KLIFS: a structural kinase-ligand interaction database

Albert J. Kooistra, Georgi K. Kanev, Oscar P.J. van Linden, Rob Leurs, Iwan J.P. de Esch and Chris de Graaf*

Division of Medicinal Chemistry, Amsterdam Institute for Molecules, Medicines and Systems (AIMMS), Vrije Universiteit Amsterdam, Amsterdam, 1081 HV, The Netherlands

ABSTRACT

Protein kinases play a crucial role in cell signaling and are important drug targets in several therapeutic areas. The KLIFS database contains detailed structural kinase-ligand interaction information derived from all (>2900) structures of catalytic domains of human and mouse protein kinases deposited in the Protein Data Bank in order to provide insights into the structural determinants of kinase-ligand binding and selectivity. The kinase structures have been processed in a consistent manner by systematically analyzing the structural features and molecular interaction fingerprints (IFPs) of a predefined set of 85 binding site residues with bound ligands. KLIFS has been completely rebuilt and extended (>65% more structures) since its first release as a data set, including: novel automated annotation methods for (i) the assessment of ligand-targeted subpockets and the analysis of (ii) DFG and (iii) αC-helix conformations; improved and automated protocols for (iv) the generation of sequence/structure alignments, (v) the curation of ligand atom and bond typing for accurate IFP analysis and (vi) weekly database updates. KLIFS is now accessible via a website (http://klifs.vu-compmedchem.nl) that provides a comprehensive visual presentation of different types of chemical, biological and structural chemogenomics data, and allows the user to easily access, compare, search and download the data.

INTRODUCTION

Protein kinases are enzymes that modulate the biological activity and expression of proteins by catalyzing the phosphorylation of serine, threonine or tyrosine residues. The 518 human protein kinases constitute one of the largest protein families encoded within the human genome and play essential roles in the majority of cell signal transduction pathways (1). Kinases have therefore become important drug targets for pharmaceutical intervention in several therapeutic areas, including oncology, immunology, neurology, cardiology and infectious diseases (2,3). The catalytic domains of kinases share a conserved structure, which poses a challenge for the development of small molecule drugs that can selectively target a well-defined set of kinases in order to obtain the desired (poly)pharmacological effects (4). Currently (August 12, 2015), 2899 structures of human and mouse catalytic kinase domains have been experimentally determined (2892 X-ray, 5 NMR and 2 EM structures; see Supplementary Table S1) that can offer insights into the structural determinants of kinase-ligand interaction and selectivity. We have collected, processed, annotated and analyzed all available structural kinase-ligand interaction information in a single, enriched and searchable web resource, KLIFS—Kinase-Ligand Interaction Fingerprints and Structures—database, to enable systematic comparison and analysis of the chemical and structural features of all available experimentally-determined protein kinase structures and their small molecule ligands (Figure 1).

The initial version of KLIFS (5) was created based on a systematic analysis of all human kinase domain structures that were available in the Protein Data Bank (PDB) (6,7) at that point in time (1734 in total—August 9, 2012). The foundation of KLIFS was the definition of a consistent binding site encompassing 85 pocket residues that interacted with any bound kinase inhibitor within the catalytic front cleft, gate area and/or back cleft (type I, I$\frac{1}{2}$, II and III (8)) to allow for the systematic analysis of kinase-ligand interaction fingerprints (IFPs) (9) with different residues in the kinase binding site (Figure 2A) (5) in order to identify kinase (family) specific interaction features and classify ligands according to their binding modes. Accompanying this binding site definition was the introduction of a numbering scheme, in which each binding site residue is labeled according to the following scheme: [one letter amino acid][kinase region].[binding site residue number], e.g. the aspartic acid of the xDFG motif is labeled $D^{xDFG.81}$ (Figure 2A). This numbering scheme is also used throughout this article.

Currently the amount of available human kinase structures has grown with 982 new structures to a total of 2738 in 3 years (August 12, 2015), covering 234 different kinases from all eight main kinase groups (AGC, CAMK, CK1,

*To whom correspondence should be addressed. Tel: +31 20 5987553; Fax: +31 20 5987610; Email: c.de.graaf@vu.nl
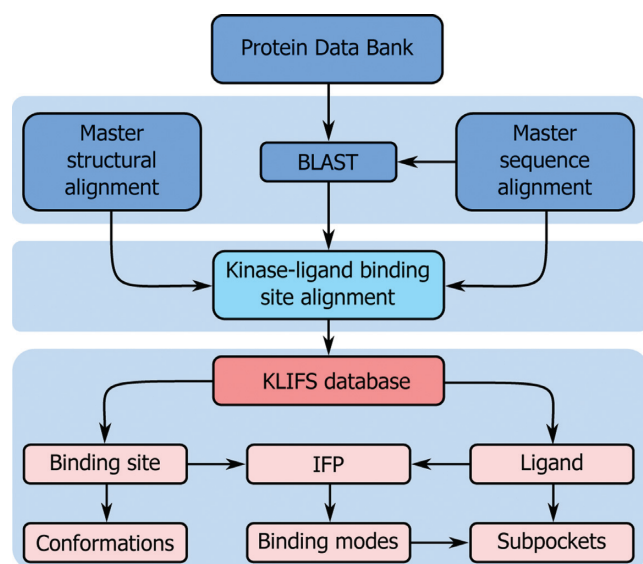
Figure 1. The data collection, processing, annotation and analysis workflow of KLIFS.

CMGC, STE, TK, TKL and other (1)), and including 1740 unique kinase inhibitors. This rapid increase emphasizes the need for KLIFS as a protein-family specific database that integrates different types of structural and chemical data. Since the first release of KLIFS we have continued our work with the aim to make our database up-to-date, and easily accessible for scientists with different backgrounds. This has resulted in a significantly extended database with new types of structural kinase-ligand interaction data, which is now publicly accessible and actively searchable via a user-friendly web interface (http://klifs.vu-compmedchem.nl). Moreover, automated weekly updates keep the database in sync with the Protein Data Bank. This new web-based version of KLIFS provides a clear and comprehensive visual presentation of different types of chemical, biological and structural chemogenomics data, and allows the user to easily access, compare and search the data to improve our understanding of the structural requirements of kinase-ligand interaction.

CREATION OF KLIFS

Data collection and preparation

All sequences from human (490) and mouse (525) eukaryotic protein kinases, the two most prevalent species in published kinase structures, are used as input for a BLAST query (10) against all structures released by the PDB (Figure 1). BLAST hits ($\geq 70\%$ sequence identity and an *e*-value ≤ 0.001) are selected for preparation and further processing. First the structures are prepared by splitting each structure into separate (alternate) models, if present, and individual chains after which each monomer is converted to atom and bond type annotated Tripos MOL2 format (to facilitate IFP analysis). Then the data in the Chemical Component Dictionary (11) are used to identify and correct errors in atom and bond types (Figure 2A) to allow for proper protonation of the molecules and accurate analysis of protein-

ligand interactions, including H-bond interaction networks (12). Manual curation of the ligand topologies resulted in additional corrections (examples shown in Supplementary Figure S1).

Data processing

A master sequence alignment (MSeq) of all human and mouse catalytic domains of eukaryotic protein kinases and a master structural alignment (MStruc) comprising three structures from each kinase group (Supplementary Table S2) were created with a focus on the binding pocket (5). The prepared monomers based on the BLAST search are subsequently processed using MOE (Chemical Computing Group Inc.) by performing a sequence alignment to the MSeq (Figure 1). The resulting sequence alignment is used to superpose each structure to the MStruc (Figure 1) based on residues of the catalytic loop, the DFG motif and the hinge region. This process gives a robust structural alignment of kinase binding sites as indicated by the low RMSD of 0.8 ± 0.1 Å for the superposing residues and an RSMD of 2.2 ± 0.2 Å for the full binding pocket (as defined by the 85 pocket residues) on average over all 2899 currently processed structures. Afterward, all individual elements (final sequence alignment, full complex, protein, pocket, orthosteric ligands (here defined as all ligands binding to the catalytic front cleft, gate area and/or back cleft), allosteric ligands (here defined as all ligands not fitting the orthosteric classification), waters, ions, cofactors, organometallics and alternative amino acids) are extracted and stored in MOL2 format.

Data annotation and analysis

The extracted structure elements are used to analyze and annotate each structure (Figure 1). One of the key elements of KLIFS is the annotation of kinase-ligand interactions, which is performed by calculating the IFPs using the FingerPrintLib developed by Marcou and Rognan (9). An IFP (Figure 2A) encodes seven different interactions types (hydrophobic contact, aromatic face-to-face, aromatic edge-to-face, H-bond donor-acceptor, H-bond acceptor-donor, ionic positive-negative, ionic negative-positive) between each of the pocket residues and the ligand in a bit string as either present (1) or absent (0). If a structure has missing residues or gaps in the pocket, the IFP is corrected by inserting seven zeroes at the respective position to allow systematic comparison of all IFPs within KLIFS.

Based on the manual annotation of the binding modes of the ligands in the previous version of KLIFS with respect to specific (sub) pockets (5,13) we were able to train an automated method to identify which pockets are occupied by each bound ligand. This method annotates whether or not a ligand addresses any of the 3 major pockets (front cleft, gate area and back cleft) and the 12 subpockets (front pockets I and II, back pockets I-A, I-B, II-in, II-A-in, II-B-in, II-out, II-B, III, IV and V) by generating clusters of probes that accurately describe these pockets (Figure 2B). If a ligand contacts any of the probes for a specific (sub) pocket, the binding mode for that ligand is annotated as such (Figure 2B). Moreover, by analyzing all water molecules interacting with
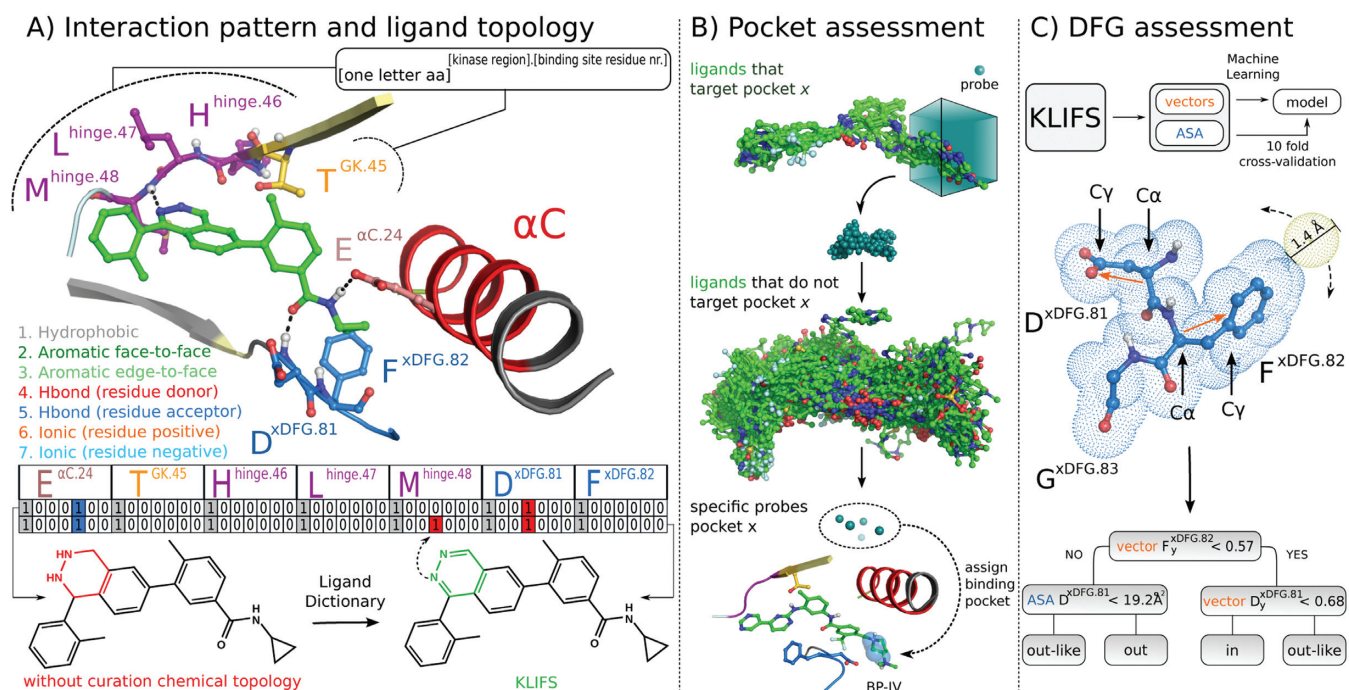
Figure 2. Annotation structural kinase-ligand interaction data in KLIFS. (**A**) Consistent structure-based kinase binding site residue annotation (top) and careful curation of the chemical topology and protonation of kinase ligands (bottom) enable the systematic analysis of kinase-ligand IFPs of kinase-ligand complexes (middle), illustrated for 7 of the 85 binding site residues of a phthalazine inhibitor (PDB ligand identifier: A17) bound p38a structure (PDB: 3DS6). (**B**) Automated assessment of ligand-targeted subpockets in protein kinase binding sites. Spatial probes (0.5 Å grid spacing) are: (i) placed around the conformations of kinase-bound ligands that are superposed according to the structural alignment of the corresponding binding sites; (ii) scored according to the ratio of close contacts (<1.0 Å) with training sets of ligands that bind a specific subpocket and ligands that do *not* bind this subpocket; (iii) highest ranked probes are selected as descriptors to assign subpockets for new kinase-ligand complexes. (**C**) A 10-fold cross-validated decision tree model derived from systematic analysis of the Accessible Surface Area (ASA) (32) of kinase binding site residues and vectors defining the directionality of $D^{xDFG.81}$ and $F^{xDFG.82}$ side chains enable the automatic assessment of the DFG conformation of kinase structures with an accuracy > 99%. The decision tree algorithm identified the ASA values of $D^{xDFG.81}$ and y-values of $D^{xDFG.81}$ and $F^{xDFG.82}$ vectors (obtained by subtracting the coordinates of Cγ by Cα) as the most predictive descriptor combination for the discrimination of DFG conformations.

a ligand we were able to define 13 conserved water clusters (I1–I11 and 2 specific DFG-out clusters O1 and O2; Supplementary Figure S2). For each structure all waters within any of the clusters are extracted and their H-bond pattern with the ligand and/or protein is evaluated using IFP.

The conformation of the DFG-motif (side chain and backbone movement) determines if a kinase is in the active (DFG-in) or inactive (DFG-out) state or somewhere in between (DFG-outlike) (13,14). Based on the manually curated first version of KLIFS, decision tree machine learning (15) was applied and resulted in a rule-based method to automatically annotate the conformation (Figure 2C). The position of the αC-helix can also be classified into αC-in, αC-outlike or αC-out based on the distance between the Cα atoms of $F^{xDFG.82}$ and $E^{αC.24}$ (5,13). The conformation of the G-rich loop is annotated based on the angle of this loop compared to hinge region and catalytic loop, the distance between the catalytic loop and the G-rich loop, as well as the rotation of the G-rich loop compared to the catalytic loop (Supplementary Figure S3). Furthermore, all missing residue atoms in the pocket are logged and the pocket residues are checked for mutations by comparing the residues to the sequence in the Uniprot database (16). Finally, the quality of the structure with respect to the binding pocket as well as the quality of the processing by KLIFS

is estimated with a so-called quality score that runs from 0 (bad) to 10 (flawless) and combines the resulting RMSDs relative to the MStruc with the number of missing residues and missing atoms (Supplementary Figure S4).

THE KLIFS WEBSITE

In order to allow researchers to easily browse, search, visualize and download data from KLIFS we have built a user-friendly and feature-rich web interface with a clear organization that gives access to all underlying data.
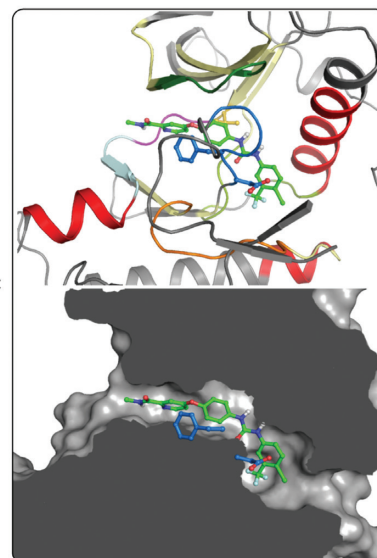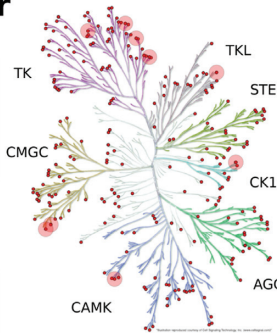
Organization

The web interface contains six main pages: *Browse*, *Search*, *Data charts*, *Favorites*, *Tutorial* and *FAQ*. *Browse* allows users to go through a complete overview of all structures. *Search* gives access to multiple search options, which can be used to focus on the most relevant structures. A detailed overview page is provided for each structure and can be accessed when browsing or searching KLIFS (Figure 3). Via *Data charts* users can compare different structure attributes and visualize them in histograms. Via *Favorites* one can keep track of specific structures throughout the website and it also enables the comparison of ligand binding modes via the integrated WebGL 3D-viewer (iview (17)). Finally,

★ Kinase insert domain receptor

Kinase structure

Kinase:	KDR
Family:	VEGFR
Group:	TK
Species:	HUMAN
PDB:	3WZE
sc-PDB:	Search
KIDFamMap:	Search
Release date:	2015-05-27
PubMed:	25589937
Chain:	A
(Alternate) Model:	A
Orthosteric ligand:	BAX

Kinases: 237
Families: 84
Groups: 8
Species: 2
PDB entries: 2899

Chains: 4235
Alt. models: 3400
Monomers: 5972
Unique ligands: 1788

Structural information

DFG conformation:	out
αC-helix conformation:	out
G-rich loop angle:	58°
G-rich loop rotation:	38.3°
Quality Score:	7.6
Resolution:	1.9 Å
Missing Residues:	0
Missing Atoms:	0

DFG: in, out, out-like
αC: in, out, out-like

Quality Score: 0-10
Resolution: 0-15

Render options:
- Ligand
- Pocket
- Protein
- Waters
- Protein surface
- Save image
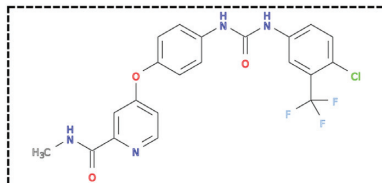
Compound name,
Structure:
- Identity
- Substructure
- Similarity
- Dissimilarity

Ligand binding mode

Pockets	Subpockets
front	BP-I-B
gate	BP-II-out
back	BP-III

Waters

Cluster	Ligand	Protein
I1	No	Yes
I9	No	Yes

Main pockets:
Front cleft, Gate area, Back cleft

Subpockets:
FP-I, FP-II, BP-I-A, BP-I-B, BP-II-in,
BP-II-A-in, BP-II-B-in, BP-II-out,
BP-II-B, BP-III, BP-IV, BP-V

Pocket waters:
I1 to I11, O1 and O2

Ligand affinity

ChEMBL ID: CHEMBL1336   **Bioaffinities:** 133 records

Species ▼	Kinase ▼	Median ↕	Min ↕	Max ↕	Type ↕	Records ↕
Human	BRAF	7.1	5.1	7.8	pIC50	13
Human	ABL1	6.8	6.8	6.8	pKi	1
⋮						
Human	VEGF2	7.7	7.4	10.7	pKi	4
Human	p38α	6.4	5.5	7.4	pIC50	6

Pocket alignment

Uniprot sequence: KPLGRGAFGQVIEVAVKMLALMSELKILIHIGLNVVNLLGAMVIVEFCKFGNLSTYLRSFLASRKCIHRDLAARNILLICDFGLA
Sequence structure: KPLGRGAFGQVIEVAVKMLALMSELKILIHIGLNVVNLLGAMVIVEFCKFGNLSTYLRSFLASRKCIHRDLAARNILLICDFGLA

Kinase-ligand interaction pattern

1-85 positions: amino acid and/or interaction type
(Any interaction, H-bond, Aromatic, Ionic, No interaction)

• Hydrophobic • Aromatic face-to-face • Aromatic face-to-edge • H-bond donor • H-bond acceptor • Ionic positive • Ionic negative

Figure 3. KLIFS entry example (sorafenib (BAX) bound KDR, PDB: 3WZE), schematic overview of KLIFS website search options (gray boxes) and 3D kinase-ligand binding mode visualization (overlay sorafenib gray carbon atoms, ligand from search example PDB: 2BAK magenta carbon atoms). All kinases with an entry in KLIFS are indicated in the interactive human kinome (red dots), and search results can also be highlighted. Top right panels: sorafenib binding mode views showing (color-coded) secondary structure elements, sliced binding site surface and user-customizable 3D-rendering settings (blue box). Further details regarding structural information, ligand search/binding mode/affinity, pocket alignment/kinase-ligand interaction pattern browse/search options are explained in the text.

Tutorial provides an overview of the website and several usage examples, while *FAQ* contains answers to (anticipated) frequently asked questions.

Features of the website

Searching. There are several search options embedded within the KLIFS website that can be used simultaneously to create a highly versatile way of focusing the available data (Figure 3). On the home page the latest entries are listed and the interactive HTML5 human kinome (based on Kinome Render (18)) can be used to search for specific kinases. From any page the *PDB search* option is available that can be used to quickly find a specific structure or ligand using their PDB-codes (e.g. structure 2XYU or ligand Q9G). The options on the main search page have been divided into nine categories: (i) ligand structure, (ii) kinase classification, (iii) structural annotation, (iv) binding mode, (v) pocket waters, (vi) structure properties, (vii) ligand properties, (viii) binding pocket composition (sequence) and (ix) ligand-kinase interaction pattern (Figure 3). The *ligand structure* search option allows the users to provide the molecular SMILES, draw a chemical structure or retrieve a molecular structure based on the name (via the HTML5 Molecule editor, MolSoft LLC). This ligand structure can subsequently be used to search for structures containing this exact ligand (or tautomers thereof (19)), a (dis)similar ligand (based on Morgan (20) or MACCS (MDL Information Systems, Inc.) fingerprint similarity) or a substructure (using RDKit: Open-source chemoinformatics; http://www.rdkit.org). The kinase classification search option allows for the selection of specific kinase families, kinase groups or organisms. The structural annotation options allow filtering based on DFG and αC-helix conformation, and whether the structure is ligand-bound or not. Via the binding mode selection options the user can carefully define whether the ligand does or does not bind to a specific (combination of) subpocket(s) in the protein kinase binding site (Figures 2B and 3). Moreover, one can search for the presence or absence of water in one or more of the 13 water clusters and their H-bonding pattern. The structure properties give the user the possibility to select requirements for the G-rich loop conformation, quality score and the resolution of the kinase structure. The ligand properties search allows for the manual selection of physicochemical properties filter or the selection of one of the 4 presets: the rule-of-five, the rule-of-three, fragment-like and lead-like (21,22). The binding pocket composition feature provides the option to select specific amino acid at specific residue positions in the pocket (e.g. a threonine (T) on position 45, i.e. the gatekeeper, e.g. see Figure 2A). Finally, the interaction pattern search option gives access to the ligand-kinase interactions and enables the user to select which interactions the bound ligand should have with specific binding pocket residues. Additionally, all search results can be projected onto the interactive human kinome.

Viewing. Each structure in KLIFS has a detailed information page that presents all information within the KLIFS database regarding this specific structure (Figure 3). First all information regarding the kinase type (names as defined by the HUGO Gene Nomenclature Committee (23) and

Uniprot (16)), the (crystal) structure, and (if present) the orthosteric and/or allosteric ligand is presented including a 2D chemical representation, images of the binding pocket (created with Pymol, Schrödinger LLC.) and a link to the integrated 3D-viewer (17). Second, all structural information is presented: DFG, αC-helix and G-rich loop conformation, quality score, resolution, missing residues and atoms. Third, if a ligand is present in the binding site all targeted (sub) pockets are listed as well as all pocket waters within any of the 13 water clusters. Fourth, kinase records in ChEMBL (K_i or IC_{50}, confidence≥8) for the bound ligand are listed if present. Fifth, the sequence of binding site of the structure is compared to the sequence from Uniprot. And finally, the interaction profile of the ligand with the binding site is shown in which each interaction with each of the 85 binding site residues is highlighted. Moreover, the interaction profile of the selected structure can be used to find other kinase-ligand complexes with a similar interaction profile. To easily allow access to additional relevant information, links have been integrated to the following external databases: ChEMBL (24), PDB (6,7), sc-PDB (25), KIDFamMap (26), Uniprot (16) and PubMed (NCBI).

Downloading. The results of each search query can be exported to an Excel or a CSV file containing all details as listed in the search results overview. Moreover, via the 'Download Structures' option all annotated and structural data can be downloaded as a MOE database or as a compressed package comprising all processed MOL2 files and a CSV detailing all annotation data. This enables researchers to use their software of choice for the comparison and analysis of the structures.

Examples of search queries

Design of selective BTK inhibitors. Shi et al. (27) synthesized a purine derivative (PDB: 4NWM) with (high) affinity for BTK, but also for JAK2, Flt2 and ITK. Subsequently, the authors designed new analogues with the aim to target the gatekeeper region in order to obtain BTK selectivity as BTK has a relatively small T gatekeeper compared to the $M/F^{GK.45}$ of the off-target kinases. The most selective compound was obtained by attaching a 2-phenylethyl group to the 9 position of the purine. A search of the current KLIFS database for a $T^{GK.45}$ and a ligand targeting both the front cleft and gate area, but *not* the back cleft yielded 98 structures that can be used to design ligands with similar (BTK selective) interaction features. Limiting the search to a purine scaffold resulted in only 2 entries (PDB: 2BDJ, 4GK2), of which 2BDJ contained a bound ligand with a 2-phenolethyl moiety on the 9 position of the purine, allowing for an additional H-bond with $D^{xDFG.81}$ (see ligand overlay in Supplementary Figure S5).

Water-mediated H-bonding and pocket composition. Levinson and Boxer (28) identified a strong correlation between conserved water molecules, cavity-lining residues and the selectivity profile of bosutinib. The presence of primarily water W1 (cluster I5 in KLIFS), but also water W2 (cluster I4 in KLIFS) allowed for a water-mediated H-bond between bosutinib and the DFG-motif. A KLIFS

search for both waters I4 and I5 yielded 648 structures (92 different kinases). A more detailed search for the crucial I5 water and the key binding site residues $T^{GK.45}$ (bulky amino acids prevent water-mediated H-bonding), $V^{b.I.36}$ (a leucine substitution reduces affinity) and $A^{xDFG.80}$ (larger amino acids obstruct access to water I5) resulted in only 6 unique kinases (29 structures), which all have (a predominantly high) affinity for bosutinib.

Diverse molecules with similar interaction patterns. Finding dissimilar molecules with similar interaction patterns can be interesting from a drug design point of view (e.g. scaffold hopping) (29,30). Starting from a sorafenib-KDR complex (PDB: 3WZE, Figure 3) an interaction similarity search was performed (similarity ≥ 0.75) resulting in 51 entries (44 unique ligands in 17 different kinases). Further refinement by additionally performing a dissimilarity search to sorafenib (Morgan similarity ≤ 0.25) limited the result to 21 structures. One of the identified structures (PDB: 2BAK, IFP similarity = 0.77) was of p38a in complex with an anilinoquinazoline inhibitor (see integrated 3D viewer overlay in Figure 3), which has a comparable affinity for p38a as sorafenib (31).

CONCLUSIONS AND OUTLOOK

With this new version KLIFS has grown from a downloadable data set to a fully featured database with a user-friendly, yet versatile, web interface. On the back end the sequence alignment, structure superposing, structure processing and kinase-ligand interaction analysis protocols have been improved and new methods have been developed for the automated annotation of subpocket-targeting by ligands, DFG and αC-helix conformations and the curation of ligand topologies for accurate kinase-ligand interaction analysis. Moreover, KLIFS will be updated weekly in sync with the Protein Data Bank to keep up with the continuous growth in published catalytic kinase domain structures (averaging 380 new structures per year over the past three years). The new front end, the web interface, provides easy access to all data and has additional features like (combined) search options based on kinase-ligand interactions, binding site sequence and ligand features, as well as (3D) viewing and rendering options to interactively analyze and compare the structures of kinase-ligand complexes.

We believe that this database will prove to be a valuable resource of structural, molecular, biological and chemogenomics kinase data. For example, computational chemists and bioinformaticians can use this database for advanced and extensive structural and sequence analyses, modeling and (structure-based) machine learning approaches, structural biologists can compare kinase structures and validate their own structures, pharmacologists can related their data (e.g. mutation data) to the available structures, and medicinal chemists can use this database to gain new structural insights and for the rational design of novel (selective) kinase inhibitors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Manning,G., Whyte,D.B., Martinez,R., Hunter,T. and Sudarsanam,S. (2002) The protein kinase complement of the human genome. *Science*, **298**, 1912–1934.
2. Cohen,P. (2002) Protein kinases-the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.*, **1**, 309–315.
3. Merritt,C., Silva,L.E., Tanner,A.L., Stuart,K. and Pollastri,M.P. (2014) Kinases as Druggable Targets in Trypanosomatid Protozoan Parasites. *Chem. Rev.*, **114**, 11280–11304.
4. Apsel,B., Blair,J.A., Gonzalez,B., Nazif,T.M., Feldman,M.E., Aizenstein,B., Hoffman,R., Williams,R.L., Shokat,K.M. and Knight,Z.A. (2008) Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat. Chem. Biol.*, **4**, 691–699.
5. van Linden,O.P., Kooistra,A.J., Leurs,R., de Esch,I.J. and de Graaf,C. (2014) KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *J. Med. Chem.*, **57**, 249–277.
6. Rose,P.W., Prlic,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **43**, D345–D356.
7. Gutmanas,A., Alhroub,Y., Battle,G.M., Berrisford,J.M., Bochet,E., Conroy,M.J., Dana,J.M., Fernandez Montecelo,M.A., van Ginkel,G., Gore,S.P. *et al.* (2014) PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, **42**, D285–D291.
8. Lamba,V. and Ghosh,I. (2012) New directions in targeting protein kinases: focusing upon true allosteric and bivalent inhibitors. *Curr. Pharm. Des.*, **18**, 2936–2945.
9. Marcou,G. and Rognan,D. (2007) Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.*, **47**, 195–207.
10. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
11. Westbrook,J.D., Shao,C., Feng,Z., Zhuravleva,M., Velankar,S. and Young,J. (2015) The chemical component dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank. *Bioinformatics*, **31**, 1274–1278.
12. Bietz,S., Urbaczek,S., Schulz,B. and Rarey,M. (2014) Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J. Cheminform.*, **6**, 12.
13. Liao,J.J.L. and Andrews,R.C. (2007) Targeting protein multiple conformations: a structure-based strategy for kinase drug design. *Curr. Top. Med. Chem.*, **7**, 1394–1407.
14. Zhao,Z., Wu,H., Wang,L., Liu,Y., Knapp,S., Liu,Q. and Gray,N.S. (2014) Exploration of type II binding mode: A privileged approach for kinase inhibitor focused drug discovery? *ACS Chem. Biol.*, **9**, 1230–1241.
15. Quinlan,J.R. (1986) Induction of decision trees. *Mach. Learn.*, **1**, 81–106.

16. The UniProt Consortium. (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.
17. Li,H., Leung,K.S., Nakane,T. and Wong,M.H. (2014) iview: an interactive WebGL visualizer for protein-ligand complex. *BMC Bioinformatics*, **15**, 56.
18. Chartier,M., Chenard,T., Barker,J. and Najmanovich,R. (2013) Kinome Render: a stand-alone and web-accessible tool to annotate the human protein kinome tree. *PeerJ*, **1**, e126.
19. Kochev,N.T., Paskaleva,V.H. and Jeliazkova,N. (2013) Ambit - Tautomer: An Open Source Tool for Tautomer Generation. *Mol. Inf.*, **32**, 481–504.
20. Rogers,D. and Hahn,M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50**, 742–754.
21. van Deursen,R., Blum,L.C. and Reymond,J.L. (2011) Visualisation of the chemical space of fragments, lead-like and drug-like molecules in PubChem. *J. Comput. Aided Mol. Des.*, **25**, 649–662.
22. de Graaf,C., Vischer,H.F., de Kloe,G.E., Kooistra,A.J., Nijmeijer,S., Kuijer,M., Verheij,M.H., England,P.J., van Muijlwijk-Koezen,J.E., Leurs,R. *et al.* (2013) Small and colorful stones make beautiful mosaics: fragment-based chemogenomics. *Drug Discov. Today*, **18**, 323–330.
23. Gray,K.A., Yates,B., Seal,R.L., Wright,M.W. and Bruford,E.A. (2015) Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res.*, **43**, D1079–D1085.
24. Davies,M., Nowotka,M., Papadatos,G., Dedman,N., Gaulton,A., Atkinson,F., Bellis,L. and Overington,J.P. (2015) ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res.*, **43**, W612–W620.
25. Desaphy,J., Bret,G., Rognan,D. and Kellenberger,E. (2015) sc-PDB: a 3D-database of ligandable binding sites–10 years on. *Nucleic Acids Res.*, **43**, D399–D404.
26. Chiu,Y.Y., Lin,C.T., Huang,J.W., Hsu,K.C., Tseng,J.H., You,S.R. and Yang,J.M. (2013) KIDFamMap: a database of kinase-inhibitor-disease family maps for kinase inhibitor selectivity and binding mechanisms. *Nucleic Acids Res.*, **41**, D430–D440.
27. Shi,Q., Tebben,A., Dyckman,A.J., Li,H., Liu,C., Lin,J., Spergel,S., Burke,J.R., McIntyre,K.W., Olini,G.C. *et al.* (2014) Purine derivatives as potent Bruton's tyrosine kinase (BTK) inhibitors for autoimmune diseases. *Bioorg. Med. Chem. Lett.*, **24**, 2206–2211.
28. Levinson,N.M. and Boxer,S.G. (2014) A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity. *Nat. Chem. Biol.*, **10**, 127–132.
29. Furtmann,N., Hu,Y. and Bajorath,J. (2015) Comprehensive analysis of three-dimensional activity cliffs formed by kinase inhibitors with different binding modes and cliff mapping of structural analogues. *J. Med. Chem.*, **58**, 252–264.
30. Hu,Y., Furtmann,N. and Bajorath,J. (2015) Current compound coverage of the kinome. *J. Med. Chem.*, **58**, 30–40.
31. Mendez-Lucio,O., Kooistra,A.J., de Graaf,C., Bender,A. and Medina-Franco,J.L. (2015) Analyzing multitarget activity landscapes using protein-ligand interaction fingerprints: interaction cliffs. *J. Chem. Inf. Model.*, **55**, 251–262.
32. Shrake,A. and Rupley,J.A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, **79**, 351–371.