ACS Medicinal Chemistry Letters

Letter

pubs.acs.org/acsmedchemlett

Targeted Kinase Selectivity from Kinase Profiling Data

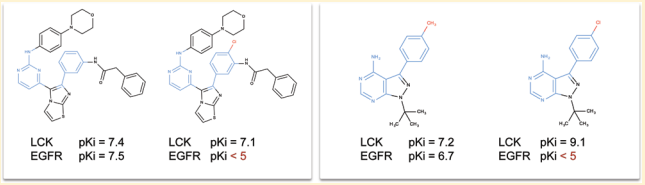
Francesca Milletti*,[†] and Johannes C. Hermann[‡]

[†]pRED Informatics, Roche, 340 Kingsland Street, Nutley, New Jersey 07110, United States

[‡]Discovery Chemistry, Roche, 340 Kingsland Street, Nutley, New Jersey 07110, United States

Supporting Information

ABSTRACT: Kinase selectivity plays a major role in the design strategy of lead series and in the ultimate success of kinase drug discovery programs. Although profiling compounds against a large panel of protein kinases has become a standard part of modern drug discovery, data accumulated from these kinase panels may be underutilized for new kinase projects. We present a method that can be used to optimize the selectivity profile of a compound using historical kinase profiling data. This method proposes chemical transformations based on pairs of very similar compounds, which are both active against a desired target kinase and differ in activity against another kinase. We show that these transformations are transferable across scaffolds, thus making this tool valuable to exploit kinase profiling data for unrelated series of compounds.



KEYWORDS: *Kinase inhibitors, kinase selectivity, kinase panels, activity cliffs, matched pairs, computational method*

In recent years much research has been done in the area of kinase inhibitors. Even though the numerous clinical successes have proven that kinase inhibitors can be sufficiently selective, achieving selectivity is still challenging.

Selectivity is often achieved by targeting specific inactive states of the kinase, such as the DFG-out or the αC-helix-out state.[1,2] These states did not evolve to recognize ATP and are often sufficiently different across the kinome.[3] Selectivity can also be increased by targeting specific subpockets such as the hydrophobic pocket behind the "gatekeeper" residue or solvent exposed residues outside the ATP binding site.[4,5]

Selectivity within a specific family of kinases is often critical to the success of a project. For example, several pharmaceutical companies have developed selective inhibitors within the JAK family of kinases, which includes JAK1, JAK2, JAK3, and Tyk2. However, for these kinases with highly similar binding sites, it is very hard to achieve selectivity using a structure-based approach.[6−11]

Structure-based design is not optimal to discover certain selectivity mechanisms that are driven by protein dynamic effects or rearrangements of water networks. The energetics of displacing waters is not well captured in the crystal structure and is hard to predict. As a result, many selectivity mechanisms are discovered by chance and often remain unexplained. For example, Imatinib hits c-SRC with an affinity that is at least 2000-fold lower than that for ABL2, although the binding sites of Abl2 and SRC are nearly identical even for the DFG-out state.[12]

Kinase profiling data often contain many compounds from the same series. This makes it possible to discover chemical changes that lead to better selectivity. For example, Figure 1 shows one compound (a) active on both KIT and ABL1 and another (b) active on KIT only. Note that the two compounds only differ by one chemical change (F → CH_3).

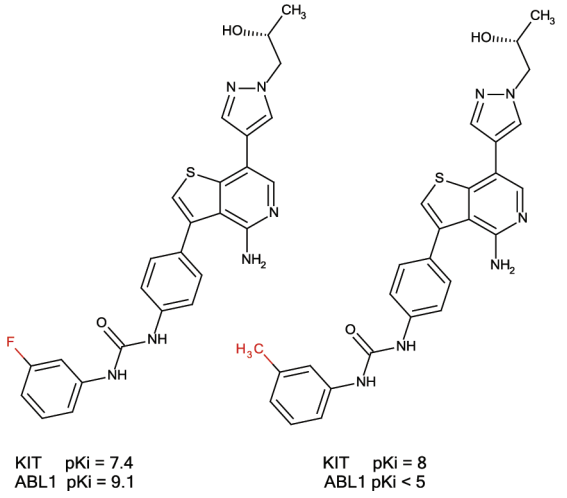


Figure 1. Chemical transformation proposed to increase selectivity for KIT with respect to ABL1 (the two compounds were taken from the database published by Metz et al.).

With more than 8000 K_d and more than 600,000% inhibition data points for kinases, the Roche internal kinase profiling data is very extensive. However, many data are also available in the public domain. For example, the data set published by Metz et al.[13] contains about 150,000 kinase inhibitory values, and the Kinase SARfari data set,[14] which is freely available from the web, has more than 430,000 kinase bioactivity data points.

Received: January 10, 2012
Accepted: March 14, 2012
Published: March 14, 2012

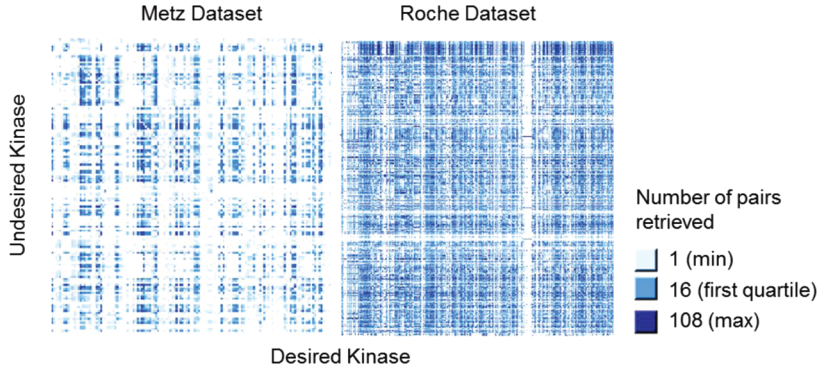


Figure 2. Heat map for the 172 kinases in the Metz public panel and for the 400 kinases in the Roche kinase inhibitor profile database representing the number of chemical transformations available for each pair of kinase. Darker colors reflect a higher number of compound pairs found, with white indicating that no pair was found.

To use the information contained in these databases, we developed a program that identifies chemical transformations to achieve selectivity against a specific unwanted kinase while maintaining activity for the target kinase. The method presented has similarities with the concept of "matched molecular pairs"[15,16] and "activity cliffs",[17,18] which center the analysis on pairs of molecules which are very similar (or have a common substructure) but have a large change in activity. In this paper we express binding affinities as pK_i (inhibition constant) and we consider active all compounds with $pK_i > 7$ and inactive all compounds with $pK_i < 5$.

The purpose of the program is to generate a list of compound pairs with the type of activity cliff shown in Figure 1. The program first clusters compounds by molecular similarity. Then it identifies pairs of very similar compounds within the same cluster. Depending on the similarity threshold, the pairs are separated by one (or rarely more) chemical transformation.

This type of approach is much faster than a maximal common substructure (MCS) search method. In addition, it can retrieve valuable pairs that would otherwise be excluded by MCS. However, a similarity approach is sensitive to the similarity threshold selected. Lowering the Tanimoto threshold increases the number of pairs found to a point where the pair list is too long and less meaningful.

The identified molecular pairs are valuable, especially when the binding mode of the compounds compared is known. This is the case for many public and in-house kinase inhibitors. If the binding mode is known, the position of the chemical transformation suggests where an analogous transformation on a different series may also be useful to improve selectivity.

We illustrate this approach by using examples derived from a public data set published by Metz et al.[13] which contains more than 3800 compounds tested against 172 protein kinases. Although this data set contains only a fraction of all kinase inhibition data available within large pharmaceutical companies such as Roche, its coverage is still good for illustration purposes. Here we define coverage as the number of target/undesired pairs with at least one chemical transformation divided by the total number of pairs.

The heat map in Figure 2 shows the coverage for the Metz data set and for the Roche data set, with darker colors corresponding to more pairs of compounds found for a given undesired kinase and target kinase. The heat map is based on 172 kinases for the Metz data set and on 400 kinases for the Roche data set (1919 compounds).[19] Taking into account the full set of 517 kinases, the coverage is 2.5% for the Metz database and 45% for the Roche database.

It should be noted that these results depend on the Tanimoto similarity cutoff chosen for a pair (in this example, we used 0.5). Although 0.5 is quite low and suggests that very different structures might be treated as pairs, we found that this is an acceptable threshold because useful pairs are retrieved also at low Tanimoto similarity. The next section presents three case studies based on different pairs of kinases. Since a similarity cutoff of 0.5 produced a very long list of pairs, higher cutoffs were used for practical reasons.

Case Study 1: CDC2 (Undesired Kinase)/CDC7 (Target Kinase). Considering only pairs with Tanimoto similarity >0.75, we retrieved 27 pairs of compounds (Supporting Information Table 1) where one is active on both CDC2 and CDC7, while the other is active only on CDC7 (target kinase). The 27 pairs correspond to three different clusters, as shown in Supporting Information Table 1. Even though no PDB structure is available for either CDC2 or CDC7, the binding modes can be predicted with reasonable certainty from other protein kinases cocrystallized with ligands with the same or very similar scaffolds (Figure 3). As shown in Figure 3, the binding mode for each scaffold shows that the R-group that drives

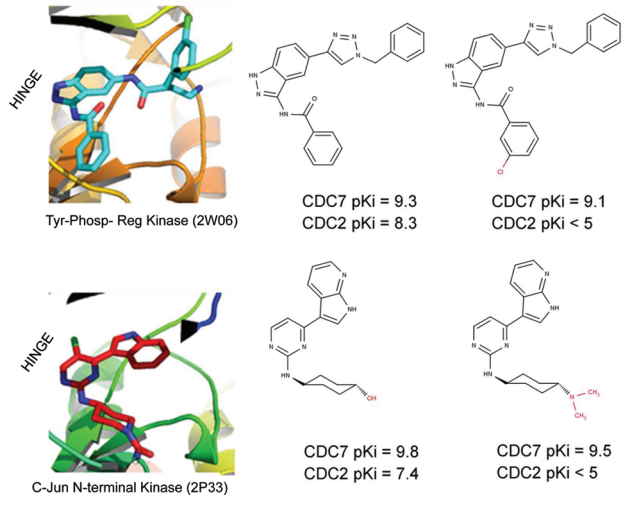


Figure 3. Two examples with chemical transformation proposed to increase selectivity for CDC7 with respect to CDC2. The PDBs shown suggest the binding mode for these compounds based on the binding of their cocrystallized ligands.

dx.doi.org/10.1021/ml300012r | ACS Med. Chem. Lett. 2012, 3, 383−386

selectivity lies in the front pocket. The front pocket contains flexible residues that interact with water molecules and that are challenging to consider for predicting selectivity.

Case Study 2: LCK (Target Kinase)/EGFR (Undesired Kinase). For this pair of kinases, we found examples where the same chemical transformation leads to better selectivity in two different scaffolds. Figure 4 shows the 3D superimposition of

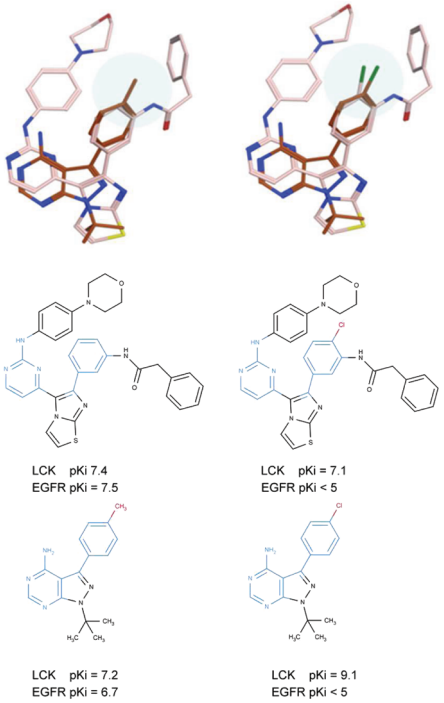


Figure 4. The chlorine atom leads to better selectivity for LCK with respect to EGRF in two completely different scaffolds. The superimposed structures were obtained by 3D molecular alignment using MOE.

compounds active on both LCK and EGFR (left) and of compounds active on LCK only (right). The chlorine atom in the analogous position leads to LCK selectivity for the two different scaffolds.

Case Study 3: CSFR1 (Target Kinase)/PDGFRA (Undesired Kinase). Setting CSF1R as target kinase and PDGFRA as undesired kinase yielded 16 pairs with similarity above 0.75, with some representatives shown in Figure 5. The binding mode proposed here suggests that interactions in the front pocket may drive selectivity between these two kinases. Transformations in R1 show that a more bulky substituent might be important for selectivity.

Although the pairs found may highlight differences that are difficult to transfer to another scaffold, the concept to achieve selectivity can be extracted and applied to a new scaffold. For example, the pairs in Figure 5 suggest that bulkier substituents tend to increase selectivity.

CONCLUSIONS

Structure-based design is traditionally used to improve kinase selectivity. However, chemical transformations that increase selectivity are often found by chance, and in many cases the mechanism remains unexplained.

The increasing amount of kinase profiling data contains a vast number of nonobvious transformations that lead to

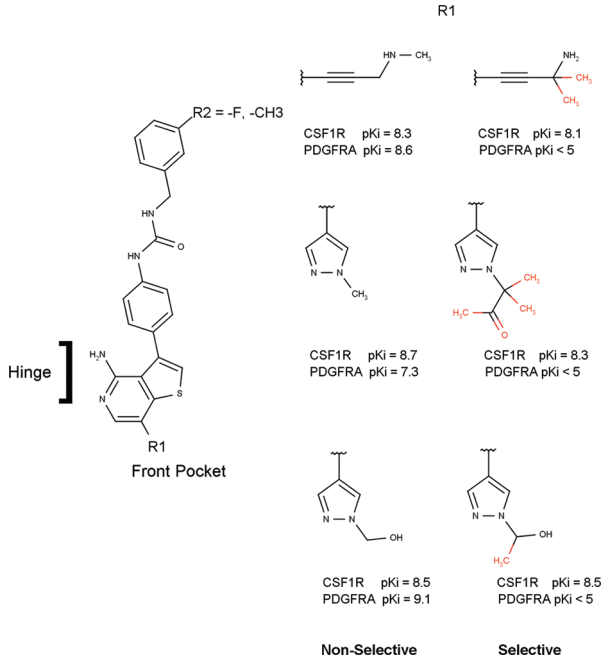


Figure 5. Transformations leading to higher selectivity for CSFR1 (over PDGFRA). The proposed binding mode is based on the crystal structure 2WGJ, where the ligand is similar to the scaffold shown here.

improved selectivity, and data mining efforts may help to discover these hidden design principles. The method we propose may help to design more selective compounds by suggesting chemical transformations—in the form of compound pairs—that lead to improved selectivity with respect to a specific undesired kinase.

Chemists routinely extract selectivity rules from various series within a project. The method presented complements this effort by systematically mining and analyzing large kinase data sets.

The examples presented show that this method enables the design of more selective compounds by extracting information for different, but superimposable scaffolds. This is especially useful for discovering nonobvious changes that increase selectivity. The method we propose can capture these changes, thereby complementing structure based design approaches for kinase selectivity.

EXPERIMENTAL PROCEDURES

K_i values corresponding to more than 3800 compounds tested on 172 kinases were extracted from the kinase inhibitor database published by Metz et al. A pipeline pilot protocol was developed to identify all chemical transformations that cause at least a 100-fold drop in activity against a kinase, while they leave the activity for the other kinase substantially unchanged. The protocol retrieves the chemical transformations, with the names of target kinase and undesired kinase; it analyses all compounds within a cluster—the compounds in the Metz database are already divided into clusters on the basis of their structural similarity—and for each cluster extracts the n compound pairs with a Tanimoto similarity higher than a selected threshold, generally higher than 0.5. This similarity score was obtained by using the Fingerprint Similarity NXN with Tanimoto as similarity coefficient and ECFP_6 as descriptors. Lastly, the protocol implements a filter that retains only pairs of compounds where one is active ($K_i > 7$) against both the target and the undesired kinase, while the other is only active on the target kinase ($K_i > 7$) but inactive on the undesired kinase ($K_i < 5$). These thresholds can be modified in the program. Because the compounds in each pair are very similar and belong to a

dx.doi.org/10.1021/ml300012r | ACS Med. Chem. Lett. 2012, 3, 383−386

common cluster, they contain a common substructure. It should be noted that the number of pairs retrieved depends on the similarity index chosen. For example, by limiting results to pairs with a similarity index > 0.75, with respect to the 517 kinases, the Metz database would have only 1% coverage and the Roche database only 22%.

■ ASSOCIATED CONTENT

Supporting Information

The pipeline pilot protocol in xml format, and the list of pairs found for the three case studies. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: francesca.milletti@roche.com.

Author Contributions

J.C.H. conceived the idea of using compound pairs for predicting kinase selectivity. F.M. designed and executed the computational methodology. J.C.H. and F.M. wrote the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Matthew Lucas and Michael Soth for their helpful feedback and discussion, and the reviewers for their help in improving this manuscript.

■ REFERENCES

(1) Liu, Y.; Gray, N. S. Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2* (7), 358−364.

(2) Jacobs, M. D.; Caron, P. R.; Hare, B. J. Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: Structure of lck/imatinib complex. *Proteins* **2008**, *70* (4), 1451−1460.

(3) Rabiller, M.; Getlik, M.; Klüter, S.; Richters, A.; Tückmantel, S.; Simard, J. R.; Rauh, D. Proteus in the World of Proteins: Conformational Changes in Protein Kinases. *Archiv. Pharm.* **2010**, *343* (4), 193−206.

(4) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. Protein Kinase Inhibitors: Insights into Drug Design from Structure. *Science* **2004**, *5665* (303), 1800−1805.

(5) Goldstein, D. M.; Kuglstatter, A.; Lou, Y.; Soth, M. J. Selective p38a Inhibitors Clinically Evaluated for the Treatment of Chronic Inflammatory Disorders. *J. Med. Chem.* **2010**, *53* (6), 2345−2353.

(6) Garber, K. Pfizer's JAK inhibitor sails through phase 3 in rheumatoid arthritis. *Nat. Biotechnol.* **2011**, *29* (6), 467−468.

(7) Lin, T. H.; et al. Selective functional inhibition of JAK-3 is sufficient for efficacy in collagen-induced arthritis in mice. *Arthritis Rheum.* **2010**, *62* (8), 2283−2293.

(8) Malerich, J. P.; Lam, J. S.; Hart, B.; Fine, R. M.; Klebansky, B.; Tanga, M. J.; D'Andrea, A. Diamino-1,2,4-triazole derivatives are selective inhibitors of TYK2 and JAK1 over JAK2 and JAK3. *Bioorg. Med. Chem. Lett.* **2010**, *20* (24), 7454−7457.

(9) Pardanani, A.; Lasho, T.; Smith, G.; Burns, C. J.; Fantino, E.; Tefferi, A. CYT387, a selective JAK1/JAK2 inhibitor: in vitro assessment of kinase selectivity and preclinical studies using cell lines and primary cells from polycythemia vera patients. *Leukemia* **2009**, *23* (8), 1441−1445.

(10) Tsui, V.; Gibbons, P.; Ultsch, M.; Mortara, K.; Chang, C.; Blair, W.; Pulk, R.; Stanley, M.; Starovasnik, M.; Williams, D.; Lamers, M.; Leonard, P.; Magnuson, S.; Liang, J.; Eigenbrot, C. A new regulatory switch in a JAK protein kinase. *Proteins* **2011**, *79* (2), 393−401.

(11) Harikrishnan, L. S.; et al. Pyrrolo[1,2-f]triazines as JAK2 inhibitors: Achieving potency and selectivity for JAK2 over JAK3. *Bioorg. Med. Chem. Lett.* **2011**, *21* (5), 1425−1428.

(12) Seeliger, M. A.; Nagar, B.; Frank, F.; Cao, X.; Henderson, M. N.; Kuriyan, J. c-Src Binds to the Cancer Drug Imatinib with an Inactive Abl/c-Kit Conformation and a Distributed Thermodynamic Penalty. *Structure* **2007**, *15* (3), 299−311.

(13) Metz, J. T.; Johnson, E. F.; Soni, N. B.; Merta, P. J.; Kifle, L.; Hajduk, P. J. Navigating the kinome. *Nat. Chem. Biol.* **2011**, *7* (4), 200−202.

(14) Kinase SARfari, https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari (Accessed Dec 2011).

(15) Leach, A. G.; Jones, H. D.; Cosgrove, D. A.; Kenny, P. W.; Ruston, L.; MacFaul, P.; Wood, J. M.; Colclough, N.; Law, B. Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties; a Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure. *J. Med. Chem.* **2006**, *49* (23), 6672−6682.

(16) Papadatos, G.; Alkarouri, M.; Gillet, V. J.; Willett, P. Lead Optimization Using Matched Molecular Pairs: Inclusion of Contextual Information for Enhanced Prediction of hERG Inhibition, Solubility, and Lipophilicity. *J. Chem. Inf. Model.* **2010**, *50* (10), 1872−1886.

(17) Guha, R.; Van Drie, J. H. Structure−Activity Landscape Index: Identifying and Quantifying Activity Cliffs. *J. Chem. Inf. Model.* **2008**, *48* (3), 646−658.

(18) Wassermann, A. M.; Bajorath, J. Chemical Substitutions That Introduce Activity Cliffs Across Different Compound Classes and Biological Targets. *J. Chem. Inf. Model.* **2011**, *50* (7), 1248−1256.

(19) Ambit KINOMEscan—High-Throughput Kinase Selectivity Profiling, www.kinomescan.com (Accessed Dec 2011).

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published on the Web on March 14, 2012, with an error in the Supporting Information. Changes were made to the pipeline pilot protocol, to fix an error that caused the same image for both columns, and to the PDF files, which contain the output from the pipeline pilot protocol. The corrected version was reposted on March 26, 2012.