Published in final edited form as:

J Proteome Res. 2010 September 3; 9(9): 4433–4442. doi:10.1021/pr100662s.

Protein kinase–inhibitor database: Structural variability of and inhibitor interactions with the protein kinase P-loop

Ronak Y. Patel¹ and Robert J. Doerksen^{*,1,2}

¹Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, University, MS, 38677-1848

²Research Institute of Pharmaceutical Sciences, University of Mississippi, University, MS, 38677-1848

Abstract

Structure based drug design of protein-kinase inhibitors has been facilitated by availability of an enormous number of structures in the Protein Databank (PDB), systematic analyses of which can provide insight into the factors which govern ligand–protein kinase interactions and into the conformational variability of the protein kinases. In this study, a non-redundant database containing 755 unique, curated and annotated PDB protein kinase–inhibitor complexes (each consisting of a single protein kinase chain, a ligand and water molecules around the ligand) was created. With this dataset, analyses were performed of protein conformational variability and interactions of ligands with 11 P-loop residues. Analysis of ligand–protein interactions included ligand atom preference, ligand–protein hydrogen bonds and the number and position of crystallographic water molecules around important P-loop residues. Analysis of variability in the conformation of the P-loop considered backbone and side-chain dihedral angles, and solvent accessible surface area (SASA). A distorted conformation of the P-loop was observed for some of the protein kinase structures. Lower SASA was observed for the hydrophobic residue in $\beta 1$ of several members of the AGC family of protein kinases. Our systematic studies were performed amino-acid by amino-acid, which is unusual for analyses of protein kinase–inhibitor complexes.

Keywords

Distorted P-loop; Hydrophobic motif; AGC kinase [cAMP-dependent protein kinase (PKA)/protein kinase G/protein kinase C (PKC)]; Kinase structure space; Atom preference; Conformational plasticity

Introduction

Transfer of the γ -phosphate from adenosine tri-phosphate (ATP) to a serine, threonine or tyrosine residue of a protein is catalyzed by protein kinases. Kinases play key roles in a variety of pathways that mediate essential physiological processes such as transcription, metabolism, cell cycle progression, apoptosis and development. More than 500 genes of protein kinases have been shown to be present in the human genome^{1–3} and ~2% of the

*CORRESPONDING AUTHOR: Dr. Robert J. Doerksen, University of Mississippi, Department of Medicinal Chemistry, 421 Faser Hall, University, MS 38677, Office: 662-915-5880, Fax: 662-915-5638, rjd@olemiss.edu.

Supporting Information Available: Three tables showing the residue names of ATP-like ligands; common names and UniProt IDs, synonyms and family information for the structures in our database; and the PDB ID and name of ligands for each protein in our database. Two figures showing the ligand–protein interactions for notable structures are also given. This material is available free of charge via the Internet at <http://pubs.acs.org>.

proteome of eukaryotes consists of protein kinases.^{4, 5} Selective inhibition of kinases is a well accepted therapeutic approach for treatment of various diseases.^{6, 7} The US Food and Drug Administration (FDA) has approved several protein kinase inhibitors (small molecules imatinib, gefitinib, pegaptanib, sorafenib, dasatinib, sunitinib, erlotinib, nilotinib, lapatinib, pazopanib; macrolides sirolimus, temsirolimus, and everolimus; and even antibodies which act through kinase inhibition, such as bevacizumab, ranibizumab, cetuximab and panitumumab).

Human protein kinases share a protein kinase domain (Pfam ID: PF00069) containing 12 conserved motifs.⁸ These motifs fold into a 3D structure having two lobes, the N- and C-terminal lobes, which are interconnected by a hinge region (Fig. 1). The conserved structural regions in a kinase domain span 17 secondary structure elements: 8 helices (α B- α I) and 9 strands (β 1- β 9) forming three β -sheets. The N-terminal lobe has an anti-parallel β -sheet composed of five strands and two helices in the order β 1- β 2- β 3- α B- α C- β 4- β 5. The C-terminal lobe starts with a small helix α D and a long helix α E. This is followed by strand β 6. The C-terminal lobe has small β -sheets each made of two strands (β 6, β 9 and β 7, β 8). An activation loop, which needs to be activated before catalysis can occur for some of the kinases, is located after β 9. The activation loop is followed by four helices (α F- α I). One or more residues in or around β 1- β 5, α C and β 7- β 8 form a binding site for ATP and ATP-competitive inhibitors (Fig. 1).

The Protein Databank (PDB)⁹ is flooded with structures of protein kinases in complex with ligands. The number of deposited structures containing at least one chain with a kinase domain as a function of year is shown in Fig. 2. As of July 2009, there were 1248 distinct PDB entries with at least one kinase domain. In 2007 and 2008, 239 and 269 structures, respectively, were deposited in the PDB, or approximately 2 structures every 3 days, making it difficult to keep pace with the growing amount of kinase structure data. Automated tools and systematic analysis are urgently needed to extract information from this huge source of knowledge.

There have been some efforts to analyze the structures of protein kinases all together¹⁰⁻²³ or in a specific group.²⁴⁻²⁹ Overall the studies can be categorized into analysis of hydrogen bonds,^{22, 23, 30} analysis of active-site hydration,²⁹ overall binding mode of ligands and 2D representation,^{15-18, 21} pharmacophore models of kinases bound to inhibitors,¹² kinase binding-site similarities and/or differences,^{14, 31, 32} correlation of properties of ligands and conformations of proteins,^{27, 28} and analysis and modeling of sequence and/or structural variability in protein kinases.^{11, 13, 19, 20, 30} In addition to these, there are outstanding review articles about protein kinase-inhibitor interactions and on the conformational space accessed by protein kinases.^{3, 6, 7, 33, 34} Altogether, these studies provide an excellent understanding of kinase binding sites, especially for interactions with the hinge region³⁰ and with the Asp-Phe-Gly (DFG) motif in the activation loop. However, interpretations are sometimes difficult to follow because of the use of numbering and nomenclature of some particular kinase, making it difficult to correlate with the residue numbering of other kinases which may be of interest to the reader, a concern which has recently been described.³⁰ Frequently, proteins are brought to a common frame of reference by structure superposition before systematic analysis is performed. There is a significant chance that some regions could be misaligned because of the conformational variability in kinase structures and the large number of structures available.

In order to complement previous reported studies on large scale analysis of protein kinases, herein we aimed to analyze in detail the conformational variability in amino-acid residues in the protein kinase P-loop and their interactions with ligands. As a first step of this analysis, a non-redundant database of human protein kinases in complex with inhibitors was created.

The database contains 755 unique protein kinase–inhibitor complexes collected from the PDB, each with a single protein kinase chain, a non ATP-like ligand and water molecules around the ligand. This dataset was curated, including removal of co-crystallized chains or other peptides/proteins, assignment of ligand bond order from the information available in Ligand Expo, addition of hydrogen atoms to ligands and removal of redundancy. None of the analysis performed in the present study depends on kinase 3D structural alignment, and hence our database should avoid errors of incorrect structural alignment which could arise from conformational variability of one or more regions in certain structures. We catalogued this database using common nomenclature and family categorization of protein kinases as found at <http://www.kinase.com>.¹

After preparation of the database, we have started analysis of kinase structure and kinase–ligand interactions with a focus on the phosphate loop (P-loop). The P-loop, often referred to as the Glycine rich loop (G-loop/G-rich loop), is a part of the N-terminal β -sheet region and is located between strands $\beta 1$ and $\beta 2$. One hydrophobic residue of $\beta 1$ and one hydrophobic residue of $\beta 2$ are part of the A-pocket (adenine-binding pocket) of the front binding cleft and provide non-polar interactions to the adenine ring of ATP.¹⁵ Phosphorylation of the P-loop has been shown to be a regulatory switch for various protein kinases.^{35, 36} Several mutations in the P-loop were shown to be responsible for resistance to imatinib and the subsequent death of patients with chronic myeloid leukemia.^{37, 38} Considering the importance of the P-loop and the fact that it marks the start of the kinase domain, herein we study the conformational space accessed by the P-loop and its interactions with ligands, using the protein kinase–inhibitor database we have prepared. We restricted our analysis to 11 residues of the P-loop and made use of a relative residue reference system in order to avoid using the nomenclature of any one particular kinase.

Analysis of ligand–protein interactions included ligand atom preference, ligand–protein hydrogen bonds and the number and position of crystallographic water molecules around important P-loop residues. Analysis of variability in the conformation of the P-loop considered backbone and side-chain dihedral angles, and solvent accessible surface area. While the analysis of conformational variability in the P-loop provided insights into additional structural space accessed by a few of the kinases, the analysis of ligand–protein interaction provides overall clues about ligand–P-loop interactions. We have prepared a web interface for our results through which it is easy not only to identify protein structures having particular interactions with individual residues of the P-loop but also to comprehend inhibitor/water–protein kinase interactions for individual amino-acids of the P-loop. In addition to this, the curated database can be used for a variety of purposes such as drug design, lead optimization, kinase selectivity studies, benchmarking/cross-docking experiments and/or further analysis of binding site similarities/differences.

Methods

Database

Amino-acid sequences of proteins in the Protein Databank (PDB)⁹ were downloaded from ftp://ftp.wwpdb.org/pub/pdb/derived_data/pdb_seqres.txt on 20th July, 2009 (a total of 147,231 sequences). The sequences were searched for the protein kinase domain (Pfam ID: PF00069.17) using the *hmmsearch* module of HMMER.³⁹ It should be noted that although PF00069.17 can fetch all the protein kinases having typical protein-kinase domains, it cannot retrieve atypical protein kinases. Since our analysis is focused on structurally conserved regions that are common to typical protein kinases, we preferred not to include atypical protein kinases in our analysis. The structures of sequences containing a protein kinase domain were then downloaded from the PDB.

The structures were split into three parts: protein, ligands other than water, and water. The protein part was further split into different chains. Only those chains containing a protein kinase domain as identified by *hmmsearch* were retained. ATP and ATP-like ligands were removed (the identification names of the residues that were removed are listed in the Supporting Information). Small ligands with ≤ 15 atoms were also removed. The ligands in PDB format files do not contain hydrogen atoms or bond order information. For each ligand, a structure data file (SDF format), which contains bond order information for individual ligands, was downloaded from Ligand Expo of the Research Collaboratory for Structural Bioinformatics (RCSB). Hydrogen atoms were added to each ligand using Schrödinger software suite.⁴⁰ For some entries hydrogen atoms were not added properly by the software and problems were found with the geometry of the ligand in the modified SDF file. Such ligands were identified by visual inspection after which the correct bond order and hydrogen atoms were assigned by referring to the literature.

For each protein chain with a kinase domain, neighboring ligands were identified within a distance of 6 Å. Any crystallographic water molecules were retained if they were found within 9 Å of the centroid of ligands. This was followed by combining protein, ligand and water molecules. This procedure sometimes resulted in having more than one ligand per chain. In such a case, after visual inspection of each structure the ligand in either the ATP binding site or the allosteric site was retained, and any other ligands were removed.

If for two different entries the PDB IDs and ligand residue names were matching, then only one of the chains was retained in the database. In this way, only one chain was retained from multimeric proteins containing the same kinase and same ligand. The UniProt accession ID of each remaining chain was taken from the DBREF line of the PDB file. This was followed by removal of redundant entries having both the same UniProt ID and the same ligand residue name. After removal of redundancy, the database was considerably filtered. Hydrogen atoms were added to the protein using PyMol.⁴¹ Each entry in the filtered database has a protein with a single chain, one ligand (with hydrogen atoms added using information from Ligand Expo) in the kinase/allosteric binding site and crystal water molecules around the ligand. After the curation, there were 755 entries in the database.

Derived information

Information about the source organism was found in each PDB file. For each human and mouse protein kinase, the common name of the kinase, the common abbreviation used for the kinase, and the family to which it belongs was manually obtained from <http://www.kinase.com>,¹ by matching with the UniProt names/IDs.⁴² The amino-acid sequence of each protein in the database was derived from its PDB structure. Hence the sequence will not contain any amino-acid missing in the PDB structure. This procedure was used in order to have a strict match between structure and sequence. The sequences were then aligned using *hmmalign*, which employs the protein kinase domain as a reference. Since there are so many sequences in the database (755), some of which may have missing regions, it is likely that a few sequences are misaligned despite our great care.

Selection of structures for analysis of the P-loop

The P-loop links $\beta 1$ and $\beta 2$ of the kinase domain. Considering the residues that interact with inhibitors and are located together in a single region, the P-loop in our analysis is defined by 11 residues, four from $\beta 1$, two from the loop connecting $\beta 1$ and $\beta 2$, and five from $\beta 2$ (Figs. 1 and 3). The actual region used in the alignment included 18 positions, because some of the sequences contained regions of misalignment. A stepwise procedure was followed. First, structures containing all 11 residues in these regions were selected for analysis of the P-loop, and other structures were removed from the analysis. Structures with a broken backbone or

missing atoms in side-chains were also removed from the analysis of the P-loop. The 532 structures which passed these two filtration criteria were used for the analysis of conformational changes in the P-loop and for study of interactions of the P-loop with ligands.

Nomenclature of residues

The most conserved glycine in the P-loop was given residue number zero (G0). Neighboring residues towards the N-terminal end from G0 were given negative residue numbers (such as G-2) and towards the C-terminal were given positive residue numbers (such as G3), in sequential order (Fig. 3B). In our nomenclature, the residue number is always preceded by the most conserved residue, if any. If it was not possible to identify a highly conserved residue then the letter X was used instead (such as X-3).

Analysis

Sequence alignments were performed using *hmmalign* of *HMMER* 2.3.2.³⁹ Sequence logos were generated using *WebLogo*.⁴³ Secondary structure assignment and backbone dihedral angles (ϕ , ψ) were computed using *Stride*.⁴⁴ *Stride* represents α -helix, extended structure (β -sheets), isolated β -bridge, turn, coil, 3-10 helix and π -helix as H, E, B/b, T, C, G and I, respectively. The solvent accessible surface areas were computed using *NAccess*.⁴⁵ Side-chain dihedral angles were computed using *Dangle* (<http://kinemage.biochem.duke.edu/software/dangle.php>). Visualization and rendering of structures were done using *PyMol*,⁴¹ *RasMol*⁴⁶ and *Maestro* from the Schrödinger software suite.⁴⁰ For visualization of alignments *BioEdit*⁴⁷ was used. Protein-ligand hydrogen bonds were computed using the Schrödinger software suite.⁴⁰ After adding hydrogens, further minimization was not performed. A distance criterion of 2.5 Å between donor hydrogen and acceptor atom was used to define hydrogen bonds. Graphs were created using *grace* (<http://plasma-gate.weizmann.ac.il/Grace/>).

Results and discussions

Dataset

The number of crystal structures in our database after processing (described in Methods) is 755 (Supporting Information). The database consists of protein kinases from 13 different organisms. Table 1 summarizes the number of structures per organism in the dataset. The kinase which appears most frequently in the database is CDK2 of human origin, with 134 structures (Table 2). At least 25 structures each are also included for MAP kinase p38 alpha (p38 α /MK14_Human), checkpoint Kinase 1 (CHK1/CHK1_Human) and proto-oncogene serine/threonine-protein kinase 1 (PIM1/PIM1_Human) of human and protein kinase A c-alpha (PKACa/KAPCA_BOVIN) of bovine origin. Among the eight families of human protein kinases, AGC, CAMK, CMGC, CK1, STE, TK, TKL and Other, the greatest number of unique structures in our database is from the TK family and the fewest are from the CK1 family (only two unique structures belong to the latter). A summary of the total number of structures, unique structures and names of each member is given in Table 3. The database is accessible through a web interface (http://www.pharmacy.olemiss.edu/medicinal_chemistry/kinases/pkinase_db.html).

The most common ligand in the dataset is staurosporin (STU), which is present in complex with 15 different kinases. Information about each member of the dataset is given in the Supplementary Information along with ligand ID, PDB ID, protein name and common abbreviations. A typical ligand makes contact with an average of 21 protein kinase residues (residues with even a single atom within 4 Å of the ligand were considered to be in contact),

with the least being with 12 residues (2hwp_A and 3h30_A with ligands DJK and RFZ, respectively) and the greatest being with 34 residues (3ctq_A with ligand 337).

Ligand–P-loop interactions

Out of the total of 755 structures, 532 had a complete backbone, side-chain and residues properly aligned in the P-loop alignment. The P-loop was defined by 11 residues in our analysis. These residues span across two strands, β 1 and β 2, as well as the loop connecting them (Fig. 1). The sequence conservation in this region and the variation in secondary structure elements are described in Figs. 3A and 3B, respectively. Figure 3A shows that there are six physicochemically conserved residues in this region: a hydrophobic residue in β 1 (L–3) which is most commonly leucine or isoleucine, two glycines separated by one residue (G–2 and G0), an aromatic residue (X2) which is most commonly phenylalanine or tyrosine, a glycine (G3) and a hydrophobic residue in β 2 (V5), which is mostly valine. Figure 3B shows, as expected, that the P-loop residues are primarily part of an extended β -sheet structure except for the central residues which are primarily turn, though there is considerable variability for G0 and G2 which also sometimes are found to participate in the turn.

The two hydrophobic residues in the P-loop (L–3 and V5) are almost always in contact with inhibitors (Fig. 4A). These residues are part of the adenine pocket (A-pocket) and provide hydrophobic interactions with the adenine ring of ATP.¹⁵ In addition to these two residues, ligands were observed to make contact most frequently with the first (G–2) and second (G0) conserved glycine and with the variable residue in between them (X–1) (Fig. 4A). It is of great interest to medicinal chemists to design ligands that interact with the maximum possible number of amino-acids to provide the best activity and selectivity. For interactions between inhibitors and individual amino-acids of the P-loop, the maximum number of contacts made by a ligand to the entire P-loop was observed for ligand ARX in the structure of PKACa/KAPCA_BOVIN (3BWJ) (Figure S1). In this structure, the ligand interacts with eight residues of the P-loop, which is significantly more compared to most other structures (for 396 of the structures, only 2–4 P-loop residues are involved in interactions with ligands). One can obtain clues from structures like 3BWJ of the requirements for design or optimization of inhibitors which bind well to the P-loop which may lead to improved activity and/or selectivity for particular protein kinases.

In general carbon atoms were the ligand atoms most commonly found to be in the neighborhood of any of the residues (Fig. 4B). The presence of carbon atoms in the neighborhood of these amino-acids indicates that the primary nature of ligand–P-loop interaction is hydrophobic. The residues most frequently near ligand carbons were the hydrophobic L–3 and V5, which would be expected to help stabilize the binding of hydrophobic regions of ligands, followed by G–2, G0 and X2. However, for some complexes, oxygen and nitrogen atoms were present, especially for the residue V5 in β 2, indicating polar interactions with the backbone or unfavorable interactions with the hydrophobic side-chains. A few examples of various types of inhibitor atoms interacting with protein kinases are described in Figure S2.

There were only a few complexes in which a ligand was observed to make one or more hydrogen bonds with a protein (Fig. 4C). Eleven structures out of 532 formed hydrogen bonds with a ligand through residue (X2) in the P-loop (Fig. 4C), in general with the backbone and occasionally with a Tyr side-chain. We have provided an overview of the important ligand-protein interactions in the present manuscript, and have made the database accessible to readers to explore through our web page for more detailed analysis relevant to their individual interest.

Crystallographic water molecules in the neighborhood of the P-loop

Almost all residues of $\beta 1$ are on the protein surface. According to the distance criterion we used, a significant number of water molecules were observed around each residue except X6 (Fig. 4D). These water molecules might not always be intermediate between ligand and kinase. Analysis of the water molecules around the P-loop residues showed that mostly those around the first glycine (G-2) and hydrophobic residue (L-3) are in the region exposed to the surface (Fig. 1C). However, water molecules around the hydrophobic residue in the $\beta 2$ strand (V5) are instead surprisingly positioned between the ligand and hydrophobic residues of the binding site in a few structures (Fig. 1C). The water molecules near V5 are not characteristic of a specific protein, but are found in several different proteins.

Conformational space of the P-loop

In order to obtain a picture of the conformational space accessed by important residues in the binding site of protein kinases, the backbone and side-chain dihedral angles were evaluated. Analysis of the backbone dihedral angles of the P-loop residues making significant contact with ligands showed that for some structures of ABL1 of human and mouse, p38 α (p38 α /MK14_HUMAN), basic fibroblast growth factor receptor 1 (FGFR1/FGFR1_HUMAN), casein kinase II subunit alpha (CK2 α 1/CSK21_HUMAN), and c-Jun N-terminal kinases 3 (JNK3/MK10_HUMAN) of human and Aurora kinase A (AurA/STK6_MOUSE) of mouse there were one or more significantly different backbone dihedral angles when compared to other kinases for residues L-3 and G-2 (Fig. 5). For the majority of protein kinases, ϕ of amino-acid L-3 ranges from -60° to -150° , while ψ ranges from -60° to 60° . For a few protein kinases listed above, although ϕ remains more or less the same as for the other protein kinases ψ is quite different, with values between 120° to 180° for L-3. Similarly, for these selected proteins, G-2 takes on a different conformation ($\phi = 30^\circ$ to 120° and $\psi = -60^\circ$ to 60°) than that of the majority of other protein kinases (where both ϕ and ψ access conformations either between -120° and -180° or between 120° and 180°). When the proteins with different conformation of residues L-3 and G-2 were visually inspected, these proteins were found to have a distorted/collapsed P-loop relative to that of other structures (Fig. 6). This P-loop conformational difference is not a property of any particular kinase or group of kinases. In spite of variations in this region, a remarkable similarity in backbone dihedral angle was observed for V5 (Fig. 5).

The P-loop is known to be a highly flexible region of kinases. A different conformation of the P-loop has been reported previously for various kinases such as FGFR1,⁴⁸ ABL1,⁴⁹⁻⁵² Casein kinase 2 alpha, CK2 α /CSK2A_HUMAN,⁵³ Rho-associated protein kinase 1, ROCK1/HUMAN,⁵⁴ p38 α ⁵⁵ and CDK5⁵⁶. This different conformation of the P-loop is referred to as a collapsed,⁵³ distorted,⁵¹ caged⁵⁰ or inactive conformation.⁵⁰ The altered conformation of the P-loop has been suggested to improve ligand-kinase shape complementarity and to provide more interactions with inhibitors.^{49, 52} Since different conformations are observed only for some kinase-inhibitor complexes, this conformation may be induced by the physicochemical nature of the inhibitor. Imatinib, an approved kinase inhibitor drug, stabilizes the collapsed conformation of the P-loop,^{50, 52} and the mutations of ABL1 that cause resistance to imatinib are suggested to act by preventing stabilization of the altered conformation of the P-loop.⁵⁰ Apart from reports of ligand-induced change in the conformation of the P-loop, there is at least one study in which an altered conformation of the P-loop was observed for an inactive protein without bound inhibitor, CK2 α .⁵³ In addition to this, a surprising observation about an altered conformation of the P-loop comes from the crystal structure of ROCK1 in complex with fasudil (PDB ID: 2F2U). In this structure, one subunit has a collapsed P-loop in the active site, while the other does not. The ligand fasudil was also observed to have a different conformation in each subunit.⁵⁴ Considering all of the above-listed studies in which different conformations of the P-loop

were observed for different systems (same ligand–same protein, protein without ligand, same protein–different ligand), the exact mechanism and factors responsible for these conformational changes cannot be conclusively stated and merit further study.

Analysis of side-chain dihedral angles for L–3 showed that χ_1 is found mainly in -60° and $+60^\circ$ conformations and χ_2 accesses *trans* conformations (Fig. 7). There are some variations from this behavior in which intermediate conformers were observed for some structures. For the hydrophobic residue in β_2 (V5), the most conserved residue is valine, which has only one side-chain dihedral angle. This dihedral accesses mostly *trans* conformations; however, deviation from this trend is observed for some structures (Fig. 7). The χ_1 dihedral for some of the p38 α structures deviated from -60° or $+60^\circ$ for L–3. The variability in Fig. 7 which shows the flexibility of these two residues which provide interactions to most protein-kinase inhibitors in almost all the crystal structures highlights the importance of the use of flexible side-chains, induced fit effects or ensemble (multiple reference structures) methods for ligand docking.

Solvent accessibility of residues in the P-loop

In order to analyze the difference in relative exposure to solvent of residues in different protein kinases, solvent accessible surface areas (SASA) of three important residues in the P-loop were computed (Fig. 8). The SASA of G–2 (average, $27.48 \text{ \AA}^2$; minimum, $13.77 \text{ \AA}^2$ and maximum, $39.37 \text{ \AA}^2$) and V5 (average, $31.84 \text{ \AA}^2$; minimum, $17.72 \text{ \AA}^2$ and maximum, $48.31 \text{ \AA}^2$) were almost the same for all structures in the database. However, the SASA for L–3 (the hydrophobic residue in β_1 of the P-loop) was on average $68.73 \text{ \AA}^2$ but varied greatly to a maximum of $103.33 \text{ \AA}^2$ for human PIM1. SASA for L–3 was significantly lower for a few proteins which belong to the AGC family of protein kinases: human Rho-associated protein kinase 1 (ROCK1/ROCK1_HUMAN, $21.69 \text{ \AA}^2$); human RAC- α serine/threonine-protein kinase beta (Akt2/AKT2_HUMAN; $13.33 \text{ \AA}^2$); and bovine and mouse cAMP-dependent protein kinase catalytic subunit alpha (PKACa/KAPCA_BOVIN, $12.94 \text{ \AA}^2$ and KAPCA_MOUSE, $16.44 \text{ \AA}^2$, respectively). Visual inspection of the exceptional structures showed that for these proteins a C-terminal extension passes across the β_1 strand (Fig. 9), hiding L–3 and hence making it less solvent-accessible.

The AGC family of kinases are known to have a unique conserved hydrophobic motif, HM (FxxF[S,T,D]Y), at the C-terminal of the protein, and its importance in activation of members of the AGC family has been documented.^{57–60} In Akt2, ROCK1 and PKACa, this motif fits into the groove formed by α_B and α_C . For this binding the disordered C-terminal segment crosses over the N-terminal β -sheet rich domain of the kinase. On its way, the segment interacts with the hydrophobic residue in β_1 (L–3), hiding it from the solvent, which results in a decrease in the solvent accessible surface area for these structures.

Conclusions

In the present study, a non-redundant database of human protein kinases in complex with inhibitors was created. The database contains 755 unique protein kinase–inhibitor complexes. Statistical analyses of the interactions of each amino-acid (11 residues) of the P-loop with ligands and the variability of structures within the P-loop were performed for 532 complexes with a complete backbone and all atoms in side-chains of amino-acids in the P-loop. The analyses undertaken in the present study provide in-depth insight into the structure and interactions of the P-loop in protein kinases. The created, curated and annotated dataset of protein kinase–inhibitor complexes may prove useful for drug design, lead optimization, kinase selectivity studies, benchmarking/cross-docking experiments and/or further analysis of binding site similarities/differences. Analysis of other kinase domain regions is underway and will be published in due course.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this work came from National Science Foundation EPS-0556308 and EPS-0903787. This investigation was conducted in a facility constructed with support from Research Facilities Improvements Program (C06 RR-14503-01) from the NIH National Center for Research Resources.

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002; 298(5600):1912–34. [PubMed: 12471243]
2. Eswaran J, Knapp S. Insights into protein kinase regulation and inhibition by large scale structural comparison. *Biochim Biophys Acta*. 2009; 1804(3):429–432. [PubMed: 19854302]
3. Cherry M, Williams DH. Recent kinase and kinase inhibitor X-ray structures: mechanisms of inhibition and selectivity insights. *Curr Med Chem*. 2004; 11(6):663–73. [PubMed: 15032722]
4. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LS, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJ, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vossell LB, Zhang J, Zhao Q, Zheng XH, Lewis S. Comparative genomics of the eukaryotes. *Science*. 2000; 287(5461):2204–15. [PubMed: 10731134]
5. Huse M, Kuriyan J. The conformational plasticity of protein kinases. *Cell*. 2002; 109(3):275–82. [PubMed: 12015977]
6. Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure. *Science*. 2004; 303(5665):1800–5. [PubMed: 15031492]
7. Thaimattam R, Banerjee R, Miglani R, Iqbal J. Protein kinase inhibitors: structural insights into selectivity. *Curr Pharm Des*. 2007; 13(27):2751–65. [PubMed: 17897021]
8. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*. 1988; 241(4861):42–52. [PubMed: 3291115]
9. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol*. 2003; 10(12):980. [PubMed: 14634627]
10. Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J Med Chem*. 2004; 47(2):337–44. [PubMed: 14711306]
11. Kufareva I, Abagyan R. Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *J Med Chem*. 2008; 51(24):7921–32. [PubMed: 19053777]
12. McGregor MJ. A pharmacophore map of small molecule protein kinase inhibitors. *J Chem Inf Model*. 2007; 47(6):2374–82. [PubMed: 17941626]
13. Knight JD, Qian B, Baker D, Kothary R. Conservation, variability and the modeling of active protein kinases. *PLoS One*. 2007; 2(10):e982. [PubMed: 17912359]
14. Niwa T. Elucidation of characteristic structural features of ligand binding sites of protein kinases: a neural network approach. *J Chem Inf Model*. 2006; 46(5):2158–66. [PubMed: 16995746]
15. Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem*. 2007; 50(3):409–24. [PubMed: 17266192]
16. Liao JJ. Molecular targeting of protein kinases to optimize selectivity and resistance profiles of kinase inhibitors. *Curr Top Med Chem*. 2007; 7(14):1332–5. [PubMed: 17692023]
17. Liao JJ, Andrews RC. Targeting protein multiple conformations: a structure-based strategy for kinase drug design. *Curr Top Med Chem*. 2007; 7(14):1394–407. [PubMed: 17692028]

18. Naumann T, Matter H. Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J Med Chem.* 2002; 45(12): 2366–78. [PubMed: 12036347]
19. Kornev AP, Taylor SS, Ten Eyck LF. A helix scaffold for the assembly of active protein kinases. *Proc Natl Acad Sci U S A.* 2008; 105(38):14377–82. [PubMed: 18787129]
20. Ten Eyck LF, Taylor SS, Kornev AP. Conserved spatial patterns across the protein kinase family. *Biochim Biophys Acta.* 2008; 1784(1):238–43. [PubMed: 18067871]
21. Ghose AK, Herbertz T, Pippin DA, Salvino JM, Mallamo JP. Knowledge based prediction of ligand binding modes and rational inhibitor design for kinase drug discovery. *J Med Chem.* 2008; 51(17):5149–71. [PubMed: 18710211]
22. Voth AR, Ho PS. The role of halogen bonding in inhibitor recognition and binding by protein kinases. *Curr Top Med Chem.* 2007; 7(14):1336–48. [PubMed: 17692024]
23. Panigrahi SK. Strong and weak hydrogen bonds in protein-ligand complexes of kinases: a comparative study. *Amino Acids.* 2008; 34(4):617–33. [PubMed: 18180869]
24. Verkhivker GM. Imprint of evolutionary conservation and protein structure variation on the binding function of protein tyrosine kinases. *Bioinformatics.* 2006; 22(15):1846–54. [PubMed: 16720585]
25. Verkhivker GM. Exploring sequence-structure relationships in the tyrosine kinome space: functional classification of the binding specificity mechanisms for cancer therapeutics. *Bioinformatics.* 2007; 23(15):1919–26. [PubMed: 17537753]
26. Verkhivker GM. In silico profiling of tyrosine kinases binding specificity and drug resistance using Monte Carlo simulations with the ensembles of protein kinase crystal structures. *Biopolymers.* 2007; 85(4):333–48. [PubMed: 17167796]
27. Subramanian J, Sharma S, BR C. A novel computational analysis of ligand-induced conformational changes in the ATP binding sites of cyclin dependent kinases. *J Med Chem.* 2006; 49(18):5434–41. [PubMed: 16942017]
28. Subramanian J, Sharma S, B Rao C. Modeling and selection of flexible proteins for structure-based drug design: backbone and side chain movements in p38 MAPK. *ChemMedChem.* 2008; 3(2): 336–44. [PubMed: 18081134]
29. Kriz Z, Otyepka M, Bartova I, Koca J. Analysis of CDK2 active-site hydration: a method to design new inhibitors. *Proteins.* 2004; 55(2):258–74. [PubMed: 15048820]
30. Brooijmans N, Chang YW, Mobilio D, Denny RA, Humblet C. An enriched structural kinase database to enable kinome-wide structure-based analyses and drug discovery. *Protein Sci.*
31. Kuhn D, Weskamp N, Hullermeier E, Klebe G. Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem.* 2007; 2(10):1432–47. [PubMed: 17694525]
32. Kinnings SL, Jackson RM. Binding site similarity analysis for the functional classification of the protein kinase family. *J Chem Inf Model.* 2009; 49(2):318–29. [PubMed: 19434833]
33. Liu Y, Gray NS. Rational design of inhibitors that bind to inactive kinase conformations. *Nat Chem Biol.* 2006; 2(7):358–64. [PubMed: 16783341]
34. Scapin G. Structural biology in drug design: selective protein kinase inhibitors. *Drug Discov Today.* 2002; 7(11):601–11. [PubMed: 12047871]
35. Gu Y, Rosenblatt J, Morgan DO. Cell cycle regulation of CDK2 activity by phosphorylation of Thr160 and Tyr15. *EMBO J.* 1992; 11(11):3995–4005. [PubMed: 1396589]
36. Zukerberg LR, Patrick GN, Nikolic M, Humbert S, Wu CL, Lanier LM, Gertler FB, Vidal M, Van Etten RA, Tsai LH. Cables links Cdk5 and c-Abl and facilitates Cdk5 tyrosine phosphorylation, kinase upregulation, and neurite outgrowth. *Neuron.* 2000; 26(3):633–46. [PubMed: 10896159]
37. Branford S, Rudzki Z, Walsh S, Parkinson I, Grigg A, Szer J, Taylor K, Herrmann R, Seymour JF, Arthur C, Joske D, Lynch K, Hughes T. Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis. *Blood.* 2003; 102(1): 276–83. [PubMed: 12623848]
38. Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, Sawyers CL. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor

- imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell*. 2002; 2(2):117–25. [PubMed: 12204532]
39. Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998; 14(9):755–63. [PubMed: 9918945]
 40. Maestro, Version 9.0. Schrödinger LLC; New York, NY: 2009.
 41. DeLano, WL. The PyMOL Molecular Graphics System. DeLano Scientific; Palo Alto, CA, USA: 2002.
 42. UniProt Consortium, The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res*. 2009; 37(Database issue):D169–74. [PubMed: 18836194]
 43. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004; 14(6):1188–90. [PubMed: 15173120]
 44. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins*. 1995; 23(4):566–79. [PubMed: 8749853]
 45. Hubbard, SJ.; Thornton, JM. NACCESS. Department of Biochemistry and Molecular Biology, University College London; 1993.
 46. Sayle RA, Milner-White EJ. RASMOL: biomolecular graphics for all. *Trends Biochem Sci*. 1995; 20(9):374. [PubMed: 7482707]
 47. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999
 48. Mohammadi M, McMahon G, Sun L, Tang C, Hirth P, Yeh BK, Hubbard SR, Schlessinger J. Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. *Science*. 1997; 276(5314):955–60. [PubMed: 9139660]
 49. Zhou T, Parillon L, Li F, Wang Y, Keats J, Lamore S, Xu Q, Shakespeare W, Dalgarno D, Zhu X. Crystal structure of the T315I mutant of Abl kinase. *Chem Biol Drug Des*. 2007; 70(3):171–81. [PubMed: 17718712]
 50. Cowan-Jacob SW, Fendrich G, Floersheimer A, Furet P, Liebetanz J, Rummel G, Rheinberger P, Centeleghe M, Fabbro D, Manley PW. Structural biology contributions to the discovery of drugs to treat chronic myelogenous leukaemia. *Acta Crystallogr D Biol Crystallogr*. 2007; 63(Pt 1):80–93. [PubMed: 17164530]
 51. Nagar B, Bornmann WG, Pellicena P, Schindler T, Veach DR, Miller WT, Clarkson B, Kuriyan J. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res*. 2002; 62(15):4236–43. [PubMed: 12154025]
 52. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science*. 2000; 289(5486):1938–42. [PubMed: 10988075]
 53. Raaf J, Issinger OG, Niefind K. First inactive conformation of CK2 alpha, the catalytic subunit of protein kinase CK2. *J Mol Biol*. 2009; 386(5):1212–21. [PubMed: 19361447]
 54. Yamaguchi H, Kasa M, Amano M, Kaibuchi K, Hakoshima T. Molecular mechanism for the regulation of rho-kinase by dimerization and its inhibition by fasudil. *Structure*. 2006; 14(3):589–600. [PubMed: 16531242]
 55. Liu C, Lin J, Pitt S, Zhang RF, Sack JS, Kiefer SE, Kish K, Doweyko AM, Zhang H, Marathe PH, Trzaskos J, McKinnon M, Dodd JH, Barrish JC, Schieven GL, Leftheris K. Benzothiazole based inhibitors of p38alpha MAP kinase. *Bioorg Med Chem Lett*. 2008; 18(6):1874–9. [PubMed: 18296051]
 56. Mapelli M, Massimiliano L, Crovace C, Seeliger MA, Tsai LH, Meijer L, Musacchio A. Mechanism of CDK5/p25 binding by CDK inhibitors. *J Med Chem*. 2005; 48(3):671–9. [PubMed: 15689152]
 57. Gold MG, Barford D, Komander D. Lining the pockets of kinases and phosphatases. *Curr Opin Struct Biol*. 2006; 16(6):693–701. [PubMed: 17084073]
 58. Yang J, Cron P, Good VM, Thompson V, Hemmings BA, Barford D. Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. *Nat Struct Biol*. 2002; 9(12):940–4. [PubMed: 12434148]

59. Yang J, Cron P, Thompson V, Good VM, Hess D, Hemmings BA, Barford D. Molecular mechanism for the regulation of protein kinase B/Akt by hydrophobic motif phosphorylation. *Mol Cell*. 2002; 9(6):1227–40. [PubMed: 12086620]
60. Davies TG, Verdonk ML, Graham B, Saalau-Bethell S, Hamlett CC, McHardy T, Collins I, Garrett MD, Workman P, Woodhead SJ, Jhoti H, Barford D. A structural comparison of inhibitor binding to PKB, PKA and PKA-PKB chimera. *J Mol Biol*. 2007; 367(3):882–94. [PubMed: 17275837]

Figure 1.

(A) Typical location of the P-loop in a representative protein kinase structure (1A9U) is shown as anti-parallel magenta strands $\beta 1$ and $\beta 2$ connected by a magenta loop. Secondary structure elements which interact with ATP-competitive inhibitors are labeled. (B) The P-loop is magnified and individual C_{α} atoms are shown as magenta spheres. The residue nomenclature is shown, with positive and negative numbering relative to G0. Side-chains are shown as sticks. (C) The locations of water molecules (oxygen atoms) within 3 Å of the P-loop residues L-3 (orange spheres), G-2 (blue spheres) and V5 (cyan spheres) are shown for all 532 structures. The amino-acid backbone of each structure was superimposed on the backbone of the corresponding amino-acid of structure 1A9U. Ligand SB2 (co-crystallized in 1A9U) is shown with green sticks. The protein backbone and side chains of 1A9U are shown as magenta sticks. The side-chains of residues L-3 and V5 of other protein structures are shown as thin magenta sticks.

Figure 2. Number of structures containing at least one kinase domain submitted to RCSB Protein Data Bank (PDB) per year. The date of submission was derived from the PDB file HEADER records.

Figure 3.

(A). Residue conservation in the P-loop represented using sequence logos. The sequence logos were created using the sequences of 532 structures containing a complete P-loop in terms of backbone and side-chain atoms. The Y-axis for both panels is the residue position from panel A, which is determined relative to the central glycine position in the P-loop (G0). (B) Conservation of secondary structure, represented using sequence logos. The secondary structure was computed using *Stride*, where E, C and T represent the secondary structure type of the amino-acid as extended configuration of a β -sheet, coil, and turn, respectively.

Figure 4.

Interactions of the P-loop residues (L-3 to X7) of protein kinases with ligands. (A) Number of protein structures having at least one non-water atom within 3 Å of amino-acid residues. (B) Number of various atoms (carbon, C; nitrogen, N; oxygen, O; and halogens, F, Cl, Br and I) within 3 Å of protein residue. (C) Number of ligand-protein hydrogen bonds as a function of protein residue. (D) Number of structures with at least one water molecules within 3 Å of protein residue.

Figure 5. Backbone dihedral angles, ϕ and ψ (in $^{\circ}$), computed using *stride* for the three residues (L-3, G-2 and V5) which make contact with inhibitors in the most structures in the dataset. Data is shown only for 485 proteins with >2 structures in the original dataset of 755 structures.

Figure 6.

(A) Usually observed fully-extended conformation (PDBs: ABL1/ABL1_HUMAN, 2hz0; JNK3/MK10_HUMAN, 1pmn; p38 α /MK14_HUMAN, 1w82) and (B) Rarer distorted conformation of the P-loop, rendered as ribbons (PDBs: ABL1/ABL1_HUMAN: 2hzi, 2f4j, 1opl, 3cs9, 2hiw, 2e2b, 2hyy; ABL1/ABL1_MOUSE: 1m52, 1opk, 3dk3, 2qoh, 1iep and 1fpu; p38 α /MK14_HUMAN: 1w84, 2i0h, 3d7z, 1w7h, 3bx5, 1wbw, 1zzl, 3e92, 3c5u, 1yqj, 3ctq; ABL2/ABL2_HUMAN: 3gvu and 3hmi; AurA/STK6_MOUSE: 3dj5 and 3dj6; FGFR1/FGFR1_HUMAN: 3c4f; CK2 α 1/CSK21_HUMAN: 2zjw; JNK3/MK10_HUMAN: 2r9s). The following color coding was used for individual proteins: green, ABL1; blue purple, JNK3; spring green, p38 α ; blue, ABL2; orange, AurA; magenta, FGFR1 and pink, CK2 α 1. This snapshot was created using the *maestro* module of the Schrödinger software suite.

Figure 7. Conformational flexibility of amino-acid side chains represented by side chain dihedral angles χ_1 , χ_2 and χ_3 for two residues (A) L-3 and (B) V5. The side chain dihedral angles were calculated using the *dangle* program.

Figure 8.

Solvent accessible surface area (SASA) calculated using *NAccess* for the different protein kinase groups for three important P-loop residues making significant contacts with inhibitors (L-3, G-2 and V5). The single SASA value plotted for each protein represents the average over all the structures in the database for that particular protein; the error bars show the standard deviation in SASA for all the structures for that protein.

Figure 9. Representative example of C-terminal extension (rendered as black ribbons) of AGC-family members, 2c1b, PKaCA (*left panel*); 2jdo, AKT2 (*middle panel*) and 3d9v, ROCK1 (*right panel*) interacting with the L-3 residue (rendered as magenta spheres) of the P-loop. The P-loop is rendered as orange ribbons. This interaction is responsible for reduction in solvent accessible surface area of residue L-2 for the crystal structures of these proteins.

Table 1

Total number of protein kinase structures and number of unique protein kinases represented in those structures from different organisms found in our database.

Organism	Number of protein kinase-ligand structures	
	Total	Unique
Human	607	101
Cow	49	2
Mouse	31	5
Chicken	23	2
Maize	16	1
Zebra fish	6	1
Rat	5	2
<i>Cryptosporidium parvum</i>	4	4
<i>Xenopus</i>	4	2
<i>Mycobacterium</i>	3	2
<i>Plasmodium falciparum</i>	4	2
<i>Schizosaccharomyces pombe</i>	2	1
<i>Saccharomyces cerevisiae</i>	1	1
Total	755	126

Table 2

The ten protein kinases having the highest number of structures in our database of 755 ligand–kinase complexes.

Protein kinase (UniProt ID)	Number of structures
CDK2 (CDK2_ HUMAN)	134
p38 α (MK14_ HUMAN)	57
PKAC α (KAPCA_BOVIN)	47
CHK1 (CHK1_ HUMAN)	35
PIM1 (PIM1_ HUMAN)	25
JNK3 (MK10_ HUMAN)	23
KDR (VGFR2_ HUMAN)	19
SRC (SRC_CHICK)	18
LCK (LCK_HUMAN)	16
AurA (STK6_HUMAN)	16

Table 3

Family coverage of human protein kinase–inhibitor space in our database.

Family	Number of protein kinase–ligand structures		Protein kinases
	Total	Unique	
CMGC	256	15	CDK2, p38 α , JNK3, GSK3B, Erk2, JNK1, CK2 α 1 ^a , CDK6, CDK5, CLK1, p38 β , JNK2, DYRK1A, CK2 α 2 ^a , CDK9
TK	153	35	KDR, LCK, MET, EGFR, ABL1, EphB4, SRC, SYK, HCK, JAK2, PYK2, INSR, IGF1R, FGFR1, FMS, TIE2, RET, ACK, KIT, JAK1, FAK, ErbB4, ABL2, ZAP70, FLT1, MER, JAK3, ITK, FYN, FGFR2, FES, EphA7, EphA3, CSK, MET
TKL	15	5	TGF β R1, BRAF, IRAK4, MLK1, ALK2
AGC	34	10	PDK1, AKT2, ROCK1, SGK1, RSK1, PKC ϵ , AKT1, PKC δ , PKC ζ , DMPK1
CAMK	84	16	CHK1, PIM1, MAPKAPK2, DAPK3, DAPK1, CaMK2d, CHK2, PIM2, MNK2, CaMK4, CaMK2b, CaMK2a, CaMK1g, CaMK1d, DAPK2, CaMK2g
STE	22	11	MAP2K1, SLK, PAK1, MST3, LOK, PAK5, PAK4, MST4, MAP2K6, MAP2K2, MAP3K5
CK1	6	2	CK1g3, CK1g1
Other	35	7	AurA, Wee1, TTK, PLK1, Haspin, MPSK1, NEK2

^a Ambiguous between CMGC and Other family.