J Comput Aided Mol Des (2014) 28:13–23
DOI 10.1007/s10822-013-9700-4

Scaffold mining of kinase hinge binders in crystal structure database

Li Xing · Brajesh Rai · Elizabeth A. Lunney

Received: 20 September 2013 / Accepted: 16 December 2013 / Published online: 29 December 2013

Abstract Protein kinases are the second most prominent group of drug targets, after G-protein-coupled receptors. Despite their distinct inhibition mechanisms, the majority of kinase inhibitors engage the conserved hydrogen bond interactions with the backbone of hinge residues. We mined Pfizer internal crystal structure database (CSDb) comprising of several thousand of public as well as internal X-ray binary complexes to compile an inclusive list of hinge binding scaffolds. The minimum ring scaffolds with directly attached hetero-atoms and functional groups were extracted from the full compounds by applying a rule-based filtering procedure employing a comprehensive annotation of ATP-binding site of the human kinase complements. The results indicated large number of kinase inhibitors of diverse chemical structures are derived from a relatively small number of common scaffolds, which serve as the critical recognition elements for protein kinase interaction. Out of the nearly 4,000 kinase-inhibitor complexes in the CSDb we identified approximately 600 unique scaffolds. Hinge scaffolds are overwhelmingly flat with very little sp3 characteristics, and are less lipophilic than their corresponding parent compounds. Examples of the most common as well as the uncommon hinge scaffolds are presented. Although the most common scaffolds are found in complex with multiple kinase targets, a large number of them are uniquely bound to a specific kinase, suggesting certain scaffolds could be more promiscuous than the others. The compiled collection of hinge scaffolds along with their three-dimensional binding coordinates could serve as basis set for hinge hopping, a practice frequently employed to generate novel invention as well as to optimize existing leads in medicinal chemistry.

Keywords Kinase domain · Kinase hinge · Crystal structure database · Hinge scaffold · ATP-binding site · Kinase inhibition · Selectivity · Lead optimization

Abbreviations

KD	Kinase domain
CSDb	Crystal structure database
SAR	Structure–activity relationship
PSA	Polar surface area
3D	Three-dimension

Introduction

Protein kinases are a leading class of drug targets, with applications to treat a number of diseases including inflammation, neurodegenerative disorders, cardiovasclular diseases and cancer [1]. The 518 protein kinases encoded in the human genome are involved in regulating essentially all aspects of cellular processes, from metabolism and cell cycle progression to differentiation and apoptosis [2]. The

Electronic supplementary material The online version of this article (doi:10.1007/s10822-013-9700-4) contains supplementary material, which is available to authorized users.

L. Xing (✉)
Pfizer Worldwide Research and Development,
200 CambridgePark Drive, Cambridge, MA 02140, USA
e-mail: li.xing@pfizer.com

B. Rai
Pfizer Worldwide Research and Development, 1 Eastern Point Road, Groton, CT 06340, USA

E. A. Lunney
Pfizer Worldwide Research and Development, 10777 Science Center Drive, San Diego, CA 92121, USA

complexity of protein kinase regulation offers many potential routes to their inhibition which have been actively pursued by the pharmaceutical industry. The majority of protein kinase inhibitors described to date occupy the ATP binding site. Despite initial concerns that the high ATP concentration within cells and the number of kinases present would confound endeavors to identify potent and selective inhibitors, significant advances have been made over the past 20 years [3–5]. To date, 25 drugs have been approved (including antibodies) targeting this class of enzymes, and another 200 or so are in clinical development for a wide variety of therapeutic areas.

Many kinases are believed to exist in a state of equilibrium between their active and inactive forms [6]. When the activation loop of a kinase is phosphorylated, protein kinases shift toward the catalytically active and productive conformational form. Numerous X-ray crystal structures have elucidated that catalytically active kinases share almost identical active site conformations, whereas unactivated kinases have diverse structures. Two frequently observed inactive conformations expose additional pockets that are accessible to kinase inhibitors, commonly termed DFG-out and αC-helix-out. In the DFG-out conformation, the conserved Asp-Phe-Gly (DFG) motif at the beginning of the activation segment is rotated out of the catalytic site, creating a pocket that is otherwise unavailable in the active kinase conformation. The approved Abl inhibitor Imatinib, RAF inhibitor Sorafenib, the p38 kinase inhibitor BIRB796 and the VEGFR/PDGFR inhibitor TAK-593 all exploit this deep pocket [7–10]. In the αC-helix-out conformation, a conserved salt bridge that anchors the αC-helix in the active kinase structure is disrupted, and the αC-helix rotates away from the ATP site, creating an additional pocket for compound such as Lapatinib, another FDA approved compound [11]. To the best of our knowledge, crystal structures are available for about 200 distinct protein kinases to date. Kinases with displaced αC-helices or DFG-out conformation of the activation segment each account for approximately 30 % of the total structures [12, 13]. A few kinases, including SRC and ABL, appear to have unlimited versatility of adopting both DFG-out and αC-helix-out capabilities [14–17]. Most of the time the inactive kinase conformation is stabilized by the complementary ligand binding. However, some kinases have been shown to naturally rest in the inactive states. For example, in complex with a number of compact ligands well confined in the adenine pocket, the DFG motif and activation loop of cFMS are found to be in the canonically inactive conformation [18]. Additionally, the crystal structure of the unphosphorylated BTK kinase domain shows that its αC-helix rotates out in an apo state [19].

Independent of the conformational state that a particular kinase inhibitor targets, strong hydrogen bond interaction with the kinase hinge is usually required in order to accomplish potent inhibition. Whereas the selectivity can be attained by boundless medicinal chemistry modifications beyond the adenosine binding pocket, the hinge binding core serves as the primary anchor to the ATP pocket for the kinase inhibitors. Their chemical structures along with their 3D positioning and orientation in the ATP pocket are of paramount importance to compound design. Here we describe a protocol to extract the kinase hinge binding scaffolds from large body of X-ray structural data. With the aid of 3D coordinates, the compiled collection can be applied as substrates to hinge hopping to discover novel chemical matters for several kinase programs.

Methods

Crystal structure database (CSDb)

Kinase protein/ligand crystal complexes were retrieved from the in-house crystal structure database, a large collection of X-ray structure repository that contains both internally solved structures and those selected and imported from the Protein Data Bank [20]. A total of 3,980 crystal structures of kinases in complex with small, drug-like ligands exemplifying a variety of ligand binding modes were pulled from the database. In such binary associations, a total of 167 distinct protein kinase across nine subfamilies are represented out of the 518 protein kinase complements, accounting for nearly 30 % of the human kinase genome [2]. These structures were then utilized to mine and extract the hinge interacting ring fragments. An important aspect of this database was that all structures from a given family were aligned onto a common frame of reference based on a superposition of preselected residues within the ATP site, making the subsequent analysis of the data derived from this database particularly efficient.

Annotation of kinase hinge residues

Kinase sequence alignment was performed by PFAAT using Blosum62 similarity matrix. A similarity analysis is performed based on residues in the kinase domain. Results are sorted first by ATP site similarity then by kinase domain similarity. Residues in the ATP binding sites making close contacts with the endogenous ligand are annotated by their respective positions. A backbone representation of the kinase annotation map is provided in Fig. 1. The amino acids on kinase hinge are fully annotated from H1 to H13, with H3 corresponding to the gatekeeping residue. In the C-terminal domain, the catalytic loop, the activation loop and the intervening β-strand β4 also serve

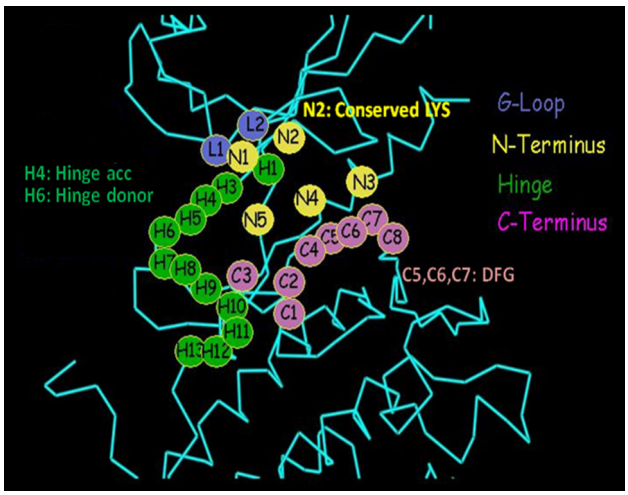


Fig. 1 Backbone representation of the catalytic domain of a kinase structure. Important ATP-site residues are shown using an internally-developed, consistent numbering scheme. Hinge residues used for hydrogen bond detection included H3, H4, H5, H6 and H7

as a structural elements that compose the ATP binding site, of which the key residues are denoted by C1 to C8, with C5, C6 and C7 representing the DFG motif. In the N-terminal domain, a number of important residues are annotated around the αC-helix, of which the conserved glutamate residue is N3. Additionally, the L1 and L2 annotations refer to the beginning and the end of the glycine-rich loop, which typically defines the ceiling of the ATP binding pocket. As a reference, for AKT kinase, these residues correspond to M227 for H3, E228 for H4, Y229 for H5, A230 for H6 and N231 for H7.

Extraction of hinge scaffolds

A set of criteria were defined to guide the hinge extraction which is depicted in the flowchart (Fig. 2). The first step in this process was to determine whether or not a given bound ligand formed a hydrogen bond with one or more hinge residues. Specifically, we considered hydrogen bonds between ligand and the backbone carbonyl or amine of the five hinge residues, labeled as H3, H4, H5, H6 and H7 in Fig. 1. Kinase domains of human protein complements are aligned by sequence homology, based on which the corresponding hinge residues are annotated for each kinase. The geometric criteria of a hydrogen bond interaction X–H···Y–Z are: X···Y distance <3.4 Å, and both XHY and HYZ angles >90°.

Once a ligand atom was found to form a hydrogen bond to a hinge residue, it was traversed along the connectivity map toward the rest of the molecule to complete the

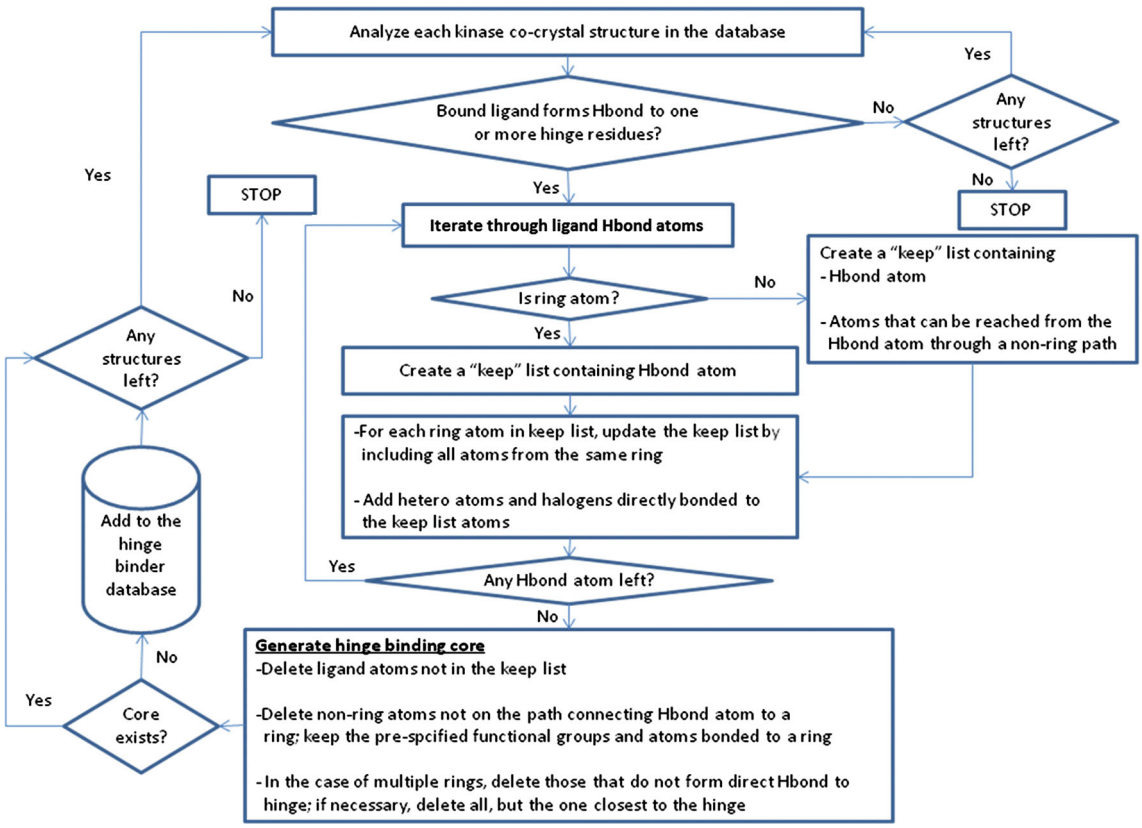


Fig. 2 Flowchart describing the major steps involved in generating hinge binder database

extraction of the core structure. Specifically, the hinge binding scaffold was defined using the following rules: if a ring fragment directly formed one or more hydrogen bonds to the above-specified hinge residues, all ring atoms, including any fused-ring systems, were extracted; in cases where the hydrogen bond was formed through a non-ring atom, the algorithm found the nearest ring fragment in addition to the non-ring atoms that forms the hydrogen bond, and the hinge core was defined by the aforementioned elements and includes all acyclic atoms in the path of the hydrogen bond atom to the closest ring. Furthermore, in case of multiple ring systems, precedence was given to the one that formed direct hydrogen bond interaction with the hinge over the one where the hinge interacting group was an acyclic atom. When necessary, another precedence criterion was applied to select ring system that was closest to the hinge region. Finally, due to their unique polar attributes and/or hydrogen bond capacities, hetero-atoms directly connected to the ring were considered part of the ring feature and preserved. For similar considerations, common functional groups were also preserved as part of the ring system. These included cyano, nitro, nitroso, carbonyl, amide, urea, carboxylate, sulfone, sulfonamide, sulfoxide, sulfinic and sulfonic acids (structures are given in Fig. S1).

The rest of the structural substituents comprising atoms and bonds that were farther away from the hinge were then removed. Hydrogen atoms were added to fill the open valences. In the end, the unique hinge scaffolds were identified from the nearly four thousand kinase inhibitors in CSDb.

Kinase selectivity data

Kinase selectivity data were extracted from multiple screening panels measured both in-house as well as by commercial suppliers, for compounds that exist in crystal structure database. For each selectivity panel, compounds were tested at two concentrations of 1 and 10 μM respectively. ATP concentrations are at the K_m level for the corresponding kinases. Compound activity was measured as % inhibition of the total signal based upon the control wells. The Gini coefficient at each screening concentration were computed using the percent inhibition data [21].

Results and discussion

Annotation of the ATP site

Multiple sequence alignments are routinely used to identify homologous sequences that are similar in structure and function. When combined with structural analysis this type of endeavor is particularly productive for discerning trends and patterns. The protein kinases bind a common endogenous ligand. Annotating the multiple sequence alignment of the human protein kinase complements provides quick access to composing information of the ATP active site of each kinase (Fig. 1). Knowledge is specified by annotation regarding conservation and/or mutation of a particular residue which could be critical to binding of a specific chemotype. Residue annotations are searchable, and allow selection of alignment columns for further sequence analysis, e.g. computing percent identities of the ATP sites. It can also be used to calculate the sequence identity of a particular structural element, e.g. the activation loop. The sequence annotation is hyperlinked to CSDb, which allows rapid retrieval of crystal structures. The annotated residues can be mapped quickly to the 3D crystal structures shall they exist, providing detailed graphical visualization of the ATP binding pocket. Critical structural elements including kinase hinge, glycine-rich loop, αC-helix and activation loop are effectively highlighted along with their residue compositions. Ligand–protein interaction patterns can be derived from the sequence annotation. As described herein, the ATP site annotation of the complete human kinase genome is a powerful knowledge base which could afford in-depth analysis to help elucidate ATP-competitive compound design.

The ATP binding pocket is situated at the folding cleft of the N- and C-lobes, which are connected by the hinge region. Interactions are formed with the hinge backbone through two hydrogen bonds: the adenine N1 is an acceptor pairing with the backbone NH (H6), and the 6-amino group is a donor to the backbone CO (H4). Furthermore, the carbonyl group of the same H6 residue is oriented toward the pocket for additional hydrogen bond formation with an exogenous ligand. Effective interaction in the ATP pocket that complements the topological and the electrostatic characteristics of the kinase hinge is essential for potent inhibition of protein kinases [5]. In the current study we extract the hinge binding scaffolds based on hydrogen bond interactions with the annotated kinase hinge residues (Fig. 2).

Analysis of crystal structures

The number of structures represented in the CSDb varies substantially across different kinase targets. As illustrated in Fig. 3, p38a, CHK1, CDK2 and JAK2 are the most abundant kinases investigated by X-ray crystallography, each exceeding hundreds of crystal structures. They have historically been extensively targeted for different therapeutic indications, and their expression and purification systems are relatively well established through years of investigation by structural biologists. Correspondingly the

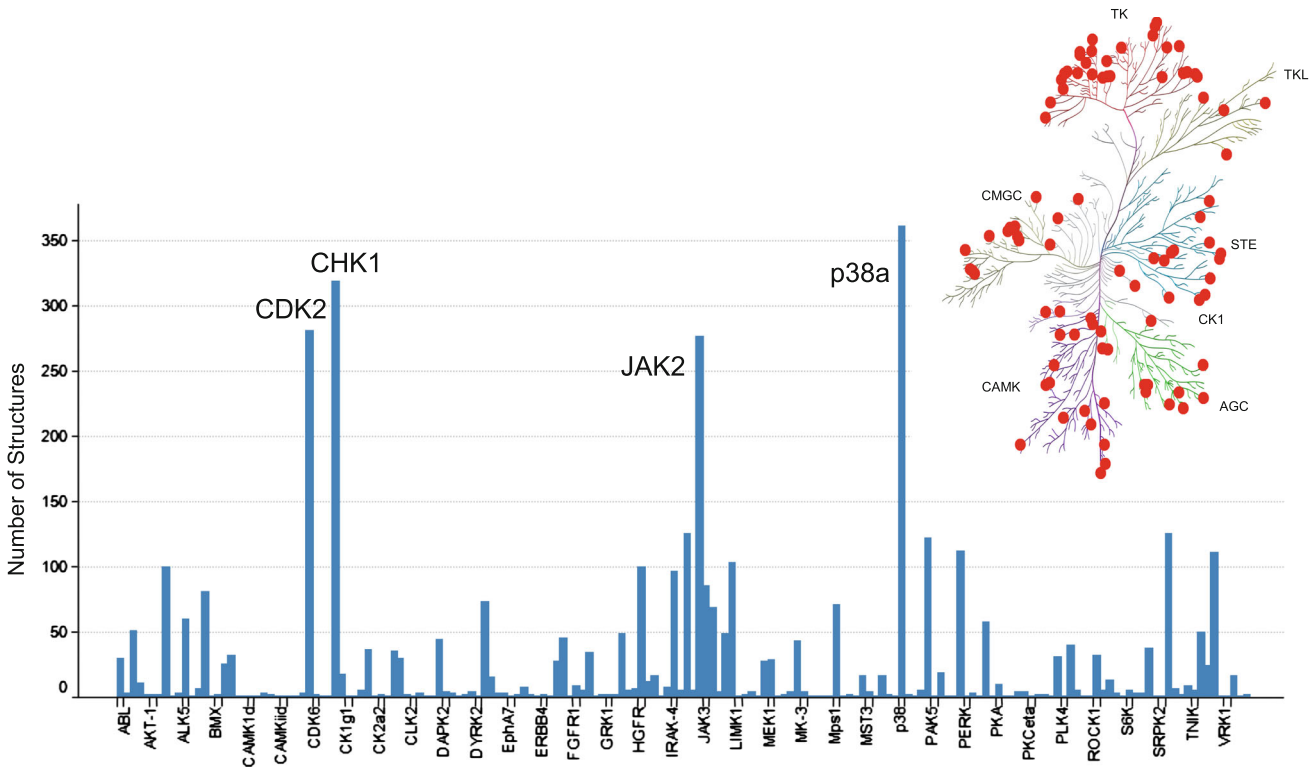


Fig. 3 Distribution of crystal structures by kinase targets

kinase groups they each belong to, namely MAPK, CAMKL, CDK and JAK, lead the population of the rest of the groups. As depicted in Fig. S2, the most prevailing kinase groups contain several hundred of crystal structures. In terms of kinase families, the tyrosine kinases (TK) are the most extensively studied, with more than 1,600 protein–ligand complexes produced to date. Following the TK family, CDK/MAPK/GSK3/CLK (CMGC) and calcium/calmodulin-dependent kinases (CAMK) are also heavily investigated, each consisting of 903 and 547 crystallographic experiments respectively (Fig. S3). Oncology has been the therapeutic area on which the great majority of kinase drug discovery programs have been focused, and where it has been most successful. Except for the recently approved Tofacitinib for the treatment of rheumatoid arthritis, the rest of the marketed kinase inhibitors are all chemotherapies for various malignant cancer types. The tyrosine family kinases play essential roles in cell proliferation and differentiation, and a great number of their inhibitors have been approved or in late stages of development for clinical use [5].

Properties of hinge scaffolds

A plethora of kinase inhibitors of varying degrees of structural diversity have been synthesized to date. However, the hinge recognition elements from which they are derived overlap to a great extent. Out of the nearly 4,000 crystal ligands we identified 595 unique hinge scaffolds, reflecting an average degeneration of about sevenfold. From a lead progression perspective it is of interest to understand how a fully developed kinase inhibitor compares to its corresponding hinge scaffold, the structural element that is minimally required for binding to kinases in general. In Fig. 4 the distributions of molecular weight, ClogP, three-dimensional polar surface area (PSA-3D) and fraction of sp3 hybridized carbons (Fsp3) are plotted for the hinge scaffolds (a) and compared with the full compounds (b).

It is noted that more than 90 % of the hinge scaffolds are small fragments of molecular weight between 100 and 200 Dalton. The very few scaffolds that tail at 400 Dalton and above are staurosporin analogs. The smallest hinge elements are various aromatic five-membered heterocycles including (is)oxazole, oxadiazole, pyrazole and triazole. The pyrazole and triazole could donate their NH proton and simultaneously accept a proton via their nitrogen lone pair, forming two hinge contacts. The other aromatic rings have the capacity to form one hinge contact by accepting via their heteroatom the NH of H6 residue of kinase hinge.

The scaffolds display a very desirable ClogP distribution, centering between -1 and 3. Another important parameter that is moderately related to lipophilicity is polar surface area (PSA), which is a conformation- dependent

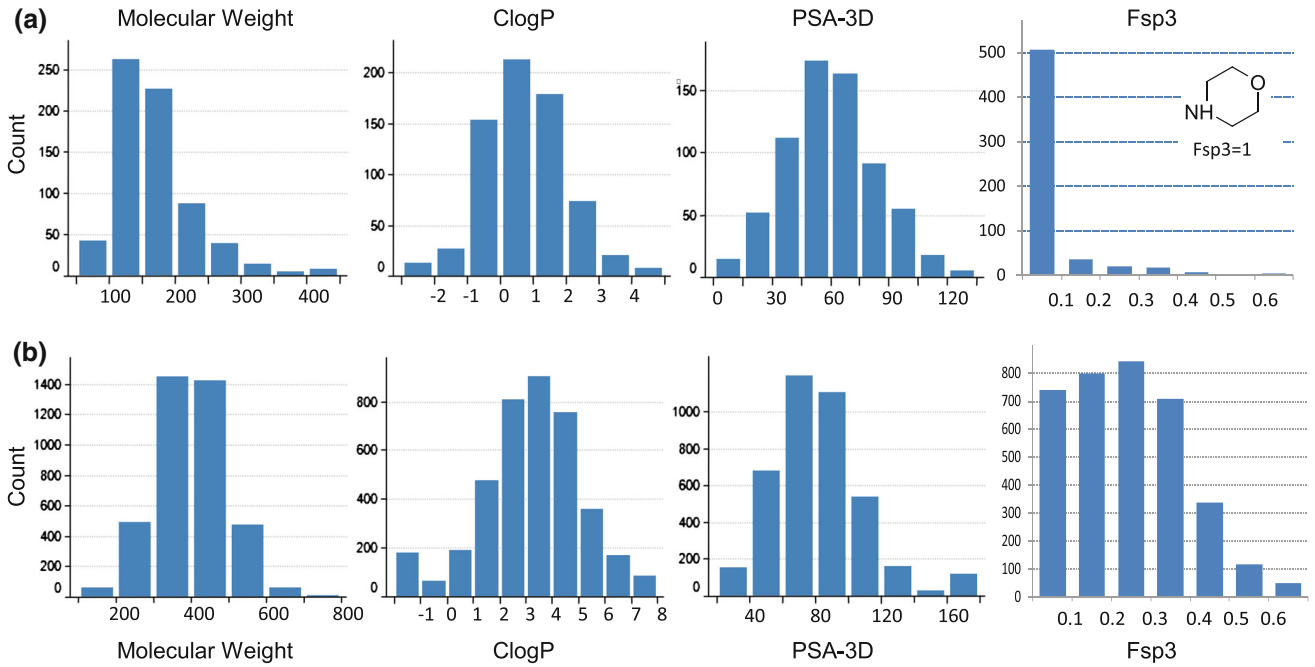


Fig. 4 Comparison of physicochemical properties of **a** hinge scaffolds and **b** full compounds

property. It has significant implication in passive diffusion across the lipid bilayer of cell membrane [22]. For good oral absorption a compound's PSA is typically less than 120. In lacking of 3D structures, many reports have adopted the topological polar surface area (tPSA) [23], a rule-based methodology, to quickly predict drug transport properties. In our study the 3D coordinates from the crystal structures allowed fast and accurate computation of PSA. The results show that PSA peaks between 30 and 90 for the scaffolds.

In comparison with the full compounds, several molecular properties exhibit monotonic increases upon the evolution of medicinal chemistry optimization toward delivering a drug candidate. Along with structural growth comes the lipophilic addition (ClogP), which exhibits a significant shift toward higher range of 2–5. PSA histograms are also right shifted, peaking between 40 and 120. As the hinge binding cores are elaborated into full compounds, localized hydrophobic interactions that complement the spatial subtleties of target proteins are exploited to augment potency. In parallel to growing compound to better potency and specificity, close attention should be paid to maintaining the rule-of-five compliance so that the pharmacokinetic properties of an oral drug candidate is not compromised [24].

Recently the fraction of sp3 hybridized carbons in a given molecule is found to have implication in improving clinical success [25]. Compounds that are highly saturated in nature are more likely to succeed in transitions from discovery through clinical trials to drugs. In this context, it is interesting that the vast majority (508 of 595) of the hinge scaffolds are completely devoid of saturation, accounting for 85 % of the total population (Fig. 4). The ensuing planarity of the hinge elements is likely imposed by the restriction that the binding clefts sandwiched between the kinase lobes are characteristically flat. However, there is one exception. Through this exercise the morpholine moiety was identified as the hinge recognition element, which is exclusively composed of sp3 centers (Fsp3 = 1). This scaffold is employed by the p38 kinase inhibitor BIRB-796, one of the first non-classical kinase inhibitors that bind to the inactive DFG-out state, indirectly competing with ATP [7]. Additional thirty or so scaffolds of Fsp3 > 0.3 are listed in Fig. S4. In the process of lead optimization, saturated elements and sometimes chiral centers are added beyond the hinge binding motif to attain three-dimensional characteristics. This has been a common as well as necessary practice to impart specificity as well as to improve potency. The current collection of inhibitors display a broad range of Fsp3, with a small fraction (6.5 %) approaching the drug-like space (e.g. Fsp3 $\geq$ 0.5) [25].

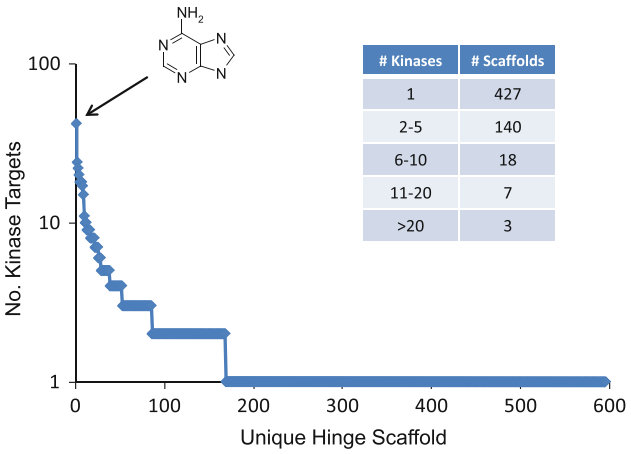
Most common and uncommon hinge scaffolds

The majority of small molecular kinase inhibitors that have been developed to date target the ATP binding site, with the kinase adopting a conformation almost identical to that used to bind ATP. These are termed the type I inhibitors. They typically bind to the active kinase conformations

through formation of hydrogen bonds to the hinge residues in combination with hydrophobic interactions around the region occupied by the adenine moiety of ATP. As exception to this rule, ATP competitive inhibitors that do not engage hinge hydrogen bond have also been reported. Another type of kinase inhibitors is termed Type II, which typically uses the ATP binding site as well. Additionally they exploit the allosteric sub-pockets made possible by DFG residues folding away from the conformation required for ATP phosphate transfer (DFG-out), and/or the C-helix swinging out to break the conserved Lys-Glu ion pair required for coordination of α and β phosphates of ATP (C-helix-out). Albeit sometimes less reliant on the hinge interactions compared to type I inhibitors, the majority of the type II compounds extends to the adenine pocket and preserves the hydrogen bonds with the kinase hinge, including the successful drugs Imatinib, Sorafenib and Lapatinib [10, 11]. Therefore to the large extent, the present discussions on hinge interacting scaffolds have implication on both type I and type II kinase inhibitors. The caveat is, as advantageous as affording the 3D binding information, the CSDb may be biased by scaffolds that were lead templates of projects where a large number of closely related compounds were synthesized and crystallized.

The number of kinases that each scaffold complexes with, as elucidated by crystallographic experiments, could suggest the degree of binding versatility of the distinct core elements that recognize the kinase hinge. Figure 5 depicts the number of different kinase targets as a function of hinge scaffolds. Of the nearly 600 unique scaffolds, a small number of them demonstrate the capacity to bind to multiple kinases, whereas the majority of others are more specific to just one or a small number of targets. Adenine is the most prevalent scaffold, being complexed in 42 distinct kinases. In the second place is staurosporine core, which is found in complex of 24 different kinases. These are not surprising, as one is in the endogenous substrate, and the other is from the well known nonselective inhibitor. It is noted that the number of targets by each scaffold tails off very rapidly. The overwhelming majority of the scaffolds (567 out of 595, or 95 %) were identified to interact with less than five kinases, and amongst them 427 scaffolds (72 %) are unique to a specific kinase.

It is intriguing that such large body of hinge scaffolds are relatively uncommon or less known to the kinase community. The observation that many compounds sharing the same scaffold are only in complex with one protein kinase raises an interesting question: are certain scaffolds more selective than the others? Or are they simply under-explored by medicinal chemistry and/or crystallography at an immense scale? Admittedly, the number of crystallographic experiments may be biased by a dominating lead



# Kinases	# Scaffolds
1	427
2-5	140
6-10	18
11-20	7
>20	3

Fig. 5 Distribution of number of kinase targets in complex with each hinge scaffolds. The table shows the number of scaffolds in each range of kinase targets. Adenine is found in complexes with 42 different kinases

series, and the absence of a crystal structure does not directly translate to the lack of inhibitory activity. On the other hand, it is typical that large numbers of compounds are derived from one singular scaffold. Diverse substitution patterns make them structurally dissimilar, assigning them into different chemical series. More often than not, they are developed for different protein kinase targets. Given these considerations, the singular association of a scaffold-protein pair may to certain extent manifest the relative specificity of the scaffolds.

Some examples of the unique hinge scaffolds are shown in Fig. 6, and are labeled by the kinase targets they are associated with in the crystal structure. Recognizing the possibility that these scaffolds are under-explored by medicinal chemistry giving rise to a bias in the number of crystal complexes, we wonder if there is any suggestion that these scaffolds are less promiscuous than the others. To address that we mined our research informatics database for kinome selectivity data for any compounds that are derived from these scaffolds. Gini coefficient is calculated from percent inhibition data of panel screening to estimate selectivity. Nonselective inhibitors are characterized by Gini values close to zero, e.g. 0.15 for Staurosporine [21]. Highly selective compounds exhibit Gini values close to 1. In Fig. 6, the CKI scaffold pyrido-dihydropyrrolone has reported kinase selectivity screen from two compounds, showing an average Gini coefficient of 0.72 at 1 μM and 0.74 at 10 μM, respectively. Another example, the tricyclic thiazinedioxide core for p38, exhibits Gini coefficients of 0.64 (1 μM) and 0.75 (10 μM) from one tested compound. Unfortunately we do not have panel selectivity data for most of the exemplified scaffolds. Despite the few data points, the Gini's of these two scaffolds, afforded by their underlying compounds, appear to indicate relatively high specificity.

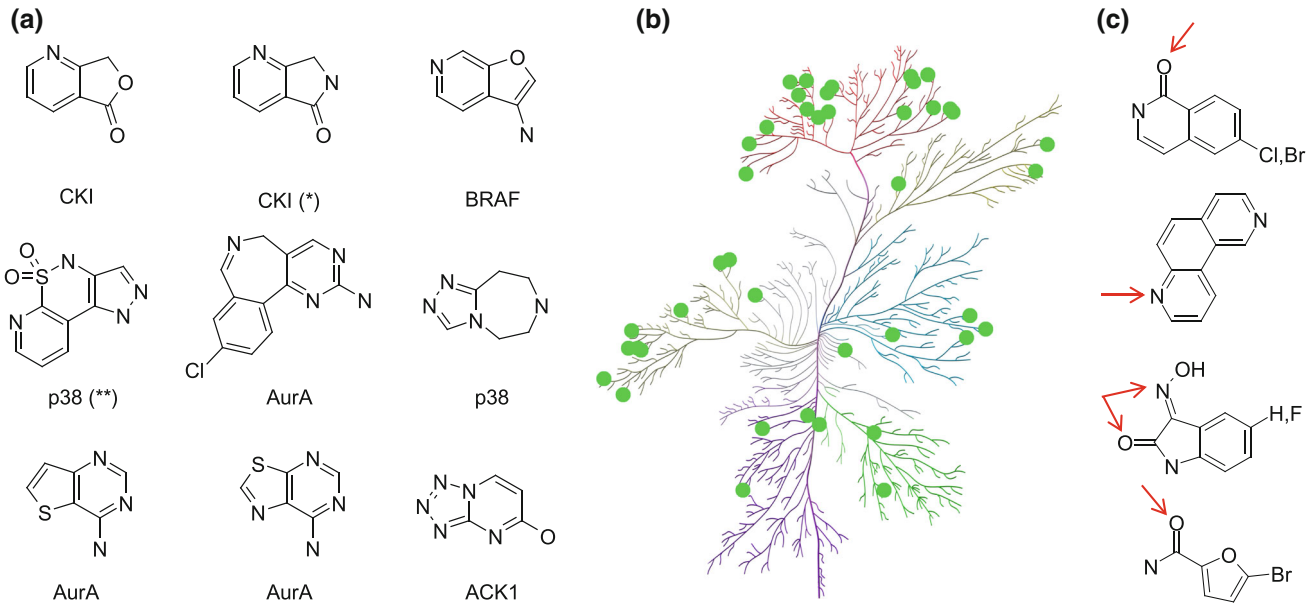


Fig. 6 **a** Examples of unique hinge scaffolds. Labeled are the kinase targets they are complexed in. The average Gini coefficients are 0.72 (1 μM) and 0.74 (10 μM) for the CKI scaffold (*), and 0.64 (1 μM) and 0.75 (10 μM) for the p38 scaffold (**). **b** Kinome distribution of protein targets for unique scaffolds. **c** Unique scaffolds in binary complexes with JNK3 kinase in the CSDb. Atoms that form hinge hydrogen bondsare highlighted by *arrows*

The kinome distribution of the protein targets for unique hinge scaffolds are plotted in Fig. 6b. The more or less even distribution on the dendrogram suggests that the unique hinge scaffolds can be designed for diverse kinase targets, and are not dominated by any specific kinase family. For many kinases, there are multiple unique scaffolds of specific structures. One example is given for JNK3 kinase in Fig. 6c, where six distinct hinge scaffolds were found to uniquely bind JNK3 kinase in the CSDb. They all partner with Met149 backbone NH of JNK3 hinge via either carbonyl oxygen or aromatic nitrogen as hydrogen bond acceptors, which are highlighted by pointing arrows. This set of structurally unrelated scaffolds confers the diverse opportunities for medicinal chemistry design to target the conserved hinge interactions. Whereas certain cores are most commonly employed in kinase programs, the kinase hinge can accommodate a lot more structural variation than that we are familiar with.

The ten most common scaffolds found in the CSDb are illustrated in Fig. 7. Notably the adenine (Core4) is the most non-discriminative core, crossing over to 42 different kinases in the CSDb. Most of these compounds are either ATP analogs or its derivatives. Except for the pseudo lead compound epitomized by the natural ligand, adenine is rarely employed in a chemical series as kinase inhibitors. Adenine has very few available valences for chemistry derivatization. Specifically only its C2 and C8 positions can be substituted. In its binding complex to kinase, either and/or both of these positions may be in close contact with protein, thus any substitution is prohibited. For such reasons adenine is limited for being used in kinase inhibitors. After adenine the second most common scaffold is exemplified by staurosporine-type compounds (Core2). A prototypical ATP-competitive kinase inhibitor that binds ubiquitously to many kinases with high affinity, staurosporine has been widely used as a research tool to study kinase functions and develop assay technologies. After its original discovery as an antifungal agent, several derivatives of bisindolyl maleimide have advanced to human clinical trials as protein kinase C (PKC) and cell-cycle checkpoint-control kinase (CHK1) inhibitors [26, 27]. In the current dataset about a dozen different staurosporine analogs were identified. Next in the ranking are aminopyrazole (Core88) and aminopyrimidine (Core3) cores, replicating the same donor–acceptor patterns of ATP in some of their binding orientations. Notably, a single hinge scaffold can adopt multiple binding orientations in the ATP pocket, especially when alternative hydrogen bond donating and accepting patterns are presented.

Pyrrolopyridine (Core124) and its close permutation, pyrrolopyrimidine (Core285), are also common hinge elements for kinase inhibitors. Pyrrolopyrimidine is the hinge binding scaffold of Tofacitinib, a successful pan-JAK inhibitor recently approved for rheumatoid arthritis in US, and currently under phase III development for treating a number of additional inflammatory conditions [28]. The 4-aminopyrrolopyrimidine (Core64) is a close analog of adenine, and has been employed in a number of advanced

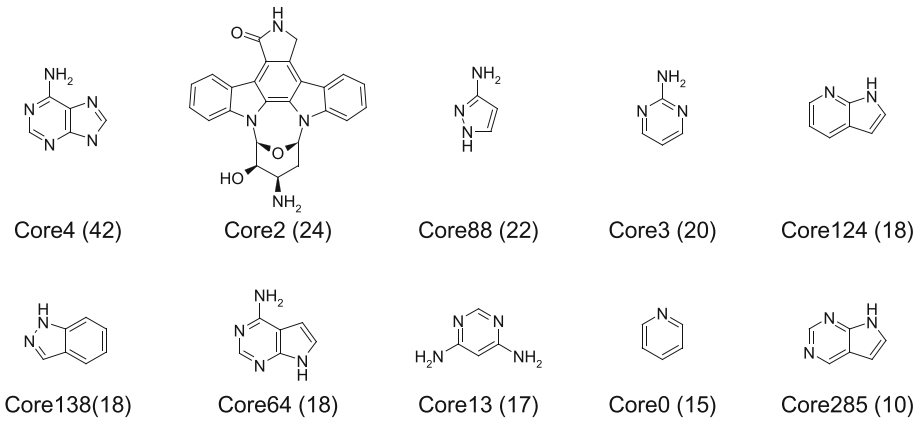


Fig. 7 Ten most common hinge scaffolds. In parenthesis are the number of different kinases found in the binary complexes. Pyridine has an average Gini coefficient of 0.58 (1 μM) and 0.47 (10 μM)

Bruton Tyrosine Kinase (BTK) inhibitors [29]. Being the smallest fragment among the most promiscuous scaffolds, the pyridine (Core0) appears to exhibit unlimited versatility of being decorated in numerous defferent ways for targeting various kinases. Despite its one-point hinge contact, it has been utilized in a number of kinase programs as an effective element from which potent inhibitors were derived, including p38, ABL, BRAF, FGFR and others [30–33].

Clustering of ring scaffolds

The scaffolds were further clustered by their common ring structures irrespective of substitution patterns as summarized in Table 1. The pyridine and pyrimidine are the most common as hinge binders, with each moiety complexed with more than thirty distinct protein kinases in X-ray diffraction experiments. They represent the broadest distribution on the kinome tree, covering kinase families other than casein kinase 1 (CK1) (Fig. 8). The quinoline and quinazolines are also abundantly employed in kinase inhibitors, but less populating than the pyridine/pyrimidine clusters. They delineat coverage of kinases in TK, TKL, CMGC and CAMK families. Quinoline additionally exemplifies kinases from protein kinase A, G and C families (AGC), homologs of yeast Sterile kinases (STE) and CK1 families. Purine and imidazo[1,2-a]pyridine are clusters of 5- and 6-membered fused ring system. Due to different arrangements of aromatic nitrogen atoms, their patterns of hydrogen bonds are entirely divergent. Interestingly the purine fragment is mainly seen with two kinase families, specifically CAMK and STE. There are a few examples for TKL, CMGC and AGC families as well. Interestingly purine is not seen with the largest TK family. It is also noted that none of the major clusters are complexed with the CK1 kinase family. We only found nine X-ray binary complexes for CK1 family of kinases in CSDb, and they are different isoforms of CK1 and the

Table 1 Six major clusters of hinge scaffolds

Cluster	# of scaffolds	# of crystal complexes	List of kinases
Pyridine	17	140	ABL1, ABL2, ACK1, ALK1, ALK2, ALK5, BRAF, CDK2, CHK1, CKID, DAPK1, EGFR, EPHA7, FAK, FGFR1, FMS, GSK-3B, HGFR, IRAK-4, ITK, JAK2, JNK3, MPS1, NEK2, P38, PAK4, PDK1, RON, SYK, TNIK, TYK2
Pyrimidine	19	276	ABL1, ACK1, ALK, AURA, BRAF, BTK, CDK2, CHK1, CKID, DAPK1, EGFR, EPHB4, ERK2, FAK, FGFR2, FMS, GSK-3B, IGFR, ITK, JAK2, JAK3, JNK1, JNK3, MLCKCA, MPS1, P38, PAK4, PDK1, PERK, SYK, TYK2
Quinoline	16	78	ACK1, ACTRIIB, ALK5, AURA, CAMKII, CDK2, CHK1, CKIT, CLK2, EGFR, GSK-3B, HASPIN, HGFR, JAK2, JNK3, MPS1, P38, PAK4, PAK6, PDK1, RON, SYK, TYK2
Quinazoline	13	57	ACK1, ALK5, AURA, CDK2, CHK1, EGFR, EGFR, EPHB4, IRAK-4, ITK, JNK1, MPS1, P38, SYK
Purine	15	109	CHK1, DAPK1, GSK-3B, MLK1, PAK1, PAK4, PDK1, PIM1, PIM2, TNIK, WEE1
Imidazo[1,2-a]pyridine	15	93	ACK1, ALK5, CDK2, CHK1, CKIT, FMS, IRAK-4, ITK, JAK3, JNK3, PDK1, SYK

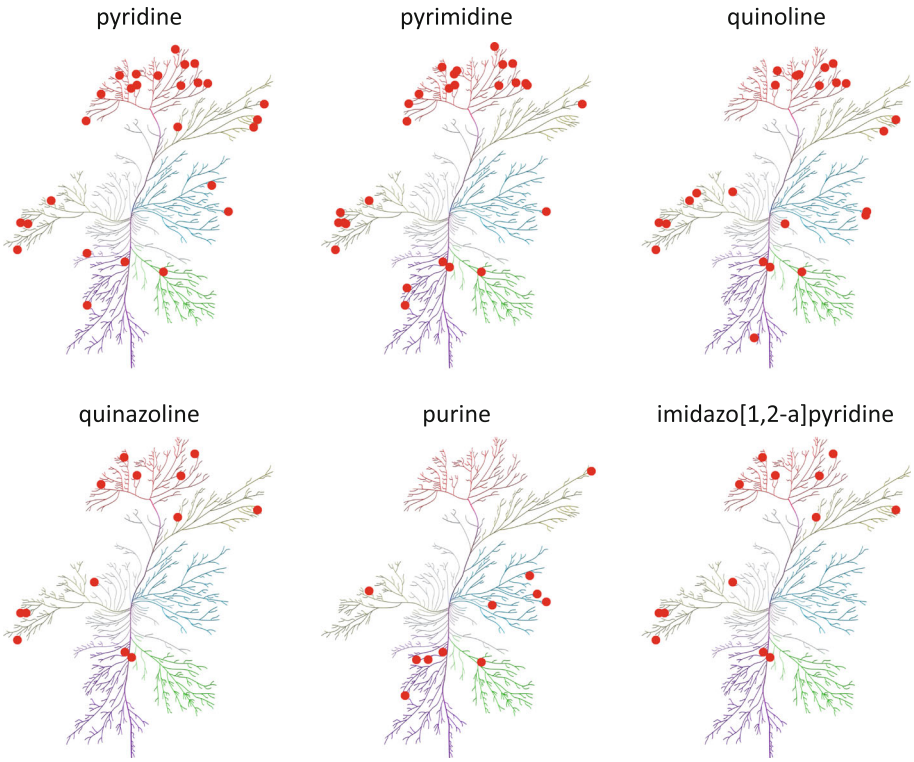


Fig. 8 Human kinome distribution of kinase inhibitors clustered by the hinge binding ring systems. Each *dot* represents crystal structures of ligand–protein complexes of which the ligands contain the corresponding ring system to form hydrogen bonds with the kinase hinge

vaccinia related kinase 1(VRK1). Very few of them use the ring scaffolds clustered here for hinge recognition.

Hinge hopping

In drug discovery research, the accessible training set data are often associated with challenges to intellectual property, potency, selectivity, pharmacokinetic properties, and/or toxicity of a specific lead series. Hence the need arises to find additional chemical space. One efficient way to discover novel lead compounds is to retain the desirable properties of the original series, but remove the liabilities associated with that series. For kinase inhibitors, specificity toward the primary target is usually accomplished by distinct decorations on the hinge scaffold. On the other hand, since the hinge hydrogen bonds are more or less conserved, the hinge binding elements are considered fairly exchangeable in theory across different kinase targets. The boundary condition is, the original selectivity fragments and the replacing hinge scaffolds need to afford reasonable match in 3D space, so that the binding of all critical elements are not perturbed in any significant way by the construction of the hybrid molecule. Apparently, the 3D properties such as hydrogen bonds and conformational rearrangements in the ATP binding site are of vital importance to the successful design and selection of hinge elements. The results of such analysis will be discussed in a separate report.

In hinge hopping, the scaffold is able to be joined with the selectivity fragment by a single bond formation, while maintaining its own hydrogen bond with the kinase hinge. Such adequacy needs to be screened for the pool of candidates. With the aid of 3D coordinates this can be done by a quick geometry check: (1) the connecting atoms are within bonding distance; (2) the newly formed bond matches bond angle of the open valence, this is bidirectional for both connecting atoms. Once the criteria are passed a new hinge scaffold is identified to swap out the existing one to form a novel chemical matter. We have applied this strategy to a couple of kinase programs to successfully generate novel lead series by replacing the hinge binding core of kinase inhibitors. Their three-dimensional properties such as hydrogen bonds and conformational rearrangements in the ATP binding site are analyzed. The complete compilation of the hinge scaffolds along with their 3D binding coordinates makes possible the hinge hopping experiment.

Conclusion

Kinase inhibitors can exert their modulation of target function by elaborate mechanisms. Type I inhibitors form H-bonds with the kinase hinge and occupy the adenine binding pocket. Type II inhibitors occupy the adenosine pocket, but also capture conformational changes that is

typically either DFG-out or C-helix out, or both. The scope of the current investigation covers both type I and type II kinase inhibitors. Nearly 600 different hinge scaffolds were extracted from 4,000 inhibitor-kinase complexes from CSDb composed of Pfizer internal and public structures. Our analysis reveals that about 80 % of the scaffolds are completely flat devoid of any sp3 centers. Three dimensional characteristics are introduced beyond the hinge binding region upon evolution to produce the full compounds, during which process more lipophilicity and polar surface areas are added. Although some of the most common scaffolds are found in complex with multiple kinase targets, more than two-thirds of the cores uniquely bind to a specific kinase as evidenced by crystal structures. These under-explored hinge scaffolds may present valuable opportunities for further investigation of developing novel inhibitors in the crowded kinase field. The compiled collection of hinge scaffolds along with their binding coordinates could serve as basis set for hinge hopping, a practice frequently employed to generate novel invention as well as to optimize existing leads in medicinal chemistry.

Acknowledgments We thank Jacquelyn Mcleod-Klug for extracting kinase selectivity data and Gini coefficient, Kieth Burdick from Accelrys for technical support. We also thank the reviewers for their insightful suggestions. This study was sponsored by Pfizer Inc.

References

1. Cohen P (2002) Nat Rev Drug Discov 1(4):309
2. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) Science 298(5600):1912
3. Cohen P, Alessi DR (2013) ACS Chem Biol 8(1):96
4. Dar AC, Shokat KM (2011) Annu Rev Biochem 80:769
5. Noble ME, Endicott JA, Johnson LN (2004) Science 303(5665):1800
6. Johnson LN, Noble ME, Owen DJ (1996) Cell 85(2):149
7. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J (2002) Nat Struct Biol 9(4):268
8. Iwata H, Imamura S, Hori A, Hixon MS, Kimura H, Miki H (2011) Biochemistry 50(5):738
9. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R, Cancer Genome P (2004) Cell 116(6):855
10. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, Clarkson B, Kuriyan J (2002) Cancer Res 62(15):4236
11. Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, Ellis B, Pennisi C, Horne E, Lackey K, Alligood KJ, Rusnak DW, Gilmer TM, Shewchuk L (2004) Cancer Res 64(18):6652
12. Jacobs MD, Caron PR, Hare BJ (2008) Proteins 70(4):1451
13. Brooijmans N, Chang YW, Mobilio D, Denny RA, Humblet C (2010) Protein Sci 19(4):763
14. Levinson NM, Kuchment O, Shen K, Young MA, Koldobskiy M, Karplus M, Cole PA, Kuriyan J (2006) PLoS Biol 4(5):e144
15. Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, Superti-Furga G, Kuriyan J (2006) Mol Cell 21(6):787
16. Dar AC, Lopez MS, Shokat KM (2008) Chem Biol 15(10):1015
17. Seeliger MA, Nagar B, Frank F, Cao X, Henderson MN, Kuriyan J (2007) Structure 15(3):299
18. Schubert C, Schalk-Hihi C, Struble GT, Ma HC, Petrounia IP, Brandt B, Deckman IC, Patch RJ, Player MR, Spurlino JC, Springer BA (2007) J Biol Chem 282(6):4094
19. Mao C, Zhou M, Uckun FM (2001) J Biol Chem 276(44):41435
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28(1):235
21. Graczyk PP (2007) J Med Chem 50(23):5773
22. Palm K, Stenberg P, Luthman K, Artursson P (1997) Pharm Res 14(5):568
23. Ertl P, Rohda B, Selzer P (2000) J Med Chem 43:3714
24. Lipinski CA (2000) J Pharmacol Toxicol Methods 44(1):235
25. Lovering F, Bikker J, Humblet C (2009) J Med Chem 52(21): 6752
26. Graves PR, Yu L, Schwarz JK, Gales J, Sausville EA, O'Connor PM, Piwnica-Worms H (2000) J Biol Chem 275(8):5600
27. Engel GL, Farid NA, Faul MM, Richardson LA, Winneroski LL (2000) Int J Pharm 198(2):239
28. Kremer JM, Bloom BJ, Breedveld FC, Coombs JH, Fletcher MP, Gruben D, Krishnaswami S, Burgos-Vargas R, Wilkinson B, Zerbini CA, Zwillich SH (2009) Arthritis Rheum 60(7):1895
29. Honigberg LA, Smith AM, Sirisawad M, Verner E, Loury D, Chang B, Li S, Pan Z, Thamm DH, Miller RA, Buggy JJ (2010) Proc Natl Acad Sci USA 107(29):13075
30. Simard JR, Getlik M, Grutter C, Pawar V, Wulfert S, Rabiller M, Rauh D (2009) J Am Chem Soc 131(37):13286
31. Perry JJ, Harris RM, Moiani D, Olson AJ, Tainer JA (2009) J Mol Biol 391(1):1
32. Namboodiri HV, Bukhtiyarova M, Ramcharan J, Karpusas M, Lee Y, Springman EB (2010) Biochemistry 49(17):3611
33. Hansen JD, Grina J, Newhouse B, Welch M, Topalov G, Littman N, Callejo M, Gloor S, Martinson M, Laird E, Brandhuber BJ, Vigers G, Morales T, Woessner R, Randolph N, Lyssikatos J, Olivero A (2008) Bioorg Med Chem Lett 18(16):4692