BMC Bioinformatics

HKPocket: human kinase pocket database for drug design

Huiwen Wang, Jiadi Qiu, Haoquan Liu, Ying Xu, Ya Jia and Yunjie Zhao[*]

Abstract

Background: The kinase pocket structural information is important for drug discovery targeting cancer or other diseases. Although some kinase sequence, structure or drug databases have been developed, the databases cannot be directly used in the kinase drug study. Therefore, a comprehensive database of human kinase protein pockets is urgently needed to be developed.

Results: Here, we have developed HKPocket, a comprehensive Human Kinase Pocket database. This database provides sequence, structure, hydrophilic-hydrophobic, critical interactions, and druggability information including 1717 pockets from 255 kinases. We further divided these pockets into 91 pocket clusters using structural and position features in each kinase group. The pocket structural information would be useful for preliminary drug screening. Then, the potential drugs can be further selected and optimized by analyzing the sequence conservation, critical interactions, and hydrophobicity of identified drug pockets. HKPocket also provides online visualization and pse files of all identified pockets.

Conclusion: The HKPocket database would be helpful for drug screening and optimization. Besides, drugs targeting the non-catalytic pockets would cause fewer side effects. HKPocket is available at http://zhaoserver.com.cn/HKPocket/HKPocket.html.

Keywords: Pocket database, Human kinase proteins, Drug discovery, Side effects

Background

Kinase proteins are considered as one of the most attractive drug targets for drug discovery targeting cancer, chronic neurodegenerative or other diseases [1–4]. Previous studies have highlighted two major strategies targeting kinases: ATP-binding inhibitors (type I and II) and non-ATP inhibitors (type III and IV) [3, 5]. Currently, most developed drugs are ATP-competitive inhibitors [6, 7]. Andrea et al. performed a systematic analysis of catalytic ATP-binding pockets. Their results showed that ATP-binding pockets are highly conserved [8]. Therefore, the ATP-competitive drugs may inhibit most of the kinase proteins and cause side effects, such as hypertension, hand-foot skin reaction and acute renal failure [9–11]. Type III and type IV inhibitors are usually very selective and have fewer side effects because their targeted binding sites are usually unique to a particular kinase [3, 5, 12]. Thus, there is an urgent need to

develop new drugs targeting non-catalytic pockets to reduce side effects.

Computer-aided drug design is widely used in drug development to shorten the time and reduce the cost of experiments [13–22]. There are several existing kinase databases with sequence, structure or drug information. For example, (1) kinase protein databases (the Kinase.com, the Protein Kinase Resource, the Target Informatics Platform and the KinG database) explore the genomics, evolution and function of protein kinases [23–26]; (2) experimental information databases (the Kinase Validation Set, the KINOMEscan data, the PhosphoBase, the KinMutBase, and the Kinase Pathway Database) contain compound bioactivity, phosphorylation and mutation experimental data [27–32]; (3) kinase catalytic pocket databases (the Kinase Knowledgebase and the Kinase-Ligand Interaction Fingerprints and Structure database) studied the structural and sequence features of ATP-binding and closely nearby pockets [33–35]. However, most of the drugs in these databases are ATP-competitive leading to many side effects. In addition, the available kinase information cannot be directly used in

* Correspondence: yjzhaowh@mail.ccnu.edu.cn
Department of Physics, Central China Normal University, Wuhan 430079, China

Wang *et al. BMC Bioinformatics*   (2019) 20:617

Page 2 of 11

the kinase drug study. The well-analyzed kinase structures are still limited. Thus, a comprehensive and updated human kinase pocket database is urgently needed especially for inhibitors targeting non-catalytic pockets with fewer side effects.

Recently, the kinase family is very well covered by tertiary structures, making it possible to perform a systematic analysis of potential selective binding pockets. Here, we performed a systematic analysis of binding pockets from 255 available human kinase structures to provide potential selective binding pockets and developed HKPocket database with sequence, structure, hydrophilic-hydrophobic and druggability information for kinase drug design.
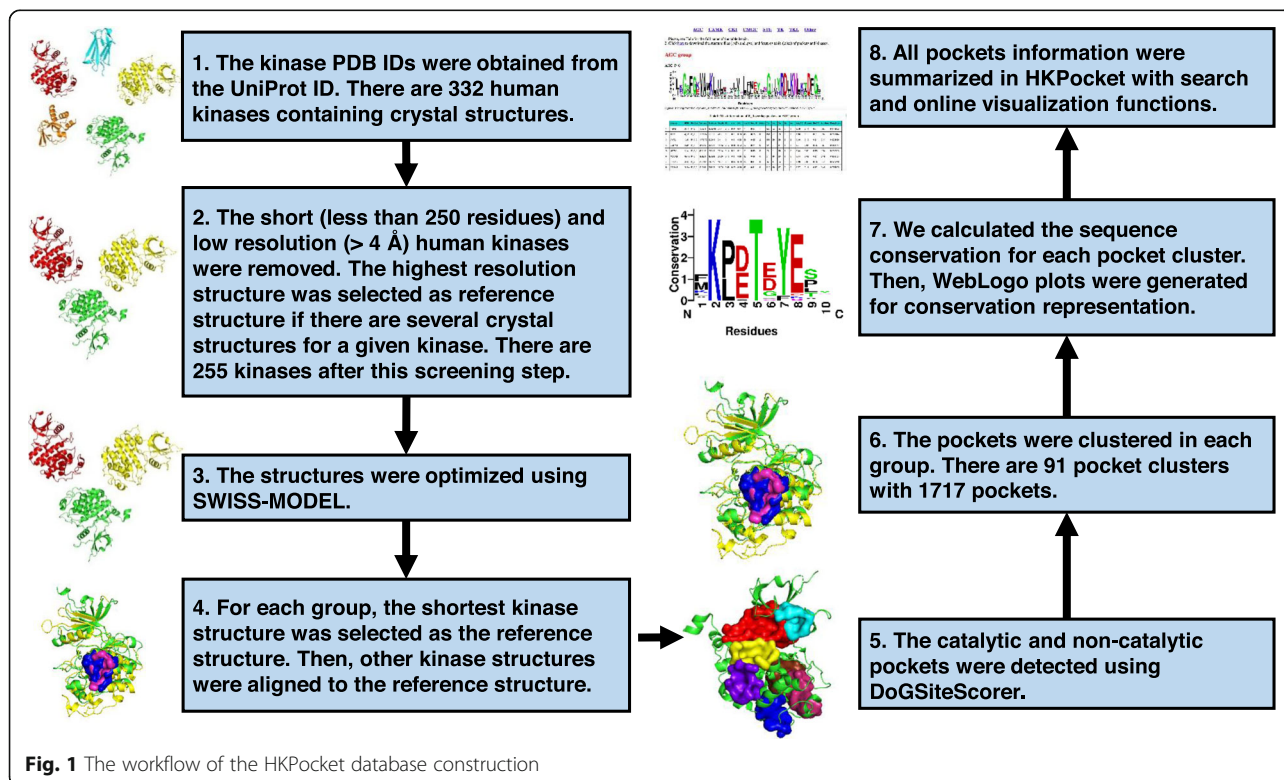
Construction and content
HKPocket database construction
The whole human kinome contains a total of 518 kinases with 478 typical kinases and 40 atypical kinases. The 478 typical kinases were divided into nine groups (AGC: 63, CAMK: 74, CK1: 12, CMGC: 61, RGC: 5, STE: 47, TK: 90, TKL: 43, Other: 83) [36]. A workflow of constructing the HKPocket database is shown in Fig. 1.

(1) We extracted structures from the PDB (Protein Data Bank) database [37] based on the human kinase UniProt ID [38]. There are 313 human kinase structures (AGC: 41, CAMK: 43, CK1: 10, CMGC: 37, RGC: 0, STE: 31, TK: 73, TKL: 34, Other: 44).

(2) We obtained the kinase protein structures by keeping high-resolution structures (resolution < 4 Å) and removing the short proteins with the length fewer than 250 residues. We considered the length cut off based on the following considerations: (i) 90% of kinase proteins are larger than 250 residues [39, 40]. (ii) it is very difficult to determine the pocket information using short kinase proteins with many missing residues. For example, pocket information cannot be extracted from short NEK2 protein kinase (PDB ID: 6H0O) without Cα-helix and other residues (Fig. 2) [41]. The structure with the highest resolution was selected if there are several structures for the same kinase. There are remaining 255 kinase proteins showed in Fig. 3 (AGC: 27, CAMK: 36, CK1: 10, CMGC: 37, RGC: 0, STE: 25, TK: 60, TKL: 26, Other: 34).

(3) All the 255 kinase proteins were optimized to fill in the missing atoms using the template-based structure modeling tool SWISS-MODEL [42].

(4) The protein with the shortest sequence length was selected as the reference structure in each group. The remaining kinase structures were aligned to the reference structure in the corresponding group.

Fig. 1 The workflow of the HKPocket database construction

Fig. 2 The structure of NEK2 protein kinase. NEK2 (PDB ID: 6H0O) is an incomplete protein with only 219 residues. Therefore, NEK2 does not contain Cα-helix and many loop residues. It is very difficult to identify a pocket using this incomplete structure

(5) All kinase pockets were detected by DoGSiteScorer which uses a Gaussian filter to detect drug pockets and define drug pocket features [43, 44]. There are 6347 identified pockets from 255 available human kinase structures.

(6) The pse files were generated by PyMOL (www.pymol.org) for kinases and their pockets visualization. The 91 cluster pse files (AGC: 11, CAMK: 9, CK1: 13, CMGC: 10, STE: 12, TK: 8, TKL: 14, Other: 14) contain 1717 detected pockets.

(7) The multi-sequence alignment of kinase protein sequences in each group was performed. The sequences of pockets were extracted from the aligned kinase protein sequences. The sequence conservation of pocket was analyzed and generated by WebLogo [45]. The overall height of the sequence symbol indicates the sequence conservation at the particular position.

(8) HKPocket database is developed in classical MVC (Model-View-Controller) architecture. The model layer is the database containing sequence, structure, and other pocket information. The

View layer is the online visualization application implemented by JSmol. The Controller layer provides search function access to the pocket data designed using REST API.

The HKPocket database will be updated annually and provides sequence, structure, and other information, such as volumes, depths, surface, hydrophilic-hydrophobic, and drug score. All the information can be downloaded from the HKPocket database website. In addition, the HKPocket database provides an online visualization module. Users can scale and rotate the structures by cartoon or spacefill representations.

Content
The differences between HKPocket and existing databases
HKPocket database is a comprehensive human kinase pocket database for drug study against kinase-related diseases. The following differences distinguish the HKPocket database from the existing kinase-related databases (Table 1).

1. The human kinase protein databases

The Kinase.com provides the sequences and evolutionary trees of 15 kinomes, such as human kinome, mouse kinome, and drosophila kinome [36, 46, 47]. The Protein Kinase Resource includes aligned sequences of 390 eukaryotic protein kinases and a description of 50 protein kinase structures [23]. The Target Informatics Platform (TIP) provides more than 195,000 high-resolution protein structures, covering every major drug target family including proteases, kinases, nuclear receptors, phosphatases, phosphodiesterases, and GPCRs [24]. The KinG database is a comprehensive collection of Ser/Thr/Tyr specific kinases and their similar sequences and provides the sequences, functional domain assignments of kinases [25, 26]. These databases simply provide the sequence, structure and evolutionary information but cannot be directly used in the kinase drug study.

2. The human kinase experiment databases

The Kinase Validation Set contains over 3880 molecule structures and corresponding pIC50 data across three kinase targets (ABL1, SRC, and AURKA) [27]. The KINOMEscan data is a table of all small molecules in the HMS LINCS collection that profiled by KINOMEscan, including links to the raw binding data [48]. The PhosphoBase is a eukaryotic phosphorylation site database [28, 29]. The KinMutBase contains 251 mutations representing 621 patients in protein kinase domains [30, 31]. The Kinase Pathway Database provides functional conservation information, protein-gene/
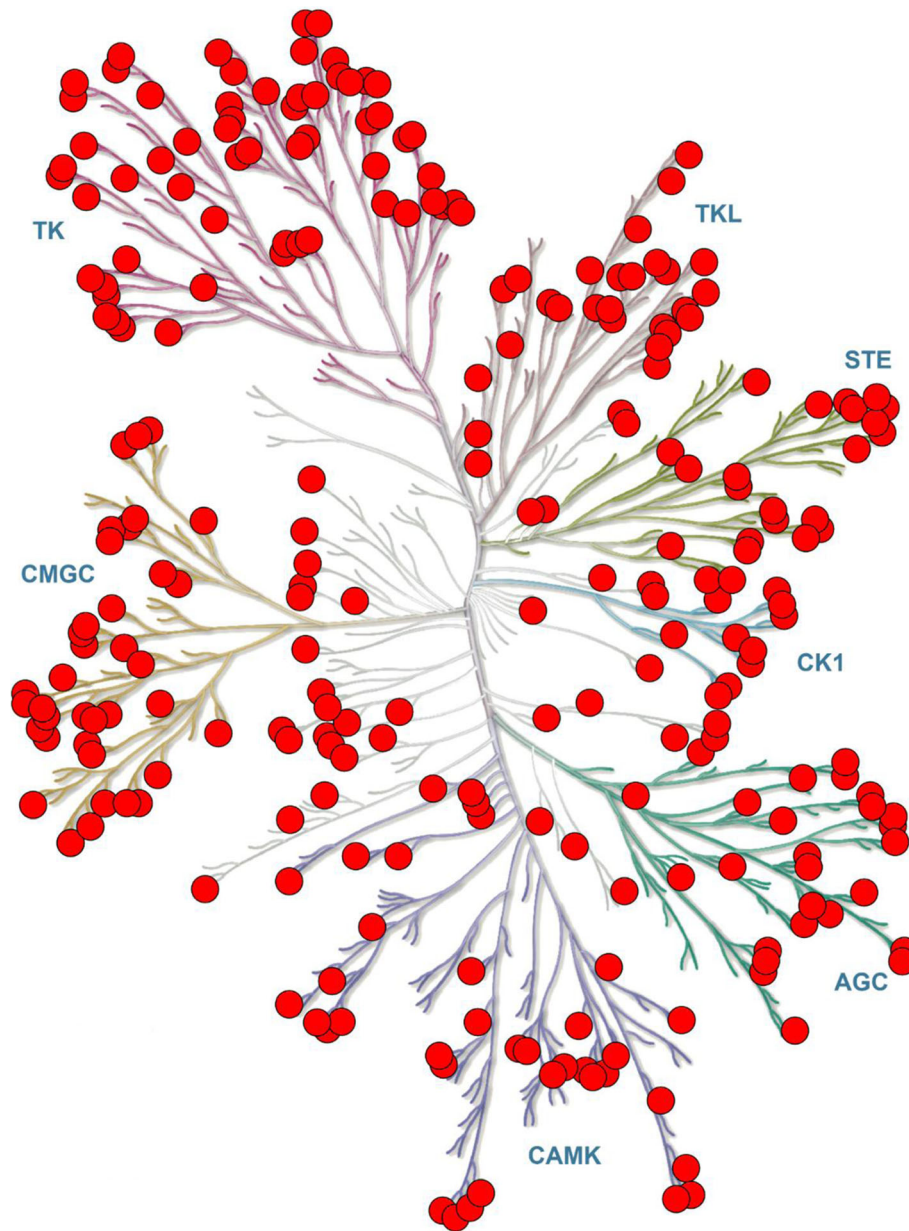
Wang *et al. BMC Bioinformatics*   (2019) 20:617

Page 4 of 11

Fig. 3 The distribution of 255 human kinases. There is the distribution of 255 human kinases (AGC: 27, CAMK: 36, CK1: 10, CMGC: 37, RGC: 0, STE: 25, TK: 60, TKL: 26, Other: 34) in HKPocket database on human kinome tree. The red dots represent each kinase structure

protein/compound interactions in existing databases and papers [32]. These databases provide the phosphorylation, mutation, and binding affinity data but without pocket structural information.

3. The human kinase catalytic pocket databases

The Kinase Knowledgebase (KKB) is a database of kinase structure-activity and chemical synthesis data. This database contains all crystallized catalytic domain structures [35]. The Kinase-Ligand Interaction Fingerprints and Structure database (KLIFS) contains kinase-ligand interaction information, ligand and catalytic pocket structures of kinase proteins [34]. Current databases focus on catalytic pockets (ATP-binding pockets or the pockets closely located at ATP-binding pockets). However, the information on ATP pockets is very limited to drug design.

To bridge this gap, we performed a large-scale analysis of 255 available human kinase structures by systematic pocket detection and comparison. HKPocket contains

Wang *et al. BMC Bioinformatics*   (2019) 20:617

Page 5 of 11

Table 1 The differences distinguish HKPocket database from the existing kinase-related databases

Types of Database	Databases	Links	Kinase Protein			Experiment				Pocket Information	
			Sequences	Structures	Evolutionary trees	Molecules	Affinities	Phosphorylation sites	Mutation sites	catalytic	Non-catalytic
Human kinase protein databases	Kinase.com	http://www.kinase.com/	✓		✓						
	Protein Kinase Resource	http://www0.nih.go.jp/mirror/Kinases/pk_home.html	✓	✓							
	Target Informatics Platform (TIP)	http://www.eidogen.com/tcc.php		✓							
	KinG	http://king.mbu.iisc.ernet.in/	✓								
Human kinase experiment databases	Kinase Validation Set	https://www.eidogen.com/kinasedhld.php				✓	✓				
	KINOMEscan data	http://lincs.hms.harvard.edu/kinomescan/				✓					
	PhosphoBase	http://www.cbs.dtu.dk/databases/PhosphoBase/pbase2/						✓			
	KinMutBase	http://structure.bmc.lu.se/idbase/KinMutBase/							✓		
Human kinase pocket databases	Kinase Pathway Database	http://kinasedb.ontology.ims.u-tokyo.ac.jp/	✓				✓				
	Kinase Knowledgebase (KKB)	http://www.eidogen.com/kinasekb.php					✓			✓	
	Kinase-Ligand Interaction Fingerprints and Structure database (KLIFS)	http://www.vu-compmedchem.nl/klifs		✓		✓	✓			✓	
	HKPocket	http://zhaoserver.com.cn/HKPocket/HKPocket.html	✓	✓	✓					✓	✓

1717 identified pockets which 85% are non-ATP pockets. A clustering of non-ATP pockets provides a framework to decipher pockets for further study. The major difference between HKPocket and previous work is that we have performed systematic pocket detection, comparison, annotation and visualization of non-ATP pockets.
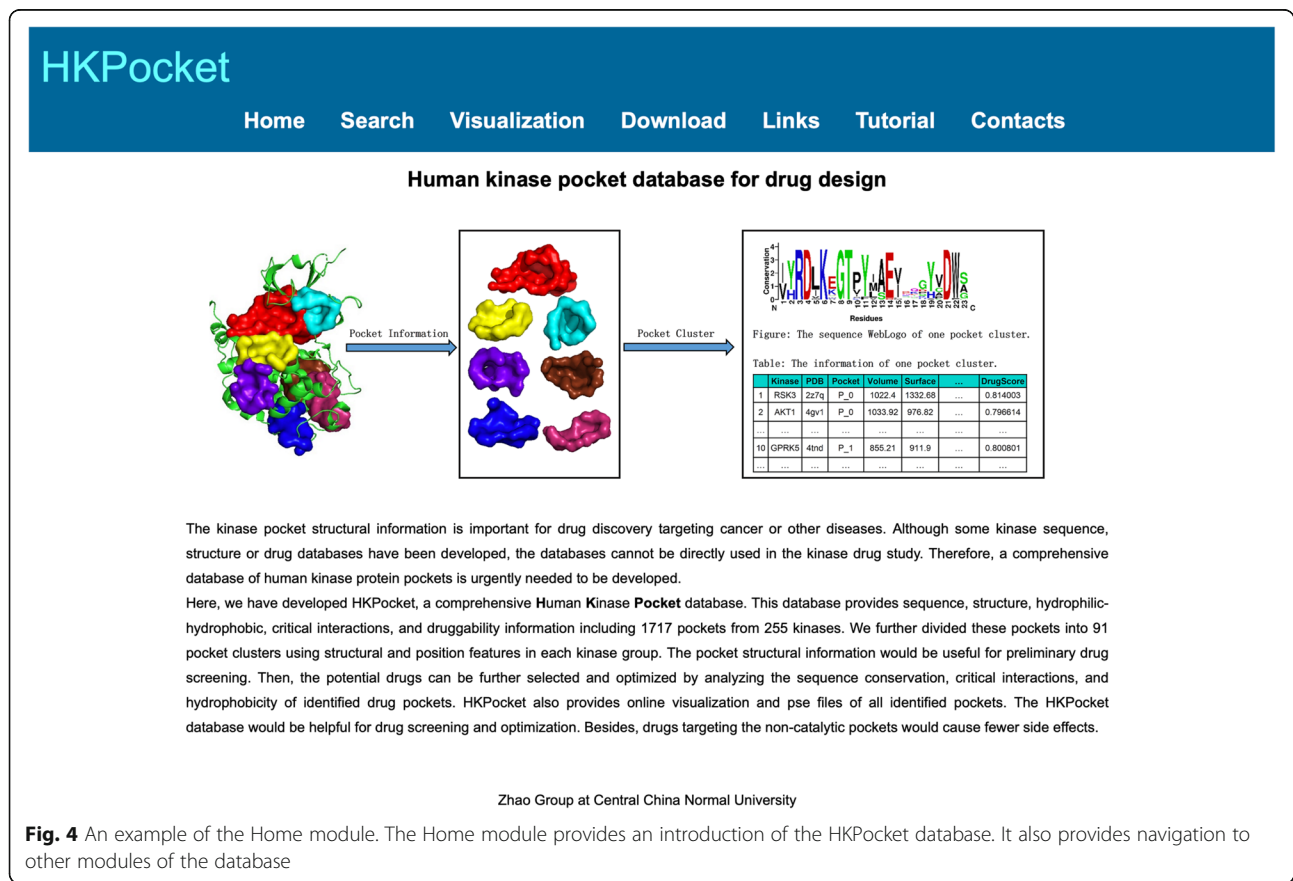
The features of HKPocket database

We have developed a human kinase pocket database for kinase drug design study. Currently, it contains 1717 pockets from 255 kinases.

(1) HKPocket database provides the tertiary structures and the structural topology information (volume, surface, and depth) of 91 pocket clusters. In addition, we also provide other quantitative information such as enclosure, ratios between ellipsoid main axes. Most drug discovery development approaches are based on the lock and key model [49, 50]. The pocket topology information would be useful for preliminary drug screening. For example, Volkamer et al. [8] studied the conservation of ATP-competitive pocket in the human kinome by analyzing the volume, and depth of the ATP-competitive pocket. Therefore, the specificity drug pocket study will promote the development of specific drugs to reduce drug side effects.

(2) Second, HKPocket provides sequence conservation analysis, the number of metals and specific elements (carbon, nitrogen, sulfur, oxygen and other atoms). The sequence conservation analysis results of detected pockets are shown in WebLogo format. The overall height of the sequence indicates the frequency and conservation at the corresponding position. The pocket sequences and atomic level information would play important roles for further drug screening.

(3) HKPocket also provides interaction information of pockets containing hydrophobic interactions as well as the ratio of apolar, polar, positive, negative amino acids and hydrophobicity. These detail interactions would be helpful for drug optimization, especially for side chain or group optimization.

(4) Moreover, the drug scores were calculated using a Support Vector Machine (SVM) model [51, 52]. The drug score represents the druggability of pocket ranging from 0 to1 which the higher score indicating a more druggable pocket.

(5) HKPocket provides an online visualization module. Users can scale or rotate the pocket tertiary structures by cartoon or spacefill representations.

Fig. 4 An example of the Home module. The Home module provides an introduction of the HKPocket database. It also provides navigation to other modules of the database

The key residues can be labeled and highlighted in different colors.
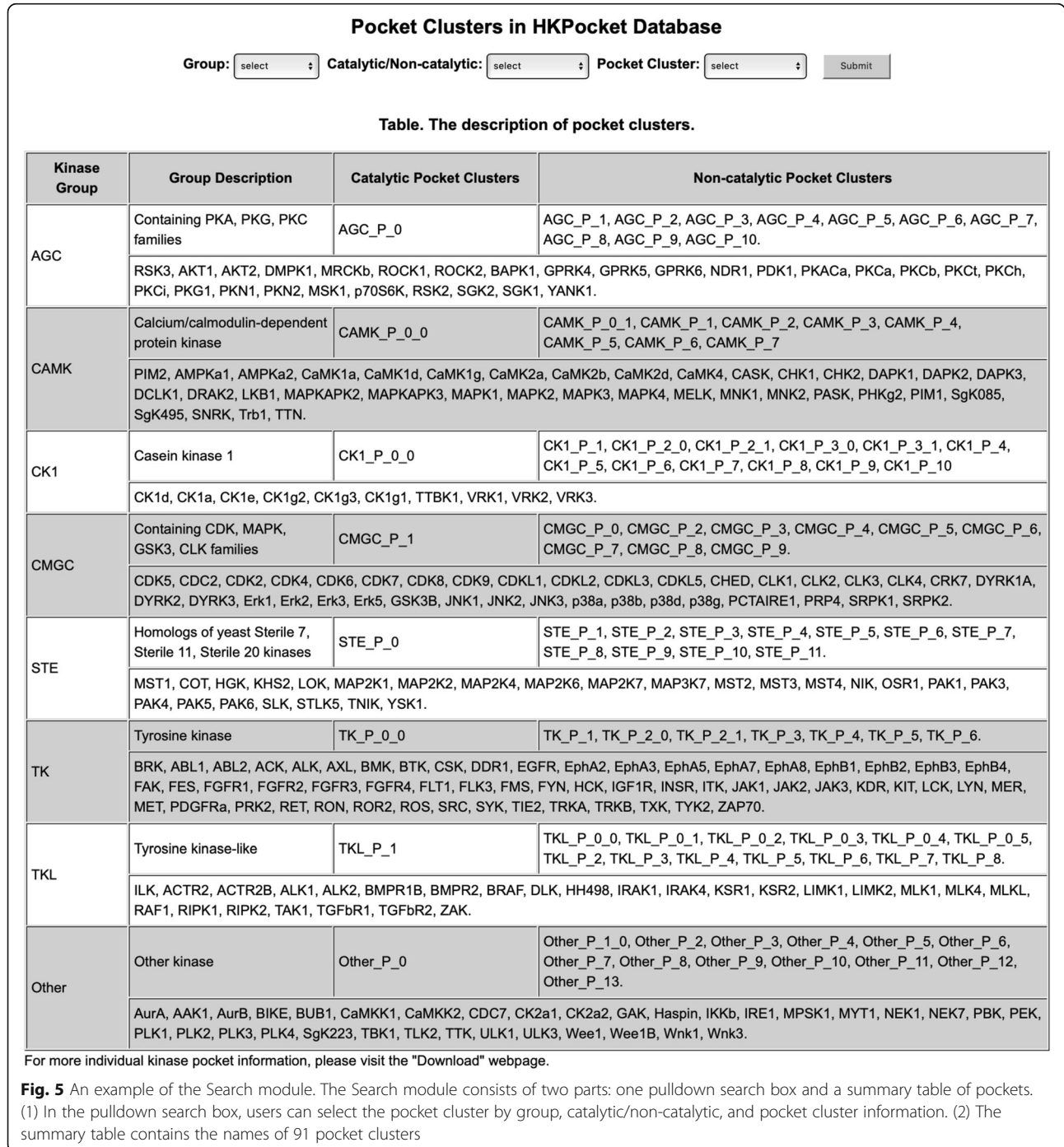
Utility and discussion

HKPocket provides a user-friendly online server. The server contains seven modules: Home, Search, Visualization, Download, Links, Tutorial, and Contacts. The detail information for each module is as follows.

Home module

The HKPocket Home module (Fig. 4) provides an introduction to the HKPocket database. It also provides navigation to other HKPocket modules.

Search module

The Search module (Fig. 5) consists of two parts: one pulldown search box and a summary table of pocket clusters. In the pulldown search box, users can select

---

Pocket Clusters in HKPocket Database

Group: select   **Catalytic/Non-catalytic:** select   **Pocket Cluster:** select   Submit

Table. The description of pocket clusters.

Kinase Group	Group Description	Catalytic Pocket Clusters	Non-catalytic Pocket Clusters
AGC	Containing PKA, PKG, PKC families	AGC_P_0	AGC_P_1, AGC_P_2, AGC_P_3, AGC_P_4, AGC_P_5, AGC_P_6, AGC_P_7, AGC_P_8, AGC_P_9, AGC_P_10.
AGC	RSK3, AKT1, AKT2, DMPK1, MRCKb, ROCK1, ROCK2, BAPK1, GPRK4, GPRK5, GPRK6, NDR1, PDK1, PKACa, PKCa, PKCb, PKCt, PKCh, PKCi, PKG1, PKN1, PKN2, MSK1, p70S6K, RSK2, SGK2, SGK1, YANK1.		
CAMK	Calcium/calmodulin-dependent protein kinase	CAMK_P_0_0	CAMK_P_0_1, CAMK_P_1, CAMK_P_2, CAMK_P_3, CAMK_P_4, CAMK_P_5, CAMK_P_6, CAMK_P_7
CAMK	PIM2, AMPKa1, AMPKa2, CaMK1a, CaMK1d, CaMK1g, CaMK2a, CaMK2b, CaMK2d, CaMK4, CASK, CHK1, CHK2, DAPK1, DAPK2, DAPK3, DCLK1, DRAK2, LKB1, MAPKAPK2, MAPKAPK3, MAPK1, MAPK2, MAPK3, MAPK4, MELK, MNK1, MNK2, PASK, PHKg2, PIM1, SgK085, SgK495, SNRK, Trb1, TTN.		
CK1	Casein kinase 1	CK1_P_0_0	CK1_P_1, CK1_P_2_0, CK1_P_2_1, CK1_P_3_0, CK1_P_3_1, CK1_P_4, CK1_P_5, CK1_P_6, CK1_P_7, CK1_P_8, CK1_P_9, CK1_P_10
CK1	CK1d, CK1a, CK1e, CK1g2, CK1g3, CK1g1, TTBK1, VRK1, VRK2, VRK3.		
CMGC	Containing CDK, MAPK, GSK3, CLK families	CMGC_P_1	CMGC_P_0, CMGC_P_2, CMGC_P_3, CMGC_P_4, CMGC_P_5, CMGC_P_6, CMGC_P_7, CMGC_P_8, CMGC_P_9.
CMGC	CDK5, CDC2, CDK2, CDK4, CDK6, CDK7, CDK8, CDK9, CDKL1, CDKL2, CDKL3, CDKL5, CHED, CLK1, CLK2, CLK3, CLK4, CRK7, DYRK1A, DYRK2, DYRK3, Erk1, Erk2, Erk3, Erk5, GSK3B, JNK1, JNK2, JNK3, p38a, p38b, p38d, p38g, PCTAIRE1, PRP4, SRPK1, SRPK2.		
STE	Homologs of yeast Sterile 7, Sterile 11, Sterile 20 kinases	STE_P_0	STE_P_1, STE_P_2, STE_P_3, STE_P_4, STE_P_5, STE_P_6, STE_P_7, STE_P_8, STE_P_9, STE_P_10, STE_P_11.
STE	MST1, COT, HGK, KHS2, LOK, MAP2K1, MAP2K2, MAP2K4, MAP2K6, MAP2K7, MAP3K7, MST2, MST3, MST4, NIK, OSR1, PAK1, PAK3, PAK4, PAK5, PAK6, SLK, STLK5, TNIK, YSK1.		
TK	Tyrosine kinase	TK_P_0_0	TK_P_1, TK_P_2_0, TK_P_2_1, TK_P_3, TK_P_4, TK_P_5, TK_P_6.
TK	BRK, ABL1, ABL2, ACK, ALK, AXL, BMK, BTK, CSK, DDR1, EGFR, EphA2, EphA3, EphA5, EphA7, EphA8, EphB1, EphB2, EphB3, EphB4, FAK, FES, FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLK3, FMS, FYN, HCK, IGF1R, INSR, ITK, JAK1, JAK2, JAK3, KDR, KIT, LCK, LYN, MER, MET, PDGFRa, PRK2, RET, RON, ROR2, ROS, SRC, SYK, TIE2, TRKA, TRKB, TXK, TYK2, ZAP70.		
TKL	Tyrosine kinase-like	TKL_P_1	TKL_P_0_0, TKL_P_0_1, TKL_P_0_2, TKL_P_0_3, TKL_P_0_4, TKL_P_0_5, TKL_P_2, TKL_P_3, TKL_P_4, TKL_P_5, TKL_P_6, TKL_P_7, TKL_P_8.
TKL	ILK, ACTR2, ACTR2B, ALK1, ALK2, BMPR1B, BMPR2, BRAF, DLK, HH498, IRAK1, IRAK4, KSR1, KSR2, LIMK1, LIMK2, MLK1, MLK4, MLKL, RAF1, RIPK1, RIPK2, TAK1, TGFbR1, TGFbR2, ZAK.		
Other	Other kinase	Other_P_0	Other_P_1_0, Other_P_2, Other_P_3, Other_P_4, Other_P_5, Other_P_6, Other_P_7, Other_P_8, Other_P_9, Other_P_10, Other_P_11, Other_P_12, Other_P_13.
Other	AurA, AAK1, AurB, BIKE, BUB1, CaMKK1, CaMKK2, CDC7, CK2a1, CK2a2, GAK, Haspin, IKKb, IRE1, MPSK1, MYT1, NEK1, NEK7, PBK, PEK, PLK1, PLK2, PLK3, PLK4, SgK223, TBK1, TLK2, TTK, ULK1, ULK3, Wee1, Wee1B, Wnk1, Wnk3.		

For more individual kinase pocket information, please visit the "Download" webpage.

Fig. 5 An example of the Search module. The Search module consists of two parts: one pulldown search box and a summary table of pockets. (1) In the pulldown search box, users can select the pocket cluster by group, catalytic/non-catalytic, and pocket cluster information. (2) The summary table contains the names of 91 pocket clusters
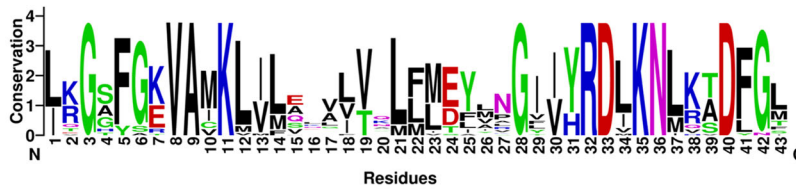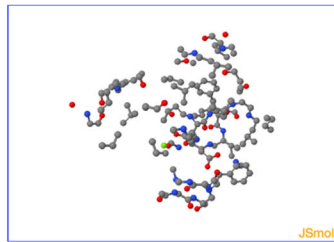
AGC_P_0

1. Sequence WebLogo

Figure. The sequence WebLogo of P_0 pocket of RSK3 kinase(PDB ID:2Z7Q) was generated by sequences of 27 kinase pockets in AGC group.

2. Pocket Visualization

Save [image]
Display [large] [small]
Color [rainbow] [gray]
Show [spacefill] [wire] [ball&stick] [cartoon]

3. Pocket Information

Table. The information of P_0 pocket cluster in AGC group.

	Kinase	PDB	Pocket	Volume	Surface	Depth	Enc	c/a	b/a	Hyd_I	Hyd-R	Metal	Cs	Ns	Os	Ss	Xs	NegAA	PosAA	PolAA	ApolAA	DrugScore
1	RSK3	2z7q	P_0	1022.40	1332.68	21.47	0.12	0.14	0.50	71	0.47	1	132	23	26	0	1	0.09	0.16	0.20	0.52	0.814003
2	AKT1	4gv1	P_0	1033.92	976.82	20.90	0.10	0.20	0.36	45	0.35	0	151	27	39	3	0	0.16	0.16	0.31	0.37	0.796614
3	AKT2	1o6l	P_0_0	1173.76	1200.40	0.40	0.00	0.10	0.28	56	0.35	2	151	29	39	2	2	0.16	0.16	0.30	0.34	0.829669
4	DMPK1	2vd5	P_2	456.45	423.66	12.63	0.17	0.15	0.22	27	0.51	0	62	9	11	1	0	0.20	0.05	0.25	0.50	0.654464
5	MRCKb	5ote	P_0_0	439.56	397.21	13.54	0.12	0.20	0.50	21	0.48	0	63	5	10	1	0	0.24	0.05	0.19	0.52	0.608955
6	ROCK1	5wne	P_0	802.24	824.83	20.96	0.17	0.10	0.38	42	0.46	0	97	16	20	1	0	0.18	0.06	0.27	0.48	0.835703
7	ROCK2	4l6q	P_0	836.48	889.19	19.27	0.19	0.13	0.39	36	0.40	0	95	17	23	2	0	0.19	0.06	0.28	0.47	0.822279
8	BARK1	3v5w	P_0_0	813.79	769.22	18.53	0.07	0.33	0.46	42	0.41	0	118	20	27	0	0	0.17	0.12	0.22	0.47	0.774085
9	GPRK4	4yhj	P_0_0	511.19	565.63	12.55	0.14	0.11	0.20	22	0.39	0	69	12	11	1	0	0.12	0.08	0.33	0.46	0.534980
10	GPRK5	4tnd	P_1	855.21	911.90	17.11	0.11	0.21	0.25	27	0.30	1	113	26	25	1	1	0.12	0.12	0.30	0.43	0.800801
11	GPRK6	2acx	P_1	451.09	512.44	17.68	0.17	0.09	0.14	16	0.33	0	60	9	10	0	0	0.09	0.09	0.27	0.55	0.771950
12	NDR1	6bxi	P_0_0	914.75	932.84	0.40	0.00	0.16	0.32	30	0.28	1	140	30	29	3	1	0.15	0.15	0.23	0.45	0.789109
13	PDK1	5lvo	P_1	456.96	468.54	11.33	0.19	0.10	0.22	19	0.40	0	59	8	9	1	0	0.15	0.10	0.30	0.45	0.611691
14	PKACa	4wb8	P_0	763.90	902.42	20.69	0.04	0.11	0.76	32	0.30	2	120	19	27	2	2	0.17	0.07	0.36	0.36	0.825996
15	PKCa	4ra4	P_2	376.51	380.46	16.15	0.17	0.16	0.20	18	0.46	0	51	6	9	2	0	0.18	0.06	0.12	0.65	0.718212
16	PKCb	2i0e	P_2	476.67	487.09	14.98	0.16	0.20	0.50	33	0.54	0	63	5	10	2	0	0.20	0.05	0.20	0.55	0.727855
17	PKCt	1xjd	P_0	573.57	527.01	16.14	0.12	0.16	0.55	23	0.43	0	69	11	17	1	0	0.19	0.08	0.35	0.38	0.760312
18	PKCh	3txo	P_0	735.68	717.02	19.36	0.09	0.09	0.29	31	0.42	0	95	12	20	1	0	0.19	0.06	0.19	0.56	0.831068
19	PKCi	5li9	P_0	752.90	752.26	23.30	0.15	0.09	0.33	36	0.44	0	112	16	17	0	0	0.14	0.06	0.31	0.50	0.852769
20	PKG1	6c0t	P_0_0	854.59	728.45	20.38	0.09	0.14	0.28	35	0.38	0	111	18	23	4	0	0.16	0.08	0.24	0.53	0.776585
21	PKN1	4oth	P_0_0	897.28	930.44	27.88	0.06	0.11	0.25	44	0.39	0	143	23	27	1	0	0.15	0.07	0.24	0.54	0.798518
22	PKN2	4crs	P_0_0	545.02	571.97	13.94	0.03	0.26	0.84	16	0.30	1	82	12	14	1	1	0.19	0.11	0.19	0.48	0.630973
23	MSK1	1vzo	P_0_0	617.73	772.55	17.83	0.07	0.20	0.69	14	0.18	0	85	16	21	0	0	0.17	0.22	0.22	0.39	0.429762
24	p70S6K	4l42	P_0_0	608.38	669.97	13.87	0.06	0.15	0.37	41	0.47	0	90	16	19	1	0	0.14	0.14	0.31	0.41	0.599929
25	RSK2	4nw6	P_0_0	613.69	560.32	12.25	1.00	0.13	0.72	24	0.40	0	71	12	15	0	0	0.24	0.08	0.28	0.40	0.605276
26	SGK1	2r5t	P_0_0	718.59	875.06	14.96	0.04	0.16	0.42	33	0.37	1	90	14	22	0	1	0.17	0.13	0.27	0.40	0.600948
27	YANK1	4fr4	P_0_0	618.05	671.56	21.36	0.09	0.11	0.23	26	0.35	0	96	15	20	2	0	0.16	0.06	0.19	0.58	0.633597

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 The pocket information of AGC_P_0. The information of AGC_P_0 pocket cluster contains three parts: Sequence WebLogo, Pocket Visualization, and Pocket Information. (1) A WebLogo plot was generated to show the sequence conservation of the pocket. The overall height of the residues in WebLogo indicates the frequency and conservation at the corresponding position. For AGC_P_0 pocket, the G3, V8, A9, K11, G28, R32, D33, K35, N36, and D40 residues are highly conserved. (2) Users can scale and rotate the AGC_P_0 pocket structure. HKPocket provides four representations: "spacefill", "wire", "ball&stick", and "cartoon". The key residues can be highlighted in different colors. Users can also generate and save the picture. (3) A pocket information table contains the structural shape (volume, surface, depth, etc.), sequence (negative amino acid ratio, polar amino acid ratio, etc.), atom (the number of metals, carbons, etc.), hydrophilic-hydrophobic, critical interactions and druggability information of cluster pockets

the pocket cluster by group, catalytic/non-catalytic, and pocket cluster information. For example, Fig. 6 shows the detail information for AGC_P_0. (1) A WebLogo plot was generated to show the pocket sequence conservation. The overall height of the sequences in WebLogo indicates the frequency and conservation at the corresponding position. For AGC_ P_0 pocket, the G3, V8, A9, K11, G28, R32, D33, K35, N36, and D40 residues are highly conserved. (2) Users can scale and rotate the pocket structures. HKPocket provides four representations: "spacefill", "wire", "ball&stick", and "cartoon". The key residues can be highlighted in different colors. Users can also generate and save the picture. (3) A pocket information table contains the structural shape (volume, surface, depth, etc.), sequence (negative amino acid ratio, polar amino acid ratio, etc.), atom (the number of metals, carbons, etc.), hydrophilic-hydrophobic, critical interactions, and druggability information.

The summary table contains the information of 91 pocket clusters including 8 catalytic pocket clusters and 83 non-catalytic pocket clusters. The identified pockets of a given kinase were sequentially numbered by P_0, P_ 1, P_2, etc. The pocket can be further divided into several small sub-pockets. For example, the pocket P_0 can be divided into two sub-pockets P_0_0 and P_0_1. Therefore, there are 11, 9, 13, 10, 14, 12, 8, 14 clusters of pockets in AGC, CAMK, CK1, CMGC, Other, STE, TK and TKL groups, respectively.

Visualization module

In the visualization module, users can upload and investigate the pocket structure. The pocket structure will be visualized in four representations: "spacefill", "wire", "ball&stick", and "cartoon". The key residues can be highlighted in different colors. Users can scale and rotate the pocket structures. Users can also generate and save the picture.

Download module

A summary table was provided in the download module. Users can download the individual group or all the

pocket information. The download data consists four files: (1) The tables (xlsx format) with sequence and topology features for 91 pocket clusters; (2) The structure files (PDB format) with structure information of all pockets from 255 kinase proteins; (3) The pse files for structural detail visualization for 255 kinase proteins and their pockets. (4) The sequences files (txt format) of 255 kinase proteins in the HKPocket database.

Links module

The Links module provides the other useful links of protein 3D structure resources, sequence alignment, molecular modeling, molecular dynamics, molecular dynamics, molecular visualization/analysis, and kinase-related database websites. These useful websites would be helpful to the kinase-related drug design.

Tutorial module

The Tutorial module provides the introduction to use the HKPockt and the abbreviation for the HKPocket database.

Contacts module

The Contacts module provides emails for users to comment or ask questions.

Discussion

The kinase protein contains one N-terminal and one C-terminal lobe. The two lobes form the ATP-binding pocket. During the cell cycle, the kinase switches between the active (open) and inactive (closed) states due to the conformational transition of the DFG-loop. Previously, Kornev et al. analyzed the active and inactive of CDK2, SRC, and IRK structures [53, 54]. The results showed that there are some conformational changes in the catalytic region while fewer changes in the non-catalytic region. We analyzed the non-catalytic pockets of CDK2, SRC, and IRK kinase proteins in both active and inactive states. The CDK2, IRK, and SRC contain 9, 9, and 7 non-catalytic pockets in active states (Fig. 7). The results show 77% (6, 7 and 6) of non-catalytic pockets are very similar between active and inactive

Fig. 7 The structural analysis of non-catalytic pockets of CDK2, SRC, and IRK in active and inactive states. The active structure is colored in green. The inactive structure is colored in yellow. The results show 77% of non-catalytic pockets are very similar between active and inactive states
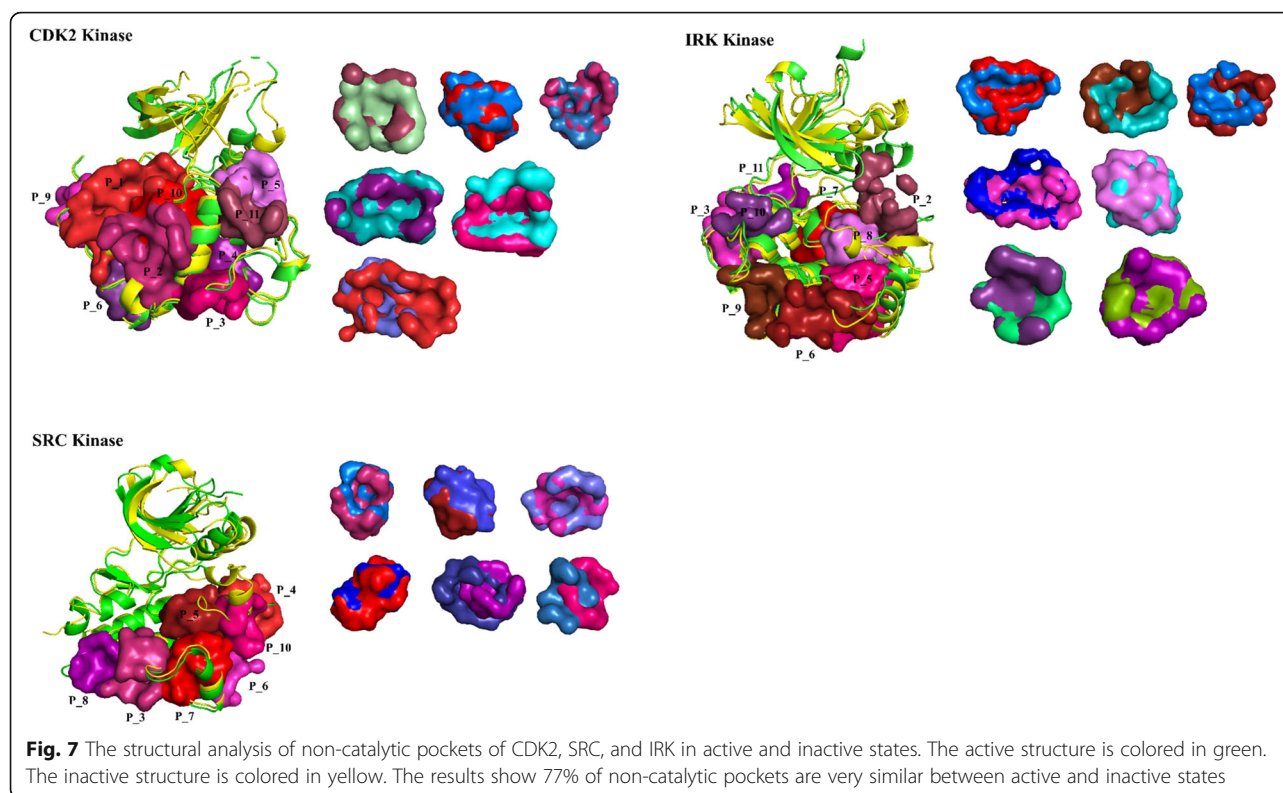
states. Therefore, the pocket information in HKPocket would be useful for allosteric drug design.

Conclusions

The precision medicine initiative in kinase drug design is needed urgently due to the abnormal kinase activity could cause unexpected diseases. Müller et al. raised this question in 2015 and pointed out that the human kinome is now very well covered by the tertiary structure, making it possible to perform a comprehensive analysis of potential drug binding pockets for developing specific kinase drugs. In summary, we developed a well-analyzed human kinome pocket database with quantitative information of sequence, structure, interaction, and drug score. The HKPocket allows users to perform a systematic analysis of human kinase pockets for specific drug design. We hope the HKPocekt database will be useful for drug screen and optimization if the targeted pocket is known.

References
1. Reiterer V, Eyers PA, Farhan H. Day of the dead: pseudokinases and pseudophosphatases in physiology and disease. Trends Cell Biol. 2014;24:489–505.
2. Matthieu C, Thierry C, Jonathan B, Rafael N. Kinome render: a stand-alone and web-accessible tool to annotate the human protein kinome tree. Peerj. 2013;1:e126.

3.   Ferguson FM, Gray NS. Kinase inhibitors: the road ahead. Nat Rev Drug Discov. 2018;17:353–77.

4.   Sonoshita M, Scopton AP, Ung PMU, Murray MA, Silber L, Maldonado AY, et al. A whole-animal platform to advance a clinical kinase inhibitor into new disease space. Nat Chem Biol. 2018;14:291–8.

5.   Müller S, Chaikuad A, Gray NS, Knapp S. The ins and outs of selective kinase inhibitor development. Nat Chem Biol. 2015;11:818–21.

6.   Jänne PA, Nathanael G, Jeff S. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. Nat Rev Drug Discov. 2009;8:709–23.

7.   Comess KM, Sun C, Abad-Zapatero C, Goedken ER, Gum RJ, Borhani DW, et al. Discovery and characterization of non-ATP site inhibitors of the mitogen activated protein (MAP) kinases. ACS Chem Biol. 2011;6:234–44.

8.   Andrea V, Sameh E, Samo T, Sabrina J, Friedrich R, Simone F. Pocketome of human kinases: prioritizing the ATP binding sites of (yet) untapped protein kinases for drug discovery. J Chem Inf Model. 2015;55:538–49.

9.   Yang CH, Lin WC, Chuang CK, Chang YC, Pang ST, Lin YC, et al. Hand-foot skin reaction in patients treated with sorafenib: a clinicopathological study of cutaneous manifestations due to multitargeted kinase inhibitor therapy. Brit J Dermatol. 2010;158:592–6.

10. Wood LS. Management of vascular endothelial growth factor and multikinase inhibitor side effects. Clin J Oncol Nurs. 2009;13:13–8.

11. Zhao Y, Zeng C, Tarasova NI, Chasovskikh S, Dritschilo A, Timofeeva OA. A new role for STAT3 as a regulator of chromatin topology. Transcription. 2013;4:227–31.

12. Zhao Y, Zeng C, Massiah MA. Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain. PLoS One. 2015;10:e0124377.

13. Aqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. Protein Engi Design Selection. 1994; 7:385–91.

14. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;432:862–5.

15. Mazalan L, Bell A, Sbaffi L, Willett P. Cross-classified multilevel Modelling of the effectiveness of similarity-based virtual screening. ChemMedChem. 2018; 13:582–7.

16. Öztürk H, Ozkirimli E, Özgür A. DeepDTA: deep drug-target binding affinity prediction. Bioinformatics. 2018;34:I821–9.

17. Wu J, Zhang Q, Wu W, Pang T, Hu H, Chan W, et al. WDL-RF: predicting bioactivities of ligand molecules acting with G protein-coupled receptors by combining weighted deep learning and random Forest. Bioinformatics. 2018;34:2271–82.

18. Zhao Y, Chen H, Du C, Jian Y, Li H, Xiao Y, et al. Design of tat-activated Cdk9 inhibitor. Int J Pept Res Ther. 2019;25:807–11.

19. Wang K, Jian Y, Wang H, Zeng C, Zhao Y. RBind: computational network method to predict RNA binding sites. Bioinformatics. 2018;3:3131–6.

20. Wang HW, Wang KL, Guan ZY, Jian YR, Jia Y, Kashanchi F, et al. Computational study of non-catalytic T-loop pocket on CDK proteins for drug development. Chinese Phys B. 2017;26:128702.

21. Chen H, Zhao Y, Li H, Zhang D, Huang Y, Shen Q, et al. Break CDK2/Cyclin E1 interface allosterically with small peptides. PLoS One. 2014;9:e109154.

22. Zhao Y, Jian Y, Liu Z, Liu H, Liu Q, Chen C, et al. Network analysis reveals the recognition mechanism for dimer formation of bulb-type Lectins. Sci Rep. 2017;7:2876.

23. Niedner RH, Buzko OV, Haste NM, Taylor A, Gribskov M, Taylor SS. Protein kinase resource: an integrated environment for phosphorylation research. Proteins. 2006;63:78–86.

24. Hambly K, Danzer J, Muskal S, Debe DA. Interrogating the druggable genome with structural informatics. Cheminform. 2006;38:273–81.

25. Krupa A, Abhinandan KR, Srinivasan N. KinG: a database of protein kinases in genomes. Nucleic Acids Res. 2004;32:153–5.

26. Dardick C, Chen J, Richter T, Ouyang S, Ronald P. The Rice kinase database. A Phylogenomic database for the Rice Kinome. Plant Physiol. 2007;143:579–86.

27. Sharma R, Schürer SC, Muskal SM. High quality, small molecule-activity datasets for kinase research. F1000res. 2016;5:1366.

28. Francesca D, Gould CM, Claudia C, Allegra V, Gibson TJ. Phospho.ELM: a database of phosphorylation sites--update 2008. Nucleic Acids Res. 2008;36:240–4.

29. Diella F, Cameron S, Gemünd C, Linding R, Via A, Kuster B, et al. Phospho. ELM: A database of experimentally verified phosphorylation sites in eukaryotic proteins. Bmc Bioinformatics. 2004;5:79.

30. Csaba O, Jouni VL, Kaj S, Mauno V. KinMutBase: a registry of disease-causing mutations in protein kinase domains. Hum Mutat. 2005;25:435–42.

31. Stenberg KA, Riikonen PT, Vihinen M. KinMutBase, a database of human disease-causing protein kinase mutations. Nucleic Acids Res. 2000;28:369–71.

32. Koike A, Kobayashi Y, Takagi T. Kinase pathway database: an integrated protein-kinase and NLP-based protein-interaction resource. Genome Res. 2003;13:1231–43.

33. Volkamer A, Eid S, Turk S, Rippmann F, Fulle S. Identification and visualization of kinase-specific subpockets. J Chem Inf Model. 2016;56:335–46.

34. Kooistra AJ, Kanev GK, van Linden OP, Leurs R, de Esch IJ, De GC. KLIFS: a structural kinase-ligand interaction database. Nucleic Acids Res. 2016;44: D365–71.

35. 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. 2010;19:763–74.

36. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. Science. 2002;298:1912–34.

37. Parasuraman S. Protein data bank. J Pharmacol Pharmacother. 2012;3:351–2.

38. Consortium U. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43:204–12.

39. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J. 1995;9:576–96.

40. Bick MJ, Lamour V, Rajashankar KR, Gordiyenko Y, Robinson CV, Darst SA. How to switch off a histidine kinase: crystal structure of Geobacillus stearothermophilus KinB with the inhibitor Sda. J Mol Biol. 2009;386:163–77.

41. Byrne MJ, Cunnison RF, Bhatia C, Bayliss RW. Crystal structure of a Nek2/ inhibitor complex. To be published.

42. Marco B, Stefan B, Andrew W, Konstantin A, Gabriel S, Tobias S, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42:W252–8.

43. Andrea V, Daniel K, Thomas G, Friedrich R, Matthias R. Combining global and local measures for structure-based druggability predictions. J Chem Inf Model. 2012;52:360–72.

44. Andrea V, Axel G, Thomas G, Matthias R. Analyzing the topology of active sites: on the prediction of pockets and subpockets. J Chem Inf Model. 2010;50:2041–52.

45. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–100.

46. Sean C, Glen C, Sucha S, Tony H, Gerard M. The mouse kinome: discovery and comparative genomics of all mouse protein kinases. P Natl Acad Sci USA. 2004;101:11707–12.

47. Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci. 2002;27:514–20.

48. Fabian MA, Biggs WH, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, et al. A small molecule-kinase interaction map for clinical kinase inhibitors. Nat Biotechnol. 2005;23:329–36.

49. Cramer F. Biochemical correctness: Emil Fischer's lock and key hypothesis, a hundred years after — an essay. Pharm Acta Helv. 1995;69:193–203.

50. Awino JK, Hu L, Zhao Y. Molecularly responsive binding through co-occupation of binding space: a lock-key story. Org Lett. 2016;18:1650–3.

51. William SN. What is a support vector machine? Nat Biotechnol. 2006;24:1565–7.

52. Jian Y, Wang X, Qiu J, Wang H, Liu Z, Zhao Y, et al. DIRECT: RNA contact predictions by integrating structural patterns. BMC Bioinformatics. 2019;20:497.

53. Kornev AP, Haste NM, Taylor SS, Eyck LFT. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. PNAS. 2006;103:17783–8.

54. Xie Q, Fulton DB, Andreotti AH. A selective NMR probe to monitor the conformational transition from inactive to active kinase. ACS Chem Biol. 2015;10:262–8.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.