BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 2 2010, pages 198–204
doi:10.1093/bioinformatics/btp650

Structural bioinformatics

Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis

Danzhi Huang[1,*], Ting Zhou[1], Karine Lafleur[1,2], Cristina Nevado[2] and Amedeo Caflisch[1,*]

[1]Department of Biochemistry and [2]Department of Organic Chemistry, University of Zürich, Winterthurerstrasse 190

Received on July 30, 2009; revised on November 2, 2009; accepted on November 12, 2009

Advance Access publication November 26, 2009

Associate Editor: Alex Bateman

ABSTRACT

Motivation and method: Small-molecule inhibitors targeting the adenosine triphosphate (ATP) binding pocket of the catalytic domain of protein kinases have potential to become drugs devoid of (major) side effects, particularly if they bind selectively. Here, the sequences of the 518 human kinases are first mapped onto the structural alignment of 116 kinases of known three-dimensional structure. The multiple structure alignment is then used to encode the known strategies for developing selective inhibitors into a fingerprint. Finally, a network analysis is used to partition the kinases into clusters according to similarity of their fingerprints, i.e. physico-chemical characteristics of the residues responsible for selective binding.

Results: For each kinase the network analysis reveals the likelihood to find selective inhibitors targeting the ATP binding site. Systematic guidelines are proposed to develop selective inhibitors. Importantly, the network analysis suggests that the tyrosine kinase EphB4 has high selectivity potential, which is consistent with the selectivity profile of two novel EphB4 inhibitors.

Contact: dhuang@bioc.uzh.ch; caflisch@bioc.uzh.ch

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Protein kinases control and modulate a wide variety of biological processes through their catalytic activity, which is the transfer of a phosphate group from adenosine triphosphate (ATP) to other proteins (Manning *et al.*, 2002). Several among the 518 human kinases and mutants thereof are causally involved in a large variety of diseases, and most importantly in cancer (Cohen, 2002; Dibb *et al.*, 2004; Sawyers, 2004), so that small molecule inhibitors of pharmacologically relevant kinases have good potential as drug candidates. Since the ATP binding site is similar in different kinases, one of the most important challenges in kinase inhibitor discovery and development is the optimization of selectivity (Traxler and Furet, 1999), as poor specificity undermines the clinical effectiveness of some drugs targeting the ATP binding site of protein kinases (Grunwald *et al.*, 2007; Wolter *et al.*, 2008).

The majority of small-molecule inhibitors of kinases target the ATP binding site, with the conserved Phe residue of the Asp-Phe-Gly (DFG) motif buried in a hydrophobic pocket in the groove between the two lobes of the kinase (type I inhibitors) (Liu and Gray, 2006; Pargellis *et al.*, 2002). Another set of inhibitors, so-called type II, occupy not only the ATP binding site, but exploit also an allosteric pocket which appears to be accessible only when the Phe side chain of the DFG motif moves out from the hydrophobic pocket (Bogoyevitch and Fairlie, 2007; Liu and Gray, 2006; Pargellis *et al.*, 2002). Type II inhibitors are in general less promiscuous than type I which is revealed by several selectivity profile data on panels of kinases ranging from 60 to 318 (Bain *et al.*, 2007; Bamborough *et al.*, 2008; Fabian *et al.*, 2005; Fedorov *et al.*, 2007; Goldstein *et al.*, 2008; Karaman *et al.*, 2008). However, some type I inhibitors can be very specific, e.g. VX-745 and CI-1033 (**1** and **2** of Fig. S1, Supplementary Material) for p38α mitogen-activated protein kinase (p38α) and epidermal growth factor receptor (EGFR; Karaman *et al.*, 2008), respectively, and both of them have entered clinical development. The success of these type I inhibitors demonstrates that, despite the highly conserved ATP binding site, it is feasible to optimize selectivity for certain kinases by following appropriate strategies.

As the majority of kinase inhibitors belong to type I, it is useful to investigate which kinases have a high likelihood to be inhibited selectively and if there are systematic guidelines to identify and/or design selective inhibitors for them. To answer both questions, it is more useful to compare the key residues of the ATP binding site, rather than using an overall similarity measure like the percentage identity of the whole sequence (Hopkins, 2008). Here, first we accurately identify the ATP binding site residues for most human kinases via structure and sequence-based approaches. For each of them, key residues are then selected based on the available knowledge for developing selective type I inhibitors. These residues are then encoded into a 9-bit fingerprint and analyzed by a network approach. The network analysis reveals large differences in the selectivity potential in the human kinome.

2 METHODS

Structural alignment using the ATP-binding pocket residues The 491 human protein kinase sequences used in this work were taken from the published data of Manning *et al.* (2002). The non-redundant PDB structures of protein kinases with a resolution >3.0 Å have been considered. There were 116 unique structures (as of May 2008) with ≥95% sequence identity

*To whom correspondence should be addressed.

to at least one of the 491 human kinase sequences (Manning *et al.*, 2002). In cases of multiple structures of the same kinase (e.g. with different inhibitors), the first in chronological order was chosen (see Supplementary Material for the PDB codes). The Ser/Thr kinase CDK2 (cyclin-dependent kinase 2, PDB code 1HCK), tyrosine kinase ABL1 (PDB code 1FPU) and calcium/calmodulin-dependent protein kinase CaMK1a (PDB code 1A06) were arbitrarily selected as templates. Their binding site residues were identified manually based on the definition shown in Figure 1. Structural superposition of the other 113 kinases was then performed by applying a modified version of the Gerstein–Levitt procedure (Gerstein and Levitt, 1998). To obtain an initial alignment needed for the Gerstein–Levitt procedure, a sequence alignment algorithm (based on dynamic programming) was used in which the scoring matrix is a function that rewards the alignment of C_α pairs with similar local backbone geometry. After the Gerstein–Levitt alignment step, the list of aligned pairs is purged by removing pairs (i,j) for which the distance between the C_α atoms in the structural overlap exceeds a given threshold.

The structural superposition was performed in two steps. First, all Cα atoms were taken into account to identify the rough binding site residues, i.e. those within 2 Å of the binding site of CDK2, ABL1 or CaMK1a. Then the Cα and Cβ atoms of these residues were used to further refine the fitting of ATP binding sites using a threshold of 3 Å. In addition, fitting mistakes due to structural variabilities, e.g. crystal packing or conformational changes upon inhibitor binding, were corrected by comparing with the three templates. Finally, the residues of the ATP binding site for 116 kinases were identified and mapped onto the respective kinase sequences.

For other 375 kinases of unknown three-dimensional (3D) structure, sequence alignments to each of 116 kinase with known structure was performed using ClustalW v1.8 (Thompson *et al.*, 1994) and BLOSUM30 matrix (Henikoff and Henikoff, 1992). Among them, 239 kinases have alignment scores >40% of maximal score to at least one of the 116 kinases, 80 kinases have scores between 30% and 40% and the remaining 56 kinases have scores <30%. The alignment results were checked by visual analysis of conserved residues (Hanks and Hunter, 1995), e.g. glycine loop, DFG motif. Finally, the ATP binding sites of 469 kinases were identified and 22 kinases were eliminated from further analysis due to low alignment scores and missing conserved residues.

Construction of the kinome network based on selective features To construct the network, each kinase was encoded by a 9-bit string. Every bit of the string stands for a strategy (Supplementary Material) for designing selective type I inhibitor and was assigned with an integer value depending on the corresponding residue. Edges were generated between nodes (kinases) with difference <3 bits and weighted with 9, 4 and 1 for difference in 0, 1 and 2 bits, respectively. Different weighting schemes change only marginally the layout of the network but not the clustering into subsets (Supplementary Material). There are 469 kinases (nodes) and 71 558 edges in the network. Nodes were colored according to the scheme shown in the bottom and placed according to the Fruchterman–Reingold algorithm. The figure was made using igraph (cneurocvs.rmki.kfki.hu).

Enzymatic assay of selectivity profile These experiments were carried out at the National Centre for Protein Kinase Profiling of University Dundee. All assays (25.5 µl volume) were carried out robotically at room temperature and were linear with respect to time and enzyme concentration under the conditions used. Assays were performed for 30 min using Multidrop Micro reagent dispensers (Thermo Electron Corporation, Waltham, MA, USA) in a 96-well format. The concentration of magnesium acetate in the assays was 10 mM and [γ-33P]ATP (800 c.p.m./pmol) was used at 5, 20 or 50 µM as indicated in the Supplementary Material, in order to be at or below the K_m for ATP for each enzyme. The assays were initiated with MgATP, stopped by the addition of 5 µl of 0.5 M orthophosphoric acid and spotted on to P81 filter plates using a unifilter harvester (PerkinElmer, Boston, MA, USA). The data is presented as mean percentage activity at a 3 µM concentration of inhibitor in duplicate assays, and compared with DMSO controls. It has been reported that percentage activity at single concentration of inhibitor is a cost-effective proxy of the dissociation constant or the inhibitor concentration required for half-maximal activity (Bamborough *et al.*, 2008; Lafleur *et al.*, 2009).

3 RESULTS AND DISCUSSIONS

Structure of the ATP binding site in protein kinases The kinase catalytic domain is a single polypeptide chain which folds into two lobes joined by a segment referred as hinge loop (Fig. 1A) (Hanks and Hunter, 1995). The N-terminal lobe consists of mostly antiparallel β-sheets and a conserved α-helix, while the C-terminal lobe is mainly helical (Hanks and Hunter, 1995). ATP binds in a cavity formed between the two lobes. The ATP binding site is made up of five pockets (Traxler and Furet, 1999) lined by 36 residues that are in van der Waals contact and/or involved in hydrogen bond(s) with one or more inhibitors of type I (Fig. 1).

Several analyses of kinase residues surrounding the ATP binding site have been published (Chen *et al.*, 2007; Hopkins, 2008; Sheinerman *et al.*, 2005; Vieth *et al.*, 2004, 2005; Vulpetti and Bosotti, 2004). The number of 3D structures of kinase domain deposited in the PDB databank (www.rcsb.org) has increased substantially in the past 2 years. The growing amount of structural information calls for a sequence alignment of the human kinome which should lead to an accurate characterization of the ATP binding site residues.

Here, it was first verified that in the structural alignment of 116 kinases of known 3D structure the ATP binding site is similar for different kinases. Then, the binding site residues of 469 kinases were identified by mapping the sequences of 353 kinases onto the structural alignment (details are presented in Section 2). Finally, the residues of the ATP binding site were renumbered from 1 to 36 according to their occurrence along the sequence (Fig. 1B). Remarkably, the amino acid distribution of the gatekeeper residue in some of the kinase groups differ significantly from the whole (i.e. 469 kinases) distribution (Fig. 2). In particular, Thr and Phe are much more frequent among the TK (see legend of Fig. 2 for abbreviations of kinase groups) and CMGC group, respectively, than in the overall distribution.

The network of kinase selectivity potential Effective strategies for the design of selective type I inhibitors mainly consist of targeting sequence variability at a given position of the ATP binding site or exploiting a cavity of variable size in different kinases (Liao, 2007; Noble *et al.*, 2004; Traxler and Furet, 1999). These strategies take into account the gatekeeper residue (Traxler and Furet, 1999), eventual covalent bonds with cysteine side chains (positions 9, 24 and 33 in Fig. 1B) (Blair *et al.*, 2007; Cohen *et al.*, 1994; Smaill *et al.*, 2000; Torrance *et al.*, 2000), the hydrogen bonding ability of the side chain at position 24 (Knight and Shokat, 2005; Laird *et al.*, 2000; Liao, 2007; Richardson *et al.*, 2006; Swahn *et al.*, 2005), the intrinsic flexibility of the hinge loop (Fitzgerald *et al.*, 2003; Xing *et al.*, 2009) and the size of the adenine binding pocket (Liao, 2007). Nine selectivity features, encoded into nine bits (Fig. 3), are used here to recapitulate the aforementioned strategies, and each kinase is then scrutinized for these features. Bit 1 describes the size of the gatekeeper residue which directly affects the accessibility of the hydrophobic pocket (Blencke *et al.*, 2003; Eyers *et al.*, 1998; Gorre *et al.*, 2001; Liu *et al.*, 1999; Shah *et al.*, 1997). About 20% of the

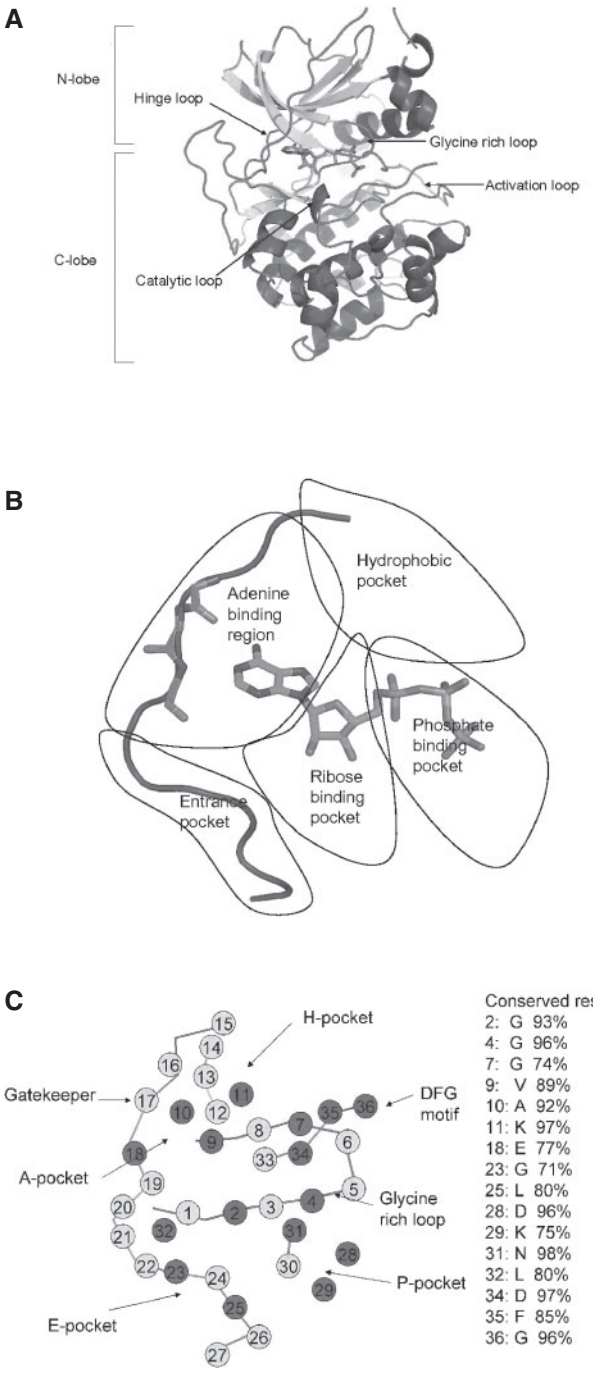

Fig. 1. 3D structure of the kinase catalytic domain and residues in the ATP binding site. (**A**) X-ray crystal structure of ATP bound to cAMP-dependent protein kinase (PDB code 1ATP). (**B**) Schematic of ATP binding site and its five pockets (Traxler and Furet, 1999). (**C**) 2D projection of the ATP binding site viewed down the axis from the N-terminal lobe to the C-terminal lobe. The residues of the ATP binding site are renumbered from 1 to 36 according to their occurrence along the sequence. Red, blue and cyan circles represent residues which are most (>90%), medium(70–90%) and not conserved (<70%), respectively. The positions conserved in >70% of the 469 kinases are listed in the right column. [Color figure can be viewed in the online issue, which is available at http://bioinformatics.oxfordjournals.org.]

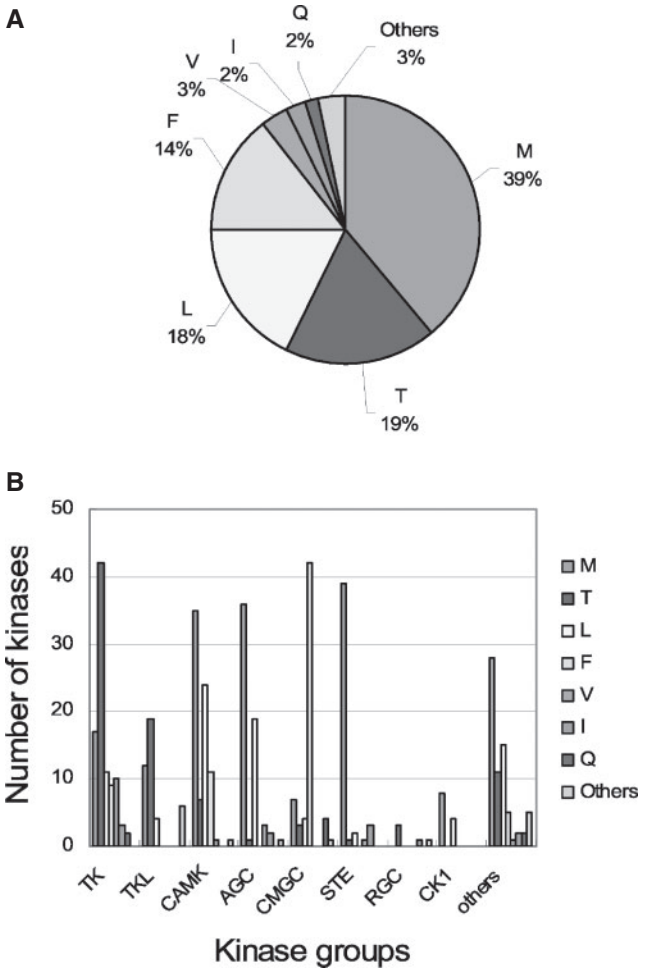

Fig. 2. Amino acid distribution of the gatekeeper residue. (**A**) The distribution for 469 protein kinases. (**B**) The amino acid distribution of gatekeeper for kinases in different kinase groups. Threonine is most frequent in TK and TKL. Phenylalanine is most frequent in CMGC. The kinase groups have been defined as follows (Manning *et al.*, 2002): STE: homologs of Steriles 7, 11 and 20 kinases; CMGC: CDK, MAPK, GSK3 and CLK; TK: tyrosine kinase; TKL: tyrosine kinase-like; ACG: PKA, PKG and PKC; CAMK: calcium/calmodulin-dependent protein kinase; RGC: receptor guanylate cyclase; CK1: casein kinase 1; others: protein kinases not belonging to the major groups. [Color figure can be viewed in the online issue, which is available at http://bioinformatics.oxfordjournals.org.]

human protein kinases have a small gatekeeper, i.e., glycine, valine, alanine, or threonine. Bit 2 reflects the hydrogen bonding ability of the gatekeeper. Nearly 20% protein kinases have a gatekeeper that can act as hydrogen bond donor and/or acceptor (Deininger and Druker, 2003; Talpaz *et al.*, 2006; Torrance *et al.*, 2000). Bits 3, 4 and 5 report on cysteine at position 9, 24 and 33, respectively. Around 14, 14 and 10% of kinases have this feature, respectively. Bit 6 describes the hydrogen bonding ability at position 24. Bit 7 relates to the residue at position 21 in the hinge loop (Fig. 1). In the case of glycine, backbone conformational changes that lead to differences in hydrogen bonding patterns with the hinge have been observed (Fitzgerald *et al.*, 2003). About 10% protein kinases have glycine at this position. Bits 8 and 9 report on the size of the adenine pocket

Strategy:	Gatekeeper		Covalent bond with Cys			Position 24	Hinge loop	Size of A-pocket	
Position(s):	17	17	9	24	33	24	21	10	HM motif
Fingerprint:	0/1	0/1/2	0/1	0/1	0/1	0/1/2/3	0/1/2	0/1	0/1

Selectivity features:

1. Position 17: 0 for valine, threonine, alanine, or glycine and 1 for others.
2. Position 17: 0 for residues with apolar side chain, 1 for threonine, serine, or cysteine, and 2 for glutamine or asparagine.
3. Position 9: 1 for cysteine and 0 for others.
4. Position 24: 1 for cysteine and 0 for others.
5. Position 33: 1 for cysteine and 0 for others.
6. Position 24: 0 for aspartate or glutamate, 1 for threonine, serine, cysteine, asparagine, or glutamine, 2 for lysine or arginine, and 3 for others.
7. Positions 21: 1 for glycine at position 21, 2 for three PIM kinases, and 0 for others.
8. Position 10: 0 for alanine or glycine and 1 for others.
9. HM motif: 0 and1 for kinases without and with HM motif, respecively.

Fig. 3. Definition of 9-bit selectivity fingerprint. Each protein kinase is encoded by a 9-bit fingerprint based on the strategies for designing selective type I inhibitors (see text). Every bit stands for a strategy and is assigned with an integer value. The positions within the 36 residues of the ATP binding site are listed in Figure 1B. [Color figure can be viewed in the online issue, which is available at http://bioinformatics.oxfordjournals.org.]

(A-pocket) (Battistutta *et al.*, 2000; Biondi, 2004; Goldsmith *et al.*, 2007; Liao, 2007; Yde *et al.*, 2005). About 10% protein kinases have a relatively small adenine pocket. Position 10 (i.e. bit 8) is alanine in 92% of the protein kinases. Note that the aforementioned strategies for designing selective inhibitors of type I, employed in this work to derive the 9-bit fingerprint, can be deduced based on sequence information alone. On the other hand, those strategies that require the knowledge of the 3D structure (Bikker *et al.*, 2009; Wood *et al.*, 2004) are not used in this study.

It is useful to consider the 469 kinases (116 and 353 kinases fitted by structure- and sequence-based alignment, respectively, as described in Section 2) and the similarity of their 9-bit fingerprints as nodes and links of a network, respectively (Fig. 4). Interestingly, kinases belonging to different groups [i.e. widely separated in the dendrogram of overall sequence similarity (Manning *et al.*, 2002)] are mixed together upon projection into the plane by a spring-embedded algorithm (Fruchterman and Reingold, 1991) that clusters the kinases according to pairwise similarity of the key residues in the ATP binding site. This result shows that full-sequence similarity does not correlate with the similarity of the ATP binding site residues that are relevant for selectivity. In contrast, the 9-bit fingerprint seems appropriate to determine the selectivity potential. Particularly, the human kinome is divided into two major sets, A and B (triangles and circles, respectively, in Fig. 4), depending on the size of the gatekeeper's side chain, and each of these two sets is further partitioned mainly into six clusters and a few outliers. The most populated cluster belongs to set A and contains the 139 kinases with the most frequent selectivity features. Importantly, the kinase selectivity potential network is useful for devising optimization strategies for selective inhibitors. As an example, the small gatekeeper existing in about 100 kinases of set B, especially threonine, should be first utilized to gain selectivity over about 400 kinases in set A, and then one could look for other differences to achieve further selectivity (see below).

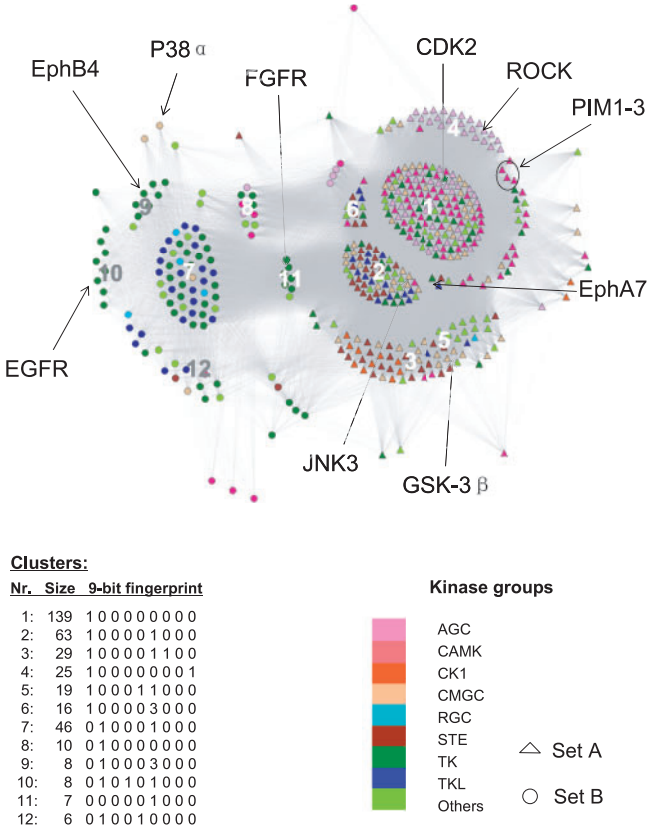

Fig. 4. The network of kinase selectivity potential. Pairs of protein kinases (nodes) are considered to have similar selectivity features (i.e. are connected by an edge) if they have <3 bits difference in the 9-bit fingerprint of Figure 3. Edges are weighted quadratically, i.e. with weight 9, 4 and 1 for 0-, 1- and 2-bit difference, respectively. The weighting is used for the projection onto the plane, which is obtained by a spring-embedder (i.e. force-directed placement) algorithm (Fruchterman and Reingold, 1991). Similar projections result by using the same weighting factor for 0-, 1- and 2-bit difference, or linear weighting (Supplementary Fig. S2). There are 469 kinases and 71 558 edges in this network. The nodes can be divided into two sets, A (triangles) and B (circles). Set A contains kinases with bulky gatekeeper residue, whereas set B contains kinases with small gatekeeper. The fingerprints of the six largest clusters of sets A and B are listed in the bottom left. Nodes are colored according to the kinase group (Manning *et al.*, 2002) (scheme shown in the bottom right). Kinases mentioned in the text are labeled while the kinase name of all nodes can be found in the Supplementary Material. The variable colors in each cluster show that full-sequence similarity does not reflect the similarity of key residues in the ATP binding site. [Color figure can be viewed in the online issue, which is available at http://bioinformatics.oxfordjournals.org.]

Comparison of the kinase selectivity potential network with the small molecule–kinase interaction map The usefulness of the network of kinase selectivity potential can be assessed by comparing it with the recently published inhibitor profile data (Karaman *et al.*, 2008) which can also be treated as a complex network (Fig. 5) (Newwan, 2003). The nodes represent 277 kinases [upon exclusion of the mutants from the original set of 317 (Karaman *et al.*, 2008)], and the 9394 edges reflect the profile similarity (PSim) which is a measure of how similarly two kinases

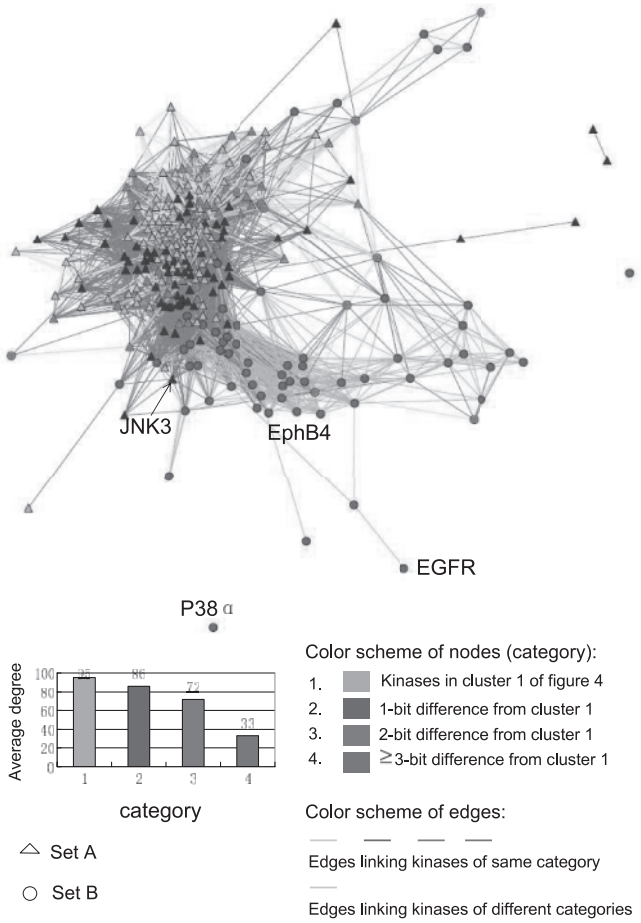

Fig. 5. The network of kinase inhibitor selectivity. The network is constructed using PSim which is a pairwise measure of kinase similarity based on experimentally determined binding affinity values of a set of 24 inhibitors of type I (Karaman *et al.*, 2008). Edges connect pairs of nodes (kinases) with similar profile of inhibition by the 24 inhibitors. The PSim weighting is also used for the projection onto the plane by a spring-embedder algorithm (Fruchterman and Reingold, 1991). The degree of node k is the number of edges incident to node k, i.e. a measure of how many kinases share a very similar inhibitory profile as kinase k. The network was plotted using igraph (cneurocvs.rmki.kfki.hu). [Color figure can be viewed in the online issue, which is available at http://bioinformatics.oxfordjournals.org.]

interact with a set of 24 type I inhibitors (Karaman *et al.*, 2008). The value of PSim is calculated for any pair of kinases k_1, k_2 by $\mathrm{PSim}_{k_1k_2} = \left[\frac{1}{24}\sum_{i=1}^{24}(\Delta G_{ik_1} - \Delta G_{ik_2})^2\right]^{-\frac{1}{2}}$, where ΔG_{ik} is the affinity of inhibitor i for kinase k, and the index i runs over the 24 inhibitors of type I. Edges are generated between kinases with PSim $\geq$ 1/(1.5 kcal/mol), i.e. two kinases are considered similar and connected by an edge if the root mean square deviation of their binding free energies for the 24 inhibitors is <1.5 kcal/mol, corresponding to about 10 times difference in affinity. The number of edges incident to a given node, so-called degree, indicates the similarity of a kinase to others with respect to inhibitor profile. Most kinases are grouped together forming a relatively large cluster and the remaining ones surround this cluster sparsely (Fig. 5).

To compare with the network of kinase selectivity potential, the 277 kinases are divided into four categories based on their selectivity fingerprints. Kinases of category 1 are those of cluster 1 in Figure 4. It is interesting to note that these kinases mostly locate in the center of the large cluster (Fig. 5) and they have the highest number of edges (degree). Therefore, kinases of cluster 1 in general have similar inhibitor profiles, which agrees with the ratiocination of the network of kinase selectivity potential. The remaining kinases are divided into three categories according to their similarity to cluster 1 with respect to the 9-bit selectivity fingerprint. Kinases of categories 2 and 3 differ from category 1 by 1 and 2 bits, respectively. They mainly locate around the border of the large cluster (Fig. 5) with average degree of 86 and 72 for categories 2 and 3, respectively. As they are similar to kinases of category 1, their selectivity potential against type I inhibitors is expected to be relatively low. The 63 kinases of category 4 (listed in the Supplementary Material) differ by at least 3 bits compared with those of category 1. They belong mainly to set B of Figure 4, and locate outside of the large cluster sparsely (Fig. 5). Their inhibitor profile for the set of 24 inhibitors is distinct from the majority of the other kinases as indicated by their average degree of only 33. This observation provides further evidence on the usefulness of the network analysis based on the selectivity fingerprint. In other words, the set B kinases, and in particular the 63 kinases differing by $\geq$3 bits from cluster 1, have a fairly high potential to be selective against type I inhibitors.

It is interesting to directly compare the kinase selectivity potential network and the small molecule–kinase interaction map (Karaman *et al.*, 2008) for certain specific classes. One interesting case is CDK2, a kinase belonging to the largest cluster of the selectivity potential network. Four compounds were developed using CDK2 as the primary target (Karaman *et al.*, 2008). However, these four compounds are rather promiscuous according to the selectivity score (relative to their primary target). This result is consistent with the network analysis which suggests that kinases of the largest cluster are difficult to be inhibited selectively. On the other hand, the selectivity potential network is based on nine features and therefore it may be difficult to explain the selectivity of a specific kinase inhibitor if it does not fully utilize these features. A typical cases is p38α. Of the three compounds in advanced development as p38α inhibitors, only VX-745 is significantly selective, while SB-202190 and SB-203580 are not (Karaman *et al.*, 2008). The main reason is that the latter two compounds were developed without considering the selectivity feature concerning the conformational change of the hinge loop (Fitzgerald *et al.*, 2003).

Usefulness of kinase selectivity potential network: target choice and strategies to design selective kinase inhibitors The network classification provides useful information for prioritizing targets since the selectivity potential can be an important property besides the pharmaceutical relevance. In general, specific inhibitors are more likely to be identified for kinases in small clusters and with few links. One such example is p38α. The unique fingerprints combination particularly of bits 1, 2 and 7 locates p38α in a cluster of only three kinases (Fig. 4), and thus very specific type I inhibitors can be developed by utilizing these features (Fitzgerald *et al.*, 2003; Laufer *et al.*, 2008; Xing *et al.*, 2009). Another example is Rho-associated coiled–coiled containing protein kinase (ROCK) which belongs to a cluster of only 25 kinases (cluster 4). The main hallmark of cluster 4 kinases is the hydrophobic motif (HM) in the adenine binding pocket

(bit 9). A characteristic phenylalanine residue at the HM can be used to design selective inhibitors for kinases in this cluster (Jacobs *et al.*, 2006; Liao, 2007). In contrast, kinases in the large clusters, e.g. cluster 1, are likely to be recalcitrant to selective inhibition because a large number of them have similar selectivity features. For these targets, inhibitor development should follow rules different from those mentioned above. For instance, targeting allosteric pockets may be a more appropriate strategy.

The network classification also provides guidelines to develop selective kinase inhibitors, especially for those kinases in the small clusters. For example, the eight kinases in group 10 have a threonine as gatekeeper, and cysteine at position 24. The combination of these two features is rather unique according to the network analysis. Therefore, one should try to first make use of these residues, e.g. forming hydrogen bond to threonine, occupying the hydrophobic pocket, and also forming a covalent bond to the residue at position 24. In fact, inhibitors having these three features are indeed rather selective, e.g. CI-1033 (an inhibitor of EGFR, compound **2** in Supplementary Fig. S1). Similar strategies can be applied for six kinases in cluster 12, which also have a threonine as gatekeeper and a cysteine at position 33. Another example is the seven kinases that are in class 11 and have a valine for gatekeeper and a polar residue at position 24 which can act as hydrogen bond donor and/or acceptor. Based on these features, one can design inhibitors that can fully explore the hydrophobic pocket and also form hydrogen bonds to the residue at position 24.

Application of kinase selectivity potential: selective inhibitors of EphB4 The kinases domain of the Ephrin receptor B4 (EphB4) is an attractive target to develop type I inhibitors as it belongs to cluster 9, a small cluster of only eight members (Fig. 4). In 2008, Bardelle and coworkers reported a 2,4-bis-anilinopyrimidine derivative as a remarkably selective inhibitor of EphB4 (Bardelle *et al.*, 2008). Recently, using high-throughput docking (Kolb *et al.*, 2008) and hit improvement by medicinal chemistry (Lafleur *et al.*, 2009), we have discovered the single-digit nanomolar EphB4 inhibitor **10** (see Supplementary Material, Fig. S1). Notably, the selectivity profile of compound **10** indicates that only 5 of 85 protein kinases are inhibited significantly at 1 μM (<20% activity remained compared with a DMSO control). Three of them belong to cluster 7 (1 bit different from EphB4), and two of them belong to cluster 9. Furthermore, compound **11** (Supplementary Material) is a low micromolar inhibitor of EphB4 discovered in a recent high-throughput docking campaign (T.Zhou *et al.*, manuscript in preparation). Its selectivity profile was tested against 85 protein kinases at a single inhibitor concentration of 10 μM. Only for three kinases <50% activity remained (compared with a DMSO control) indicating that compound **11** is selective. These results together with the example of p38α provide evidence that the choice of a kinase in a small cluster can simplify significantly, or even bypass, the selectivity optimization process.

Fingerprint extension The current network analysis of selectivity potential was performed using only nine selectivity features, which is far from complete. However, a general trend emerges that by utilizing only a few key descriptors it is feasible to develop selective inhibitors targeting the ATP binding site for a subset of the kinases. Note that the profile of inhibition of a subset of kinases belonging to the same cluster can still be different due to several reasons, e.g. kinase-dependent binding site flexibility or other selectivity features not used in the current clustering. The former is usually difficult to explore but can be used to devise unique strategy for certain kinases, examples include Lapatinib bound to the C-helix-out from of EGFR (Bikker *et al.*, 2009; Wood *et al.*, 2004) and JNK3 whose methionine gatekeeper can be rearranged to allow access to the hydrophobic pocket (Scapin *et al.*, 2003; Swahn *et al.*, 2005). These features could be used to extend the fingerprint to obtain a more fine-grained classification.

4 CONCLUSIONS

The human kinases are aligned based on the primary and tertiary structures of their ATP binding site. Using this alignment the strategies for designing selective inhibitors targeting the ATP binding site are encoded in a 9-bit fingerprint. In this way, the kinases and their ATP binding site similarity can be treated as nodes and links of a network, respectively. The network analysis reveals the potential for selective inhibition of each kinase providing useful information for target prioritization and structure-based drug design. Finally, the analysis suggests that the tyrosine kinase EphB4 has high selectivity potential, which is consistent with the selectivity profiles of two novel inhibitors of EphB4 (compounds **10** and **11** in Supplementary Fig. S1).

ACKNOWLEDGEMENTS

We thank Dr Stefanie Muff for critical reading of the manuscript, and Armin Widmer (Novartis Pharma, Basel) for providing the modeling program Witnotp which was used for structural superpositions. We thank Dr Stefan Knapp for providing a list of unique kinase PDB structures.

Funding: Swiss National Science Foundation (grant 31003A_122442 to D.H.); Sino-Swiss Science and Technology Cooperation Joint Research Projects No. IZL CZ3 123945.

Conflict of Interest: none declared.

REFERENCES

Bain,J. *et al*. (2007) The selectivity of protein kinase inhibitors: a further update. *Biochem J.*, **408**, 297–315.

Bamborough,P. *et al*. (2008) Assessment of chemical coverage of kinome space and its implications for kinase drug discovery. *J. Med. Chem.*, **51**, 7898–7914.

Bardelle,C. *et al*. (2008) Inhibitors of the tyrosine kinase EphB4. Part 1: structure-based design and optimization of a series of 2,4-bis-anilinopyrimidines. *Bioorg. Med. Chem. Lett.*, **18**, 2776–2780.

Battistutta,R. *et al*. (2000) The replacement of ATP by the competitive inhibitor emodin induces conformational modifications in the catalytic site of protein kinase CK2. *J. Biol. Chem.*, **275**, 29618–29622.

Bikker,J.A. *et al*. (2009) Kinase domain mutations in cancer: implications for small molecule drug design strategies. *J. Med. Chem.,* **52**, 1493–1509.

Biondi,R.M. (2004) Phosphoinositide-dependent protein kinase 1, a sensor of protein conformation. *Trends Biochem. Sci.*, **29**, 136–142.

Blair,J.A. *et al*. (2007) Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol.*, **3**, 229–238.

Blencke,S. *et al*. (2003) Mutation of threonine 766 in the epidermal growth factor receptor reveals a hotspot for resistance formation against selective tyrosine kinase inhibitors. *J. Biol. Chem.*, **278**, 15435–15440.

Bogoyevitch,M.A. and Fairlie,D.P. (2007) A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug Discov. Today,* **12**, 622–633.

Chen,J. *et al*. (2007) Molecular basis for specificity in the druggable kinome: sequence-based analysis. *Bioinformatics,* **23**, 563–572.
Cohen,F.E. *et al*. (1994) Structural clues to prion replication. *Science,* **264**, 530–531.
Cohen,P. (2002) Protein kinases–the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.*, **1**, 309–315.
Deininger,M.W.N. and Druker,B.J. (2003) Specific targeted therapy of chronic myelogenous leukemia with imatinib. *Pharmacol. Rev.*, **55**, 401–423.
Dibb,N.J. *et al*. (2004) Switching on kinases: oncogenic activation of BRAF and the PDGFR family. *Nat. Rev. Cancer*, **4**, 718–727.
Eyers,P.A. *et al*. (1998) Conversion of SB 203580-insensitive MAP kinase family members to drug-sensitive forms by a single amino-acid substitution. *Chem. Biol.*, **5**, 321–328.
Fabian,M.A. *et al*. (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.*, **23**, 329–336.
Fedorov,O. *et al*. (2007) A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc. Natl Acad. Sci. USA*, **104**, 20523–20528.
Fitzgerald,C.E. *et al*. (2003) Structural basis for p38alpha MAP kinase quinazolinone and pyridol-pyrimidine inhibitor specificity. *Nat. Struct. Biol.*, **10**, 764–769.
Fruchterman,T.M.J. and Reingold,E.M. (1991) Graph drawing by force-directed placement. *Software Pract. Exper.,* **21**, 1129–1164.
Gerstein,M. and Levitt,M. (1998) Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci.*, **7**, 445–456.
Goldsmith,E.J. *et al*. (2007) Substrate and docking interactions in serine/threonine protein kinases. *Chem. Rev.*, **107**, 5065–5081.
Goldstein,D.M. *et al*. (2008) High-throughput kinase profiling as a platform for drug discovery. *Nat. Rev. Drug Discov.*, **7**, 391–397.
Gorre,M.E. *et al*. (2001) Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science*, **293**, 876–880.
Grunwald,V. *et al*. (2007) Managing side effects of angiogenesis inhibitors in renal cell carcinoma. *Onkologie*, **30**, 519–524.
Hanks,S.K. and Hunter,T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.*, **9**, 576–596.
Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA.*, **89**, 10915–10919.
Hopkins,A.L. (2008) Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.*, **4**, 682–690.
Jacobs,M. *et al*. (2006) The structure of dimeric ROCK I reveals the mechanism for ligand selectivity. *J. Biol. Chem.*, **281**, 260–268.
Karaman,M.W. *et al*. (2008) A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.*, **26**, 127–132.
Knight,Z.A. and Shokat,K.M. (2005) Features of selective kinase inhibitors. *Chem. Biol.*, **12**, 621–637.
Kolb,P. et al. (2008) Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Prot. Struct. Funct. Bioinform.*, **73**, 11–18.
Lafleur,K. *et al*. (2009) Structure-based optimization of potent and selective inhibitors of the tyrosine kinase Ephb4. *J. Med. Chem.*, **52**, 6433–6446.
Laird,A.D. *et al*. (2000) SU6668 is a potent antiangiogenic and antitumor agent that induces regression of established tumors. *Cancer Res.*, **60**, 4152–4160.
Laufer,S.A. *et al*. (2008) Design, synthesis, and biological evaluation of novel Tri- and tetrasubstituted imidazoles as highly potent and specific ATP-mimetic inhibitors of p38 MAP kinase: focus on optimized interactions with the enzyme's surface-exposed front region. *J. Med. Chem.*, **51**, 4122–4149.
Liao,J.J.L. (2007) Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.*, **50**, 409–424.
Liu,Y. *et al*. (1999) Structural basis for selective inhibition of Src family kinases by PP1. *Chem. Biol.*, **6**, 671–678.
Liu,Y. and Gray,N.S. (2006) Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.*, **2**, 358–364.
Manning,G. *et al*. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
Newwan,M. (2003) Graph drawing by force-directed placement. *SIAM Rev.*, **45**, 167–256.
Noble,M.E.M. *et al*. (2004) Protein kinase inhibitors: insights into drug design from structure. *Science,* **303**, 1800–1805.
Pargellis,C. *et al*. (2002) Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.*, **9**, 268–272.
Richardson,C.M. *et al*. (2006) Triazolo[1,5-a]pyrimidines as novel CDK2 inhibitors: protein structure-guided design and SAR. *Bioorg. Med. Chem. Lett.*, **16**, 1353–1357.
Sawyers,C. (2004) Targeted cancer therapy. *Nature,* **432**, 294–297.
Scapin,G. *et al*. (2003) The structure of JNK3 in complex with small molecule inhibitors: structural basis for potency and selectivity. *Chem. Biol.,* **10**, 705–712.
Shah,K. *et al*. (1997) Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl Acad. Sci. USA,* **94**, 3565–3570.
Sheinerman,F.B. *et al*. (2005) High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J. Mol. Biol.*, **352**, 1134–1156.
Smaill,J.B. *et al*. (2000) Tyrosine kinase inhibitors. 17. Irreversible inhibitors of the epidermal growth factor receptor: 4-(phenylamino)quinazoline- and 4-(phenylamino)pyrido[3,2-d]pyrimidine-6-acrylamides bearing additional solubilizing functions. *J. Med. Chem.*, **43**, 1380–1397.
Swahn,B.M. *et al*. (2005) Design and synthesis of 6-anilinoindazoles as selective inhibitors of c-Jun N-terminal kinase-3. *Bioorg. Med. Chem. Lett.*, **15**, 5095–5099.
Talpaz,M. *et al*. (2006) Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias. *N. Engl. J. Med.*, **354**, 2531–2541.
Thompson,J.D. *et al*. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
Torrance,C.J. *et al*. (2000) Combinatorial chemoprevention of intestinal neoplasia. *Nat. Med.* **6**, 1024–1028.
Traxler,P. and Furet,P. (1999) Strategies toward the design of novel and selective protein tyrosine kinase inhibitors. *Pharmacol. Ther.*, **82**, 195–206.
Vieth,M. *et al*. (2004) Kinomics-structural biology and chemogenomics of kinase inhibitors and targets. *Biochim. Biophys. Acta,* **1697**, 243–257.
Vieth,M. *et al*. (2005) Kinomics:characterizing the therapeutically validated kinase space. *Drug Discov. Today*, **10**, 839–846.
Vulpetti,A. and Bosotti,R. (2004) Sequence and structural analysis of kinase ATP pocket residues. *Farmaco*, **59**, 759–765.
Wolter,P. *et al*. (2008) The clinical implications of sunitinib-induced hypothyroidism: a prospective evaluation. *Br. J. Cancer*, **99**, 448–454.
Wood,E.R. *et al*. (2004) 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.*, **64**, 6652–6659.
Xing,L. *et al*. (2009) Structural bioinformatics-based prediction of exceptional selectivity of p38 MAP kinase inhibitor PH-797804. *Biochemistry*, **48**, 6402–6411.
Yde,C.W. *et al*. (2005) Inclining the purine base binding plane in protein kinase CK2 by exchanging the flanking side-chains generates a preference for ATP as a cosubstrate. *J. Mol. Biol.*, **347**, 399–414.