UC San Diego
UC San Diego Previously Published Works

Title
Druggable exosites of the human kino-pocketome

Permalink

Journal

ISSN

Authors
Nicola, George
Kufareva, Irina
Ilatovskiy, Andrey V
et al.

Publication Date

DOI

Druggable Exosites of the Human Kino-Pocketome

George Nicola, **Irina Kufareva**, **Andrey V. Ilatovskiy**, **Ruben Abagyan**

Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego

Abstract

Small molecules binding at any of the multiple regulatory sites on the molecular surface of a protein kinase may stabilize or disrupt the corresponding interaction, leading to consequent modulation of the kinase cellular activity. As such, each of these sites represents a potential drug target. Even targeting sites outside the immediate ATP site, the so-called *exosites*, may cause desirable biological effects through an allosteric mechanism. Targeting exosites can alleviate adverse effects and toxicity that is common when ATP-site compounds bind promiscuously to many other types of kinases. In this study we have identified, catalogued, and annotated all potentially druggable exosites on the protein kinase domains within the existing structural human kinome. We then priority-ranked these exosites by those most amenable to drug design. In order to identify pockets that are either consistent across the kinome, or unique and specific to a particular structure, we have also implemented a normalized representation of all pockets, and displayed these graphically. Finally, we have built a database and designed a web-based interface for users interested in accessing the 3-dimensional representations of these pockets. We envision this information will assist drug discovery efforts searching for untargeted binding pockets in the human kinome.

Introduction / Scientific Background

Protein Kinases as Drug Targets

Protein kinases are enzymes catalyzing transfer of a phosphate from ATP to either an alcohol group of Ser/Thr or a phenol group of a Tyr residue on their substrate proteins. They constitute about 2%, but modify up to 30% of the mammalian proteomes, which makes them an attractive class of drug targets. Most cellular signal transduction events are mediated by protein kinases [3]. Consequently, kinase defects, such as mutations, fusion, over-expression etc., are implicated in a wide variety of human pathologies including cancer, cardio-vascular diseases, inflammation, and diabetes (http://www.cellsignal.com/reference/kinase_disease.html)[4]. These factors make protein kinases an attractive class of drug targets [5]. A number of small molecule inhibitors of protein kinase activity have already become successful therapeutics (see Roskoski 2016 and www.brimr.org/PKI/PKIs.htm), and others are at various stages of clinical research [6,7].

Kinase Exosites and Exosite Compounds

Of the 518 proteins in the human genome with annotated protein kinase activity, 507 share a common structurally conserved catalytic domain called the protein kinase (PK) domain [3]. For most proteins, this domain is flanked by extended C- and/or N-termini or other functional globular domains all of which play a role in kinase regulation. They do so by

stabilizing a particular conformation of the catalytic domain, or by mediating multimerization, membrane translocation or substrate binding. Some kinases are additionally regulated by binding to other proteins (e.g. regulatory subunit of PKA and CK2). A molecular surface of a typical kinase has multiple sites of regulatory interactions, some within and others outside the protein kinase domain. Such interactions may inhibit or activate the kinase, or even change its substrate preferences (e.g. CK2 [8–10]). Small molecules binding at any of these regulatory sites may stabilize or disrupt the corresponding interaction leading to consequent modulation of the kinase cellular activity. As such, each of these sites represents a potential drug target. Moreover, even targeting sites without a recorded functional role may cause desirable biological effects through an allosteric mechanism.

The vast majority of existing kinase inhibitors target the ATP binding site located in a cleft between two lobes of the protein kinase domain. These compounds ultimately need very high affinity in order to compete with intracellular ATP. Moreover, as all kinases bind ATP, the atomic composition of binding site is highly conserved among them. As a result, ATP-site compounds are often plagued by lack of specificity: they bind promiscuously to many types of kinases, leading to adverse effects and toxicity.

This problem can be somewhat alleviated by targeting regions other than the ATP binding site, the so-called exosites. Such regions are less conserved or sometimes even unique for a kinase of interest. The first successful kinase-targeting therapeutic, imatinib, employs an additional site on the PK domain of its targets, the so-called hydrophobic (selectivity) pocket. Unlike typical ATP-site compounds, imatinib acts by stabilizing a particular inactive ATP-incompatible conformation of the target kinase domain by binding outside the ATP-site [11]. Therefore, along with other compounds of this class, called type-II inhibitors [12], imatinib exemplifies the concept of exosite inhibition of kinase activity. Further work led to discovery of type-III compounds [13,14], which bind concurrently with ATP and act by disrupting a conserved Lys-Glu salt bridge and stabilizing yet a new type of inactive conformation of the kinase domain.

Type-II, type-III, and other exosite-targeting compounds have attracted increasing interest in both academia and the pharmaceutical industry in recent years [15,16]. In addition to the advantages of being non-ATP-competitive and more specific, some of these compounds possess the ability to modulate subtly the kinase activity by changing its substrate preferences or subcellular localization. Examples of such compounds include the myristoyl-site inhibitors of BCR-ABL-dependent cell proliferation [17], substrate-selective inhibitors of p38α MAP kinase [18], and the substrate competitive inhibitors of ERK1/ERK2 [19], JNK [20], and IKK [21]. Other exosite compounds modulate the cellular activity of their target kinases by binding outside the PK domain, e.g. by disrupting the recruitment of the kinase to the cellular membrane by binding to its PH domain [22].

While some kinase exosites such as the ones mentioned above have been thoroughly studied, characterized, and targeted with small-molecule therapeutics, others are still elusive. This is in part due to the lack of structural information in the form of a complex of the kinase with its endogenous exosite partners. Another reason has a purely structural basis; there is usually

only one active site on the surface of the protein and this typically forms a deep cavity [23]. Exosites are often numerous and, as sites of protein-protein interactions, tend to be flat and conformationally variable. This presents a challenge to drug discovery efforts which seek to identify unique compounds that are specific for a particular site.

Structure-based methods, including Virtual Ligand Screening (VLS) have the potential to reduce dramatically the number of candidates for experimental validation while preserving their chemical diversity. VLS techniques were found to be successful in a wide variety of applications (e.g. [24–26]), especially combined with improved scoring functions [27,28]. However, the recipe for successful application of this approach starts from a 3D structure (or an ensemble of 3D structures) of a protein target with a specified, well-defined ligand binding pocket.

To overcome this obstacle, a large-scale screening for druggable exosites in the structural kinome is needed. We have previously developed a computational structure-based method for the identification of druggable pockets on the surface of any protein with a 3D structure. This method, called ICM PocketFinder, has demonstrated very high specificity and sensitivity for ligand binding pockets in a comprehensive screen of existing protein structures [29]. It was also shown that the size/shape characteristics of the ICM PocketFinder envelopes are predictive of the size and shape of potential ligands. In particular, insufficiently large PocketFinder envelopes characterize challenging, poorly druggable sites [30].

There have been considerable efforts by other groups to organize the large amount of data being published on protein kinases. The KinBase resource hosted at the Salk Institute, for example, has a database of kinases from a range of sequenced genomes [31]. Other groups, such as Susan Taylor's at UCSD, have assembled protein kinases into an interactive tree display, which is useful for navigating the relationships between the known functions of kinases [32]. Others have curated biological and chemical information from scientific literature and patents. The Kinase Knowledgebase (KKB) from Eidogen-Sertanty is a database of kinase structure-activity and chemical synthesis data [33]. A recent study by another group has described an approach to identify subpockets between two sets of protein kinase structures [34]. This method helps distinguish specificity between closely related kinases solely based on the 3-dimensional structures of their binding sites. In another study, researchers have used molecular interaction fingerprints based on known binding site residues to create a database of kinase-ligand information, called KLIFS [35]. This resource helps offer insights into the structural determinants of ligand interaction and selectivity.

The goal of the present work is to identify, catalog and annotate the potential druggable exosites on the protein kinase domains within the existing structural kinome. It seeks to aggregate all potential exosites across the human kinome, including those that have not been previously characterized. We envision this information will have a twofold purpose: to help biologists researching the modulation of a particular kinase in identifying allosteric binding sites, and to assist drug discovery efforts searching for untargeted binding pockets in the human kinome. To understand better a particular kinase, a structural perspective of the

functional binding sites is required. A comprehensive screening of all potential binding sites within the kinome will assist in these efforts.

Methods

Assembling a Database of Kinase Domain Sequences and Structures

Amino-acid sequences of known human protein kinases were downloaded from Swiss-Prot [36,37]. Of the 507 sequences, only 320 have at least one structure of the human kinase domain (KD) or its closest species homolog (95% sequence identity) available in the PDB (as of this writing). We use non-human kinase structures as surrogates when no human kinase crystal structure is available. Most (197/320) KDs have more than one, and 48 KDs have more than 20 associated PDB entries (Figure 1). In many cases, an asymmetric unit of a PDB structure contains several KDs forming a crystallographic or a biological multimer; these domains were separated and treated as individual entries. The resulting database contained 4169 KD structures of the 320 unique Swiss-Prot sequences (Table 1).

All kinase domains were superimposed into a single template using the key residues in the ATP-binding cleft between the two lobes of the kinase domain. The residue mapping for the superimposition was identified from careful sequence-structure alignment. The superimposition was performed using a previously developed algorithm [30,38]. Briefly, all protein structures in the kinase ensembles were superimposed using only the backbone atoms in the immediate vicinity of the ligand binding pocket. The algorithm uses an iterative procedure that starts from two equivalent atom arrays, establishes atomic equivalences and weights for each atom pair, performs a superimposition and RMSD evaluation, calculates deviations for each atom pair, and calculates new weights.

The kinase domain sequences used in our study are captured in the name of the Swiss-Prot identifier. For example, the kinase domain of human STK10 is delimited by residues 36–294, and is represented in our nomenclature as STK10_HUMAN_36_294.

Structure-based Identification and Characterization of Kinase Domain Pockets

The automated ICM "PocketFinder" algorithm was run on all 4169 KD structures to produce a list of drug-like pockets for each. The algorithm produces potential grid maps according to a method described previously [29]. The method has been used successfully to identify druggable binding sites in proteins [39,40]. Briefly, the algorithm first creates a grid potential map of the van der Waals force field using a probe atom within surrounding receptor atoms. The potential maps are then smoothened and contoured to produce ligand envelopes that graphically represent putative small-molecule binding sites. The final step sorts the created envelopes by their volumes and filters out those smaller than 100 Å^3. The algorithm does not take into account the existing functional, binding, or inhibitor data, thus producing unbiased predictions. A database of these predicted envelopes was made whereby each row corresponds to a druggable pocket (Supplemental Table 1).

Annotation of Potential False Positives

The inherent plasticity of proteins in general and kinases in particular frequently leads to selected regions not being visible in the crystallographic density and thus being partially absent from the final model. Thus, in a number of cases, the PocketFinder algorithm generated a pocket or cavity which resulted from disordered or missing fragments in the structure. Sometimes the fragments are present but are annotated as having high degree of thermal mobility via the assignment of the so-called B factors, or partial (or even zero) occupancy. During the scan for pockets in this study, the gaps in the protein may falsely indicate a larger cavity. To alert the user to regions in the protein structure that do not have full occupancy in the PDB coordinates, residues lining a pocket with occupancy assignments lower than 1.0 are listed in a separate "LowOcc" column (Supplemental Table 1). The "Gaps" column lists breaks in the PDB chain. The "Suspect" column lists those residues from the "LowOcc" or "Gaps" column that intersect the "Lining Residues" column and thus indicates the flexible regions that are lining the particular pocket.

Another source of false positives in structure-based identification of pocket cavities is non-native, crystallographic contacts. Pockets adjacent to crystal neighbors may be only artificially supported by a crystallographic partner and thus of lower significance *in vivo*. Whether due to flexibility of protein or missing due to X-ray structural artifact, the B-factor of the residues lining these pockets should thus be carefully checked and kept in mind.

Finally, in order to identify and exclude ATP binding pockets, a "PocketType" column was added. The distance between the center of the pocket and the geometric mean of all ATP binding sites is measured; if this distance is less than 8.0 Å, and thus the pocket is within close proximity to the known ATP site, "ATP" is indicated in the column, to represent that it is not an exosite.

Pre-existing Ligands within Pockets

Many kinase domains analyzed in this study are co-crystallized with small molecules and polypeptides. In developing drug targets, it is often instructive to analyze existing binding properties of a particular site. As such, all ligands within the structures were retained in order to compare any *holo* and *apo* versions of the same structure. In this context, a *ligand* is any small molecule appearing within the graphical pocket density, excluding common buffer ingredients. It is important to note, however, that the PocketFinder routine did not use these ligands in its search for pockets. The existence of a ligand in a pocket was identified with an 'X' in the "LigExist" column of the database.

In addition to indicating whether a ligand exists in each pocket, a routine was performed to determine if a hetero group or a fragment of a polypeptide chain is present in the given pocket in at least one structure of its kinase domain. The method for this pocket classification, which we call "X-druggable", first checks whether two pockets in different PDBs of the same protein overlap (with a similarity threshold of 0.5Å rmsd), and if so, whether there is a drug-like ligand in one of them. If these criteria are met, an "X" is added to column "LigExists" for both pockets (Supplemental Table 1). This is useful when

screening for pockets that have already been targeted by drug-like, molecular binding studies, as described in the Results Section below.

Ranking Predicted Kinase Domain Pockets with Importance Score

An importance score was determined for each pocket in order to quantify its druggability in terms of size and surface area, and to take into account the resolution of the crystal structure. This so-called "ImpScore" column determines values for each pocket using the following formula:

$$S = (V - V_0)^2 + (A - A_0)^2 + R$$

Here S denotes the score, R is the crystallographic resolution, V and A are the envelope volume and surface area, respectively; V_0 (450 Å^3) and A_0 (450 Å^2) represent the mean for all known pockets targeted by small molecule drugs [29]. A similar function developed for bacterial pocketomes has been published previously [41]. The resulting 32274 pocket entries were ranked according to this ImpScore, with lower scores being better.

Pocket Consistency Analysis to Visualize Most Common Exosites

In the superimposed kinase domain structural ensembles, a normalized representation of all pockets across the structural human kinome was created in two steps. In the first step, the pocket grid potential maps of individual structures were averaged across the structural ensembles, resulting in a single aggregated pocket map per kinase domain. In the second step, the resulting 256 kinase domain pocket maps were averaged to identify pockets that are persistently present, in identical locations, across all or most kinase domains.

Common exosites proximal to known conserved functional fragments were labeled according to their annotation in the PKC kinase [42]: activation segment (loop), ATP (substrate)-binding pocket, Glycine flap, Helix C, Catalytic (DFG) loop.

Designing Online Database Entries for Kinase Ensembles with Exosites

Database entries and corresponding web links were created for each individual kinase domain represented by an ensemble of corresponding PDBs superimposed onto a common reference frame (Figure 6). These entries are formatted as Molsoft ICM Project Files [43] and available for download and facile viewing using free Molsoft ICM-Browser software [44]. For each kinase ensemble entry, interactive checkboxes were coded for every protein chain and ligand structure, to display or undisplay these objects from the view. The project files make use of ICM technology which allows interactive viewing of a molecular image in 3D. Implementing this technology provides the ability to view the kinase domain structures and their exosites by panning, zooming, and rotating around the default view.

Computing

All calculations were performed on either an Intel Core™ 2 Quad or AMD Phenom™ II processor with 8GB RAM. Computation time was generally several hours. All algorithms

were written and executed in ICM. All figures were also made with ICM. Tables were compiled with Microsoft Excel.

Results & Discussion

Analysis of the Predicted Pockets across the Structural Kinome

A scan of all drug-like binding regions across the human kinome revealed a list of 32274 total pockets of varying degrees of size and shape. The complete database of these pockets can be found in the supplemental material (Supplemental Table 1). A visual representation of these pockets has also been developed and categorized by kinase family (see below).

Of the 518 human kinases known to exist, 320 (61.8%) are represented by crystal structures in the PDB (unique Swiss-Prot IDs). However, the number of domain structures for each kinase varies widely (Figure 1). There are typically crystal structures of one or more variants of each kinase. Some kinases (e.g. MP2K2) only have one deposited PDB structure, whereas others (e.g. CDK2) may have at least 486 PDB structures associated with them. Thus, the human structural kinome is represented by 4169 unique PDB identifiers. Enumerating each of these structures by their relevant chains reveals 5990 unique kinase domains. Thus, the resulting 32274 pockets identified in this study represent an average of 5 pockets for every kinase domain.

The value of this study is in providing a possibility for a structural biologist to quickly glance at the kinome and search for interesting pockets that are amenable for drug design. This is an area of increasing importance as existing target sites have been exhausted. Our results highlight exosites that are worthy of exploration for small molecule binding experiments, including regions that have already been co-crystallized with a molecule. However, drug discovery professionals may wish to identify regions on a kinase that have not been previously targeted. Thus, the absence of ligands may potentially serve as a criterion for selecting a new pocket.

It is important to note the fact that a small molecule or a peptide being crystallized, intentionally or not, in a given pocket, is circumstantial evidence of its druggability. On the other hand, the fact that no ligands are identified in a crystal structure cavity does not necessarily signal poor druggability.

When eliminating pockets lined with zero or low occupancy ('Suspect') residues, 23121 pockets remain, representing 256 of the original 320 kinases (Table 1). Of these pockets, 5993 are ATP binding sites. To help remedy these examples of false positives and negatives, we use the concept of "X-druggable", as described in Methods. When eliminating ATP-binding and pockets that contain one or more ligand in either their pocket or their X-druggable pocket, 10281 remain. (There are 586 non-ATP exosites that contain one or more ligand in their pocket). Some of these untargeted exosites may function as allosteric or alternative regulation sites. The pocket densities that are less spherical and instead, larger and flatter, may be indicative of protein-protein interaction sites.

Predictive value of the Importance Score

The ImpScore takes into account characteristics of the structure and pocket in an attempt to priority-rank the results. It is widely discussed in the literature that structure resolution, pocket size and shape are important but not the only parameters to describe druggability [45]. We investigated the use of B-factors and R-factors in our ImpScore calculation, but these are either inconsistently generated or not included in the PDB upon structure deposition, and thus could not be used reliably in our score. Additionally, B-factors are representative of the behavior of the structure fragments in the crystal but not necessarily in solution (e.g. loops may be artificially stabilized by crystal packing resulting in low B-factors, whereas in solution they may be highly dynamic). It is worth noting that our ImpScore is not meant to be a comprehensive characterization of the druggability of a protein kinase. But rather, it serves as a guide to rank order the list of exosites we identified in this study, given the information that is available for each protein kinase. Sorting the ImpScore column by the lowest (best) values gives the user the ability to screen for exosites that have the most amenable drug-like properties. Figure 2 reveals the distribution of ImpScore across all exosites identified. The ImpScore ranges from 12 to 2378850, with an average of 143553, and a median of 162397. Of the 4168 kinase domain structures represented in the entire pocket list, the ATP-binding pocket ranks first in 67%. The top ranking pocket in the ensemble is an ATP-binding pocket for 73% of the 256 kinase domains.

Using the concept of "X-druggability" as evidence of real druggability (although as we realize, rather circumstantial and incomplete), we sought to determine whether ImpScore correlates with it in any way. We first sorted pockets by increasing ImpScore, and then constructed a ROC curve by plotting the fraction of true positives vs the fraction of false positives under each ImpScore cutoff. As Figure 3 shows, pockets with lower ImpScores tend to be more X-druggable; for example, 85.4% of the top 1% pockets contain a ligand either in the structure in question or in another structure of the same kinase, and this also holds for 84.5% of the top 2% of the pockets and 82.6% of the top 5% of the pockets. In other words, the ImpScore is a viable method allowing separation of pockets with higher druggability.

The top-ranked exosite in this study is a pocket on chain A of PDB 5U8L (Swiss-Prot ID EGFR), which has an ImpScore of 12.6 (Table 2). This is an EGF Receptor of the Tyr protein kinase family. This exosite is unoccupied by any small molecule in the crystal structure, but appears to be well-poised for such binding as indicated by the lack of "Suspect" residues, the presence of X-druggability in the "LigExists" column, and a visual analysis of the exosite (http://abagyan.ucsd.edu/Exositome/eout/EGFR_HUMAN_712_979.icb).

The highest ranked ligand-bound exosite in this study is one of the pockets of PDB 3IW5 (Swiss-Prot MK14) which has an ImpScore of 32.5 (Table 2). This is a human p38 MAP kinase of the CMGC Ser/Thr protein kinase family. Interestingly, this pocket is occupied by two molecules of β-Octylglucoside (Figure 4), which was added to the crystallographic solution presumably to help solubilize the membrane portion of the protein [46].

Another example of an exosite with a favorable ImpScore is within the myristate binding site of BCR-ABL1 (PDB 2G2F chain A). This structure is an ABL tyrosine kinase domain in a Src-like inactive conformation [1]. The myristate pocket has been shown to be an important site that binds functional antagonists as well as agonists of ABL kinase [47]. Our results indicate this pocket is partially buried, elongated, and has the characteristics to bind a drug-like molecule (Figure 4). It is also revealed from Supplemental Table 1 and the Molsoft ICM view of this kinase (http://abagyan.ucsd.edu/Exositome/eout/ABL1_HUMAN_242_493.icb) that this pocket is occupied with a ligand in several other PDB structures of the same protein kinase. A literature search for one of these, 3K5V, reveals a study seeking pharmacological modalities to overcome resistance in the treatment of CML [2]. The authors of that study analyzed a known allosteric inhibitor of BCR-ABL, GNF-2, and demonstrated that combining it with ATP-competitive inhibitors can overcome resistance. A view of the GNF-2 allosteric inhibitor superimposed into the predicted exosite of 2G2F reveals potential for additional ligand-building around this scaffold (Figure 4), and underscores the validity of our priority-ranked table of results.

Common Exosites as a View of Pocket Consistency

Pockets and cavities in the kinase domains that are structurally conserved across the entire family may represent, from an evolutionary perspective, important sites of interaction with substrates, co-factors, or regulatory proteins. For example, the αF-helix of the kinase domain is one of three non-contiguous hydrophobic motifs that provide the basic architecture of the catalytic core and controls binding of ATP [48]. The exosite nearest this αF-helix motif, which we dubbed exosite αF, is consistently present across all structures we analyzed. Thus, we developed the concept of a common exosite as a cumulative pocket density of the exosites from all structurally characterized human kinases (Figure 5). The shape of the common exosite is determined by the signals from individual exosites in a particular region. The signal (and thus shape density) is strongest at the intersection of the highest number of exosites. This type of aggregation dramatically helped reduce noise. The common exosite graphic is a unique illustration of all pockets in the entire human kinome, visualized in one 3D image. When compared to an individual exosite within a protein of interest, this visualization serves as a reference for comparison of size and location in all kinases. For example, it may serve as a starting point when designing an inhibitor within an exosite that is specific for one kinase, but not observed in the all-kinase common exosites.

Retrieval of Kinase Exosites through a Web-based Interface

A web site was designed to assist the user in viewing the predicted kinase exosites. This service portal is available online at http://exosite.ucsd.edu. To browse for a particular kinase, the user needs only to click the 'Browse All' link. This displays all 256 kinase domain entries in the database, along with the PDB structures associated with them. Alternatively, the user may enter a query in the search box and use the drop down selector to search by Protein Name, Protein Family, or PDB code. In all cases, the output of the search represents a list of Swiss-Prot IDs for the matching kinase domains, with an indication of the entries for which 3D information is available.

Clicking on a kinase domain from the Search Results leads to a downloadable Molsoft ICM Project File, that can be opened with the free ICM-Browser (https://www.molsoft.com/icm_browser.html), following instructions available here: https://www.molsoft.com/browser/open-file.html. When opened within the software, the user will see a workspace split in two frames (Figure 6). The left frame is a pocket list of all PDB entries for the KD of interest, along with all predicted exosites in each structure. Each row corresponds to an individual exosite, and provides the following information: volume, area, nonsphericity, ImpScore, existence of known ligands, and a list of protein residues lining the pocket of the exosite. Each PDB chain and exosite has checkboxes next to them, to allow the graphical display and undisplay of each. The right frame is a 3D structural view of these entries, showing the corresponding kinase structures. To eliminate redundancy and to facilitate flexibility analysis, all kinases were superimposed in identical orientations and organized into structural ensembles. The initial view shows only the first PDB in the list (arbitrarily) and all predicted exosites in this chain. The user may select any combination of protein structures and exosites, in order to visually corroborate the ImpScore with the perceived druggability of an exosite. This can be especially useful when selecting to view exosites with no bound ligands in their respective PDB structure, and then selecting other PDB structures containing ligands.

For example, searching for serine/threonine-protein kinase 10 (STK10) produces a result that when clicked on, saves an ICM file to local disk. Opening with Molsoft ICM-Browser reveals the structure of ten PDBs (2J7T, 4AOT, 4BC6, 4EQU, 4USD, 4USE, 5AJQ, 5OWQ, 5OWR, 6EIM), several with multiple structures (chains). The normalized pockets for each are identified with blue densities. Ligands associated with these structures - those identified within these pockets - are displayed in their original orientation. Each PDB chain can be individually displayed or undisplayed, as can each exosite within these structures. The most common exosites associated with all kinases are shown as green densities, and as such can be compared to those that are unique to STK10.

Conclusion

Currently, protein structures exist for 320 of the 518 human protein kinases. In this study, we have systematically analyzed every human kinase structure in the PDB for pockets along their surface. These sites may have a known functional role, or may yet be unknown. A binding site - distinct from the ATP site - with an unknown functional role is called an exosite. As such, binding to an exosite may have no functional relevance. In some cases, however, ligand binding at an exosite may lead to modulation of the activity of the target protein. This may occur through an allosteric mechanism or by other means. One goal of this study is to identify exosites on the surface of proteins that may affect the activity of an interesting protein. By doing this, we facilitate computational structure-based discovery of novel exosite-targeting kinase modulators. This will be especially useful to drug discovery groups seeking unique, non-conserved exosites.

Developing a better understanding of kinase flexibility and ligand binding sites distinct from the ATP site is critical to both basic biology and molecular medicine. Therefore, finding and cataloguing all allosteric pockets for chemical probes and molecular therapeutics in all

conformational states of nearly 6000 available kinase structures can give us new recipes for specific modulation of kinase activity. It can also serve as a foundation for the development of new methods of incorporating protein flexibility in exosite identification and targeting.

Finally, we have created the first online public catalog of kinase exosites. Our research will help to expand dramatically the targetable kinome, to understand biological function and genomic variations of kinases, and to design next generation molecular therapeutics using rational structure-based approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Levinson NM, Kuchment O, Shen K, Young MA, Koldobskiy M, Karplus M, Cole PA, Kuriyan J (2006) PLoS Biol 4(5)

2. Zhang J, Adrián FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE, Sim T, Powers J, Dierks C, Sun F, Guo G-R, Ding Q, Okram B, Choi Y, Wojciechowski A, Deng X, Liu G, Fendrich G, Strauss A, Vajpai N, Grzesiek S, Tuntland T, Liu Y, Bursulaya B, Azam M, Manley PW, Engen JR, Daley GQ, Warmuth M, Gray NS (2010) Nature 463(7280):501 [PubMed: 20072125]

3. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) Science 298(5600):1912 [PubMed: 12471243]

4. Technology CS. Kinase-Disease Associations <https://www.cellsignal.com/common/content/content.jsp?id=science-tables-kinase-disease>. Accessed 2019

5. Cohen P (2002) Nat Rev Drug Discov 1(4):309 [PubMed: 12120282]

6. Pytel D, Sliwinski T, Poplawski T, Ferriola D, Majsterek I (2009) Anticancer Agents Med Chem 9(1):66 [PubMed: 19149483]

7. Rask-Andersen M, Zhang J, Fabbro D, Schioth HB (2014) Trends Pharmacol Sci 35(11):604 [PubMed: 25312588]

8. Bolanos-Garcia VM, Fernandez-Recio J, Allende JE, Blundell TL (2006) Trends Biochem Sci 31(12):654 [PubMed: 17084631]

9. Poletto G, Vilardell J, Marin O, Pagano MA, Cozza G, Sarno S, Falques A, Itarte E, Pinna LA, Meggio F (2008) Biochemistry 47(32):8317 [PubMed: 18636746]

10. Buchou T, Cochet C (2003) Med Sci (Paris) 19(6–7):709 [PubMed: 12942442]

11. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J (2000) Science 289(5486):1938 [PubMed: 10988075]

12. Liu Y, Gray NS (2006) Nat Chem Biol 2(7):358 [PubMed: 16783341]

13. Tecle H, Shao J, Li Y, Kothe M, Kazmirski S, Penzotti J, Ding Y-H, Ohren J, Moshinsky D, Coli R, Jhawar N, Bora E, Jacques-O'Hagan S, Wu J (2009) Bioorganic & Medicinal Chemistry Letters 19(1):226 [PubMed: 19019675]

14. Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P, Yan C, McConnell P, Spessard C, Banotai C, Mueller WT, Delaney A, Omer C, Sebolt-Leopold J, Dudley DT, Leung IK, Flamme C, Warmus J, Kaufman M, Barrett S, Tecle H, Hasemann CA (2004) Nat Struct Mol Biol 11(12):1192 [PubMed: 15543157]

15. Bogoyevitch MA, Fairlie DP (2007) Drug Discov Today 12(15–16):622 [PubMed: 17706543]

16. Müller G, Sennhenn PC, Woodcock T, Neumann L (2010) IDrugs 13(7):457 [PubMed: 20582870]

17. Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, Liu Y, Zhang G, Hur W, Ding S, Manley P, Mestan J, Fabbro D, Gray NS (2006) Nat Chem Biol 2(2):95 [PubMed: 16415863]

18. Davidson W, Frego L, Peet GW, Kroe RR, Labadia ME, Lukas SM, Snow RJ, Jakes S, Grygon CA, Pargellis C, Werneburg BG (2004) Biochemistry 43(37):11658 [PubMed: 15362850]

19. Hancock CN, Macias A, Lee EK, Yu SY, MacKerell AD, Shapiro P (2005) Journal of Medicinal Chemistry 48(14):4586 [PubMed: 15999996]

20. Stebbins JL, De SK, Machleidt T, Becattini B, Vazquez J, Kuntzen C, Chen L-H, Cellitti JF, Riel-Mehan M, Emdadi A, Solinas G, Karin M, Pellecchia M (2008) Proceedings of the National Academy of Sciences 105(43):16809

21. Burke JR, Pattoli MA, Gregor KR, Brassil PJ, MacMaster JF, McIntyre KW, Yang X, Iotzova VS, Clarke W, Strnad J, Qiu Y, Zusi FC (2003) J Biol Chem 278(3):1450 [PubMed: 12403772]

22. Lindsley CW, Zhao Z, Leister WH, Robinson RG, Barnett SF, Defeo-Jones D, Jones RE, Hartman GD, Huff JR, Huber HE, Duggan ME (2005) Bioorganic & Medicinal Chemistry Letters 15(3):761 [PubMed: 15664853]

23. Nicola G, Vakser IA (2007) Bioinformatics 23(7):789 [PubMed: 17267427]

24. Totrov M, Abagyan R (1997) Proteins Suppl 1:215

25. Chen H, Lyne PD, Giordanetto F, Lovell T, Li J (2006) J Chem Inf Model 46(1):401 [PubMed: 16426074]

26. Lovell T, Chen H, Lyne PD, Giordanetto F, Li J (2008) J Chem Inf Model 48(1):246

27. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR (2008) Br J Pharmacol 153 Suppl 1:S7 [PubMed: 18037925]

28. Schapira M, Totrov M, Abagyan R (1999) J Mol Recognit 12(3):177 [PubMed: 10398408]

29. An J, Totrov M, Abagyan R (2005) Mol Cell Proteomics 4(6):752 [PubMed: 15757999]

30. Abagyan R, Kufareva I. The Flexible Pocketome Engine for Structural Chemical Genomics. Chemogenomics: Concepts and applications of a new design and screening paradigm: Wiley, 2009

31. Sudarsanam S Kinase.com <http://kinase.com/> Accessed 2010, La Jolla, 1999

32. Niedner RH, Buzko OV, Haste NM, Taylor A, Gribskov M, Taylor SS (2006) Proteins 63(1):78 [PubMed: 16435372]

33. Sharma R, Schürer S, Muskal S (2016) High quality, small molecule-activity datasets for kinase research [version 2; referees: 2 approved]. Volume 5

34. Volkamer A, Eid S, Turk S, Rippmann F, Fulle S (2016) J Chem Inf Model 56(2):335 [PubMed: 26735903]

35. Kooistra AJ, Kanev GK, van Linden OPJ, Leurs R, de Esch IJP, de Graaf C (2016) Nucleic Acids Research 44(D1):D365 [PubMed: 26496949]

36. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LS (2005) Nucleic Acids Res 33(Database issue):D154 [PubMed: 15608167]

37. SWISS-Prot <http://ca.expasy.org/sprot/>. Accessed, 2019

38. Bottegoni G, Kufareva I, Totrov M, Abagyan R (2009) J Med Chem 52(2):397 [PubMed: 19090659]

39. Sheridan RP, Maiorov VN, Holloway MK, Cornell WD, Gao YD (2010) J Chem Inf Model 50(11):2029 [PubMed: 20977231]

40. Kufareva I, Ilatovskiy AV, Abagyan R (2012) Nucleic Acids Res 40(Database issue):D535 [PubMed: 22080553]

41. Nicola G, Smith CA, Abagyan R (2008) J Comput Biol 15(3):231 [PubMed: 18333758]

42. Knighton DR, Zheng J, Eyck LFT, Ashford VA, Xuong N-h, Taylor SS, Sowadski JM, Eyck LFTEN (1991) Advancement Of Science 253(5018):407

43. Abagyan R, Totrov M, Kuznetsov DA (1994) J Comp Chem 15:488

44. MolSoft. ICM 3.8 [ICM Program Manual]. San Diego; 2019

45. Hajduk PJ, Huth JR, Fesik SW (2005) Journal of Medicinal Chemistry 48(7):2518 [PubMed: 15801841]

46. Simard JR, Grütter C, Pawar V, Aust B, Wolf A, Rabiller M, Wulfert S, Robubi A, Klüter S, Ottmann C, Rauh D (2009) Journal of the American Chemical Society 131(51):18478 [PubMed: 19950957]

47. Iacob RE, Zhang J, Gray NS, Engen JR (2011) PLoS One 6(1):e15929 [PubMed: 21264348]

48. Kornev AP, Taylor SS (2010) Biochim Biophys Acta 1804(3):440 [PubMed: 19879387]
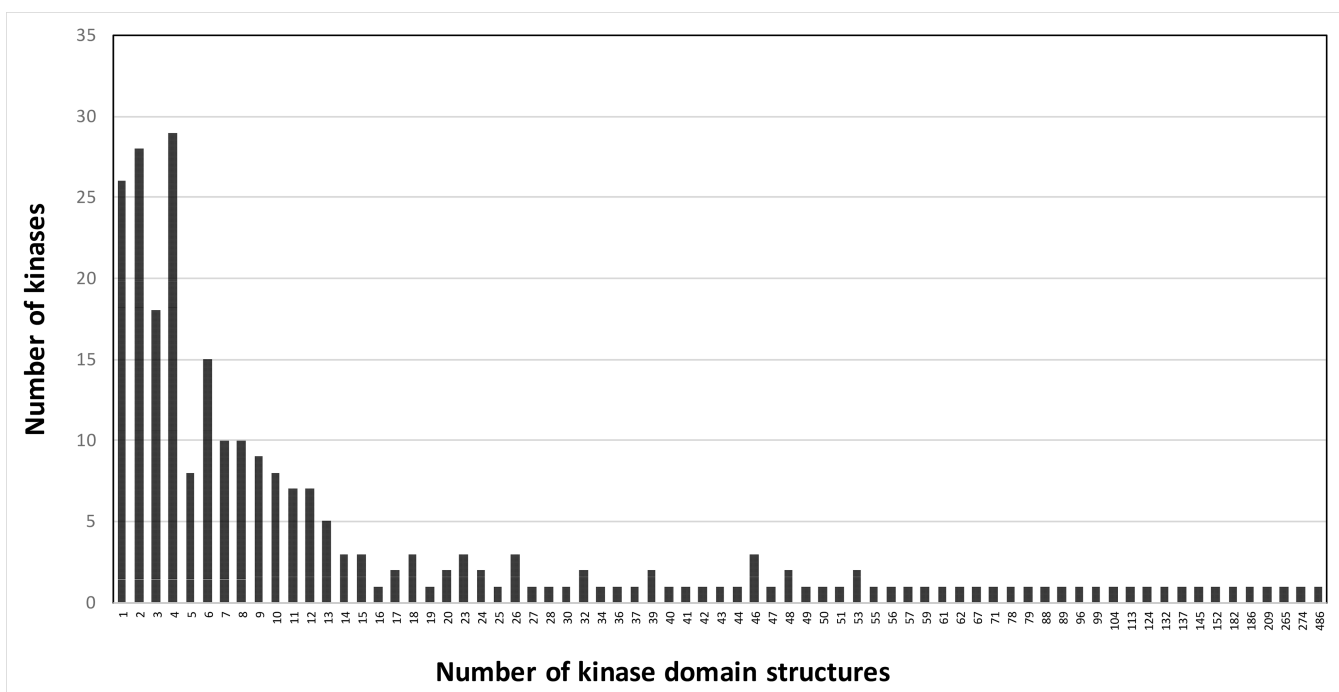
Figure 1: Structural representation of kinase domains in the PDB.
This histogram shows the number of unique kinases (as noted by Swiss-Prot IDs) and the number of structural domains published in the PDB for these kinases. For example, 29 kinase domains have only 4 structures (PDB chains) associated with them, whereas one kinase domain (that of CDK2) has 486 structures.
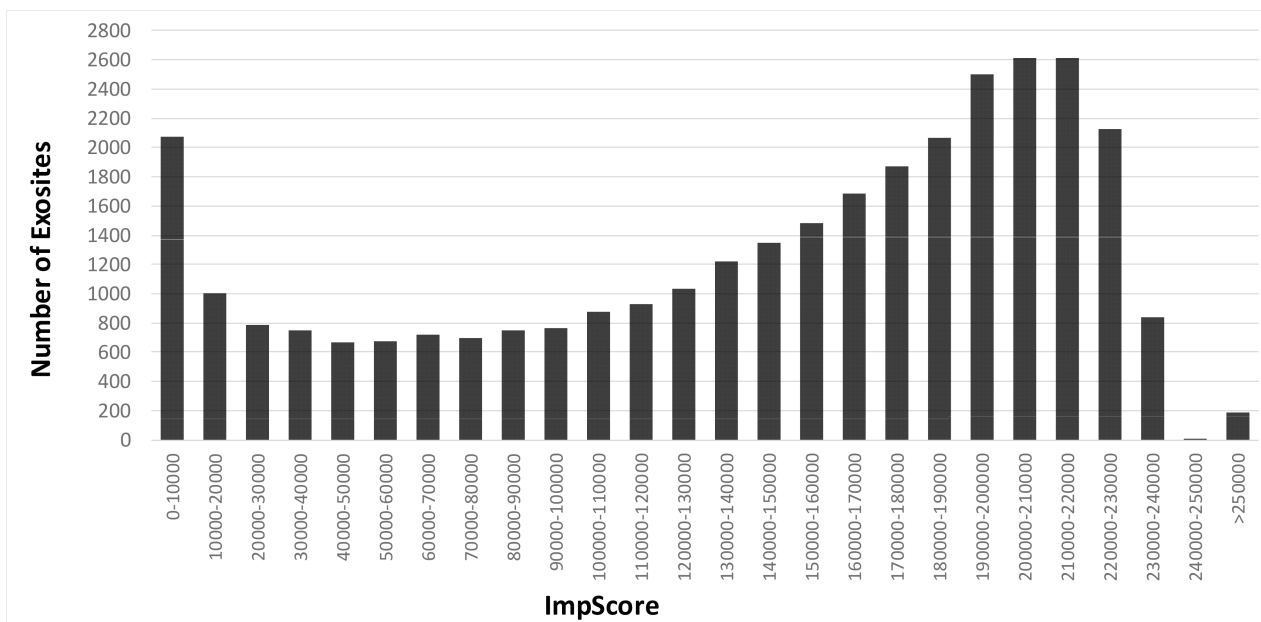
Figure 2:
Histogram showing the distribution of Importance Scores (ImpScore). The data has been truncated at a score of 250,000, as the larger numbers are scarce and unrealistic. The data has been grouped into bins of 10000 for easier viewing.
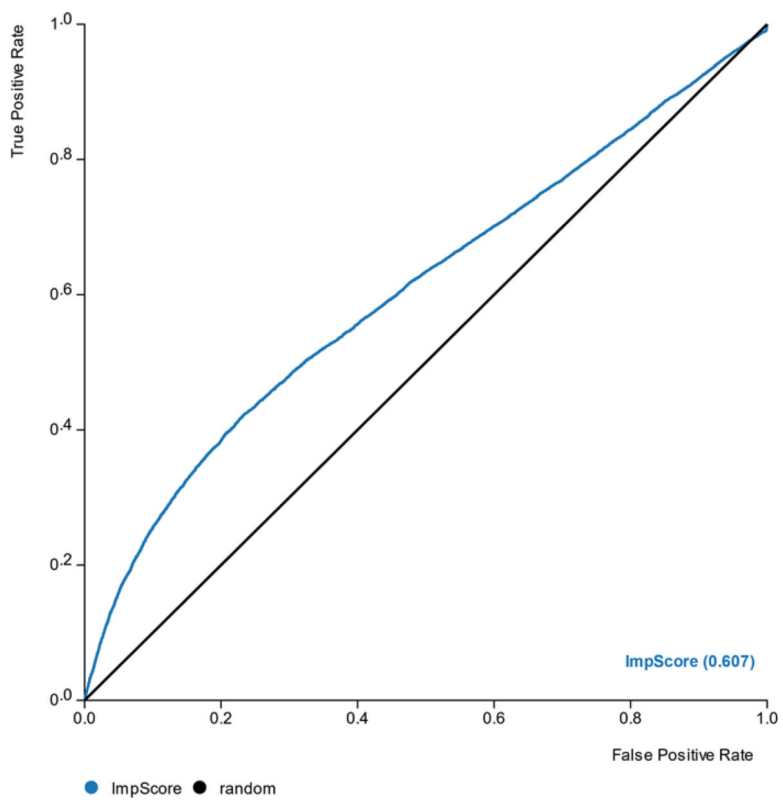
Figure 3:
Receiver Operating Characteristic (ROC) curve of LigExists vs ImpScore. The positive class value for LigExists is the presence of an 'X' (Supplemental Table 1). The area under the curve (AUC) is 0.607.
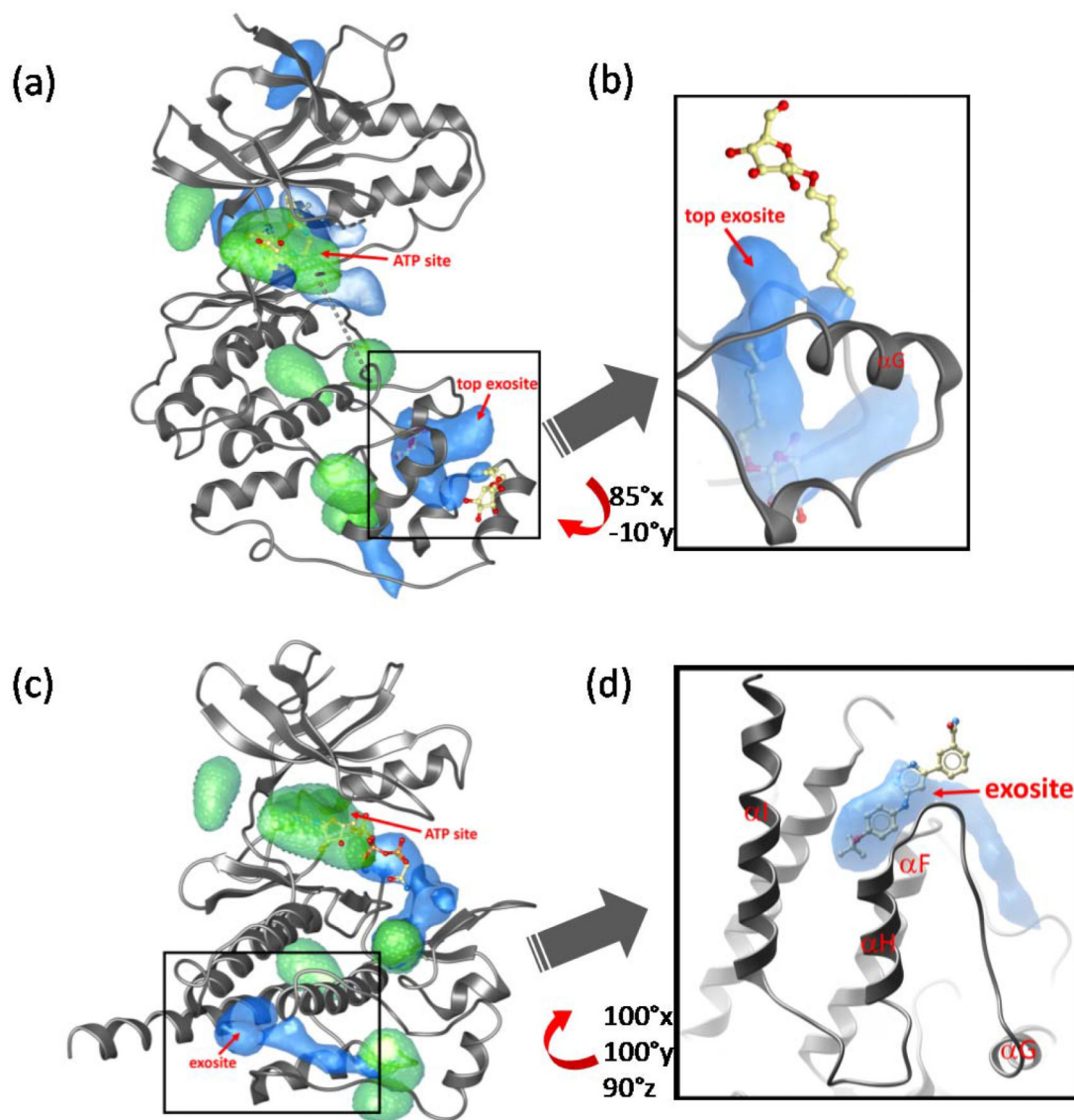
Figure 4:
A graphical representation of ligand-bound exosites in this study. Aside from the ATP binding site, the predicted binding pockets in the structure (colored blue) are unique to this kinase. As a comparison, the Common Exosites averaged across all kinase domains are displayed and colored green. Panel (a) shows the entire kinase domain of human p38 MAP kinase (PDB ID: 3IW5) along with a pocket determined to be the most druggable, with an ImpScore of 32.5. Panel (b) shows a rotated and zoomed view of this top exosite (shown in transparent blue). The pocket is partially buried, and has an elongated shape which could host a drug-like molecule. In the crystal structure, two molecules of β-Octylglucoside are visualized occupying regions of the pocket. Panel (c) shows 2G2F chain A [1] along with common- and domain-specific exosites. Panel (d) is a zoom-in of a pocket of this structure, which fits the criteria of druggable exosite: non-ATP, the residues lining it have no structural gaps or low occupancy, and has a good ImpScore (6013; 1.9% from the top). The pocket is superimposed with GNF-2 (a BCR-ABL allosteric inhibitor) known to bind at this same

position, the myristate pocket, from the 3K5V structure of the same kinase [2]. The size of the predicted exosite pocket suggests room for a larger ligand built to tunnel around helix αF.
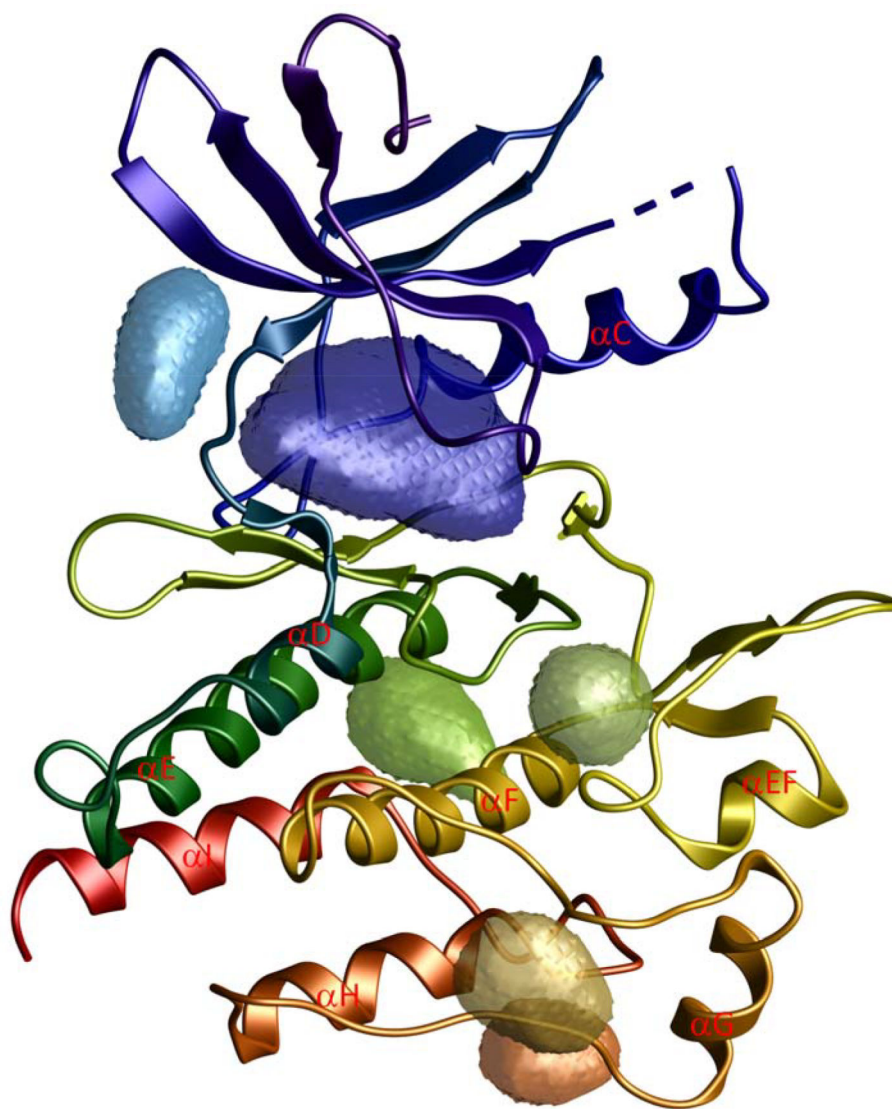
Figure 5.
A graphical view of 'Common Exosites', a normalized, averaged representation of all exosites across all kinase domains. There are six of these common exosites (colored individually) observed in druggable pockets throughout a kinase domain. The large purple pocket is the ATP-binding site. The focus of this study are the other, non-ATP pockets.

Figure 6:

Screenshot of a typical kinase entry. After downloading the Molsoft ICM project file from the online search output, the user can open with Molsoft ICM-Browser to display and control the corresponding kinase structures. These are displayed in the right-hand frame. The PDBs associated with these ensembles, any ligands bound to pockets, and exosites are listed in the left-hand frame. This service portal is available online at http://exosite.ucsd.edu.

Table 1:

Statistics regarding kinase structures and pockets. Information was collected from Swiss-Prot (http://expasy.org/sprot/), PDB (http://pdb.rcsb.org/), and from this study.

Type	Number
Human protein encoding genes	~21,000
Proteins in the human genome with annotated protein kinase activity	518
Proteins that share a common structurally conserved catalytic domain called the protein kinase (PK) domain	507
Human protein PDBs	~100,000
Kinase PDBs	~10,000
Human kinase PDBs (kinase domains)	4169
Human kinase PDB chains	5990
Unique human kinase SwissProt sequences in PDB	320
Kinase domains with more than 1 associated PDB entries	197
Kinase domains with more than 20 associated PDB entries	48
Total number of pockets	32274
Pockets excluding 'Suspect' (ie false positives)	23121
Exosites (non-ATP binding and excluding 'Suspect')	19826
ATP-binding pockets	5993
Exosites with no occupying drug-like ligand in their pocket or X-druggable pocket	10281
Pockets that contain one or more ligand in either their pocket or their equivalent (X-druggable) pocket	18275
Exosites that contain one or more ligand in either their pocket, or their equivalent (X-druggable) pocket.	9545
Exosites that contain one or more ligand in their pocket	586
Kinase domain sequences represented in the exosites list	253
Human kinase PDBs (kinase domains) represented in pocket list	4168

Table 2:

Top-ranked exosites prioritized by best ImpScore from a list of 19826 total exosites. These exosites have been purged of ATP-binding sites and low or zero occupancy residues. The first row on this list is one of the pockets of human EGF receptor (PDB ID: 5U8L) with an ImpScore of 12.6. The next two rows are exosites of AGC Ser/Thr protein kinase (PDB IDs: 3ZH8 and 3A8X).

Family	Kinase	SwissID	PDB	pdbChain	Resolution	Volume	Area	Radius	Nonsphericity	ImpScore	Ligands
Tyr protein kinase	EGF receptor	EGFR_HUMAN_712_979	5u8l	5u8l.a	1.6	446.684	450.104	0.169712	4.7421	12.6083	
AGC Ser/Thr protein kinase	PKC	KPCI_HUMAN_254_522	3zh8	3zh8.c	2.74	452.616	447.171	−0.201755	4.763	17.5848	
AGC Ser/Thr protein kinase	PKC	KPCI_HUMAN_254_522	3a8x	3a8x.b	2	450.591	445.387	−0.20788	4.75588	23.6264	
Ser/Thr protein kinase	CK2	CSK21_HUMAN_39_324	4dgl	4dgl.c	3	445.178	447.863	0.0091003	4.73676	30.8184	
CMGC Ser/Thr protein kinase	MAP kinase	MK14_HUMAN_24_308	3iw5	3iw5.a	2.5	453.14	454.483	−0.296983	4.76483	32.4522	bog, abog2
CMGC Ser/Thr protein kinase	MNB/DYRK	DYRK2_HUMAN_222_535	5lxd	5lxd.b	2.58	442.034	448.754	0.130159	4.72558	67.5865	
CAMK Ser/Thr protein kinase	CaMK	KCC1D_HUMAN_23_279	2jc6	2jc6.c	2.3	457.411	446.711	0.00444246	4.77976	68.0404	
CMGC Ser/Thr protein kinase	MNB/DYRK	DYRK2_HUMAN_222_535	5lxd	5lxd.a	2.58	440.515	448.079	0.182314	4.72017	96.2271	
CMGC Ser/Thr protein kinase	MAP kinase	MK14_HUMAN_24_308	3oef	3oef.x	1.6	445.287	441.407	0.36202	4.73715	97.6458	bog, xbog2
Tyr protein kinase	JAK	JAK1_HUMAN_875_1153	5hx8	5hx8.a	2.2	444.461	458.514	0.926701	4.73421	105.37	
CMGC Ser/Thr protein kinase	MAP kinase	MK14_HUMAN_24_308	5n65	5n65.a	2	460.791	444.124	0.218237	4.7915	152.963	8ot
STE Ser/Thr	STE20	PAK1_HUMAN_270_521	4zy5	4zy5.b	2.35	435.416	448.695	−0.111211	4.70188	216.745	

Family	Kinase	SwissID	PDB	pdbChain	Resolution	Volume	Area	Radius	Nonsphericity	ImpScore	Ligands
protein kinase											
CMGC Ser/Thr protein kinase	MAP kinase	MK14_HUMAN_24_308	3lfd	3lfd.a	3.4	463.141	443.476	−0.131782	4.79963	218.634	bog
Tyr protein kinase	JAK	JAK2_HUMAN_849_1124	4p7e	4p7e.b	2.4	453.244	435.477	0.625725	4.7652	223.837	
Tyr protein kinase	SRC	HCK_HUMAN_262_515	3vry	3vry.b	2.48	454.02	464.596	0.106762	4.76792	231.693	
Tyr protein kinase	ABL	ABL1_HUMAN_242_493	5mo4	5mo4.a	2.17	447.708	434.492	−0.0878121	4.74572	247.926	
Tyr protein kinase	SRC	HCK_HUMAN_262_515	2c0i	2c0i.a	2.3	450.807	466.464	−0.0042565	4.75664	274.011	
AGC Ser/Thr protein kinase	cAMP	KAPCA_HUMAN_44_298	3zo3	3zo3.a	2.1	429.547	454.114	0.25783	4.68066	437.353	
CMGC Ser/Thr protein kinase	MNB/ DYRK	DYRK2_HUMAN_222_535	3k2l	3k2l.a	2.36	468.092	460.999	0.314569	4.81667	450.657	