articles

molecular pharmaceutics

Cross-Reactivity Virtual Profiling of the Human Kinome by X-React^{KIN}: A Chemical Systems Biology Approach

Michal Brylinski and Jeffrey Skolnick*

Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, Georgia

Received August 31, 2010; Revised Manuscript Received September 30, 2010; Accepted October 19, 2010

Abstract: Many drug candidates fail in clinical development due to their insufficient selectivity that may cause undesired side effects. Therefore, modern drug discovery is routinely supported by computational techniques, which can identify alternate molecular targets with a significant potential for cross-reactivity. In particular, the development of highly selective kinase inhibitors is complicated by the strong conservation of the ATP-binding site across the kinase family. In this paper, we describe X-React^{KIN}, a new machine learning approach that extends the modeling and virtual screening of individual protein kinases to a system level in order to construct a cross-reactivity virtual profile for the human kinome. To maximize the coverage of the kinome, X-React^{KIN} relies solely on the predicted target structures and employs state-of-the-art modeling techniques. Benchmark tests carried out against available selectivity data from high-throughput kinase profiling experiments demonstrate that, for almost 70% of the inhibitors, their alternate molecular targets can be effectively identified in the human kinome with a high (>0.5) sensitivity at the expense of a relatively low false positive rate (<0.5). Furthermore, in a case study, we demonstrate how X-React^{KIN} can support the development of selective inhibitors by optimizing the selection of kinase targets for small-scale counter-screen experiments. The constructed cross-reactivity profiles for the human kinome are freely available to the academic community at http://cssb.biology.gatech.edu/kinomelhm/.

Keywords: X-React^{KIN}; human kinome; kinase functional space; kinome structural coverage; kinase inhibitors; drug development; drug off-targets; Chemical Systems Biology

Introduction

The human kinome, one of the largest families in the human proteome, comprises >500 genes.[1] The pivotal function of kinases is the signal transduction through a reversible phosphorylation of tyrosine, threonine and serine residues in other proteins.[2,3] The strong implication of kinase activity in numerous disease states such as cancer,[4] diabetes,[5] inflammation,[6] multiple sclerosis,[7] cardiovascular disease[8] and neurological dysfunctions[9] makes them very important drug targets. Consequently, there is a growing interest in the development of novel compounds with kinase inhibition as their mode of action;[10−12] this has resulted in over a hundred kinase crystal structures complexed with low-molecular-weight inhibitors reported in the public domain.[13]

Many therapeutic strategies have been developed to modulate kinase activity.[14] The most prevalent is kinase

* Corresponding author. Mailing address: Georgia Institute of Technology, Center for the Study of Systems Biology, 250 14th Street NW, Atlanta, GA 30318. Tel: 404 407 8989. Fax: 404 385 7478. E-mail: skolnick@gatech.edu.

(1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298* (5600), 1912–34.

(2) Hanks, S. K.; Hunter, T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **1995**, *9* (8), 576–96.

(3) Kennelly, P. J. Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS Microbiol. Lett.* **2002**, *206* (1), 1–8.

(4) Blume-Jensen, P.; Hunter, T. Oncogenic kinase signalling. *Nature* **2001**, *411* (6835), 355–65.

(5) Sasase, T. PKC - a target for treating diabetic complications. *Drugs Future* **2006**, *31* (6), 503–11.

10.1021/mp1002976
Published on Web 10/19/2010

inhibition by targeting the catalytic site of kinases with ATP-competitive inhibitors.[15] The ATP-binding site provides a compelling environment for binding a diverse range of organic molecules devised to compete with ATP, mostly by mimicking the binding interactions of the adenosine moiety.[16] Indeed, ATP-binding pockets are the primary target sites for the majority of the currently available kinase inhibitors.[17] However, the structural and chemical features of the ATP-binding site as well as the catalytic mechanism are highly conserved across the kinase family, which significantly complicates the development of kinase inhibitors with sufficient target selectivity.

To address this significant issue, a number of computational techniques have been developed to support experimental efforts directed toward the development of selective kinase inhibitors. Most employ various classification schemas for the kinase space with the underlying assumption that kinases belonging to a common category have higher potential to bind similar compounds, which may give rise to undesired cross-reactivity effects. The most straightforward approach to the classification of kinases is based on the global sequence or/and structure similarity. A comprehensive survey carried out for all available kinase sequences classified them into 30 distinct families, with 19 of them covering nearly 98% of all sequences and representing seven general structural folds.[18] Nevertheless, it has been demonstrated that a high probability of being inhibited by the same groups of compounds requires very high sequence identity thresholds, typically more than 50−60%.[19−21] However, the average pairwise global sequence identity in the human kinome is ∼25%; those kinase pairs with a sequence identity of 50−60% and less might or might not have similar pharmacological profiles.

In that regard, alternative approaches are required. A new method was proposed to classify the medicinally relevant kinase space based on structure−activity relationship, SAR, profiles.[22] Results obtained for 38 crystal structures of protein kinases and available small molecule inhibition data showed that the SAR-based dendograms differ significantly from the sequence-based clustering for distantly homologous targets. Another approach exploits structure comparison of kinases based on a feature-similarity matrix.[23] This new metric is well correlated with a pharmacological distance generated by comparing affinity fingerprints constructed from experimental cross-reactivity profiles. An interesting study reported recently employs the QSAR analysis of residue contributions to the kinase inhibition profile.[24] Using various experimental data sets, binding profiles are constructed based on the properties of 29 residues in the active site, which can be applied to predict binding similarities for untested kinases. Other chemical/structure-based classifications of ATP-binding sites in protein kinases are based on target family landscapes constructed using molecular interaction field analysis,[25] exposed physicochemical properties of the active

(6) Müller, S.; Knapp, S. Targeting kinases for the treatment of inflammatory diseases. *Expert Opin. Drug Discovery* **2010**, *5* (9), 867–81.

(7) Saarela, J.; Kallio, S. P.; Chen, D.; Montpetit, A.; Jokiaho, A.; Choi, E.; Asselta, R.; Bronnikov, D.; Lincoln, M. R.; Sadovnick, A. D.; Tienari, P. J.; Koivisto, K.; Palotie, A.; Ebers, G. C.; Hudson, T. J.; Peltonen, L. PRKCA and multiple sclerosis: association in two independent populations. *PLoS Genet.* **2006**, *2* (3), e42.

(8) Bynagari-Settipalli, Y. S.; Chari, R.; Kilpatrick, L.; Kunapuli, S. P. Protein Kinase C - Possible Therapeutic Target to Treat Cardiovascular Diseases. *Cardiovasc. Hematol. Disord. Drug Targets* **2010**. DOI: 10.2174/87121020678065529X.

(9) Mueller, B. K.; Mack, H.; Teusch, N. Rho kinase, a promising drug target for neurological disorders. *Nat. Rev. Drug Discovery* **2005**, *4* (5), 387–98.

(10) Cohen, P. The development and therapeutic potential of protein kinase inhibitors. *Curr. Opin. Chem. Biol.* **1999**, *3* (4), 459–65.

(11) Johnson, L. Protein kinases and their therapeutic exploitation. *Biochem. Soc. Trans.* **2007**, *35* (Part 1), 7–11.

(12) Weinmann, H.; Metternich, R. Drug discovery process for kinase inhibitors. *ChemBioChem* **2005**, *6* (3), 455–9.

(13) Noble, M. E.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **2004**, *303* (5665), 1800–5.

(14) McInnes, C.; Fischer, P. M. Strategies for the design of potent and selective kinase inhibitors. *Curr. Pharm. Des.* **2005**, *11* (14), 1845–63.

(15) Sawa, M. Strategies for the design of selective protein kinase inhibitors. *Mini-Rev. Med. Chem.* **2008**, *8* (12), 1291–7.

(16) Liao, J. J. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50* (3), 409–24.

(17) Stout, T. J.; Foster, P. G.; Matthews, D. J. High-throughput structural biology in drug discovery: protein kinases. *Curr. Pharm. Des.* **2004**, *10* (10), 1069–82.

(18) Cheek, S.; Zhang, H.; Grishin, N. V. Sequence and structure classification of kinases. *J. Mol. Biol.* **2002**, *320* (4), 855–81.

(19) 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* (1−2), 243–57.

(20) Vieth, M.; Sutherland, J. J.; Robertson, D. H.; Campbell, R. M. Kinomics: characterizing the therapeutically validated kinase space. *Drug Discovery Today* **2005**, *10* (12), 839–46.

(21) Bamborough, P.; Drewry, D.; Harper, G.; Smith, G. K.; Schneider, K. Assessment of chemical coverage of kinome space and its implications for kinase drug discovery. *J. Med. Chem.* **2008**, *51* (24), 7898–914.

(22) Frye, S. V. Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* **1999**, *6* (1), R3–7.

(23) Zhang, X.; Fernandez, A. In silico drug profiling of the human kinome based on a molecular marker for cross reactivity. *Mol. Pharmaceutics* **2008**, *5* (5), 728–38.

(24) Sheridan, R. P.; Nam, K.; Maiorov, V. N.; McMasters, D. R.; Cornell, W. D. QSAR models for predicting the similarity in binding profiles for pairs of protein kinases and the variation of models between experimental data sets. *J. Chem. Inf. Model.* **2009**, *49* (8), 1974–85.

(25) Naumann, T.; Matter, H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J. Med. Chem.* **2002**, *45* (12), 2366–78.

sites calculated by Cavbase,[26] geometric hashing algorithms[27] and binding site signatures created from "hot spot" residues.[28] These techniques have been shown to be relatively successful in the identification of protein kinase binding sites known experimentally to bind the same compound; however, they require high-resolution crystallographic structures of the target kinase proteins, preferably complexed with inhibitors. As a consequence, the covered kinase space remains incomplete because it is limited by the availability of experimentally solved crystal structures; this corresponds to only about 20% of the human kinome.

This gap can be bridged by protein structure prediction, particularly comparative modeling.[29,30] Current state-of-the-art protein structure prediction approaches have reached the level where they can construct protein models whose quality is often comparable to that of low-resolution experimentally determined structures.[31] Nevertheless, theoretically predicted protein structures may still have significant structural inaccuracies in their ligand binding regions;[32,33] this requires appropriate computational techniques that are different from those applicable to the crystal structures and that can accommodate structural distortions without significant loss in accuracy.

In our previous study, we described the results of the first proteome-scale structure modeling and virtual screening of the entire human kinome.[34] Using a template-based modeling procedure,[35,36] we constructed structural models for all kinase domains in humans. Subsequently, we applied a structure/evolution-based approach[37] to precisely detect target sites. These were then subjected to large-scale virtual screening against a large collection of commercially available compounds using a novel hierarchical approach that combines ligand- and structure-based filters.[38,39] Retrospective benchmarks against several commonly used ligand libraries demonstrate that predicted molecular interactions between kinases and small ligands substantially overlap with available experimental data. In this paper, we attempt to extend the modeling and virtual screening of individual protein kinases to the system level in order to construct a cross-reactivity virtual profile for the entire human kinome. To achieve this goal, we develop X-React^{KIN}, a machine learning approach that estimates the potential for cross-reactivity from sequence, structure and binding properties of the ATP-binding sites in protein kinases. We validate the results against available selectivity data from high-throughput kinase profiling experiments. Finally, we demonstrate how X-React^{KIN} can support the development of selective inhibitors by suggesting alternate targets for small-scale counter-screen experiments. The constructed cross-reactivity profiles for the human kinome are freely available to the academic community via a user-friendly web interface that can be accessed from http://cssb.biology.gatech.edu/kinomelhm/.

Methods

X-React^{KIN} Overview. Here, we use the concept of kinase family virtual profiling and compute the complete map of putative cross-interactions within the human kinome. We develop X-React^{KIN}, a machine learning approach that combines sequence, structure and ligand binding similarities of the ATP-binding sites in protein kinases to estimate the potential for cross-interactions. We note that these similarities are calculated using modeled protein structures and virtual screening ranking. We train a naïve Bayes classifier on the available inhibitor selectivity data to calculate a new probabilistic cross-reactivity score, called a CR-score. Based on the estimated similarities expressed by the CR-score values, we construct a cross-reactivity virtual profile that corresponds to the matrix of pairwise interactions within the complete human kinase family. Below, we describe the scoring functions used to construct the cross-reactivity probabilistic score, the details of the data sets and machine learning implementation including training and validation protocols.

Sequence-Based Score. For each kinase domain in the human proteome, we constructed its structural model using a state-of-the-art template-based structure prediction ap-

(26) Kuhn, D.; Weskamp, N.; Hullermeier, E.; Klebe, G. Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem* **2007**, *2* (10), 1432–47.
(27) Kinnings, S. L.; Jackson, R. M. Binding site similarity analysis for the functional classification of the protein kinase family. *J. Chem. Inf. Model.* **2009**, *49* (2), 318–29.
(28) Subramanian, G.; Sud, M. Computational Modeling of Kinase Inhibitor Selectivity. *ACS Med. Chem. Lett.* **2010**. DOI: 10.1021/ml1001097.
(29) Cozzetto, D.; Kryshtafovych, A.; Fidelis, K.; Moult, J.; Rost, B.; Tramontano, A. Evaluation of template-based models in CASP8 with standard measures. *Proteins* **2009**, *77* (Suppl. 9), 18–28.
(30) Ginalski, K. Comparative modeling for protein structure prediction. *Curr. Opin. Struct. Biol.* **2006**, *16* (2), 172–7.
(31) Moult, J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr. Opin. Struct. Biol.* **2005**, *15* (3), 285–9.
(32) DeWeese-Scott, C.; Moult, J. Molecular modeling of protein function regions. *Proteins* **2004**, *55* (4), 942–61.
(33) Piedra, D.; Lois, S.; de la Cruz, X. Preservation of protein clefts in comparative models. *BMC Struct. Biol.* **2008**, *8*, 2.
(34) Brylinski, M.; Skolnick, J. Comprehensive structural and functional characterization of the human kinome by protein structure modeling and ligand virtual screening. *J. Chem. Inf. Model.* **2010**, *50* (10), 1839–54.
(35) Zhang, Y.; Skolnick, J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. *Biophys. J.* **2004**, *87* (4), 2647–55.
(36) Skolnick, J.; Kihara, D.; Zhang, Y. Development and large scale benchmark testing of the PROSPECTOR_3 threading algorithm. *Proteins* **2004**, *56* (3), 502–18.
(37) Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (1), 129–34.
(38) Brylinski, M.; Skolnick, J. FINDSITE(LHM): a threading-based approach to ligand homology modeling. *PLoS Comput. Biol.* **2009**, *5* (6), e1000405.
(39) Brylinski, M.; Skolnick, J. Q-Dock(LHM): Low-resolution refinement for ligand comparative modeling. *J. Comput. Chem.* **2010**, *31* (5), 1093–105.

proach. This procedure, described in detail in ref 34, involves the identification of evolutionary related templates in the PDB[40] using the PROSPECTOR_3 threading algorithm,[36] followed by structure refinement/assembly by TASSER, a coarse-grained procedure guided by tertiary restraints extracted from the template structures.[35] Subsequently, modeled kinase structures were taken as targets for the prediction of ATP-binding sites by FINDSITE, a structure/evolution-based method that identifies ligand-binding sites based on binding site similarity among superimposed groups of functionally and structurally related template structures.[37] The sequence-based score corresponds to the sequence identity (fraction of identical residues) of binding residues between two protein kinases calculated using FINDSITE identified residues and structure alignments generated by TM-align.[41]

Structure-Based Score. In addition to the sequence-based scoring function, we also use a more structure-oriented measure of binding site similarity. Here, we employ a modified version of a PocketMatch score, PM-score, developed to provide a normalized similarity metric for binding site comparisons.[42] PocketMatch applies a geometric hashing algorithm to Cα atoms and side-chain geometrical centers of ligand binding residues extracted from the crystal structures of protein−ligand complexes. Each binding site is represented by a set of 90 predefined distance bins, whose populations capture its shape and chemical features. The original PocketMatch approach uses residues, one or more of whose atoms are within a distance of 4 Å from the crystallographic ligand position.[42] In our modified implementation, we use the consensus binding residues identified by FINDSITE in modeled kinase structures to populate the hash bins and calculate the PM-score.

Ligand-Based Score. Next, we introduce a new measure of binding site similarity that uses virtual screening ranks to calculate a chemical correlation. In the previous study, we carried out a large-scale virtual screening experiment for the complete human kinome.[34] Here, we use this data to calculate the correlation between compound ranks obtained for two binding pockets. The chemical correlation corresponds to the Kendall τ rank correlation coefficient[43] calculated for the average top ranked set of 10,000 ZINC compounds[44] ranked for individual target sites of the entire human kinome by structure-based virtual screening using Q-Dock^{LHM}.[39,45] Details on the docking/screening protocol are given in ref 34. Retrospective benchmarks carried out against several ligand libraries demonstrate that this collection of compounds is likely to be significantly enriched in ATP-competitive kinase inhibitors.[34] A high Kendall τ indicates that the pockets not only exhibit specific binding affinity toward similar compounds but also do not bind similar ligands. This new measure based on the similarity of virtual screening ranks complements sequence- and structure-based similarities between binding pockets.

Bioassay Data. We use three publicly available bioassay data sets to train and validate X-React^{KIN}: 28 commercially available compounds examined against a panel of 20 protein kinases (bioassay #1),[46] 38 kinase inhibitors assessed across a panel of 317 kinases representing >60% of the predicted human kinome (bioassay #2)[47] and 20 kinase inhibitors including 16 approved drugs or those in clinical development screened against a panel of 119 protein kinases (bioassay #3).[48] Bioassay #1 reports inhibitor potency as a percentage of kinase activity with respect to that in control incubations at an ATP concentration of 0.1 mM. Bioassays #2 and #3 use ATP site-dependent competition binding with each compound screened against the kinase targets at a single concentration of 10 μM and the binding efficacy reported in terms of quantitative dissociation constants, K_d. First, primary kinase targets (one per compound) are selected based on the strongest inhibition (bioassay #1) or the lowest dissociation constant (bioassays #2 and #3). Then, for each compound, we define alternate targets as kinases whose activity was inhibited to ≤25% of the control for bioassay #1 and those with $K_d \leq 10$ μM for bioassays #2 and #3. Remaining kinases are classified as nontargets. In this study, we use only compounds with at least one alternate kinase target. The list of compounds, primary target kinases and the number of alternate targets as well as nontargets is provided in SI Table 1 in the Supporting Information.

(40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–42.

(41) Zhang, Y.; Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **2005**, *33* (7), 2302–9.

(42) Yeturu, K.; Chandra, N. PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics* **2008**, *9*, 543.

(43) Kendall, M. G. A new measure of rank correlation. *Biometrika* **1938**, *30* (Part 1−2), 81–9.

(44) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–82.

(45) Brylinski, M.; Skolnick, J. Q-Dock: Low-resolution flexible ligand docking with pocket-specific threading restraints. *J. Comput. Chem.* **2008**, *29* (10), 1574–88.

(46) 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* (Pt 1), 95–105.

(47) 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* (1), 127–32.

(48) Fabian, M. A.; Biggs, W. H., 3rd; 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* (3), 329–36.

Activity-Based SAR Profiles. In addition to the bioassay data described above, we compare the virtual profiles constructed by X-React[KIN] to the experimentally derived activity-based SAR similarities on an orthogonal data set of 577 diverse compounds screened across a panel of 203 protein kinases.[21] Here, we use similarity scores expressed by a Tanimoto coefficient calculated for binding affinity fingerprints generated using an affinity threshold of 10%. Similarly to the CR-score values, kinase SAR similarity scores also range from 0 (dissimilar) to 1 (identical). For each kinase target, we assess the quality of X-React[KIN] virtual profiles calculated against the remaining kinases using the Pearson correlation coefficient between the SAR similarities and the CR-score values.

Machine Learning. In X-React[KIN], we use a naïve Bayes classifier to combine individual scoring functions: sequence-, structure- and ligand-based into a single probabilistic score. A classical naïve Bayes classification is based on estimating $P(X|Y)$, the probability or probability density of a qualitative attribute X given class Y. In our classifier, the real-value attributes are modeled by a Gaussian distribution, i.e. the classifier first estimates a normal distribution for each class by computing the mean and standard deviation of the training data in that class, which is then used to estimate $P(X|Y)$ during classification.[49] For a given pair of protein kinases, the probabilistic score from the classifier, called a CR-score, estimates the chances of the cross-reactivity from sequence, structure and binding similarities. X-React[KIN] was validated using the following leave-one-out procedure: In each round, one inhibitor and its close analogues are removed from the data set that consists of the bioassay data described above and the classifier is trained on the remaining compounds. Here, we define a close analogue as a compound that has a Tanimoto coefficient calculated using SMILES strings ≥ 0.7.[50] Then, for the excluded inhibitor and its primary target, the kinase proteins are ranked by the predicted CR-score, with the top-ranked kinases assumed to be alternate targets. We assess the accuracy of the off-target identification by a receiver operating characteristic (ROC) analysis with the CR-score used as a variable parameter. In addition to the standard ROC curves, we also calculate their distribution-free confidence bounds.[51]

Virtual Map of Kinase Cross-Reactivity. Finally, X-React[KIN] was retrained on all bioassay data and the complete map of putative cross-interactions within the human kinome was calculated. Moreover, we constructed a statistical model by fitting the distribution of the random CR-score values to a normal inverse Gaussian distribution[52] in order to calculate the associated p-values. The fitting procedure was done in R[53] using the ghyp package. The virtual cross-reactivity map is visualized using matrix2png,[54] with the kinase proteins grouped according to the subfamily classification and clustered by sequence identity using CLUTO.[55]

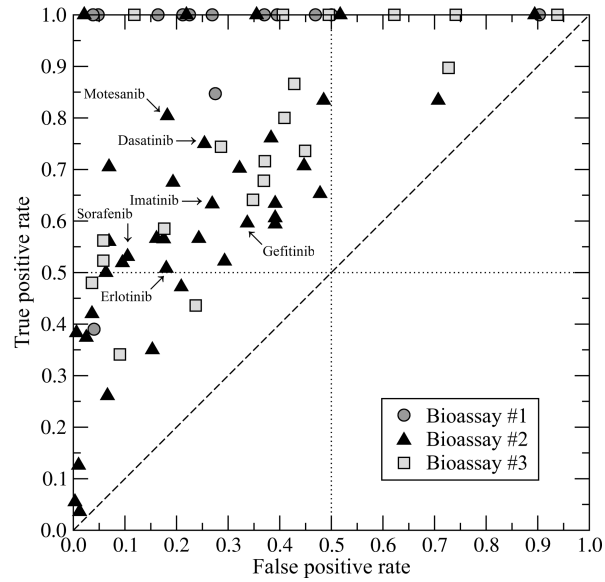

Figure 1. ROC plot for the prediction of kinase inhibitor cross-reactivity using X-React[KIN]. Compounds from bioassays #1, #2 and #3 are shown as dark gray circles, black triangles and light gray squares, respectively.

Results

X-React[KIN] Validation. Here, we use the available selectivity data from high-throughput kinase profiling experiments to train and validate X-React[KIN] in the off-target prediction. As described in the Methods section, for each kinase inhibitor and the corresponding primary target, the remaining kinases are assessed with respect to the estimated potential for cross-reactivity, i.e. ability to bind similar compounds. The results of leave-one-out validation are presented as a ROC plot in Figure 1. Encouragingly, in all cases the performance of X-React[KIN] is better than random, with a true positive rate >0.5 and a false positive rate <0.5 for almost 70% of the benchmark inhibitors. Particularly the results obtained for bioassay #2 are very promising since this panel of kinases covers >60% of the human kinome.[47] In addition, individual ROC plots for six selected compounds that include approved drugs such as Gleevec (imatinib), Iressa (gefitinib), Nexavar (sorafenib), Sprycel (dasatinib)

(49) Witten, I. H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Publishers: San Francisco, 2005.
(50) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (6), 983–96.
(51) Kestler, H. A. ROC with confidence - a Perl program for receiver operator characteristic curves. *Comput. Methods Programs Biomed.* **2001**, *64* (2), 133–136.
(52) Dagpunar, J. S. An Easily Implemented Generalised Inverse Gaussian Generator. *Commun. Stat.—Simul. Comput.* **1989**, *18* (2), 703–10.
(53) The R Development Core Team *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2008.
(54) Pavlidis, P.; Noble, W. S. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **2003**, *19* (2), 295–6.
(55) KarypisG. *CLUTO: A Clustering Toolkit*, version 2.1.1; 2003.

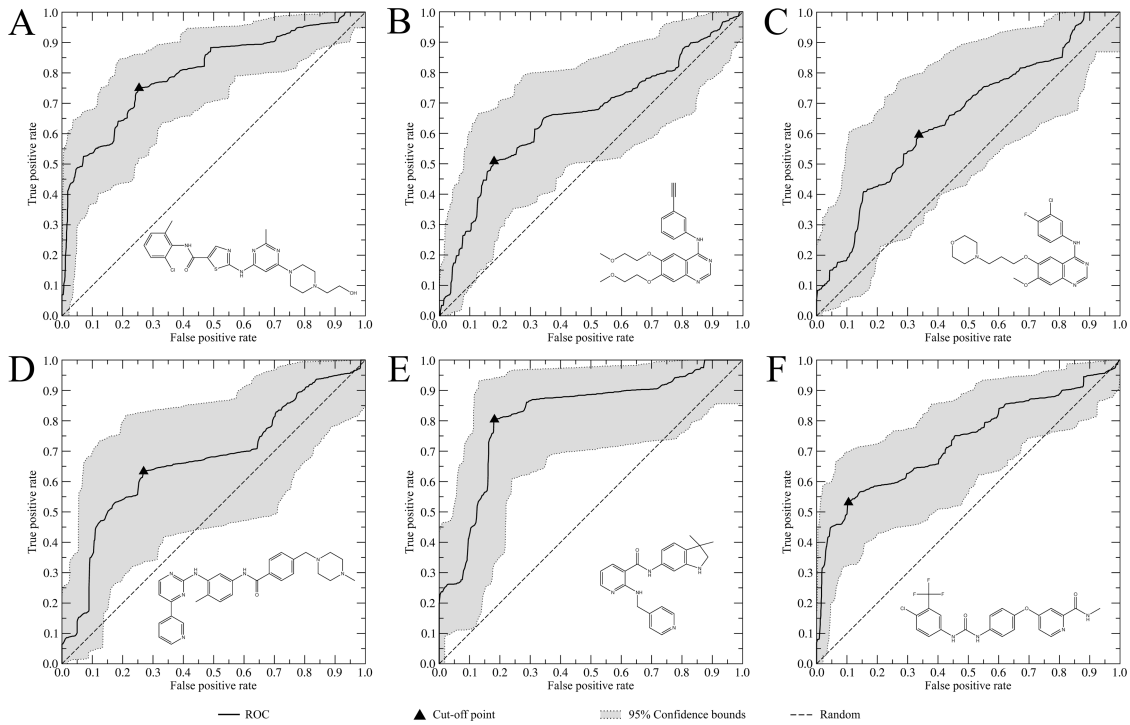

Figure 2. Individual ROC plots for selected inhibitors: (A) dasatinib, (B) erlotinib, (C) gefitinib, (D) imatinib, (E) motesanib and (F) sorafenib. In each graph, the solid black line, the gray area and the dashed line show the ROC curve for the CR-score, its 95% confidence bounds and the accuracy of a random classifier, respectively. The cutoff point that maximizes the sensitivity and specificity is represented by a black triangle. Chemical structures of the inhibitors are also displayed.

and Tarceva (erlotinib) are presented in Figure 2. In all cases, the cross-validated performance of X-React^{KIN} is significantly better than random, with tight confidence bounds particularly for dasatinib (Figure 2A), erlotinib (Figure 2B), motesanib (Figure 2E) and sorafenib (Figure 2F). The calculated cutoff points (displayed in Figure 2), which maximize the sensitivity and specificity, show that most of the cross-interacting kinases are identified at the expense of a relatively low false positive rate; the true (false) positive rate is 0.75 (0.25), 0.51 (0.18), 0.60 (0.34), 0.63 (0.27), 0.80 (0.18) and 0.53 (0.11) for dasatinib, erlotinib, gefitinib, imatinib, motesanib and sorafenib, respectively.

Human Kinome Cross-Reactivity Profile. Encouraged by the satisfactory performance of X-React^{KIN} in benchmark tests, we retrained the model on all bioassay data and constructed a complete map of putative cross-reactions within the entire human kinome. The details on the trained classifier used in X-React^{KIN} are provided in SI Table 2 in the Supporting Information. In Figure 3, for the human kinome, we compare the cross-interaction potential expressed by a sequence-based classification (Figure 3A) to the CR-score based classification (Figure 3B). In both Figures 3A and 3B, the kinases are clustered using sequence identity and the resulting dendograms are shown on the top of each plot. Comparing the sequence identity score to the CR-score, we observe many off-diagonal interactions pointed out by high CR-values (Figure 3B, blue spots). These nontrivial similarities, which are clearly the most interesting, indicate the possibility to bind similar compounds by remotely related protein kinases that belong to different groups. In particular, many potential cross-interactions are observed between kinases that belong to AGC (containing PKA, PKC and PKG protein kinases), CAMK (calcium/calmodulin-dependent protein kinases) and STE (the homologues of yeast Sterile kinases) groups. We note that, whereas the average pairwise sequence identity within these groups is relatively high, 38%, 34% and 36%, respectively, the intergroup sequence identity is notably lower, 29%, 26% and 26% for AGC/CAMK, AGC/STE and CAMK/STE, respectively. Even lower average sequence identity is seen between the TK (tyrosine kinases) group and those kinases that belong to AGC (23%), CAMK (24%) and CMGC (22%). The functional similarities indicated by the high CR-score values between these kinase proteins are undetectable on the basis of the sequence similarity alone. We have also constructed a statistical model for the CR-score distribution in order to assign statistical significance values. Here, we use a normal inverse Gaussian distribution, which fits well to the data; this is shown as histograms as well as a quantile−quantile plot in SI Figure 1 in the Supporting Information.

Comparison to SAR Profiles. For a subset of 203 protein kinases, activity-based SAR similarities have been previously reported.[21] These similarities were calculated directly from the experimental data obtained by screening the target kinases against a diverse set of >500 compounds, intended to

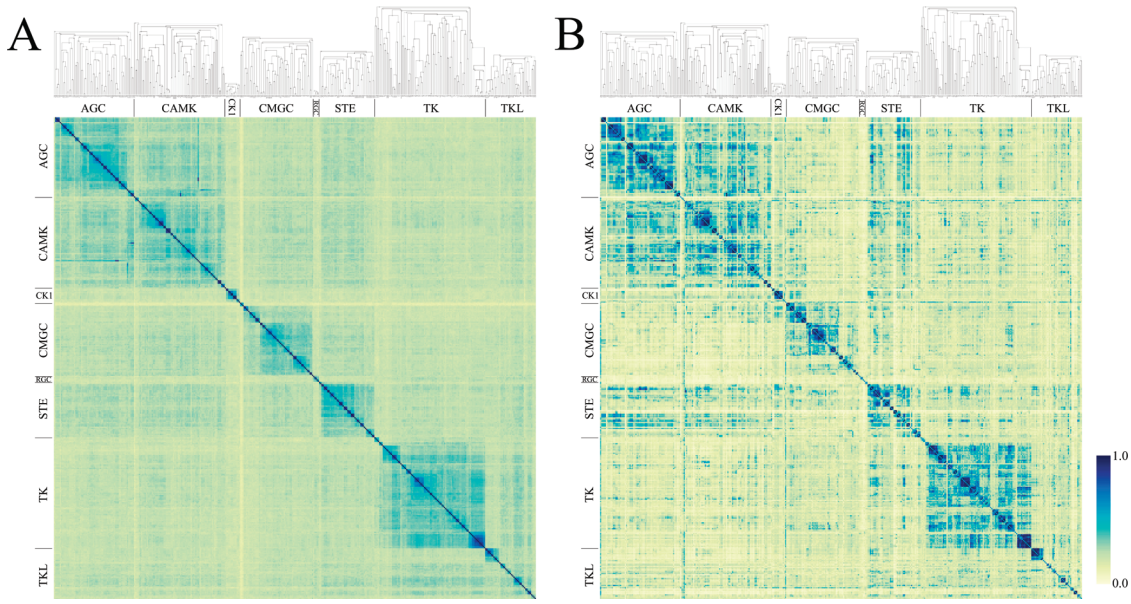

Figure 3. Classification of the human kinome by X-React^{KIN}: (A) sequence similarity matrix and (B) cross-reactivity matrix. In both plots, kinase proteins are grouped according to the subfamily classification displayed on both axes. Within each group, kinase members are clustered using sequence identity and the resulting dendograms are shown on the top of each graph. Color scale expressing the sequence similarity (A) as well as the potential cross-reactivity (B) is displayed on the right.

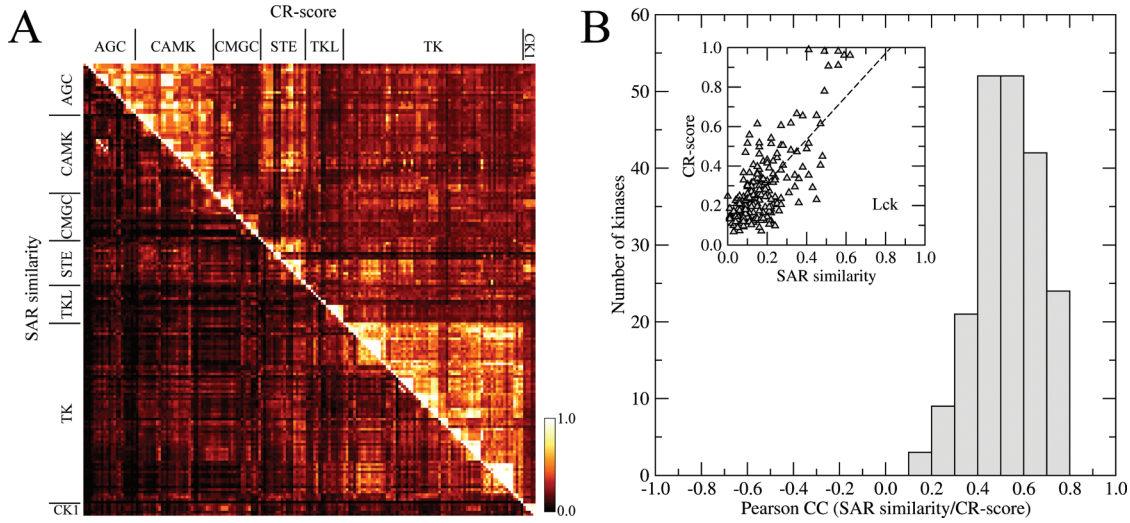

Figure 4. Comparison of the X-React^{KIN} virtual profiles to the SAR similarities on a set of 203 protein kinases. (A) Similarity between pairs of kinases ordered according to the Sugen phylogenetic tree (available at http://kinase.com). Upper right and lower left triangles represent the CR-score values and SAR similarities, respectively. The color scale expressing both similarities is displayed in the right corner. (B) Histogram of the distribution of the Pearson correlation coefficients between SAR similarities and CR-score values calculated for 203 kinase targets. Inset: Correlation between SAR similarities and CR-score values for the leukocyte-specific protein tyrosine kinase, Lck.

represent kinase inhibitor chemical space. This large-scale kinase profiling provides an orthogonal data set to validate the potential for cross-reactivity predicted by X-React^{KIN}. The results are presented in Figure 4. The direct comparison of the similarity between pairs of kinases according to the SAR profiles and the CR-score values is shown in Figure 4A. In both cases, the joint inhibition of many of these kinase pairs is observed within the TK subfamily. Moreover, good agreement between both approaches is seen for the STE subfamily, for which many predicted cross-interactions with kinases that belong to other, particularly AGC and CAMK, groups are confirmed experimentally. The distribution of the Pearson correlation coefficients between SAR similarities and CR-score values calculated for 203 kinase targets is presented in Figure 4B. This distribution is clearly shifted toward high (>0.5) values, which indicate a good overlap between experimental SAR and virtual CR-score profiles for the majority of kinase targets. The average Pearson correlation coefficient calculated across this data set is 0.53 ± 0.14. The qualitative agreement between the activity-based SAR simi-

larities and the CR-score profiles provides significant validation of the X-React^{KIN} approach.

Below, in a case study, we present a simple application of the human kinome cross-reactivity virtual profile constructed by X-React^{KIN} to demonstrate how it can be used to optimize the selection of kinase targets for small-scale selectivity counter-screens in kinase inhibitor development.

Case Study: Inhibitors of Lck. 2-Aminopyrimidine carbamates are a new class of compounds with potent and selective inhibition of the leukocyte-specific protein tyrosine kinase, Lck. Structure−activity relationship studies and extensive pharmacological tests carried out for a series of substituted 2-aminopyrimidine carbamates identified 2,6-dimethylphenyl-2-((3,5-bis(methyloxy)-4-((3-(4-methyl-1-piperazinyl)propyl)oxy) phenyl)amino)-4-pyrimidinyl(2,4-bis(methyloxy)phenyl)carbamate as a potent inhibitor of Lck, with an IC_{50} of 0.6 nM (compound **43** in the original paper).[56] Subsequently, a counter-screen against 15 other kinases that belong to TK, CMGC and AGC groups was carried out in order to characterize the selectivity profile of this compound. Here, we compare the experimental inhibition data to the *in silico* profile of Lck and demonstrate that the map of putative cross-interactions within the human kinome constructed by X-React^{KIN} can be used to suggest alternate kinase targets for the selectivity counter-screens. Figure 5 shows the selectivity profile for the pyrimidine carbamate inhibitor. Experimentally, this inhibitor was found to be highly selective with regard to the nonbinding of JAK3 (Kin. Dom. 2), MET, JNK3, PKCt, IGF1R and CDK2 (Figure 5A). With the exception of JAK3 (Kin. Dom. 2), the CR-score values (p-values) between Lck and these kinases are statistically insignificant: 0.483 (3.46×10^{-2}), 0.126 (7.03×10^{-1}), 0.182 (4.36×10^{-1}), 0.267 (1.96×10^{-1}), 0.229 (2.81×10^{-1}) and 0.162 (5.23×10^{-1}), respectively (Figure 5B). For another 8 kinase targets, the experimental IC_{50} values are in the range of 100 nM to 1 μM; here the CR-scores are higher (~0.3, p-values ~0.1 or better), with p-values <0.05 for BTK (1.39×10^{-2}) and JAK2 (Kin. Dom. 2, 4.07×10^{-2}). No selectivity was shown against SRC kinase, for which the CR-score (p-value) is 0.961 (1.55×10^{-3}).

Furthermore, the map of putative cross-interactions reveals other similarities between e.g. FGFR1 and TIE2 (CR-score = 0.856, p-value = 2.92×10^{-3}), JAK2 (Kin. Dom. 2) and TIE2 (CR-score = 0.663, p-value = 9.96×10^{-3}), BTK and ZAP70 (CR-score = 0.544, p-value = 2.23×10^{-2}), JNK3

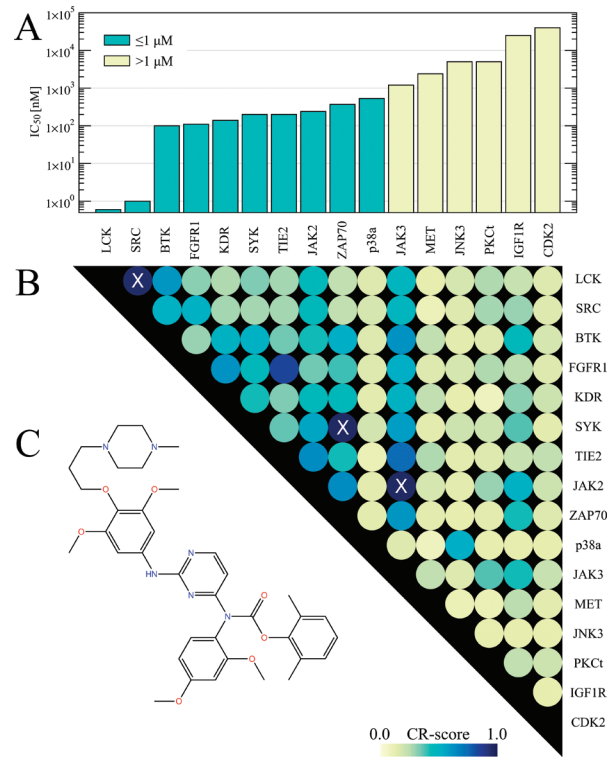

Figure 5. Selectivity profile for the pyrimidine carbamate inhibitor reported in ref 56: (A) experimental inhibition constant values in μM with the IC_{50} ≤1 μM (>1 μM) in turquoise (yellow); (B) pairwise CR-score matrix for the tested kinases, CR-score scale is given at the bottom; (C) chemical structure of the inhibitor. In panel B, kinase pairs with a pairwise sequence identity of >60% are marked with an X.

and p38a (CR-score = 0.532, p-value = 2.43×10^{-2}) or JAK3 (Kin. Dom. 2) and SYK (CR-score = 0.603, p-value = 1.49×10^{-2}), which indicate a high probability of inhibition by similar compounds. In fact, the joint inhibition of many of these kinase pairs has been already confirmed experimentally. We note that none of this information was used for the construction of the CR-score matrix; indeed we were unaware of the experimental results until after the predictions were made and we did a literature search. For example, an oral kinase inhibitor ACTB-1003 with multiple modes of action, targeting cancer mutations via FGFR1 inhibition (IC_{50} = 6 nM) and angiogenesis through inhibition of VEGFR2 (2 nM) and TIE2 (4 nM), has been recently reported.[57] Several inhibitors (compounds **10**, **11**, **12**, **13** and **14** in the original paper) were found to nonselectively inhibit JAK2 (TIE2) with the percent of enzyme activity at 1 μM concentration of 6 (35), 5 (0), 0 (1), 30 (1) and 27 (7), respectively.[58] Moreover, compound **7** in the original paper was found to be the most selective against JAK2 and TIE2 (3% and 26%) across a panel of 59 recombinant serine/

(56) Martin, M. W.; Newcomb, J.; Nunes, J. J.; McGowan, D. C.; Armistead, D. M.; Boucher, C.; Buchanan, J. L.; Buckner, W.; Chai, L.; Elbaum, D.; Epstein, L. F.; Faust, T.; Flynn, S.; Gallant, P.; Gore, A.; Gu, Y.; Hsieh, F.; Huang, X.; Lee, J. H.; Metz, D.; Middleton, S.; Mohn, D.; Morgenstern, K.; Morrison, M. J.; Novak, P. M.; Oliveira-dos-Santos, A.; Powers, D.; Rose, P.; Schneider, S.; Sell, S.; Tudor, Y.; Turci, S. M.; Welcher, A. A.; White, R. D.; Zack, D.; Zhao, H.; Zhu, L.; Zhu, X.; Ghiron, C.; Amouzegh, P.; Ermann, M.; Jenkins, J.; Johnston, D.; Napier, S.; Power, E. Novel 2-aminopyrimidine carbamates as potent and orally active inhibitors of Lck: synthesis, SAR, and in vivo antiinflammatory activity. *J. Med. Chem.* **2006**, *49* (16), 4981–91.

(57) Patel, K.; Fattaey, A.; Burd, A. ACTB-1003: An oral kinase inhibitor targeting cancer mutations (FGFR), angiogenesis (VEGFR2, TEK), and induction of apoptosis (RSK and p70S6K). *J. Clin. Oncol. (Meeting Abstr.)* **2010**, *28* (15_suppl), e13665.

threonine and tyrosine kinases. Many JNK3 inhibitors are known to also inhibit p38a; e.g. two compounds with a nanomolar activity against JNK3 (IC_{50} of 7 and 1 nM) have been reported as potent p38a inhibitors as well, with the IC_{50} of 0.2 and 4 nM, respectively.[59] Finally, *in vitro* enzymatic assays of the novel JAK3 inhibitor R348 showed potent inhibition of JAK3- and SYK-dependent pathways.[60]

Lck was also included in the large-scale assessment of the chemical coverage of the kinome space using activity-based SAR profiles.[21] In Figure 4B (inset), we compare the experimentally derived SAR similarities to the CR-score values calculated against the remaining 202 protein kinases used as targets in the high-throughput binding assay. Here, the Pearson correlation coefficient between the SAR similarities and the CR-score values is 0.73. This high correlation additionally confirms the good agreement between the potential for cross-reactivity predicted by X-React^{KIN} and the experimentally observed joint inhibition of protein kinases.

Of course, a high probability of inhibition by the same groups of compounds does not preclude a successful design of selective inhibitors. Rather, it should support the counter-screen selectivity experiments by optimizing the selection of possible off-targets, whose binding sites have the highest potential for cross-reactivity.

Discussion

Many drug candidates fail in clinical development due to their poor pharmacokinetic characteristics and because of intolerable adverse effects, which may sometimes originate in their insufficient selectivity.[61] The physicochemical similarity between highly conserved ATP-binding sites in protein kinases, one of the most important drug targets, has rendered the challenge of designing selective inhibitors difficult. Nevertheless, the discovery of selective kinase inhibitors demonstrates that there is enough conformational and chemical diversity in and around the active site that can be explored to design compounds with sufficient selectivity.[14,15] Thus, particularly in the early stages of drug development, the knowledge of alternate kinase targets with significant potential for cross-reactivity is critical. One common strategy in inhibitor design involves differential lead optimization to increase the selectivity toward a particular drug target; such efforts are typically oriented toward the development of highly specific inhibitors acting on single protein kinases. Later on, with the approval of multitarget inhibitors, such as imatinib, sunitinib or lapatinib, an alternate strategy has emerged, where drug-resistance can be overcome by simultaneously targeting multiple kinase pathways.[62] Multikinase inhibitors with highly tuned selectivity profiles are currently of particular interest in pharmaceutical research.[63] The functional classification of the entire human kinome is of paramount importance in the development of both highly selective as well as selectively unselective novel inhibitors.

Due to the sparse and nonuniformly distributed structural data,[64] cross-interactions are still poorly defined at the kinome level. To maximize the coverage of kinase functional space, we developed X-React^{KIN}, a Chemical Systems Biology approach for *in silico* cross-reactivity profiling that does not require high-resolution structural data. X-React^{KIN} employs a state-of-the-art protein structure prediction algorithm followed by the recently developed ligand homology modeling (LHM) approach to model kinase−drug interactions.[34] Subsequently, the modeling of individual kinase members is now extended to construct a cross-reactivity virtual profile for the entire human kinome. This proteome-wide analysis represents a significant improvement over other methods, which are generally confined to high-resolution structures solved by protein crystallography.

In addition to the traditional sequence and structure similarity measures, our method also uses a novel type of binding site comparison by means of virtual screening ranks. A high correlation between ligand rankings for two binding sites, referred to as a chemical correlation, indicates that these sites not only exhibit specific binding affinity toward similar molecules but also do not bind similar compounds. Here, the accuracy of ligand docking and ranking is essential. Particularly, using predicted receptor structures requires reliable docking techniques capable of dealing with structural inaccuracies in protein models. It has been demonstrated that even moderate structural distortions of the modeled binding pockets drastically interfere with the ability of the all-atom docking approaches to identify correct docking geometries and to rank ligands.[39,65] Our virtual screening protocol that provides compound ranking for the estimation of the chemical correlation employs evolution-based ligand docking[38] followed by low-resolution binding pose refinement.[39,45]

(58) Okram, B.; Nagle, A.; Adrian, F. J.; Lee, C.; Ren, P.; Wang, X.; Sim, T.; Xie, Y.; Xia, G.; Spraggon, G.; Warmuth, M.; Liu, Y.; Gray, N. S. A general strategy for creating "inactive-conformation" abl inhibitors. *Chem. Biol.* **2006**, *13* (7), 779–86.

(59) Scapin, G.; Patel, S. B.; Lisnock, J.; Becker, J. W.; LoGrasso, P. V. The structure of JNK3 in complex with small molecule inhibitors: structural basis for potency and selectivity. *Chem. Biol.* **2003**, *10* (8), 705–12.

(60) Deuse, T.; Velotta, J. B.; Hoyt, G.; Govaert, J. A.; Taylor, V.; Masuda, E.; Herlaar, E.; Park, G.; Carroll, D.; Pelletier, M. P.; Robbins, R. C.; Schrepfer, S. Novel immunosuppression: R348, a JAK3- and Syk-inhibitor attenuates acute cardiac allograft rejection. *Transplantation* **2008**, *85* (6), 885–92.

(61) Bleicher, K. H.; Bohm, H. J.; Muller, K.; Alanine, A. I. Hit and lead generation: beyond high-throughput screening. *Nat. Rev. Drug Discovery* **2003**, *2* (5), 369–78.

(62) Petrelli, A.; Giordano, S. From single- to multi-target drugs in cancer therapy: when aspecificity becomes an advantage. *Curr. Med. Chem.* **2008**, *15* (5), 422–32.

(63) Morphy, R. Selectively nonselective kinase inhibition: striking the right balance. *J. Med. Chem.* **2010**, *53* (4), 1413–37.

(64) Marsden, B. D.; Knapp, S. Doing more than just the structure-structural genomics in kinase drug discovery. *Curr. Opin. Chem. Biol.* **2008**, *12* (1), 40–5.

(65) Verdonk, M. L.; Mortenson, P. N.; Hall, R. J.; Hartshorn, M. J.; Murray, C. W. Protein-Ligand Docking against Non-Native Protein Conformers. *J. Chem. Inf. Model.* **2008**, *48* (11), 2214–25.

Such a docking/ranking procedure is well suited for virtual screening applications using modeled receptor structures since it exhibits significant tolerance to receptor structure deformation.[39]

Modern drug discovery is routinely supported by computational techniques, such as virtual screening, which prioritize drug candidates and increase the hit rate by restricting screening libraries to compounds that likely exhibit the desired bioactivity. At the system level, the functional classification of the human kinome expands our understanding of the structural, chemical and pharmacological aspects of the kinase space and provides a practical strategy that should prove useful for the design of more selective therapeutics.

Note

The cross-reactivity virtual profile of the human kinase space is available at http://cssb.biology.gatech.edu/kinomelhm/.

Acknowledgment. This work was supported in part by Grants GM-48835 and GM-37408 of the Division of General Medical Sciences of the National Institutes of Health.

Supporting Information Available: Figure depicting fitting of the asymmetric normal inverse Gaussian (NIG) to the distribution of CR-scores in the human kinome, table of bioassay data used in this study, and table of trained naïve Bayes classifier used by X-React^{KIN} to calculate the CR-score. This material is available free of charge via the Internet at http://pubs.acs.org.

MP1002976