JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Pocketome of Human Kinases: Prioritizing the ATP Binding Sites of (Yet) Untapped Protein Kinases for Drug Discovery

Andrea Volkamer,*[,†] Sameh Eid,[†] Samo Turk,[†] Sabrina Jaeger,[†] Friedrich Rippmann,[‡] and Simone Fulle*[,†]

[†]BioMed X Innovation Center, Im Neuenheimer Feld 583, 69120 Heidelberg, Germany

[‡]Global Computational Chemistry, Merck KGaA, Frankfurter Strasse 250, 64293 Darmstadt, Germany

ABSTRACT: Protein kinases are involved in a variety of diseases including cancer, inflammation, and autoimmune disorders. Although the development of new kinase inhibitors is a major focus in pharmaceutical research, a large number of kinases remained so far unexplored in drug discovery projects. The selection and assessment of targets is an essential but challenging area. Today, a few thousands of experimentally determined kinase structures are available, covering about half of the human kinome. This large structural source allows guiding the target selection via structure-based druggability prediction approaches such as DoGSiteScorer. Here, a thorough analysis of the ATP pockets of the entire human kinome in the DFG-in state is presented in order to prioritize novel kinase structures for drug discovery projects. For this, all human kinase X-ray structures available in the PDB were collected, and homology models were generated for the missing part of the kinome. DoGSiteScorer was used to calculate geometrical and physicochemical properties of the ATP pockets and to predict the potential of each kinase to be druggable. The results indicate that about 75% of the kinome are in principle druggable. Top ranking structures comprise kinases that are primary targets of known approved drugs but additionally point to so far less explored kinases. The presented analysis provides new insights into the druggability of ATP binding pockets of the entire kinome. We anticipate this comprehensive druggability assessment of protein kinases to be helpful for the community to prioritize so far untapped kinases for drug discovery efforts.

INTRODUCTION

Protein kinases are enzymes which activate a multitude of proteins through phosphorylation. They mediate signal transduction in cell growth and differentiation and are involved in a variety of diseases including cancer, inflammation, and autoimmune disorders.[1] Kinases are found to be the most frequently mutated proteins in tumors and are among the major targets in drug discovery projects.[2−4] As a consequence, structure determination efforts produced more than 3600 human kinase X-ray structures available in the Protein Data Bank (PDB).[5] Using this large data pool in a rational manner can provide crucial insights into the binding sites of protein kinases and facilitate the development of novel as well as more selective kinase inhibitors.

The human kinome holds to date nearly 560 protein kinases including 492 typical ones. Based on sequence identity, Manning and co-workers further clustered the typical protein kinases into eight main groups (TK, TKL, STE, CK1, CMGC, AGC, CAMK, RGC), whereas all remaining typical kinases were assigned to the OTHER group.[6] Kinase structures have a highly conserved adenosine triphosphate (ATP) binding site and undergo large conformational changes upon ligand binding. The active and inactive forms are also known as the DFG-in and DFG-out state (Figure 1), respectively. They are distinguishable by a flip of the Asp and Phe orientation in the activation loop as well as conformational changes in the surrounding loops.[7] In the DFG-in conformation, the Asp side chain is exposed and the Phe side chain is buried, allowing ATP binding (Figure 1 B). In the DFG-out state, the Asp side chain is buried and the Phe side chain is flipped outside the pocket, thus, interfering with ATP access to the active site (Figure 1 C). The DFG-flip is accompanied by large movements of the

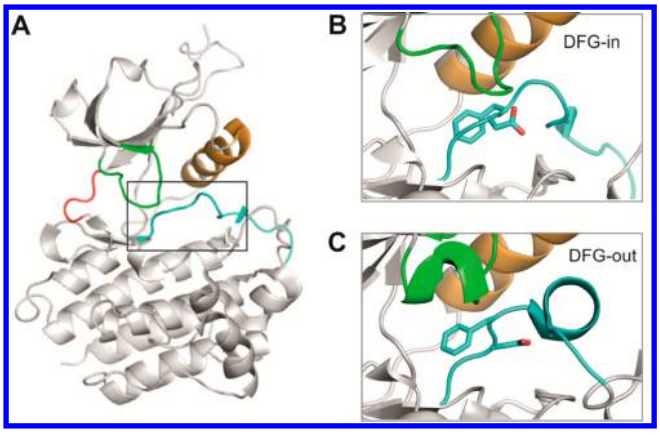

Figure 1. (A) Structure of the ABL kinase in the DFG-in state (PDB code 2GQG). Key structural parts are highlighted: Hinge region (red), G-loop (green), activation loop that includes the DFG-motif (cyan), and the αC-helix (orange). Close up of ABL kinase in DFG-in (B) (PDB code 2GQG) and DFG-out conformation (C) (PDB code 2HYY).

Received: October 16, 2014
Published: December 31, 2014

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

activation loop as well as by an outward movement of the G-loop and αC-helix.

Due to the integrative role of protein kinases in signal transduction, many drug discovery campaigns have been undertaken and successfully brought 26 small molecule kinase inhibitors as FDA approved drugs onto the market (by early 2014).[8] Due to the fact that kinases share a highly conserved ATP binding site, many inhibitors bind unselectively to a range of other kinases in addition to the key target.[9] Besides ATP competitive drugs, which bind to the active form of kinases (type I inhibitor), drugs stabilizing the inactive DFG-out conformation (type II and III) (see Figure 1), as well as allosteric inhibitors (type IV), have been discovered.[10,11]

Interestingly, most of the approved drugs have been designed to address tyrosine kinases (TK) as primary targets and only very few tackle other kinase groups,[12] showing the strong bias of kinase drug discovery—or especially cancer research in general—toward clinically validated targets.[13] In contrast, mutation studies show that kinases from almost all groups are involved in cancer.[14] Causes for a lack of diversity of approved drug targets may be a limited understanding of the enrollment of these targets in cancer signaling as well as experimental hints that they are less druggable. Comparing the fraction of established drug targets with the vast amount of targets found to be disease relevant and potentially druggable highlights the need to investigate novel drug targets.[13,15−19]

The suitability of a target for drug design is vital for the success of drug discovery projects.[20] Target druggability involves multiple parameters such as disease relevance, chemical tractability, screening feasibility, toxicological conditions, or selectivity. Computational druggability predictions[21,22] have become popular as they allow pointing to novel promising targets without the need to undertake expensive screening efforts. Druggability, as used in this article, is defined as the potential of a disease modifying target to be modulated by small, orally bioavailable compounds.[23] Most structure-based methods use calculated geometric and physicochemical features of binding sites of known drug targets to train a model or function which can subsequently be applied to predict the druggability of binding pockets from novel structures.[24−27] Recently, small to large scale druggability investigations have been published on other protein families and gathered new insights into protein family wide abilities to bind drug-like ligands.[28−31]

Structural kinase data has intensively been collected and been made available in databases such as KLIFS,[32] MOE's kinase suite,[33] and ChEMBL's Kinase SARfari.[8] Nevertheless, to date structural data is only available for less than half of the kinome. Thus, information gathered from these data sets misses over 250 kinases. Incorporating modeled structures to complete the structural kinome has recently been used for virtual screening experiments.[34] Similarly, modeled structures will also be used in the course of the current study for druggability assessments.

In this work, the druggability of the entire human kinome (i.e., all typical kinases) will be analyzed using a selected set of available kinase structures and homology models of the remaining part of the human kinome. Note that due to the small coverage of structures in the inactive conformation (DFG-out), this study concentrates on kinases in the DFG-in state. Accordingly, geometrical and physicochemical features of the ATP binding pocket of all typical human kinases (which will be referred as the "pocketome of human kinases" throughout this manuscript) were calculated and prioritized regarding their druggability using DoGSiteScorer.[26] The predicted druggability scores for the DFG-in structures vary but indicate that 75% of the pocketome of human kinases are druggable. Encouragingly, established drug targets such as ABL1 showed up in the top ranking list. The first kinase drug, Gleevec (imatinib) was actually designed to address ABL1 which is apparently, according to our analysis, an "easy-to-drug" kinase and was constituted the breakthrough in cancer treatment.[35] While imatinib blocks the DFG-out conformation of ABL1 kinase, other approved drugs such as dasatinib and bosutinib bind to the DFG-in conformation of ABL1.[36] Our analysis also assigned high druggability scores to rather unexplored kinases (e.g., MLK1, CK1g2, and CLKs), thus pointing to new opportunities for drug discovery. The analysis is supplemented by ligand-based information obtained from ChEMBL.[37]

MATERIAL AND METHODS

Generation of a Kinase Master Table. In order to collect a complete list of kinase structures from the PDB, a unified Master Table was compiled based on the information maintained at www.kinase.com and the corresponding Uniprot[38] sequences. Kinase naming used throughout this manuscript is based on the classification by Manning et al.[6] An initial list of PDB structures per kinase was obtained by mining the sequences of the catalytic domains via BLAST[39] against the PDB (as of July 2014). The mappings between PDB chains and individual human kinases was prioritized based on the alignment quality and manual verification. The resulting PDB−kinase maps were annotated regarding the DFG state using the information provided in the MOE kinase suite (annotated as DFG-in, DFG-out, or "unknown" conformation).[33] The final Master Table contained information for 492 human protein kinases (excluding atypical kinases). This includes an annotation of the corresponding kinase group and the sequence of the catalytic domain, as well as a list of available PDB structures including the annotation of their experimental source and DFG state. Based on this Master Table, kinases were separated into three sets: a nonredundant set of selected kinase structures in the DFG-in conformation and determined by X-ray crystallography (selKin), a larger redundant set of kinase structures also in the DFG-in conformation and determined by X-ray crystallography but covering the "background information" in the PDB (brKin), and those not yet experimentally determined by X-ray crystallography and, therefore, modeled by homology modeling (hmKin).

PDB Sets Compilation: brKin and selKin. Initially a total number of 1678 human protein kinase X-ray structures were collected (annotated as DFG-in in the Master Table) which cover 165 human kinases. For kinases without DFG-in structures, structures initially classified as "unknown" in the Master Table were manually checked whereby 38 further kinases structures in the DFG-in state were added. The brKin set was designed to contain several structures per kinase with highest resolution but restricting the maximal number to 10 in order to equilibrate the data set. This yielded a final brKin set of 739 structures covering 203 distinct kinases (Table 1) with a mean resolution of 2.2 Å and lower and upper bounds of 1.1 and 4.1 Å, respectively. In turn, the selKin set was designed to contain exactly one selected structure representative for each of the 203 distinct kinases (see Table 1). For this purpose, structures were checked manually with respect to the following

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

Table 1. Number of Analyzed Kinase Structures Separated by Kinase Groups

	selKin		hmKin		brKin	
group	initial set[a]	final set[b]	initial set[a]	final set[b]	initial set[a]	final set[b]
TK	49	48	45	37	231	210
TKL	16	15	27	24	34	32
STE	21	17	27	22	68	51
CK1	9	9	3	3	21	21
CMGC	34	33	29	26	134	119
AGC	21	19	42	40	80	78
CAMK	31	28	51	40	105	96
RGC	0	0	5	1	0	0
OTHER	22	20	60	41	67	62
sum	203	189	289	234	739	669

[a]Initial number of kinase structures. [b]Number of kinase structures after pocket processing and the final set used in this study.

criteria: good resolution (usually <2.5 Å), structural completeness especially in terms of having resolved DFG- and G-loops as well as favoring structures bound to ATP or ATP-analog molecules. A supplementary list of structures in the selKin and brKin sets can be obtained from the authors upon request.

Homology Models Set Compilation: hmKin. To fill all branches of the human kinome tree with structural data, the remaining 289 protein kinases (Table 1) were generated by homology modeling using the Yasara software package.[40] For template selection, sequence similarity values were calculated based on a global multiple sequence alignment of the canonical kinase domain sequences obtained using ClustalW.[41] The generation of the homology models (HMs) was done with Yasara[40] using the sequences of the catalytic domains of the query kinases and the template structures from the selKin set which yielded the highest global sequence identity, respectively. Ligands were included in the modeling step if present in the corresponding template structure. The supplementary list of used template structures can be obtained by the authors upon request. The quality of the homology models was evaluated using the WHATIF,[42] MolProbity,[43] and PROCHECK[44] programs. Furthermore, the models were checked for their ability to accommodate ATP in their binding site by docking ATP molecules into the homology models using the LeadIT software suite[45] and comparing the generated binding modes with a reference ATP structure from the PDB code 1ATP. For evaluation, the minimal RMSD between the reference ATP and the 10 top scoring poses were calculated. Since the position of the triphosphate tail of ATP is influenced by several water molecules, magnesium ions and the orientation of the substrate residue, which were not included in the docking calculation, the RMSD was only calculated based on the atoms of the purine scaffold and the ribose.

Structure Preparation. To compile a high quality data set, each kinase structure from the selKin and brKin set was prepared as follows: The respective protein structure was downloaded from the PDB. Hydrogen atoms and missing residues were added using the Structure Preparation tool within MOE.[33] An in-house Python script was used to extract the chains corresponding to the kinase domain (as annotated in the Master Table), to remove alternative side chain conformations, waters, ions, and buffers,[46] and to change modified amino acids to the corresponding standard ones (e.g., phosphate groups were removed from phosphotyrosine, phosphoserine, and phosphothreonine). Finally, the extracted structures were aligned to a kinase reference structure (PDB code: 1ATP, chain A) using PyMOL's[47] default pairwise alignment function.

Pocket Detection and Druggability Prediction. A new version of DoGSiteScorer[26,48] was used to detect the ATP binding pockets and subpockets, as well as to prioritize them based on the predicted druggability scores. In DoGSiteScorer, each protein is captured by a grid. To allow for direct comparison between the pockets, the grid box was fixed to a size of 60 Å × 67 Å × 56 Å and the grid spacing was set to 0.4 Å; otherwise default parameters were used. DoGSiteScorer employs a difference of Gaussian (DoG) filter to detect pockets and subpockets on the protein surface.[48] The resulting (sub)pockets were characterized by geometrical and physicochemical properties such as volume, depth, and hydrophobicity. Finally, this information was fed into a support vector machine (SVM) to predict the druggability of the pockets.[26]

Pocket Processing. Out of all potential pockets calculated per protein structure, the ATP binding pocket or the respective subpocket was selected. In the case of the analyzed holo structures, the bound ligand was used to select the ATP pocket whereas in the case of the apo structures, a reference ATP molecule was used for the pocket selection. Afterward, a coverage quality filter was applied in order to consider only reasonable (sub)pockets in the analysis.[26] These restrictions resulted in final set sizes of 189 for selKin, 669 for brKin, and 234 for hmKin (Table 1).

Negative and Positive Control Sets. In order to validate the resulting druggability predictions, the analysis was also performed on two control sets: The negative control set comprises protein structures which were assigned to be undruggable in the druggability data set published by Schmidtke et al.[49] The positive control set comprises kinase structures that are targeted by FDA approved drugs. For this, kinase structures which are key targets for the approved drugs were collected from the PDB and Kinase SARfari[50] as well as KLIFS[32] and filtered based on structural completeness and the DFG-state. Only pockets fulfilling the coverage criterion were included. The final control sets consist of 50 and 10 members (Table 2), respectively.

Ligand-Based Druggability Assessment of Kinases. Compounds and bioactivity data were extracted from ChEMBL

Table 2. List of Kinases of the Positive Control Set, i.e. Key Targets for FDA Approved Inhibitors Targeting the DFG-in State of Kinases[a]

key target	group	PDB code	bound drug
EGFR	TK	4HJO	erlotinib
EGFR	TK	2ITO	gefitinib
ABL1	TK	2GQG	dasatinib
EGFR	TK	1XKK	lapatinib
KDR	TK	3CJG	pazopanib
MET	TK	2WGJ	crizotinib
BRAF	TKL	3OG7	vemurafenib
RET	TK	2IVU	vandetanib
JAK3_B	TK	3LXK	tofacitinib
EGFR	TK	4G5J	afatinib

[a]Drugs are sorted by the year of approval; note that only DFG-in structures are listed that fulfill the required criteria and that all drugs are delivered orally.

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

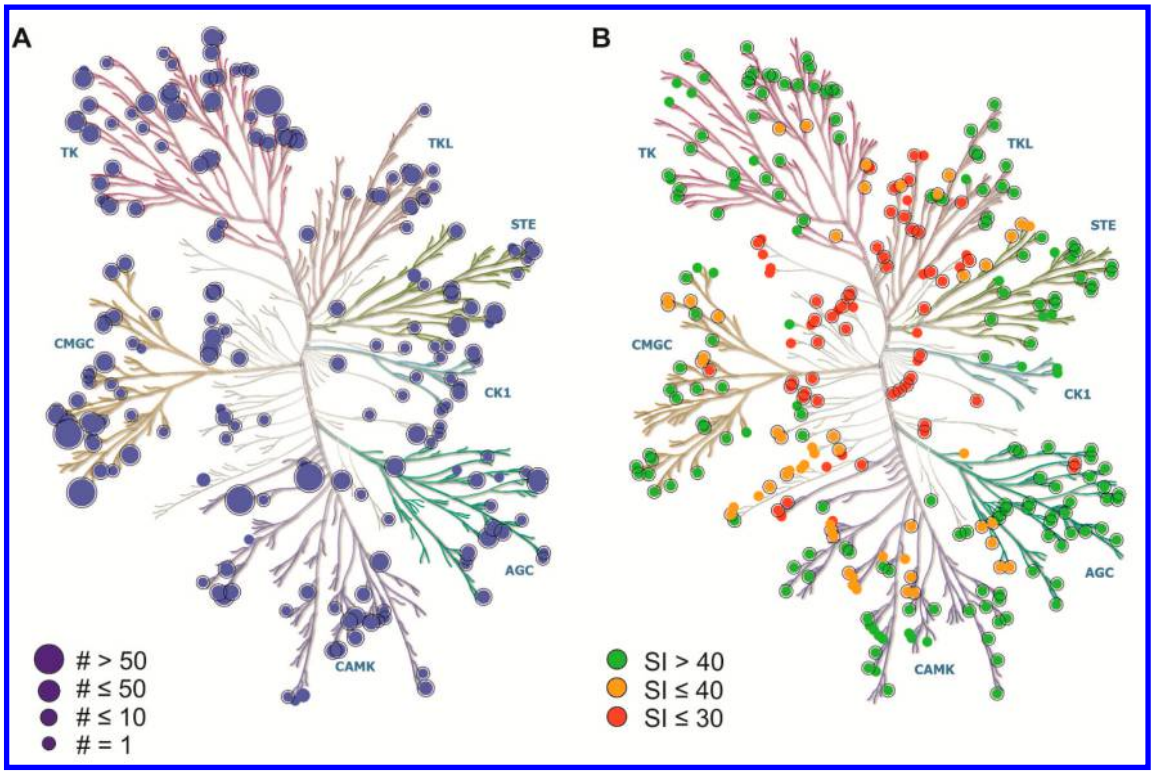

Figure 2. Distribution of PDB structures and homology models on the kinome tree. (A) PDB structures are shown in blue (size of the circles represents the number of X-ray structures in the PDB). (B) Homology models are shown in traffic light coloring depending on the sequence identity (SI) to the respective template structures (red: SI < 30%, orange: 30% ≤ SI ≤ 40%, green: SI > 40%). The final set of pockets considered in the analysis is indicated by black surrounding circles. Full names of the kinase groups are given in the abbreviation section.

version 19[8,37] by making a query via the UniProt IDs of all kinases. The resulting data was curated and filtered based on the following criteria: (1) Compounds with molecular weight between 120 and 650, clogP between −5 and 5, and which passed the PAINS filters[51] were considered. (2) All IC_{50} and K_i data from binding assays (assay type B) with target confidence score of 8 or 9, an operator "=", and measured in nanomolar units were extracted. (3) All K_i values were converted to IC_{50} using the Cheng−Prusoff[52] equation and assuming the ATP concentration to be at the value of K_m. For compounds with multiple measurements on the same kinase the median of the potencies was used. Overall, this resulted in 53 716 bioactivity data points for 26 880 compounds.

Finally, a ligand-based druggability score (LBDS) was defined as the number of bioactivities with IC_{50} < 500 nM (number of actives) divided by the number of all bioactivity data points (number of tested compounds) available per kinase.

Compilation of List of Key Targets. Furthermore, a list of all key targets was compiled (see Table 4) by querying all kinases via the "Get approved drugs for target" function in ChEMBL's web service API[37] and considering only kinase inhibitors identified using the respective USAN stems, e.g. those ending with "nib". Furthermore, the kinases FLT1, FLT4, and JAK3 were added as they are reported to be key targets for axitinib and tofacitinib.

Statistical Analysis. Values are expressed as mean (±standard deviation), unless specified otherwise. Significance was determined by the Pearson correlation coefficient as calculated using pandas[53] built-in function.

RESULTS AND DISCUSSION

Focus of this study is a comprehensive analysis of the pocketome of human kinases. Since PDB structures as well as homology models are used, the structural coverage as well as the quality of the models will be evaluated first. Next, the calculated pocket properties are investigated. Subsequently, a structure-based druggability analysis is presented for all kinases, and the outcome is compared to the kinome coverage by kinase inhibitor data. Finally, the top ranking kinase structures are discussed.

Structural Coverage of the Kinome and hmKin Set Evaluation. Human kinases are one of the most studied protein families and also the focus of structure determination efforts.[2] Due to a bias of current research to a small subset of the entire human kinome[16] as well as crystallization issues,[54] the coverage of the kinase space is unbalanced (Figure 2A). Some kinases have only one structure representative in the PDB, e.g. ABL2, while other kinases have up to hundreds of structures disclosed, e.g. CDK2. Out of the 203 kinase structures in the DFG-in state, 52% have one structure representative, 31% have between 2 and 10 structures, 14% have between 11 and 50 structures, and 3% have more than 50 structures. To balance the background data used in this study, a maximum of 10 structures per kinase (chosen based on resolution) were collected in the brKin set. In pursuit of having a nonredundant set of selected kinases structures, exactly one structure per kinase was collected in the selKin set (see Materials and Methods for further information). The final brKin and selKin sets contain 739 and 203 kinase structures in the DFG-in state, respectively, while both sets of structures have a mean resolution of 2.2 Å. Overall the available X-ray

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

Table 3. Mean Values of the Quality Assessment of the Homology Models Separated by Kinase Groups As Well As the Sequence Identities to the Template Structures Used for Homology Modelling

	internal energy and stereochemistry check[a]				
group	Z-score[b]	MolProbity[b]	PROCHECK[b]	RMSD to reference ATP[b,c]	sequence identity to respective template[b,d]
TK	−0.8 (0.6)	1.4 (0.4)	89.5 (2.8)	1.5 (1.2)	58.8 (17.4)
TKL	−1.6 (0.8)	1.9 (0.4)	86.1 (3.9)	1.2 (0.8)	48.1 (23.4)
STE	−2.1 (2.2)	2.0 (1.0)	86.2 (7.0)	1.2 (0.5)	51.8 (21.2)
CK1	−0.4 (0.1)	1.2 (0.1)	91.6 (1.7)	0.8 (0.1)	80.0 (5.6)
CMGC	−0.9 (0.6)	1.4 (0.3)	89.6 (2.1)	1.2 (1.0)	54.7 (19.6)
AGC	−0.8 (0.7)	1.3 (0.4)	90.9 (2.3)	1.3 (1.2)	62.6 (20.7)
CAMK	−1.1 (0.5)	1.4 (0.4)	90.2 (2.3)	1.5 (1.4)	52.3 (19.3)
RGC	−2.6 (0.7)	2.2 (0.6)	85.5 (3.0)	1.4 (0.9)	22.0 (−)
OTHER	−2.3 (1.2)	2.0 (0.6)	85.6 (5.4)	2.0 (1.5)	28.5 (16.2)
All	−1.4 (1.2)	1.6 (0.6)	88.4 (4.5)	1.5 (1.2)	50.9 (22.6)

[a]Evaluated using Z-score,[42] MolProbity,[43] and PROCHECK.[44] [b]Standard deviations are given in parentheses. [c]Docking has been performed using the LeadIT software suite.[45] [d]Note that values are only calculated for the final sets (those fulfilling the pocket criteria).

structures only cover less than half of the human kinome but are spread over all groups (Figure 2A).

Sequence Identities of the Template Structures Used for Homology Modeling. The structurally uncovered part of the kinome consists of 289 kinases. As a rule of thumb, 30−40% sequence identity (SI) is considered sufficient for homology modeling.[55] Encouragingly, for 60% of the kinases with unknown structure a template structure with more than 40% SI was found and for further 18% of the kinases the SI to the closest template structure is between 30 and 40% (Figure 2B). The remaining 22% of the kinases (mostly from the RGC and OTHER group, see Table 3) have lower SIs to their closest template structures. However, given that the overall fold of kinases is highly conserved, homology models were generated for all 289 kinases based on the closest template structure from the selKin set and collected in the hmKin set ("homology models").

Evaluation of the hmKin Set. Next, the generated HMs were evaluated in terms of internal energies and stereochemistry, i.e. bonds and angle constitutions, using the Z-score as calculated by the WHATIF[42] program, as well as results obtained from MolProbity[43] and PROCHECK[44] (Table 3). With a mean Z-score of −1.4 the models are slightly more than one standard deviation away from the average obtained for a set of gold standard reference structures[56] and thereby lie in the range of "good" structures. A mean MolProbity score of 1.6, which can be seen as an analogous to the resolution of X-ray structures, also underlines the quality of the models. Furthermore, PROCHECK revealed that on average 88.4% of the residues per structure lie within the most favored Ramachandran regions. Furthermore, the models were checked for their ability to accommodate ATP in their binding site by docking ATP molecules into the respective models using the LeadIT software suite.[45] The mean RMSD values per kinase group of the docked poses and the reference ATP (neglecting the triphosphate tail) was between 0.8 and 2.0 Å. Thus, the docked position resembles the binding mode in X-ray structures in most cases and the binding sites are capable of accommodating ATP-like ligands.

Overall, the underlying sequence identity of the used template structures as well as the results obtained from the stereochemistry, internal energy, and docking analysis indicate that the generated homology models can reliably be used for further binding site analysis.

Final Set Representing the Pocketome of Human Kinases. One major challenge in automatic pocket detection and description is the definition of the pocket border. Although many pocket detection methods perform reasonably well on many data sets, it proved useful to limit large scale pocket analysis to well-defined pockets (i.e., those which overlap decently with known ligand binding positions).[22] This filtering can be achieved by considering the mutual overlap between the pocket and a reference ligand. Therefore, only (sub)pockets were included in the final sets if at least 25% of the pocket was covered by the ligand and if more than 50% of the ligand was inside the pocket. This criterion resulted in final sets that cover 91%, 93%, and 81% of the initial brKin, selKin, and hmKin sets, respectively (kinases with black surrounding circles in Figure 2) and were used for all analysis described below.

Geometric and Physicochemical Features of Kinase ATP-Binding Pockets. DoGSiteScorer[26] has been used to predict potential pockets on the surface of the investigated kinase structures and the ATP pockets have been extracted for further analysis. Since the overall goal is to prioritize human kinase structures for drug design, a positive and negative control set was used to identify significant differences between the features of known druggable kinase structures as well as generally undruggable protein structures.

Druggable binding pockets are known to have a decent volume, are sufficiently enclosed, and relatively hydrophobic to accommodate a drug-like ligand with high affinity.[22] The values obtained in this study of the pocket descriptors such as volume, depth, and hydrophobicity vary between as well as within the different kinase groups but a similar trend can be observed for the three mentioned pocket descriptors (Figure 3): (I) The mean value for the volume of undruggable pockets is 313 (±213) Å^3 while the mean value of pockets of known kinase drug targets is 650 (±151) Å^3. This is in line with published values for features of known druggable and undruggable targets.[26] The mean pocket volume values of the X-ray structures from the selKin set are all lower than the mean values of the approved kinase drug targets but higher than the mean values of the undruggable pockets. The mean volumes of ATP pockets from the homology models are more spread and include larger (TKL and STE), similar sized (CK1 and AGC), as well as smaller pockets (TK, CAMK, CMGC, and OTHER) than those of the approved drug targets. However, the mean values for the different kinase groups in the selKin and hmKin

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

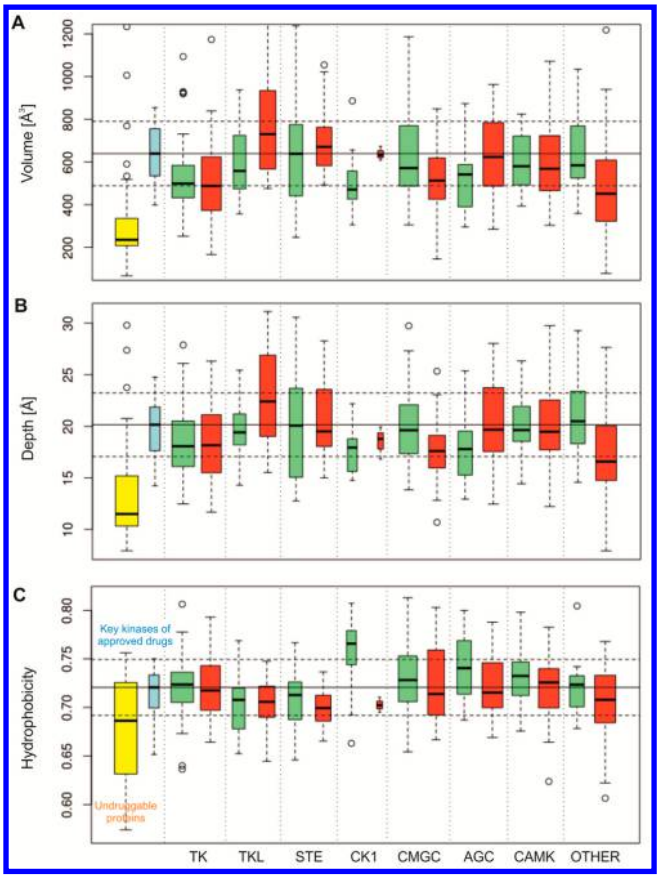

Figure 3. Boxplots showing the pocket properties volume (A), depth (B), and hydrophobicity (C), separated by kinases groups (green: PDB structures from the selKin set, red: homology models from the hmKin set) together with a negative control set of known undruggable proteins (yellow) and a positive control set of key targets of approved kinase inhibitors (blue). Box widths represent the number of pockets contained in the respective groups. The boxes indicate the median value as well as the upper and lower quartiles. The RGC group is only represented by one homology model and, thus, not shown.

sets, respectively, are all above 500 Å^3 (except for the homology models of the OTHER group which is slightly lower) and thus within the standard deviation of the approved drug targets (Figure 3A). (II) A similar conclusion can be drawn for the depth of the inspected pockets. All groups have average values between 18 (±4.5) and 23 (±4.4) Å (excluding the only modeled structure available for RGC with a depth of 16 Å) and are thereby in the range of the mean depth of the approved drug targets (20 (±3.1) Å) and considerably deeper than the pockets of undruggable targets (13.4 (±4.9) Å). (III) In terms of hydrophobicity, druggable pockets of key kinase targets have a mean value of 0.71 (±0.03), while undruggable ones have a mean value of 0.68 (±0.05). Note that the range of exhibited values here is relatively narrow. Mean values per kinase group and set range between 0.70 (±0.02) and 0.75 (±0.05), except the analyzed pocket of the modeled RGC structure which has a predicted hydrophobicity of 0.69.

Overall, these results indicate that the majority of the analyzed kinase ATP pockets have geometric and physicochemical properties similar to those of druggable kinase targets and sufficiently dissimilar to those of known undruggable targets.

Extracting the Druggable Human Kinome. For each structure of the brKin, selKin, and hmKin sets a structure-based druggability score (SBDS) was calculated using DoGSiteScorer.[26] The underlying support vector machine returns druggability scores in the range of zero to one (higher values point to higher chances of being druggable); pockets with values above 0.5 are considered as potentially druggable.[26] In line with the feature analysis presented above, 77% of the pocketome of human kinases preferentially fall in the druggability area (SBDS ≥ 0.5). As expected, the kinase structures in complex with approved drugs have a high mean SBDS of 0.7 (±0.1), while pockets of undruggable protein structures have a mean SBDS of 0.3 (±0.15) (Figure 4). Note

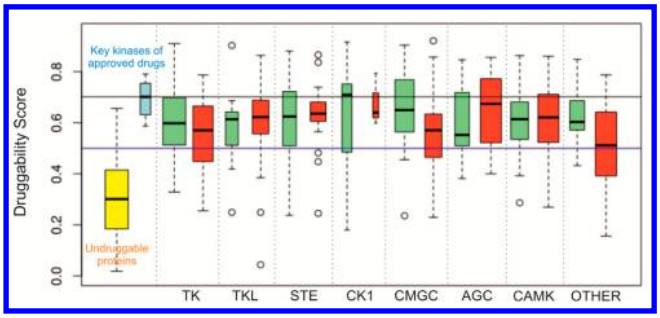

Figure 4. Boxplots showing the druggability scores separated by kinase groups (green: PDB structures from the selKin set, red: homology models from the hmKin set) together with a negative control set of known undruggable proteins (yellow) and a positive control set of key targets of approved kinase inhibitors (blue). Box widths represent the number of pockets contained in the respective sets. The boxes indicate the median value as well as the upper and lower quartiles.

that this set of kinase structures (cocrystallized with approved drugs) as well as the selKin set do not overlap with the structures used for training the DoGSiteScorer method.[26] The SBDSs of the analyzed pockets from the selKin and the hmKin sets have almost all groupwise mean druggability scores equal or larger than 0.6 (±0.1) (except for the HMs of the RGC and OTHER group). Splitting the analysis into X-ray structures and homology models, the ATP pockets of 83% of the selKin set and of 73% of the hmKin set have a SBDS ≥ 0.5. Overall, the lowest SBDSs were obtained for the homology models of the OTHER group with 49% falling into the undruggable area. Reasons for the lower SBDS values for the OTHER group might be that they are indeed less druggable or that the results are biased by the underlying homology models which have rather poor stereochemistry or internal energies (Table 3) and were generated predominantly with templates with sequence identity below 30%. Few conclusions can be drawn for kinases of the RGC group as it is only represented by one homology model with a borderline SBDS of 0.5.

Effect of Homology Models on Predicted Druggability. Results obtained for homology models have to be considered with caution due to uncertainties in the structural modeling step. However, we feel that they provide useful information especially for those cases where the underlying template structure has a high sequence identity (>40%) as present for 60% of the modeled structures. Note that the correlation between the SBDS of the homology models and the SBDS obtained for the respective template structure is weak (R = 0.3) and that no correlation between SBDSs of the homology models and the sequence identities of the respective template structures exists (R = 0.16). Thus, the modeled structures are

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

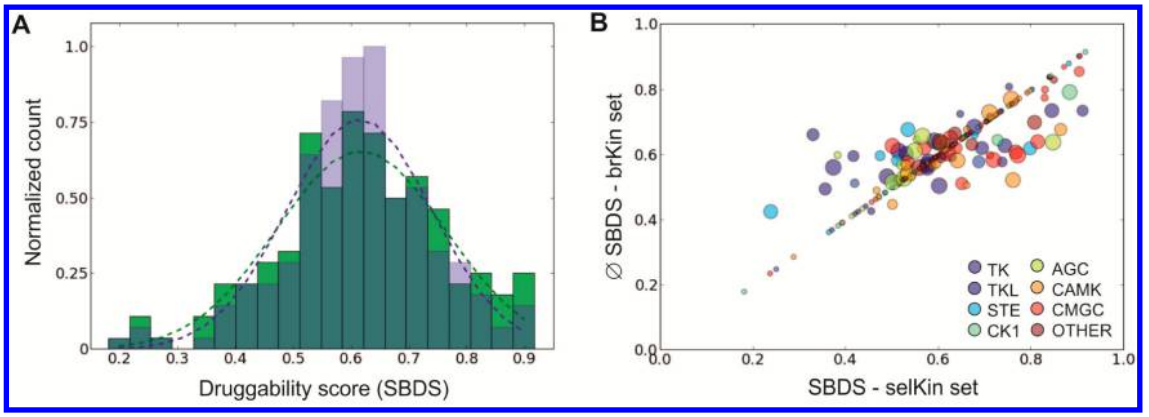

Figure 5. (A) Normalized distributions of druggability scores (SBDS) from the selKin set (green) as well as the mean SBDS from the brKin set (blue). Normal distributions of the SBDS are shown as dotted lines. (B) Correlation between SBDS of the selKin set compared to the mean SBDS of the brKin set. Points are colored by kinase groups and sizes represent the number of respective kinase structures in the brKin set.

distinguishable from their template structure and add valuable information to the analysis of the pocketome of human kinases.

A potential limitation of the presented analyses is that they are based so far on single, static structures—which represent only snapshots from otherwise flexible molecules—while even small conformational changes can impact the computationally detected pocket, the derived properties, and predicted druggability scores.[22] Accordingly, druggability scores can vary over a large range. Similar observations were obtained for other protein classes.[28−30] To test the robustness against conformational changes in the binding site, SBDSs were analyzed, first, for apo and holo structures in the selKin as well as for apo/holo kinase pairs in the brKin set, and second, SBDSs of the selected structures in the selKin set were compared to SBDSs of the respective background structures in the brKin set.

Effect of Bound State on Predicted Druggability. In the selKin set, the mean SBDS for the 163 holo structures is 0.63 (±0.13), while the mean SBDS for the 19 apo structures is 0.51 (±0.18). Note that by design of the selKin set most structures are bound to ATP-like ligands. The most prominent difference between the respective pocket properties is the calculated volume (i.e., on average the holo structures have larger pockets than the apo ones; data not shown). Due to the nonredundancy criterion of the selKin set, no pair of structures with an apo and holo structure from one kinase was present. However, in the brKin set, 42 apo/holo kinase pairs exist. In 64% of these pairs, the holo structure has a higher druggability score than the respective apo counterpart. The mean SBDSs are not significantly different for holo (0.63 (±0.15)) and apo structures (0.59 (±0.13)) for these pairs (evaluated using *t* test). The analysis of apo and holo structures indicate that the method is fairly robust against conformational changes in the binding site as induced by ligand binding. Similar conclusions were also drawn in druggability studies on other drug targets,[27,30] suggesting that also unbound structures can be used for druggability assessments.

Effect of Structure Selection on Predicted Druggability. To further test the robustness of the predicted druggability scores on the underlying protein structure, SBDSs of the individual kinase structures from the selKin set were compared to the mean SBDSs (Ø $SBDS_{br}$) obtained for all structures of the respective kinase in the brKin set. The distribution of the SBDSs of the selected structures and the Ø $SBDS_{br}$ over the respective background set are very similar (e.g., both have a mean value of 0.6 (±0.1)) and the values are correlated (R = 0.88), underlying the robustness of the calculated scores (Figure 5). Note that some kinases have only one structure available; consequently, the calculated values for SBDS and Ø $SBDS_{br}$ are the same. Excluding these data points from the analysis still results in a good correlation (R = 0.67). Overall, one could conclude that the selKin set does not only have high quality structures but also comprises representative structures for the individual kinases. Although the predicted scores are fairly robust toward small conformational changes in the binding site, it is advisable to consider more than one structure for druggability predictions if available.[27]

Kinome Coverage by Kinase Inhibitor Data. So far, we explored the potential of kinases to be druggable from a structural perspective. The experimental kinase inhibitor data as offered by ChEMBL[8] provides the opportunity to compare the druggability predictions with experimental bioactivity data. For this, we calculated the fraction of compounds that are active on the respective kinases. The latter value indicates the hit rates obtained for different kinases and is called the "ligand-based druggability score (LBDS) throughout this text in order to distinguish it from the calculated "structure-based druggability score" (SBDS). Interestingly, the analysis showed that only around 50% of the human kinome is currently investigated by publicly available bioactivity data. This again emphasizes that a large fraction of the kinome is unexplored in drug discovery projects. Both values, SBDS and LBDS, by definition imply the potential of a kinase structure to be explored by drug-like compounds. However, no correlation between SBDS and LBDS was found (Figure 6). Reasons for this can be manifold. First, research interest can steer or prevent the exploration of specific targets irrespective of the potential of the target to be addressed by low molecular weight compounds. Second, finding one compound out of a pool of thousands can already be sufficient to classify a protein as being druggable. Interestingly, we found that out of the 35 kinases with high SBDS (≥0.8), 36% have no bioactivity data available, while out of the 95 kinases with low SBDS (<0.5) 53% have no bioactivity data assigned.

Analysis of Key Kinase Targets. Next, the druggability scores of the 36 known key kinase targets for approved drugs (Table 4) are further analyzed. Please note that the underlying structures are in the DFG-in state and mostly bound to ATP-analogues in contrast to the key targets of the positive control set (Table 2) which are bound to the respective approved drug. Encouragingly, 73% of the key kinases have a SBDS ≥ 0.5 and

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

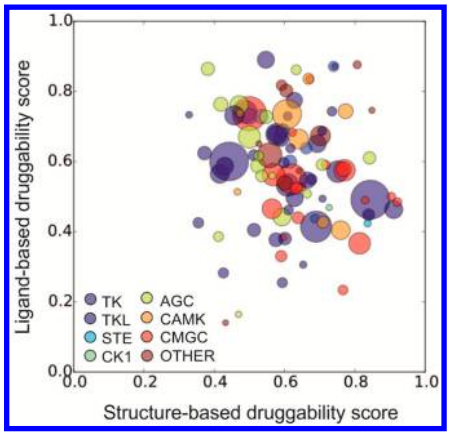

Figure 6. Scatter plot showing structure-based versus ligand-based druggability scores (i.e., SBDS vs LBDS). The plotted data include values for PDB structures from the selKin set as well as homology models from the hmKin set. Points are colored by kinase groups, and sizes represent the number of bioactivity data for the respective kinase.

70% of the key targets have a LBDS ≥ 0.5. The top ranking key targets, e.g. ABL1 (SBDS = 0.91) and EGFR (SBDS = 0.84), are among the most studied kinases in ChEMBL (i.e., large number of tested compounds). Though, several of the key targets including FRK, MAP2K2, EphA2, RET, and BTK have predicted low SBDS but also a small number of bioactivity values in ChEMBL implying that they are challenging targets. This shows that also ligand-based data has to be interpreted carefully when it comes to the prioritization of novel drug targets.

A closer inspection of the underlying structures of low SBDS scoring kinases (SBDS ≤ 0.5) revealed that one-third was represented by homology models (FRK, FLT3, PDGFRb), another third has only one X-ray structure (KDR, MAP2K2, EphA2), and that kinases in the remaining third have $ØSBDS_{br}$ in the range of druggable kinases ($ØSBDS_{br}$: JAK2 = 0.53, RET = 0.49, BTK = 0.66). Comparisons between SBDS and $ØSBDS_{br}$ values provide useful information regarding the choice of the underlying structures and the significance of the prediction. A large discrepancy between the two values points to cases where the underlying analysis should be manually inspected. For example, BTK was predicted with a low SBDS of 0.33, which differs significantly from the score calculated for the six available structures ($ØSBDS_{br}$ = 0.66). The revision of the BTK kinase structure from the selKin set (PDB code: 3PIX) revealed that its activation loop points in the ATP-binding pocket resulting in a very small pocket and thus, yielded a low (but reasonable) druggability score. The reason for the selection of this structure was that either the activation or the G-loop in all the other structures of the BTK kinase was not resolved. In general, manual inspection of the underlying targets and detected pockets is advisable when it comes to the interpretation of low ranking kinases.

Prioritizing New Kinases for Drug Discovery. Kinase drug discovery efforts are focused on a few clinically validated drug targets,[12,13] and as a result, most of the key targets of FDA approved drugs are accumulated in the TK area. The human kinome consists of 492 typical kinases of which 423 kinases were evaluated by DoGSiteScorer regarding their potential to be druggable. Remarkably, the ATP pockets of 327 of these kinases (77%) were predicted to be druggable (SBDS ≥ 0.5) and span over the complete kinome tree (Figure 7). Mutation studies as well as network analysis suggest a similar spread of the potential of many kinases other than TK for drug discovery projects.[14]

Table 4. List of Key Kinase Targets (Sorted by Descending Score) with Their SBDS Obtained for the Different Sets, LBDS, and Number of Bioactivity Data Points for Each Kinase

kinase	SBDS[a]	$ØSBDS_{br}$[b]	LBDS	no. of bioactivities
ABL1	0.91	0.74 (5)	0.46	529
EGFR	0.84	0.74 (8)	0.49	2440
TIE2	0.84	*d*	0.45	230
p38b	0.83	0.8 (2)	0.49	96
DDR2	0.75	*c*	0.57	761
ErbB2	0.75	0.80 (2)	0.82	15
LYN	0.71	*d*	0.49	83
PDGFRa	0.71	*c*	0.69	129
JAK3	0.69	0.62 (7)	0.65	514
SRC	0.69	*c*	0.41	1552
FLT1	0.67	*c*	0.55	424
FGFR3	0.66	*d*	0.64	222
RAF1	0.66	*d*	0.78	16
ErbB4	0.65	0.73 (2)	0.90	29
FGFR1	0.63	0.59 (3)	0.50	464
KIT	0.63	*d*	0.77	419
BRAF	0.62	0.63 (9)	0.64	136
FLT4	0.62	*c*	0.60	290
LCK	0.61	*c*	0.53	840
ALK	0.59	*c*	0.69	297
JAK1	0.59	0.64 (10)	0.66	377
MET	0.58	0.58 (9)	0.68	768
ITK	0.58	0.56 (8)	0.68	468
FMS	0.55	*d*	0.89	463
FGFR2	0.53	0.6 (9)	0.73	330
MAP2K1	0.53	0.68 (8)	0.69	14
JAK2	0.49	0.53 (10)	0.74	873
FRK	0.48	*c*	0.00	1
FLT3	0.46	*c*	0.73	602
KDR	0.44	*d*	0.60	2415
PDGFRb	0.43	*c*	0.59	492
MAP2K2	0.39	*d*	0.24	19
EphA2	0.37	*d*	0.43	19
RET	0.35	0.49 (6)	0.43	179
BTK	0.33	0.66 (6)	0.73	73

[a]SBDS obtained for structures from the selKin or hmKin set, respectively. [b]$ØSBDS_{br}$ obtained for the respective structures in the brKin set. The number of PDB structures is given in parentheses. If only one structure is present in the brKin set, no value is provided. [c]No X-ray structure available. [d]Only one structure present in the brKin set and thus the same value as for selKin set is obtained.

Our positive control set of 10 known key targets of FDA approved drugs obtained a mean SBDS of 0.7. A total number of 112 kinases (selKin: 54, hmKin: 58) have a druggability score equal or higher than this cutoff. Of the 54 kinases with SBDS ≥ 0.7 in the selKin set, 76% have an $ØSBDS_{br}$ > 0.7, while the remaining 24% still have an $ØSBDS_{br}$ > 0.5. Furthermore, the SBDSs of 30 kinases (selKin: 20; hmKin: 10) are between 0.8 and 0.9, and 6 kinases (selKin: 5; hmKin: 1) even have a druggability score above 0.9. In pursuit of prioritizing new targets for drug discovery, these six top ranking structures are now further investigated.

The top ranking structures include one key target of approved drugs, i.e. ABL1, as well as yet less exploited targets,

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

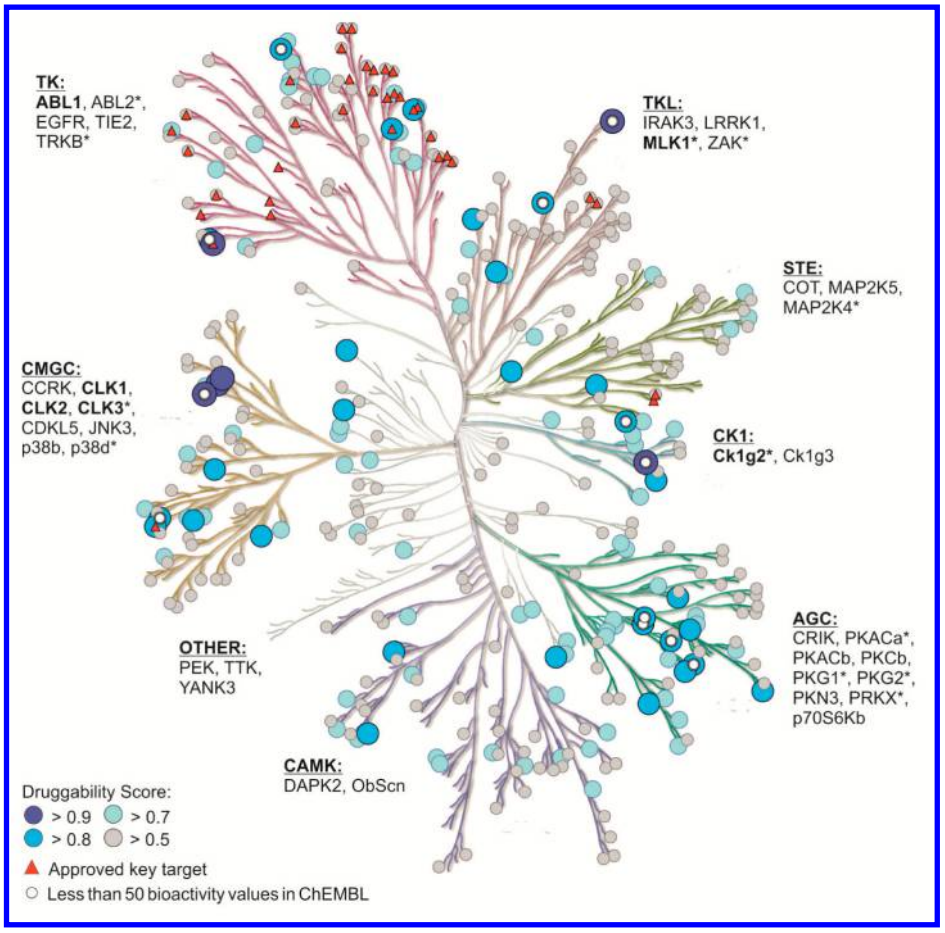

Figure 7. Distribution of kinases (including selKin and hmKin data) with a high druggability score as predicted by DoGSiteScorer. Color coding indicates the druggability score ranging from light to dark blue. Kinases which are key targets for FDA approved drugs are indicated by red triangles. Names of top ranking kinases per group (SBDS ≥ 0.8) are given of which the rather unexplored ones (i.e., those with have less than 50 bioactivity values in ChEMBL) are marked with an asterisk (Adapted with permission from www.cellsignal.com. Cell Signaling Technology, Inc.).

i.e., MLK1, CK1g2, and several CLKs (Table 5). The druggability potential of these six kinases is supported by their role in cancer formation or progression as described below.

Table 5. Kinases with SBDSs Equal or Higher than 0.9

kinase	group	SBDS (ØSBDS$_{br}$[b])	LBDS	no. of bioactivities[c]
ABL1	TK	0.91 (0.74)	0.46	529
MLK1	TKL	0.90 (−)	0.68	36
CK1g2	CK1	0.92 (−)	0.51	25
CLK1	CMGC	0.90 (−)	0.50	112
CLK3	CMGC	0.90 (0.86)	0.12	23
CLK4	CMGC	0.92 (−)[a]	0.48	126

[a]Homology model from the hmKin set. All other structures are from the selKin set. [b]ØSBDS$_{br}$ obtained for the respective structures in the brKin set. [c]Number of bioactivities in ChEMBL.

ABL1 (Abelson murine leukemia viral oncogene homolog 1; SBDS = 0.91) as well as the closely related kinase ABL2 (SBDS = 0.80) are predicted to be druggable. Notably, imatinib (Gleevec), the first FDA approved kinase drug and which was seen as the breakthrough in cancer treatment, was developed against the ABL kinase.[35] Although imatinib itself binds to the inactive DFG-out conformation, dasatinib and bosutinib bind to the active form. Interestingly, the ABL structure cocrystallized with dasatinib (PDB: 2GQG) obtained the highest SBDS (0.8) within our positive control set of key targets of approved drugs.[36] ABL transmits diverse extracellular signals to protein networks controlling proliferation, survival, migration, and invasion.[57] Several inhibitors are known to prevent the BCR-ABL complex from phosphorylating subsequent proteins and, thus, to interrupt the signaling cascade necessary for cancer growth and survival.

Another high scoring but less explored kinase is CK1g2 (casein kinase I; SBDS = 0.92). Also other members of the CK1 family were assigned with high SBDSs (SBDS of CK1g1 = 0.66; SBDS of CK1g3 = 0.88 (ØSBDS$_{br}$ = 0.79)). CK1s are involved in Wnt and Hedgehog signaling pathways and, thus, in the initiation and maintenance of many tumor types.[58] For example, CK1g2 phosphorylates the metastatic tumor antigen 1 (MTA1) short form and, thereby, modules its function. Furthermore, CK1 isoforms modulate tumor suppressor p53 and Mdm2 functions through site-directed phosphorylation.[59] Note that very few bioactivity data is available for the CK1 branch (less than 25 compounds for CK1g1 and CK1g2; none for CK1g3) and most of these compounds are not active against the respective CK1 target. The predicted SBDSs, however, point to a high potential of these kinase structures to be modulated with drug-like compounds.

Another high-scoring kinase is MLK1 (Mixed-linkage kinase 1; SBDS = 0.90). MLK1 acts upstream of Mitogen-activated

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

protein kinases (MAPKs) and is, thus, involved in the JNK pathway which is implicated in a multitude of diseases ranging from cancer to neurological and immunological/inflammatory conditions.[60] Again very few experimental data is available in ChEMBL (28) of which a large fraction of compounds are active against MKL1 (LBDS = 0.68).

The last group of top ranking kinases includes CLKs (CDC-like kinases). The whole branch of CLK kinases exhibits highly druggable pockets. CLK1, CLK3, and CLK4 (from the hmKin set) have SBDSs above 0.9, while CLK2 has still a value of 0.73. Among other roles, CLKs mediate the phosphorylated regulation of alternative RNA splicing in cancer whereby the modulation of CLKs provides a promising approach in cancer therapy.[61] While CLK1 (LBDS = 0.50) and CLK4 (LBDS = 0.48) have been tested on more than 100 compounds, CLK3 is rather unexplored (23 tested compounds and LBDS = 0.12).

Closely related kinase structures which all have high predicted druggability scores may be of high practical interest. Similar to the top ranking kinases belonging to the CK1 and CLK families, also several kinases from the branch containing the PKA, PKG, and PRK families in the AGC group have high SBDSs (i.e., PKACa, PKACb, PKG1, PKG2, PRXK SBDS > 0.8, PKACg SBDS = 0.79). Furthermore, our analysis points to yet untapped protein kinases for drug discovery projects (i.e., kinases with SBDS $\geq$ 0.8 but less than 50 tested compounds; Figure 7).

SUMMARY AND CONCLUDING REMARKS

Human kinases are among the most studied proteins in cancer research. However, the number of clinically validated kinase targets is very limited,[12,13] and most of the current drug discovery projects focus on kinases in the TK group. This focus should not be regarded as an indication of which fraction of the kinome is druggable but rather as a reflection of research bias. A reason for this might be that our insights into cellular signaling events are still limited and that about 50% of the kinases remain largely uncharacterized.[13] Mutation studies as well as network analysis indicate that a large number of kinases—which could be drug targets—might not be sufficiently explored.[14]

The structural coverage of the kinome by more than three thousand X-ray structures provides a great opportunity to prioritize the human kinome by computational methods as druggable targets. Along this line, we performed here a large and sophisticated analysis of the ATP binding pockets of the entire kinome by generating and analyzing a high-quality set of kinase structures and introducing models for the remaining part of the kinome by homology modeling. The inclusion of homology models provided valuable information which would not have been obtained by analyzing only available experimental structures.

The presented analysis revealed that from a structural perspective around 75% of the analyzed kinases have binding site characteristics that should allow the design of drug-like compounds (i.e., they are potentially druggable), of which a quarter have a druggability score above 0.7. The top ranking structures (SBDS $\geq$ 0.9) include well established drug targets such as ABL1 but also yet less explored kinases, i.e., MLK1, CK1g2, and several CLKs. Note that for the remaining less druggable quarter of the kinases the used structures or models may need individual inspections and that the scores may change toward the more druggable area once additional or more appropriate structures become available.

The presented analysis focused on a high-quality set of selected structures, which are bound to the same class of compounds (i.e., bound to ATP-like compounds). The motivation for this was to reduce the potential bias of the conformational states of the analyzed kinase structures on the resulting druggability scores. As pointed out before, conformational changes can affect the predicted druggability scores. Consequently, it is advisable to use more than one structure for druggability predictions, if available. Thus, we added the information from the background structures per kinase to support the predicted results. The BTK example nicely demonstrated that a large discrepancy between SBDS obtained for a single selected structure and the ØSBDS_{br} of the background structures can indicate that the scored conformation may have some limitation and should be considered with caution.

In addition to the structure-based druggability score (SBDS), a ligand-based druggability score was investigated in this study. Not unexpected for us, the predicted SBDSs do not correlate with the experimental hit rates (LBDS). Several reasons can account for this including the bias of drug discovery efforts on clinically validated kinases, the presence of DFG-in and DFG-out binders in the considered ligands, as well as the use of only one selected (DFG-in) structure per kinase for our main analysis.

The assessment and selection of targets is an essential but challenging area in kinase drug discovery.[13] Several aspects have to be taken into account such as the activation of the kinase in cancer cells, the frequency of mutations, and the affected pathways. However, the multitude and diversity of mutations even in similar cancer lines renders rational target selection challenging.[16,62,63] The presented prioritization of druggable ATP pockets puts the spotlight on novel and less explored kinases. We hope that this prioritization will facilitate the target selection in upcoming kinase projects, e.g., as a filter for large-scale target identification approaches or even to initiate screening campaigns.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: fulle@bio.mx (S.F.).
*E-mail: volkamer@bio.mx (A.V.).

Funding

The work of the BioMed X employees is kindly sponsored by Merck KGaA.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Mireille Krier, Daniel Kuhn, and Rebecca Wade for fruitful discussions concerning kinase specific druggability evaluation and Christian Tidona and Ulrich Betz for their general support. Furthermore, we thank Matthias Rarey and Christian Lemmen for providing an academic license of the DoGSiteScorer and the LeadIT software and Cell Signaling Technology, Inc. (www.cellsignal.com) for the courtesy of reproducing the kinome tree illustration. We also want to thank the reviewers for very useful comments.

ABBREVIATIONS

selKin, nonredundant set of selected kinase structures in the DFG-in conformation and determined by X-ray crystallog-

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

raphy; brKin, background set, holding up to 10 kinase structures in the DFG-in conformation per kinase; hmKin, set of kinase structures modeled by homology modeling; HM, homology modeling; SBDS, structure-based druggability score; LBDS, ligand-based druggability score; TK, tyrosine kinase; TKL, tyrosine kinase-like; STE, homologues of yeast sterile 7, sterile 11, sterile 20 kinases; CK1, casein kinase 1; CMGC, cyclin-dependent kinase, mitogen-activated protein kinase, glycogen synthase kinase 3, and dual specificity protein kinase containing families; AGC, containing PKA, PKG, PKC families; CAMK, calcium/calmodulin-dependent protein kinase

REFERENCES

(1) Klebl, B., Muller, G., Hamacher M., Mannhold, R., Kubinyi, H.; Folkers, G., Eds. *Protein Kinases as Drug Targets*; Wiley, 2011; Vol. *49*.

(2) Wood, L. D.; Parsons, D. W.; Jones, S.; Lin, J.; Sjöblom, T.; Leary, R. J.; Shen, D.; Boca, S. M.; Barber, T.; Ptak, J.; et al. The Genomic Landscapes of Human Breast and Colorectal Cancers. *Science* **2007**, *318*, 1108−1113.

(3) Lin, J.; Gan, C. M.; Zhang, X.; Jones, S.; Sjöblom, T.; Wood, L. D.; Parsons, D. W.; Papadopoulos, N.; Kinzler, K. W.; Vogelstein, B.; et al. A Multidimensional Analysis of Genes Mutated in Breast and Colorectal Cancers. *Genome Res.* **2007**, *17*, 1304−1318.

(4) Futreal, P. A.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M. R. A Census of Human Cancer Genes. *Nat. Rev. Cancer* **2004**, *4*, 177−183.

(5) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(6) Manning, G.; Whyte, D.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298*, 1912−1934.

(7) Brooijmans, N.; Chang, Y.-W.; Mobilio, D.; Denny, R. A.; Humblet, C. An Enriched Structural Kinase Database to Enable Kinome-Wide Structure-Based Analyses and Drug Discovery. *Protein Sci.* **2010**, *19*, 763−774.

(8) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083−D1090.

(9) 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.; et al. A Quantitative Analysis of Kinase Inhibitor Selectivity. *Nat. Biotechnol.* **2008**, *26*, 127−132.

(10) Zhang, J.; Yang, P. L.; Gray, N. S. Targeting Cancer with Small Molecule Kinase Inhibitors. *Nat. Rev. Cancer* **2009**, *9*, 28−39.

(11) Fang, Z.; Grütter, C.; Rauh, D. Strategies for the Selective Regulation of Kinases with Allosteric Modulators: Exploiting Exclusive Structural Features. *ACS Chem. Biol.* **2013**, *8*, 58−70.

(12) Zhang, L.; Daly, R. J. Targeting the Human Kinome for Cancer Therapy: Current Perspectives. *Crit. Rev. Oncog.* **2012**, *17*, 233−246.

(13) Fedorov, O.; Müller, S.; Knapp, S. The (un)targeted Cancer Kinome. *Nat. Chem. Biol.* **2010**, *6*, 166−169.

(14) Greenman, C.; Stephens, P.; Smith, R.; Dalgliesh, G. L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; et al. Patterns of Somatic Mutation in Human Cancer Genomes. *Nature* **2007**, *446*, 153−158.

(15) Patel, M. N.; Halling-Brown, M. D.; Tym, J. E.; Workman, P.; Al-Lazikani, B. Objective Assessment of Cancer Genes for Drug Discovery. *Nat. Rev. Drug Discovery* **2013**, *12*, 35−50.

(16) Manning, B. D. Challenges and Opportunities in Defining the Essential Cancer Kinome. *Sci. Signal.* **2009**, *2*, pe15.

(17) Grueneberg, D. A.; Degot, S.; Pearlberg, J.; Li, W.; Davies, J. E.; Baldwin, A.; Endege, W.; Doench, J.; Sawyer, J.; Hu, Y.; et al. Kinase Requirements in Human Cells: I. Comparing Kinase Requirements across Various Cell Types. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16472−16477.

(18) Workman, P.; Al-Lazikani, B. Drugging Cancer Genomes. *Nat. Rev. Drug Discovery* **2013**, *12*, 889−890.

(19) Bulusu, K. C.; Tym, J. E.; Coker, E. A.; Schierz, A. C.; Al-Lazikani, B. canSAR: Updated Cancer Research and Drug Discovery Knowledgebase. *Nucleic Acids Res.* **2014**, *42*, D1040−7.

(20) Egner, U.; Hillig, R. C. A Structural Biology View of Target Drugability. *Expert Opinion on Drug Discovery* **2008**, *3*, 391−401.

(21) Barril, X. Druggability Predictions: Methods, Limitations, and Applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 327−338.

(22) Volkamer, A.; Rarey, M. Exploiting Structural Information for Drug-Target Assessment. *Future Med. Chem.* **2014**, *6*, 319−331.

(23) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-Based Maximal Affinity Model Predicts Small-Molecule Druggability. *Nat. Biotechnol.* **2007**, *25*, 71−75.

(24) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **2009**, *10*, 168.

(25) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J. Chem. Inf. Model.* **2009**, *49*, 377−389.

(26) Volkamer, A.; Kuhn, D.; Grombacher, T.; Rippmann, F.; Rarey, M. Combining Global and Local Measures for Structure-Based Druggability Predictions. *J. Chem. Inf. Model.* **2012**, *52*, 360−372.

(27) Krasowski, A.; Muthas, D.; Sarkar, A.; Schmitt, S.; Brenk, R. DrugPred: A Structure-Based Approach to Predict Protein Druggability Developed Using an Extensive Nonredundant Data Set. *J. Chem. Inf. Model.* **2011**, *51*, 2829−2842.

(28) Santiago, C.; Nguyen, K.; Schapira, M. Druggability of Methyl-Lysine Binding Sites. *J. Comput. Aided. Mol. Des.* **2011**, *25*, 1171−1178.

(29) Campagna-Slater, V.; Mok, M. W.; Nguyen, K. T.; Feher, M.; Najmanovich, R.; Schapira, M. Structural Chemistry of the Histone Methyltransferases Cofactor Binding Site. *J. Chem. Inf. Model.* **2011**, *51*, 612−623.

(30) Vidler, L. R.; Brown, N.; Knapp, S.; Hoelder, S. Druggability Analysis and Structural Classification of Bromodomain Acetyl-Lysine Binding Sites. *J. Med. Chem.* **2012**, *55*, 7346−7359.

(31) Radusky, L.; Defelipe, L. a; Lanzarotti, E.; Luque, J.; Barril, X.; Marti, M. a; Turjanski, A. G. TuberQ: A Mycobacterium Tuberculosis Protein Druggability Database. *Database (Oxford)* **2014**, *2014*, bau035.

(32) Van Linden, O. P. J.; Kooistra, A. J.; Leurs, R.; de Esch, I. J. P.; de Graaf, C. KLIFS: A Knowledge-Based Structural Database to Navigate Kinase-Ligand Interaction Space. *J. Med. Chem.* **2014**, *57*, 249−277.

(33) Chemical Computing Group Inc.. Molecular Operating Environment (MOE). *Sci. Comput. Instrum.* **2004**, *22*, 32.

(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*, 1839−1854.

(35) Pray, L. A. Gleevec: The Breakthrough in Cancer Treatment. *Nat. Educ.* **2008**, *1*, 37.

(36) Treiber, D. K.; Shah, N. P. Ins and Outs of Kinase DFG Motifs. *Chem. Biol.* **2013**, *20*, 745−746.

(37) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.

(38) The UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2014**, *42*, 7486.

(39) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

(40) Krieger, E.; Vriend, G. YASARA−Yet Another Scientific Artificial Reality Application http://www.yasara.org/ (accessed Aug 27, 2014).

(41) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947−2948.

(42) Vriend, G. WHAT IF: A Molecular Modeling and Drug Design Program. *J. Mol. Graph.* **1990**, *8* (52−56), 29.

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549

(43) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: All-Atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 12−21.
(44) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.
(45) BioSolveIT GmbH. LeadIT software. http://www.biosolveit.de/LeadIT/ (accessed Sep 17, 2014).
(46) Strömbergsson, H.; Kleywegt, G. J. A Chemogenomics View on Protein-Ligand Spaces. *BMC Bioinformatics* **2009**, *10* (Suppl 6), S13.
(47) DeLano, W. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.* **2002**, No. March, 44.
(48) Volkamer, A.; Griewel, A.; Grombacher, T.; Rarey, M. Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. *J. Chem. Inf. Model.* **2010**, *50*, 2041−2052.
(49) Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53*, 5858−5867.
(50) Bellis, L. J.; Akhtar, R.; Al-Lazikani, B.; Atkinson, F.; Bento, A. P.; Chambers, J.; Davies, M.; Gaulton, A.; Hersey, A.; Ikeda, K.; et al. Collation and Data-Mining of Literature Bioactivity Data for Drug Discovery. *Biochem. Soc. Trans.* **2011**, *39*, 1365−1370.
(51) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719−2740.
(52) Cheng, Y.; Prusoff, W. H. Relationship between the Inhibition Constant (K1) and the Concentration of Inhibitor Which Causes 50 per Cent Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(53) McKinney, W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, Austin, Texas, June 28−July 3, 2010; pp 51−56.
(54) Elkins, J. M. Case Study−Structural Genomics and Human Protein Kinases. *Methods Mol. Biol.* **2014**, *1140*, 325−336.
(55) Rost, B. Twilight Zone of Protein Sequence Alignments. *Protein Eng.* **1999**, *12*, 85−94.
(56) Hooft, R. W.; Vriend, G.; Sander, C.; Abola, E. E. Errors in Protein Structures. *Nature* **1996**, *381*, 272.
(57) Greuber, E. K.; Smith-Pearson, P.; Wang, J.; Pendergast, A. M. Role of ABL Family Kinases in Cancer: From Leukaemia to Solid Tumours. *Nat. Rev. Cancer* **2013**, *13*, 559−571.
(58) Price, M. A. CKI, There's More than One: Casein Kinase I Family Members in Wnt and Hedgehog Signaling. *Genes Dev.* **2006**, *20*, 399−410.
(59) Knippschild, U.; Wolff, S.; Giamas, G.; Brockschmidt, C.; Wittau, M.; Würl, P. U.; Eismann, T.; Stöter, M. The Role of the Casein Kinase 1 (CK1) Family in Different Signaling Pathways Linked to Cancer Development. *Onkologie* **2005**, *28*, 508−514.
(60) Ola, A.; Kerkelä, R.; Tokola, H.; Pikkarainen, S.; Skoumal, R.; Vuolteenaho, O.; Ruskoaho, H. The Mixed-Lineage Kinase 1−3 Signalling Pathway Regulates Stress Response in Cardiac Myocytes via GATA-4 and AP-1 Transcription Factors. *Br. J. Pharmacol.* **2010**, *159*, 717−725.
(61) Naro, C.; Sette, C. Phosphorylation-Mediated Regulation of Alternative Splicing in Cancer. *Int. J. Cell Biol.* **2013**, *2013*, 1−15.
(62) Graves, L. M.; Duncan, J. S.; Whittle, M. C.; Johnson, G. L. The Dynamic Nature of the Kinome. *Biochem. J.* **2013**, *450*, 1−8.
(63) Duncan, J. S.; Whittle, M. C.; Nakamura, K.; Abell, A. N.; Midland, A. A.; Zawistowski, J. S.; Johnson, N. L.; Granger, D. A.; Jordan, N. V.; Darr, D. B.; et al. Dynamic Reprogramming of the Kinome in Response to Targeted MEK Inhibition in Triple-Negative Breast Cancer. *Cell* **2012**, *149*, 307−321.

DOI: 10.1021/ci500624s
J. Chem. Inf. Model. 2015, 55, 538−549