pairs of kinases with given SAR similarity. The numbers above each pie indicate the number of pairs of kinases. (b) The conservation of gatekeeper residues between pairs of kinases. Pairs with different gatekeeper residues are colored black.

It is likely that these differences between branches are caused by subtle differences within the ATP-site that are not captured in the kinase domain sequence identity scores, perhaps because of single amino acid changes. For example, the "gatekeeper" residue is important for inhibitor selectivity, as previously noted.[42]

To see the extent to which conservation of the gatekeeper residue may influence the above observations, each pair of kinases was flagged according to whether the gatekeepers are the same or different. Figure 8b shows how the proportion of pairs of kinases tested where the gatekeeper residue is invariant increases with increasing sequence identity. Unsurprisingly, this proportion increases with increasing kinase domain % identity, but the rate at which it changes is different for different branches. Overall, the variability of the gatekeeper residue within the CMGC branch is similar to that within the TK branch: just under 30% of all pairs of kinases within both branches for which experimental results were measured have the same gatekeeper residue. This is far lower than the 80% conserved within the STE branch. However, for pairs of kinases screened that are in the critical region where kinase SAR similarity transitions from low to high (40−50% sequence identity) gatekeeper residues are conserved less frequently in the CMGC branch than in any other. The lower than expected SAR similarity in this branch in the 40−50% sequence identity band could be accounted for by compounds that exploit the different gatekeeper residues. This is consistent with the literature for inhibitors of MAP kinases on the CMGC branch where gatekeeper residue size differences have been exploited to gain selectivity.[52]

As in the CMGC branch, pairs of kinases in the AGC branch also have lower SAR similarity in the 40−50% identity range than other branches (Figure 8a). The AGC kinases conserve their gatekeeper residues more often, as frequently as TK and CAMK kinases (Figure 8b). In contrast to the CMGC kinases, variation in the gatekeeper residue is unlikely to explain the greater selectivity within the AGC family. Some other explanation not apparent from sequence identity must account for this. Perhaps in the AGC family ATP-site residues other than the gatekeeper can be utilized to gain selectivity. Alternatively, compounds might exploit conformational differences between different AGC kinases that are not reflected in kinase domain similarity scores.

One final observation is that the SAR similarity between kinases with a given sequence similarity is higher in the tyrosine kinase branch than in other branches (Figure 8a). This suggests that it may be harder to develop selective inhibitors of kinases on the TK branch than others. Although it may be more difficult to find selective inhibitors of the tyrosine kinases, it is still possible, and exceptionally selective inhibitors are known. Lapatinib (*N*-(3-chloro-4-{[(3-fluorophenyl)methyl]oxy}phenyl)-6-[5-({[2-(methylsulfonyl)ethyl]amino}methyl)-2-furanyl]-4-quinazolinamine bis(4-methylbenzenesulfonate) monohydrate) was not present in this compound set but has been reported to bind to only 3 out of 290 kinases, and only to EGFR with submicromolar K_d, in a panel that did not include its other known target ErbB2.[11] This compound takes advantage of an unusual C-helix-out conformation of the EGFR kinase domain and has a slow dissociation rate that may contribute to its selectivity.[53]

Kinase SAR Similarity Tree. Trees have been used to group kinases into families using similarity matrices derived from SAR data, by analogy to evolutionary distance for phylogenetic trees.[42,43] While this involves loss of information compared to a matrix plot such as that shown in Figure 6, it is easier to view.

The tree generated from the compound binding data is shown in Figure 9, alongside the sequence-based tree (see Experimental Section). In general, the conclusions reported for a smaller data set holds true.[42] Kinases that cluster closely in the sequence similarity tree also cluster in the SAR similarity tree. For example, the ephrin family, the PDGFR family, the SRC family, and others in the TK branch are grouped together, as are most of the AGC kinases and the CMGC kinases. There are differences in the linkages between less closely related groups. However, these are the linkages that are of lower certainty in the sequence-based tree, based upon more marginal differences in sequence homology.

Other differences are more significant. GSK3α is associated more closely with the cyclin-dependent kinase sub-branch than its location on the phylogenetic tree would suggest. This has previously been noted for GSK3β.[42] A second example is the location of isoforms of p38 (α, β, and γ). They are closely

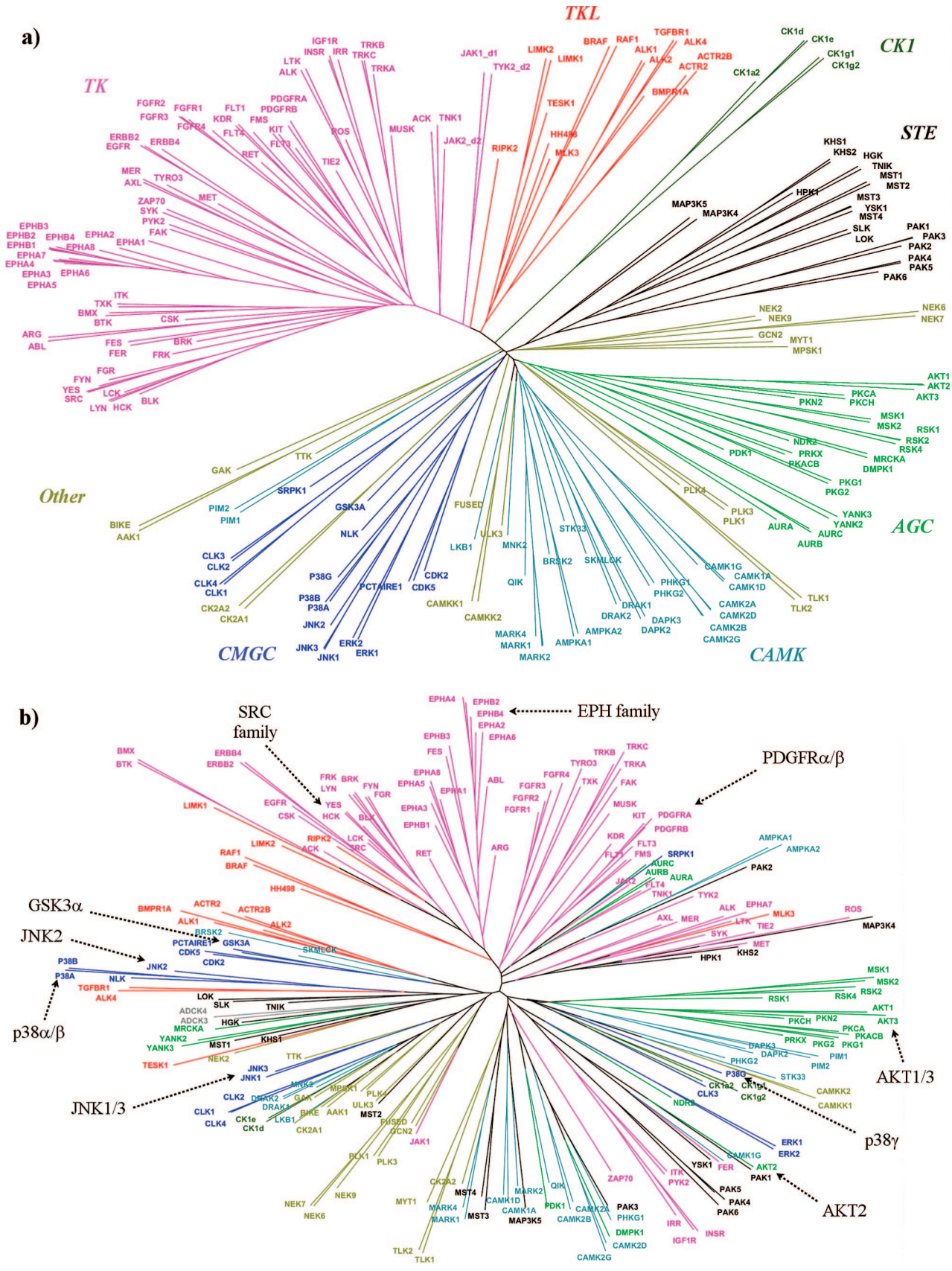

Figure 9. (a) Kinome phylogenetic tree, reconstructed using kinase domain amino acid %identity from the Sugen alignment.[5] (b) Kinase tree constructed from SAR data (see Experimental Section). Colors represent the kinome branch.

homologous over their domains (α and β share 75% identity, α and γ 67%, β and γ 65%). Nevertheless, the SAR tree places them in different locations: p38α and p38β appear alongside each other, far from p38γ. This is as expected because of the different gatekeeper residues (threonine in the α and β isoforms, methionine in γ and δ) that are known to influence inhibitor binding.[52] Kinase SAR similarity metrics for these pairs of targets are shown in Table 3.

Some other differences between the positions of kinases on the SAR and sequence trees are harder to rationalize. For example, JNK1 and JNK3 are grouped together while JNK2 is found close to p38α/β. These results differ from data from activity assays where SAR from the JNKs tends to correlate well (unpublished results). Tan10 hit-rates for the JNKs in the binding assay vary between 104 for JNK1, 54 for JNK2, and 203 for JNK3 (Table S1). Lower sensitivity for the JNK2 assay, together with the use of Tanimoto coefficients with a fixed activity threshold, could be responsible for the separation of the JNKs in the tree. The lower hit-rate for JNK2 may have arisen if it was assayed in a different state compared to JNK1/3. Many factors might explain this, including the possibility that they are in different activation states because of differential phosphorylation or because of different behavior of the kinase domain in the context of a T7 phage fusion protein.

Similarly, AKT1 and AKT3 are found together, close to other AGC family kinases, but AKT2 is found next to PAK1. Differences between the AKT isoforms may be explained by the fact that all three have a low hit-rate (Table S1). In this situation, one borderline result close to the %control threshold used to define binding can have a significant effect on the Tanimoto similarity coefficient. This underlines the importance of the choice of analysis metric and of care in interpretation of the results.

The use of other similarity metrics might help to address this, but using a larger compound set would be the best solution. Because of these uncertainties, some of the relationships between kinases, especially those with low or very different hit-rates (Table S1), are tentative.

A Closer View of the Tyrosine Kinases. The tyrosine kinases have been an especially active area of research. Patterns of high SAR-similarity within the tyrosine kinases can be seen in Figure 10, a magnification of part of Figure 6a and Figure 6c. Conserved sub-branches include the Src family, FER and FES, the FGFRs, the PDGFR branch, the IR branch, the ERBs, and the EphA/EphBs. Comparing the SAR-similarity matrix with a sequence-based similarity matrix (Figure 10) identifies regions of interest.

For example, consider the positioning of CSK (c-Src kinase) in the kinase sequence-based tree. It has roughly the same kinase domain sequence identity to ABL and to BTK, 47% and 45%, respectively (Figure 10a and Table 3), on different tyrosine kinase sub-branches. All three kinases contain SH3/SH2 domains N-terminal to the kinase domain, while BTK contains an additional pleckstrin homology domain. By kinase domain sequence similarity alone, it would be difficult to predict whether CSK would bind to compounds more like ABL or more like BTK. From Figure 10b, the SAR similarity of CSK is closer to that of the SRC/LCK family kinases than to ABL/ARG, FER/FES, or the TEC family. A more detailed analysis of the ATP sites than is possible here might explain this result. Alternatively, it may reflect the ability of ABL to bind compounds by adopting a DFG-out conformation.[54] At least in the state in which it was assayed, CSK seems to be unable to bind these compounds.

The PDGFR/FLT branch will be used as an example of a practical use of this analysis. FLT3 is a target for intervention in acute myelogenous leukemia.[55] The kinases with highest SAR similarity to FLT3 include FLT1, KIT, PDGFRα/β, and cFMS (CSF1R) (Figure 10b, Table 3). This result is as would be expected from a sequence similarity analysis of kinases phylogenetically related to FLT3, as all of these share 49−55% sequence identity with FLT3.

However, the extent to which these kinases are inhibited by the same compounds may be surprising. Eighty compounds showed %control < 1 against FLT3. Of these, 58 had %control < 1 against PDGFRβ and 62 against KIT. Out of 9 compounds whose K_d was measured against FLT3, PDGFRβ, and KIT and for which the K_d for FLT3 was below 100 nM, all nine had K_d below 100 nM against PDGFRβ and 7 against KIT. For example, the 2-aminopyrazole **16** (Figure 4) had K_d of 30 nM for FLT3, compared to 111 nM for KIT and 42 nM for PDGFRβ. The related pyrazole **16a** is claimed as a 31 nM inhibitor of FLT3 (example 37 in patent WO2005047273).[56]

From the results above, this cross-inhibition seems almost universal, suggesting that researchers interested in FLT3 inhibitors should screen inhibitors of these related kinases. This would be predictable from sequence.

Similar suggestions can be made with a lesser degree of confidence for more evolutionarily distant pairs of tyrosine kinases with high SAR similarity found in this analysis that would not be readily predictable from sequence. One example is the relatively high SAR similarity seen between the ephrin receptor family, e.g., EphA1, and homologues on the SRC and LCK branches, e.g., Fyn-related kinase FRK (marked in Figure 10b). The Tan10 SAR similarity between EphA1 and FRK is high, 0.48 (Table 3). FRK and another tyrosine kinase, SYK, have comparable kinase domain sequence identity to EphA1, 42% and 38%, respectively. However, the Tan10 SAR similarity seen between the ephrin receptor family and SYK (0.21) is much lower than that between EphA1 and FRK (0.48).

Two examples of compounds that give rise to these SAR similarities will now be given. **17** is a member of a series of literature SRC inhibitors.[57] In this study it had K_d of 5 nM against EphA1 and 2 nM against FRK, and yet showed no detectable binding to SYK at 10 μM. **18** has K_d of 122 nM against EphA3 and 46 nM against FRK and also did not bind potently to SYK. **18** belongs to a class of biphenylamide known to bind to p38α in a DFG-out conformation.[29] Several other compounds believed to require the DFG-out conformation also bound to the ephrin receptor family and to FRK but not to SYK. It is likely that the basis for selectivity of some of these compounds for FRK and the ephrin receptor family and against SYK is due to binding to a conformation that SYK cannot adopt. It must be remembered that in a cell the accessibility of such conformational binding requirements may depend on the activation state, the presence or absence of cofactors, or other influences beyond the kinase domain that are not present in the format of the binding assay used in this study.

DFG-Out Compounds. Eighteen compounds in the set were judged to have the potential to bind in the DFG-out conformation, on the basis of containing a diarylamide or -urea group positioned in the back pocket region.[58] Although it cannot be said with certainty that such compounds will bind to each of their target kinases in the DFG-out conformation, it is interesting to compare the profiles of these compounds with those of the rest of the compound set. Twenty-five out of the 36 kinases that bind to at least half of these compounds are found in the tyrosine kinase branch (Table S2). This may indicate that tyrosine kinases have a greater propensity to adopt the DFG-out conformation under these assay conditions, perhaps because they share some common modes of inactivation. Alternatively, different types of compound may be required to exploit this conformation in other kinases.

Conclusions

The ability to carry out high-throughput profiling of compounds against panels of kinases has the potential to change

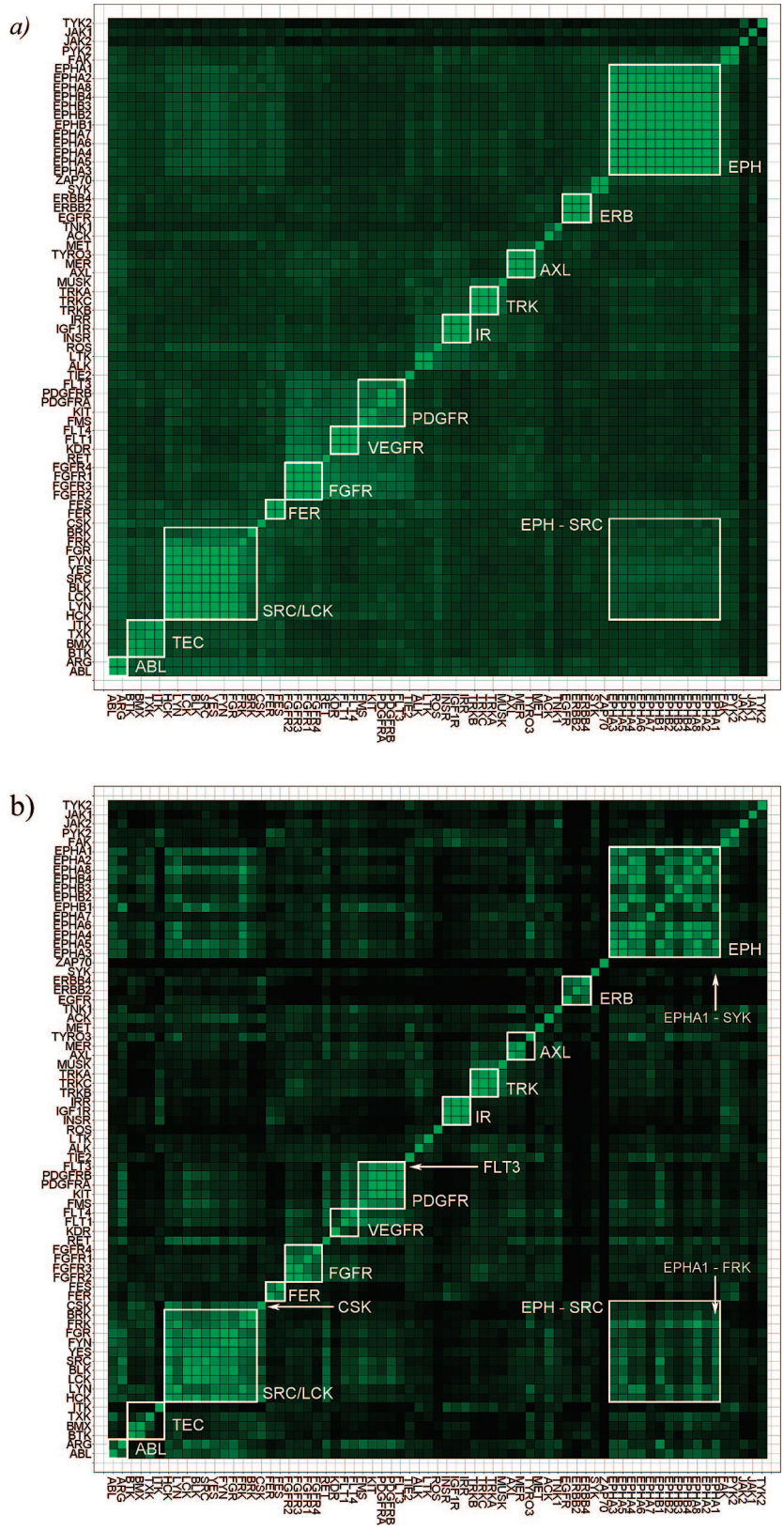

Figure 10. Similarity between pairs of tyrosine kinases: (a) sequence similarity (%identity in kinase domain) within the tyrosine kinases; (b) SAR-similarity matrix using Tan10 metric (%control = 10 threshold). Sub-branches discussed in the text are labeled.

the way that inhibitors are selected and optimized.[59] A total of 577 compounds have been screened against 203 kinases. The set of compounds is representative of a corporate kinase screening collection, accumulated over years of screening and synthesis.

(1) At least one hit was found against every kinase. For most targets, multiple active compounds covering different chemotypes were found. This hit-rate is very high compared to that normally obtained from diverse or focused screening, showing the value of cross-screening for kinase hit identification. Useful hits were found against novel kinases, for example, PIM1 and MPSK1.

(2) Different kinases have very different hit-rates. With the caveats mentioned earlier, the tractability of different kinases

(at least as defined by their ability to bind multiple inhibitors) seems to be highly variable. This information could be useful at the earliest stages of target selection, other factors being equal, to choose targets with a good chance of finding fairly selective lead molecules.

(3) Selective compounds bind to small clusters of kinases colocated on the kinome phylogenetic tree. Instead of binding to larger clusters of kinases, unselective inhibitors bind to multiple small clusters. Given the relative ease with which hits could be found for most targets by cross-screening this small compound set, a greater challenge would be to optimize them to the desired selectivity profile. It was possible to identify selective hits against some targets, and some compounds that had previously been shown to inhibit other protein kinases did not bind to any of the kinases in this panel. However, other targets only produced unselective starting points. When starting with unselective hits from this approach, at least their kinase profiles are known and attempts can be made through chemical modification to eliminate the unwanted activities. There are many examples in the literature of the use of rational approaches to dial out unwanted kinase activities in lead optimization.

The results give an indication of the number of kinases that different compounds can inhibit and how they are likely to be distributed around the kinome tree. Examples have been shown of the use of this profiling to identify off-target activities of compounds. Repeat measurements and K_d determinations would be desirable to confirm these results. It is critically important to understand the broad kinase selectivity profile of tool compounds and leads. Since this is currently expensive, it is impractical to do this for every compound in a lead optimization series. Because similar compounds tend to show similar profiles, this is probably not even necessary. However, we recommend that profiling be carried out early on and again periodically during lead optimization to check that the profile has not changed.

(4) Compounds that inhibit the same target but are structurally dissimilar are likely to have a different spectrum of off-target kinase activities. The consequences of inhibiting most kinases is poorly understood, raising questions about what to do if an off-target activity is detected. In a program with multiple lead series, one sensible strategy is to manage risk by choosing to optimize series that have different off-target activities. Even structurally diverse compounds that inhibit the same primary target are likely to inhibit the same closely related kinases unless they contain exploitable differences within the ATP site. Each compound is also highly likely to inhibit other less closely related kinases found on different branches of the kinome. Our results suggest that these remote interactions are likely to be quite unpredictable. However, our analysis also suggests that these diverse kinases are likely to be different for different chemical series. This result may influence the approach used to search for backup series in an active program. To reduce the risk of both compounds having the same undesirable off-target activity, the backup should be as chemically dissimilar as possible from the first lead series. While intuitive, this result also has important consequences for investigators selecting multiple tool compounds for use in cellular target validation experiments.

(5) Pairs of kinases sharing high sequence similarity are more likely to bind a similar range of inhibitor chemotypes than less closely related kinases. Above a kinase domain sequence identity threshold of around 40−50%, kinases become more likely than random to bind the same compounds.

(6) All kinome branches are not the same. Pairs of tyrosine kinases with a given sequence similarity to one another are more likely to bind to the same compounds than kinase pairs on other branches that have the same sequence similarity. Selectivity between tyrosine kinases may therefore be harder to achieve than between kinases on other branches.

Decisions about what selectivity profile is desirable are frequently made at the start of a program based entirely on biological knowledge. The sort of information gathered here can add pragmatism to these criteria. For example, kinases sharing more than 40−50% sequence identity have a greater chance of being inhibited by the same compounds than more divergent kinases. If selectivity regarded as essential for the efficacy or safety of a given target is likely to be unachievable because of high sequence similarity, that target may be best avoided in favor of another.

The compound set used here was limited in size and was constrained by the requirement for previously measured activity against a small subset of kinases. It is important to continue to add to the diversity of kinase compound space by preparing new compounds in order to increase the chances of finding hits with different selectivity profiles. Increasing the size of the screening data set, both in compounds and kinases screened, will allow our conclusions to be tested and hopefully confirmed with greater confidence.

We have attempted to summarize the main results and conclusions. Many of these will seem intuitive to those who have been active in the kinase field for a long time but will be new to others. The results should be valid over the subset of ∼40% of the genomic kinases covered by these assays. Indeed, conclusions could tentatively be drawn for 50% of the human kinases, assuming that those with >50% identity to one of the kinases screened behave like their homologues.

Experimental Section

Selection of Compound Set. The compound set was chosen with the aim of covering known kinase compound space as thoroughly as possible using up to 600 compounds.

Daylight SMARTS definitions ("Kinase Compound Set (KCS) Classes") have been defined to divide the GSK compound collection into manageable and structurally related subgroups (unpublished results). These are applied to lists of compounds in hierarchical order so that each compound is assigned to one class. The tightest definitions are applied before more general classifications. For example, a substructure query to identify 2,4-bis-anilinopyrimidines comes before pyrimidine substructures without the 4-aniline group. Definitions that include only the pyrimidine ring are found toward the end of the list, to capture a much more diverse range of compounds that still have some features in common. Compounds not matching any of these definitions are put into a miscellaneous group. The list of definitions is periodically updated to take account of new classes of kinase compounds.

Available compounds that had shown activity (IC_{50} < 5 μM) historically against at least one kinase were assigned to a KCS class. Additional clustering was applied using Daylight fingerprints and other clustering methods to assist in visualization. Simple molecular properties were also calculated. Each KCS class was examined in turn, and one or two representative molecules were chosen from each by medicinal chemists familiar with protein kinases and with each series. The aim was to choose simple, low molecular weight molecules to represent the class. Where examples from one KCS class showed divergent activity profiles against different protein kinases, one example from each type was picked. KCS classes showing higher diversity (multiple chemotypes or different clusters of biological behavior as judged by the chemists) were represented using more compounds. The 2,4-dianilinopyrimidine class was the most heavily represented and included 14 examples, but this is still

under 2.5% of the compound set. Most compound classes were represented by one or two molecules. A total of 577 compounds were screened at Ambit Biosciences. These covered a total of 130 KCS classes as well as 25 miscellaneous singleton compounds. A search of the AurSCOPE kinase database[60] revealed that 9% of the compounds appear in GSK publications, 5% are in publications from other sources, and the rest either are unpublished or lack published activity data.

While this manuscript was in preparation, the Structural Genomics Consortium (SGC) reported profiling results against 60 kinases of 156 inhibitors, mostly commercially available compounds.[10] While an impressively large set, this covered 30 KCS classes and was biased toward oxindoles (22 examples) and staurosporine analogues/bis-indole maleimides (17 examples). The 58 compounds classified as miscellaneous included 15 containing activated double bonds. The 577 compounds used here contained no such potentially reactive features. They are also considerably more chemically diverse. Only 51/577 compounds have Daylight fingerprint similarity of >0.6 to any of the compounds in the SGC study. The other large-scale profiling results reported to date used an updated superset of the panel reported here, including 317 kinases, several of which are disease-relevant mutants. However, the compound set was relatively small, comprising 38 literature compounds.[11]

Assays. The data presented here were generated at Ambit Biosciences, using binding assays, as opposed to activity assays, as previously described.[9,11] Kinases were expressed as fusion proteins to T7 phage. In general, full-length constructs were used for small kinases and catalytic domains for large kinases. T7-kinase-tagged phage strains were mixed with known kinase inhibitors immobilized on streptavidin-coated magnetic beads and with test compounds at a single concentration of 10 μM. Test compounds that bind to the kinase ATP site displace the immobilized ligand from the kinase/phage, which is detected using quantitative PCR. The results are reported as the percentage of kinase/phage remaining bound to the ligand/beads, relative to a control (DMSO lacking a test compound). High affinity compounds have %control = 0, while weaker binders have higher %control values. Results are reported for screening against 203 human kinases, listed in Table S1. The positions of these on the human kinome tree are shown in Figure S4.

Single-Concentration Data Quality and Interpretation. False negatives and positives inevitably occur in single-concentration screening. Confirmation of these results and quantification of affinity requires K_d measurement. Because many compounds approached complete competition for binding at 10 μM, it was necessary to select compounds for K_d determination. Priority was given to confirming results for compounds showing relatively selective inhibition of targets of special interest.

By comparison of K_d to single-concentration results for a given kinase, good correlation was found (Figure S5a). However, as there were shifts in the slope and intercept between targets, the translation between single-concentration and precise K_d values was kinase-dependent. Comparison between single-concentration data and IC_{50} values generated internally in other assay formats also showed acceptable correlation. This was especially true for binding assays but encouragingly also held for assays measuring catalytic activity. For example, Figure S5b shows good correlation between the single-concentration results for LCK and results generated for murine LCK using a catalytic assay.

K_d Determination. A total of 1180 K_d values were determined for selected compounds/assays where single concentration results had shown %control < 10. Of these, 63 returned K_d > 10 μM (~5% false positive rate). More potent single concentration results gave lower K_d values, with nearly 87% of compounds with %control < 1 at 10 μM giving submicromolar K_d values. Weaker single concentration results (%control > 10) rarely gave K_d < 1 μM (11/141, or under 8%). Since the selection of compounds and assays for K_d determination was biased toward confirming interesting or surprising selectivity results at single concentration, these values overestimate the true false negative and positive rates. As reported elsewhere, the false positive rate of this assay format was 2.4% and the false negative rate only 0.2%.[11]

On the basis of this, while individual results of interest should be confirmed by K_d determination, errors seem to be low in frequency. It is valid to use the single-concentration data to draw wider statistical conclusions.

Generation of Activity-Based SAR Similarity Scores. Bit-strings were constructed in which each bit-string represents a compound and each bit represents binding affinity for a kinase. Bits receive a value of 1 if the binding affinity is below a %control threshold (active), and 0 if it is above (inactive). Different binding affinity thresholds were tried; here, results are given for %control thresholds of 1 (Tan1) and 10 (Tan10). The use of variable thresholds for different assays based on the distribution of active compounds was considered but was not used because it cannot be assumed that two kinases should have the same hit-rate from this data set. The fingerprints of each pair of compounds were compared using the Tanimoto similarity coefficient[36] to generate "compound SAR similarity" values ranging from 0 (dissimilar) to 1 (similar). The process is represented graphically in Figure S6 (Supporting Information).

In an identical way, bit-strings were generated for each kinase, where each string represents a kinase and each bit represents binding affinity of a compound, and used to compare targets to each other. These bit-strings were constructed exactly as described before for the compound fingerprints, using the same Tan1 and Tan10 activity cutoffs, using the rows instead of the columns in the matrix. The fingerprints of each pair of kinases were compared using the Tanimoto similarity coefficient to generate "kinase SAR similarity" scores ranging from 0 to 1. Other similarity metrics were also tried. Taking the logarithm of the %control values instead of using a binary active/inactive descriptor, strings of binding data can be treated as multidimensional vectors. Similarity can then be expressed as the Euclidean distance between two vectors, or as the angle between the vectors. These metrics gave the same conclusions as the Tanimoto coefficient (results not shown).

Selectivity. The selectivity of compounds within a kinase panel can be quantified as the fraction of kinases bound at some threshold.[11] Here, the panel is a constant 203 targets, so selectivity was expressed as simply the number of kinases bound at a given %control threshold, usually <1.

Compound Structural Similarity. Structural similarity between all pairs of compounds was calculated using the DayPerl toolkit.[61] Fingerprints were generated with default parameters (minstep 0, maxstep 7, size 1024). Similarity was calculated between pairs of fingerprints using the Tanimoto coefficient.

Kinase Sequence Similarity. Sequence similarity between all pairs of kinases was expressed as the percentage of identical amino acids within the kinase domain using the Sugen alignment[5] downloaded from the http://Kinase.com Web site.[62] More subtle similarity matrices were not used, nor was the ATP site considered separately, because previous studies have shown high correlation between these metrics and the overall domain sequence identity.[42]

Generation of Trees. Tan1 and Tan10 similarity scores for all pairs of kinases were recalculated using only the more selective compounds (binding to less than 20 kinases with %control < 1). These were converted to distance matrices using the formula $100[(1 - \text{TanSim})^{1/2}]$. The GCG Growtree program was used to produce tree files using both Neighbor-Joining and UPGMA methods.[63] The TreeDyn program[64] was used to visualize these and produce the figures. The tree shown in Figure 9b is the NJ tree generated from the Tan10 similarity matrix.

Acknowledgment. The authors acknowledge Drake Eggleston for his encouragement and his support in obtaining funding for compound profiling. We thank Melissa Gomez and Bob Johnson for LC−MS analysis to test compounds for integrity and purity, Brandy Lloyd and Carol Jenkins for compound shipment, Stefan Senger for searching the Aureus database, and scientists at Ambit Biosciences for compound

profiling in the Ambit kinase panel. We are grateful to Jerry Adams, John Christopher, Chun-wa Chung, Andrew Leach, and Iain McLay as well as Patrick Zarrinkar (Ambit Biosciences) and three anonymous referees for critical reading of the manuscript.

Supporting Information Available: A pdf file containing Tables S1 and S2 and Figures S1−S6; a txt file containing 203 × 203 kinase vs kinase matrix showing SAR similarities and counts of compounds binding with different thresholds. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat. Rev. Drug Discovery* **2002**, *1*, 493–502.

(2) Giamas, G.; Stebbing, J.; Vorgias, C. E.; Knippschild, U. Protein kinases as targets for cancer treatment. *Pharmacogenomics* **2007**, *8*, 1005–1016.

(3) Gaestel, M.; Mengel, A.; Bothe, U.; Asadullah, K. Protein kinases as small molecule inhibitor targets in inflammation. *Curr. Med. Chem.* **2007**, *14*, 2214–2234.

(4) Krystal, G. W. Imatinib mesylate (STI571) for myeloid malignancies other than CML. *Leuk. Res.* **2004**, *28* (Suppl. 1), S53−S59.

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

(6) Hanks, S. K. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* **2003**, *4*, 111.

(7) 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.

(8) Bain, J.; Plater, L.; Elliott, M.; Shapiro, N.; Hastie, C. J.; McLauchlan, H.; Klevernic, I.; Arthur, J. S.; Alessi, D. R.; Cohen, P. The selectivity of protein kinase inhibitors: a further update. *Biochem. J.* **2007**, *408*, 297–315.

(9) 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.

(10) Fedorov, O.; Marsden, B.; Pogacic, V.; Rellos, P.; Muller, S.; Bullock, A. N.; Schwaller, J.; Sundstrom, M.; Knapp, S. A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 20523–20528.

(11) 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.

(12) Bachmann, M.; Moroy, T. The serine/threonine kinase Pim-1. *Int. J. Biochem. Cell Biol.* **2005**, *37*, 726–730.

(13) Pogacic, V.; Bullock, A. N.; Fedorov, O.; Filippakopoulos, P.; Gasser, C.; Biondi, A.; Meyer-Monard, S.; Knapp, S.; Schwaller, J. Structural analysis identifies imidazo[1,2-*b*]pyridazines as PIM kinase inhibitors with in vitro antileukemic activity. *Cancer Res.* **2007**, *67*, 6916–6924.

(14) Stavenger, R. A.; Witherington, J.; Rawlings, D. A.; Holt, D. A.; Chan, G. Preparation of *N*-Pyrrolopyridinyl Carboxamides as Chk1 Kinase Inhibitors for Treating Various Forms of Cancer and Hyperproliferative Disorders. Patent Application WO2003028724, 2003.

(15) Bullock, A. N.; Debreczeni, J. E.; Fedorov, O. Y.; Nelson, A.; Marsden, B. D.; Knapp, S. Structural basis of inhibitor specificity of the human protooncogene proviral insertion site in moloney murine leukemia virus (PIM-1) kinase. *J. Med. Chem.* **2005**, *48*, 7604–7614.

(16) Ohta, S.; Takeuchi, M.; Deguchi, M.; Tsuji, T.; Gahara, Y.; Nagata, K. A novel transcriptional factor with Ser/Thr kinase activity involved in the transforming growth factor (TGF)-beta signalling pathway. *Biochem. J.* **2000**, *350* (Part 2), 395–404.

(17) Ligos, J. M.; de Lera, T. L.; Hinderlich, S.; Guinea, B.; Sanchez, L.; Roca, R.; Valencia, A.; Bernad, A. Functional interaction between the Ser/Thr kinase PKL12 and *N*-acetylglucosamine kinase, a prominent enzyme implicated in the salvage pathway for GlcNAc recycling. *J. Biol. Chem.* **2002**, *277*, 6333–6343.

(18) Eswaran, J.; Bernad, A.; Ligos, J. M.; Guinea, B.; Debreczeni, J. E.; Sobott, F.; Parker, S. A.; Najmanovich, R.; Turk, B. E.; Knapp, S. Structure of the human protein kinase MPSK1 reveals an atypical activation loop architecture. *Structure* **2008**, *16*, 115–124.

(19) Podolin, P. L.; Callahan, J. F.; Bolognese, B. J.; Li, Y. H.; Carlson, K.; Davis, T. G.; Mellor, G. W.; Evans, C.; Roshak, A. K. Attenuation of murine collagen-induced arthritis by a novel, potent, selective small molecule inhibitor of IkappaB kinase 2, TPCA-1 (2-[(aminocarbonyl)amino]-5-(4-fluorophenyl)-3-thiophenecarboxamide), occurs via reduction of proinflammatory cytokines and antigen-induced T cell proliferation. *J. Pharmacol. Exp. Ther.* **2005**, *312*, 373–381.

(20) Tudhope, S. J.; Catley, M. C.; Fenwick, P. S.; Russell, R. E.; Rumsey, W. L.; Newton, R.; Barnes, P. J.; Donnelly, L. E. The role of IkappaB kinase 2, but not activation of NF-kappaB, in the release of CXCR3 ligands from IFN-gamma-stimulated human bronchial epithelial cells. *J. Immunol.* **2007**, *179*, 6237–6245.

(21) Christopher, J. A.; Avitabile, B. G.; Bamborough, P.; Champigny, A. C.; Cutler, G. J.; Dyos, S. L.; Grace, K. G.; Kerns, J. K.; Kitson, J. D.; Mellor, G. W.; Morey, J. V.; Morse, M. A.; O'Malley, C. F.; Patel, C. B.; Probst, N.; Rumsey, W.; Smith, C. A.; Wilson, M. J. The discovery of 2-amino-3,5-diarylbenzamide inhibitors of IKK-alpha and IKK-beta kinases. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3972–3977.

(22) 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.

(23) Sun, L.; Tran, N.; Tang, F.; App, H.; Hirth, P.; McMahon, G.; Tang, C. Synthesis and biological evaluations of 3-substituted indolin-2-ones: a novel class of tyrosine kinase inhibitors that exhibit selectivity toward particular receptor tyrosine kinases. *J. Med. Chem.* **1998**, *41*, 2588–2603.

(24) Hauf, S.; Cole, R. W.; LaTerra, S.; Zimmer, C.; Schnapp, G.; Walter, R.; Heckel, A.; van Meel, J.; Rieder, C. L.; Peters, J. M. The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **2003**, *161*, 281–294.

(25) Harrington, E. A.; Bebbington, D.; Moore, J.; Rasmussen, R. K.; Ajose-Adeogun, A. O.; Nakayama, T.; Graham, J. A.; Demur, C.; Hercend, T.; Diu-Hercend, A.; Su, M.; Golec, J. M.; Miller, K. M. VX-680, a potent and selective small-molecule inhibitor of the Aurora kinases, suppresses tumor growth in vivo. *Nat. Med.* **2004**, *10*, 262–267.

(26) Vertex Pharmaceuticals Press Release. http://investors.vrtx.com/releasedetail.cfm?ReleaseID=276543. November 20, 2007.

(27) Bamborough, P.; Angell, R. M.; Bhamra, I.; Brown, D.; Bull, J.; Christopher, J. A.; Cooper, A. W.; Fazal, L. H.; Giordano, I.; Hind, L.; Patel, V. K.; Ranshaw, L. E.; Sims, M. J.; Skone, P. A.; Smith, K. J.; Vickerstaff, E.; Washington, M. *N*-4-Pyrimidinyl-1*H*-indazol-4-amine inhibitors of Lck: indazoles as phenol isosteres with improved pharmacokinetics. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4363–4368.

(28) Angell, R.; Aston, N. M.; Bamborough, P.; Buckton, J. B.; Cockerill, S.; deBoeck, S. J.; Edwards, C. D.; Holmes, D. S.; Jones, K. L.; Laine, D. I.; Patel, S.; Smee, P. A.; Smith, K. J.; Somers, D. O.; Walker, A. L. Biphenyl amide p38 kinase inhibitors 3: improvement of cellular and in vivo activity. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4428–4432.

(29) Angell, R. M.; Angell, T. D.; Bamborough, P.; Bamford, M. J.; Chung, C. W.; Cockerill, S. G.; Flack, S. S.; Jones, K. L.; Laine, D. I.; Longstaff, T.; Ludbrook, S.; Pearson, R.; Smith, K. J.; Smee, P. A.; Somers, D. O.; Walker, A. L. Biphenyl amide p38 kinase inhibitors 4: DFG-in and DFG-out binding modes. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4433–4437.

(30) Cywin, C. L.; Zhao, B. P.; McNeil, D. W.; Hrapchak, M.; Prokopowicz, A. S.; Goldberg, D. R.; Morwick, T. M.; Gao, A.; Jakes, S.; Kashem, M.; Magolda, R. L.; Soll, R. M.; Player, M. R.; Bobko, M. A.; Rinker, J.; DesJarlais, R. L.; Winters, M. P. Discovery and SAR of novel naphthyridines as potent inhibitors of spleen tyrosine kinase (SYK). *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1415–1418.

(31) Witherington, J.; Bordas, V.; Gaiba, A.; Garton, N. S.; Naylor, A.; Rawlings, A. D.; Slingsby, B. P.; Smith, D. G.; Takle, A. K.; Ward, R. W. 6-Aryl-pyrazolo[3,4-*b*]pyridines: potent inhibitors of glycogen synthase kinase-3 (GSK-3). *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3055–3057.

(32) Sawyer, J. S.; Anderson, B. D.; Beight, D. W.; Campbell, R. M.; Jones, M. L.; Herron, D. K.; Lampe, J. W.; McCowan, J. R.; McMillen, W. T.; Mort, N.; Parsons, S.; Smith, E. C.; Vieth, M.; Weir, L. C.; Yan, L.; Zhang, F.; Yingling, J. M. Synthesis and activity of new aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-beta type I receptor kinase domain. *J. Med. Chem.* **2003**, *46*, 3953–3956.

(33) Sakuma, M.; Hatsushika, K.; Koyama, K.; Katoh, R.; Ando, T.; Watanabe, Y.; Wako, M.; Kanzaki, M.; Takano, S.; Sugiyama, H.; Hamada, Y.; Ogawa, H.; Okumura, K.; Nakao, A. TGF-beta type I receptor kinase inhibitor down-regulates rheumatoid synoviocytes and prevents the arthritis induced by type II collagen antibody. *Int. Immunol.* **2007**, *19*, 117–126.
(34) Bandyopadhyay, A.; Agyin, J. K.; Wang, L.; Tang, Y.; Lei, X.; Story, B. M.; Cornell, J. E.; Pollock, B. H.; Mundy, G. R.; Sun, L. Z. Inhibition of pulmonary and skeletal metastasis by a transforming growth factor-beta type I receptor kinase inhibitor. *Cancer Res.* **2006**, *66*, 6714–6721.
(35) Kawai, A.; Kawai, M.; Murata, Y.; Takada, J.; Sakakibara, M. Preparation of Pyridylpyrroles as Interleukin and Tumor Necrosis Factor Antagonists. Patent Application WO9802430, 1998.
(36) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.
(37) Kauvar, L. M.; Higgins, D. L.; Villar, H. O.; Sportsman, J. R.; Engqvist-Goldstein, A.; Bukar, R.; Bauer, K. E.; Dilley, H.; Rocke, D. M. Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **1995**, *2*, 107–118.
(38) Fliri, A. F.; Loging, W. T.; Thadeio, P. F.; Volkmann, R. A. Biological spectra analysis: linking biological activity profiles to molecular structure. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 261–266.
(39) Davies, S. P.; Reddy, H.; Caivano, M.; Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem. J.* **2000**, *351*, 95–105.
(40) Bain, J.; McLauchlan, H.; Elliott, M.; Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochem. J.* **2003**, *371*, 199–204.
(41) Frye, S. V. Structure−activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* **1999**, *6*, R3−R7.
(42) Vieth, M.; Higgs, R. E.; Robertson, D. H.; Shapiro, M.; Gragg, E. A.; Hemmerle, H. Kinomics−structural biology and chemogenomics of kinase inhibitors and targets. *Biochim. Biophys. Acta* **2004**, *1697*, 243–257.
(43) ter Haar, E.; Walters, W. P.; Pazhanisamy, S.; Taslimi, P.; Pierce, A. C.; Bemis, G. W.; Salituro, F. G.; Harbeson, S. L. Kinase chemogenomics: targeting the human kinome for target validation and drug discovery. *Mini-Rev. Med. Chem.* **2004**, *4*, 235–253.
(44) Katajisto, P.; Vallenius, T.; Vaahtomeri, K.; Ekman, N.; Udd, L.; Tiainen, M.; Makela, T. P. The LKB1 tumor suppressor kinase in human disease. *Biochim. Biophys. Acta* **2007**, *1775*, 63–75.
(45) Conner, S. D.; Schmid, S. L. Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J. Cell Biol.* **2002**, *156*, 921–929.
(46) Inohara, N.; del Peso, L.; Koseki, T.; Chen, S.; Nunez, G. RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. *J. Biol. Chem.* **1998**, *273*, 12296–12300.
(47) McCarthy, J. V.; Ni, J.; Dixit, V. M. RIP2 is a novel NF-kappaB-activating and cell death-inducing kinase. *J. Biol. Chem.* **1998**, *273*, 16968–16975.
(48) Thome, M.; Hofmann, K.; Burns, K.; Martinon, F.; Bodmer, J. L.; Mattmann, C.; Tschopp, J. Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. *Curr. Biol.* **1998**, *8*, 885–888.
(49) Chin, A. J.; Dempsey, P. W.; Cheng, G. Rip2: a key molecule that regulates both innate and acquired immunity. *Curr. Med. Chem.* **2005**, *4*, 35–42.
(50) Argast, G. M.; Fausto, N.; Campbell, J. S. Inhibition of RIP2/RIck/CARDIAK activity by pyridinyl imidazole inhibitors of p38 MAPK. *Mol. Cell. Biochem.* **2005**, *268*, 129–140.
(51) Godl, K.; Wissing, J.; Kurtenbach, A.; Habenberger, P.; Blencke, S.; Gutbrod, H.; Salassidis, K.; Stein-Gerlach, M.; Missio, A.; Cotten, M.; Daub, H. An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15434–15439.
(52) Eyers, P. A.; Craxton, M.; Morrice, N.; Cohen, P.; Goedert, M. Conversion of SB 203580-insensitive MAP kinase family members to drug-sensitive forms by a single amino-acid substitution. *Chem. Biol.* **1998**, *5*, 321–328.
(53) Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. A unique structure for epidermal growth factor receptor bound to GW572016 (lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **2004**, *64*, 6652–6659.
(54) Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. *Science* **2000**, *289*, 1938–1942.
(55) Levis, M.; Small, D. FLT3 tyrosine kinase inhibitors. *Int. J. Hematol.* **2005**, *82*, 100–107.
(56) Bold, G.; Floersheimer, A.; Furet, P.; Guagnano, V.; Masuya, K.; Vaupel, A.; Schoepfer, J. Prepatation of Thiazole and Pyrazole Derivatives as Flt-3 Kinase Inhibitors. Patent Application WO2005047273, 2005.
(57) Altmann, E.; Missbach, M.; Green, J.; Susa, M.; Wagenknecht, H. A.; Widler, L. 7-Pyrrolidinyl- and 7-piperidinyl-5-aryl-pyrrolo[2,3-*d*]pyrimidines−potent inhibitors of the tyrosine kinase c-Src. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 853–856.
(58) Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.
(59) 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.
(60) AurScope Kinase Database. Aureus Pharma. http://www.aureus-pharma.com.
(61) Daylight Chemical Systems Inc. Daylight Theory Manual. http://www.daylight.com/dayhtml/doc/theory/theory.finger.html.
(62) Salk Institute, California and Sugen Inc. Kinase.com Web site. http://kinase.com/human/kinome/phylogeny.html.
(63) Devereux, J.; Haeberli, P.; Smithies, O. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **1984**, *12*, 387–395.
(64) Chevenet, F.; Brun, C.; Banuls, A. L.; Jacq, B.; Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinf.* **2006**, *7*, 439.

JM8011036