


DATABASE

Open Access

# D3PM: a comprehensive database for protein motions ranging from residue to domain



Cheng Peng<sup>1,2†</sup>, Xinben Zhang<sup>1†</sup>, Zhijian Xu<sup>1,2\*</sup>, Zhaoqiang Chen<sup>1</sup>, Yanqing Yang<sup>1,2</sup>, Tingting Cai<sup>1</sup> and Weiliang Zhu<sup>1,2,3\*</sup> 

\*Correspondence:

zjxu@simm.ac.cn;

wlzhu@simm.ac.cn

<sup>†</sup>Cheng Peng and Xinben Zhang contributed equally to this work

<sup>1</sup> CAS Key Laboratory of Receptor Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, People's Republic of China  
Full list of author information is available at the end of the article

## Abstract

**Background:** Knowledge of protein motions is significant to understand its functions. While currently available databases for protein motions are mostly focused on overall domain motions, little attention is paid on local residue motions. Albeit with relatively small scale, the local residue motions, especially those residues in binding pockets, may play crucial roles in protein functioning and ligands binding.

**Results:** A comprehensive protein motion database, namely D3PM, was constructed in this study to facilitate the analysis of protein motions. The protein motions in the D3PM range from overall structural changes of macromolecule to local flip motions of binding pocket residues. Currently, the D3PM has collected 7679 proteins with overall motions and 3513 proteins with pocket residue motions. The motion patterns are classified into 4 types of overall structural changes and 5 types of pocket residue motions. Impressively, we found that less than 15% of protein pairs have obvious overall conformational adaptations induced by ligand binding, while more than 50% of protein pairs have significant structural changes in ligand binding sites, indicating that ligand-induced conformational changes are drastic and mainly confined around ligand binding sites. Based on the residue preference in binding pocket, we classified amino acids into “pocketphilic” and “pocketphobic” residues, which should be helpful for pocket prediction and drug design.

**Conclusion:** D3PM is a comprehensive database about protein motions ranging from residue to domain, which should be useful for exploring diverse protein motions and for understanding protein function and drug design. The D3PM is available on [www.d3pharma.com/D3PM/index.php](http://www.d3pharma.com/D3PM/index.php).

**Keywords:** Protein motion, Motion pattern, D3PM database, Amino acids preference

## Background

The conformational diversity of protein is rooted from its structure and is often a key feature of its function [1, 2]. A fundamental recognition of how protein works therefore requires knowledge of its structure and dynamism, which is also helpful to drug discovery and development. For instance, an ensemble docking strategy that tries to solve the problem of receptor flexibility has received increasing attentions on virtual screening [3, 4]. Such conformational diversity can be studied in various ways. X-ray crystallography



and nuclear magnetic resonance (NMR) are versatile experimental techniques to obtain biomolecular structures [5, 6]. In computational methods, normal mode analysis and molecular dynamics can be used to predict the conformational diversity of protein [7]. With more and more available protein structures, there is an increasing interest to relate protein structure to motion for studying its function.

As summarized in a number of reviews, most studies about the protein motion have focused on the hinge and shear motions of protein domain [8–11]. Several techniques [12–15] applied to detect dynamical protein domains have been developed such as the difference-distance method and deformation-plot analysis, and a catalog of domain motion types has been compiled. Databases of protein domain motions have been also available in recent years, for example the DynDom database [16–19]. In addition, the information of protein motions collected in recent databases involves from small loop to entire subunit besides domain region. However, in many cases, proteins have no obvious domain movement under different conditions, but show significant side-chain motion of binding pocket residue or catalytic residue [20, 21]. The side-chain motion was found to play a crucial role in responding to the access, regiospecificity, stabilization and dissociation of ligand [22, 23]. For example, the most pronounced conformational change simply occurs on the F194 of KAI2 protein with a  $\sim 90^\circ$  flip of its benzene ring when bound with inhibitor KAR<sub>1</sub> [24]. Furthermore, the dynamic residue may impact the conformation of its neighboring region [25]. Therefore, it is of significance to study side-chain motions of the residues within binding pocket.

The protein data bank (PDB) [26] contains nearly 167,000 protein entries (July 2020), and the number is growing at an exponential rate. It is therefore a useful resource for studying protein motions. Three-dimensional (3D) structures of protein are provided in the PDB, but entries of protein are redundant for structures determined under different conditions. Lots of effort has gone into collecting and analyzing the vast amount of data in the PDB, leading to many databases. The MolMovDB [18] is a dominant database containing the information of protein conformational changes. Other databases, such as the ComSin [27], AH-DB [28], PDBFlex [29], and PSCDB [30] provide structural pairs of protein in bound and free states to explore protein motions induced by ligand binding, among which the AH-DB contains the most entries (>700,000). However, protein motions are sophisticated, which are related not only to its intrinsic flexibility or experimental conditions (such as temperatures and pH), but also to external perturbations like ligand binding [31, 32]. The PCDB [33] and CoDNAS [34] provide redundant structural clusters of protein under different experimental protocols, but the PCDB has been not available for a while. The CCProf [35] is another conformational diversity database that contains 986,187 structural pairs of protein before and after ligand binding, and ten biological features are introduced in the CCProf for studying binding site dynamics. However, it is difficult to study the local motion of binding pocket residue by using these available databases.

In this study, we constructed a database that covers all kinds of protein motions ranging from overall structure to local residue, namely D3PM. The motion patterns in the database are classified into 4 types of overall structural changes and 5 types of pocket residue motions. Considering that the form of structural pairs is more convenient to analyze motion features than that of structural clusters, all the protein motions were

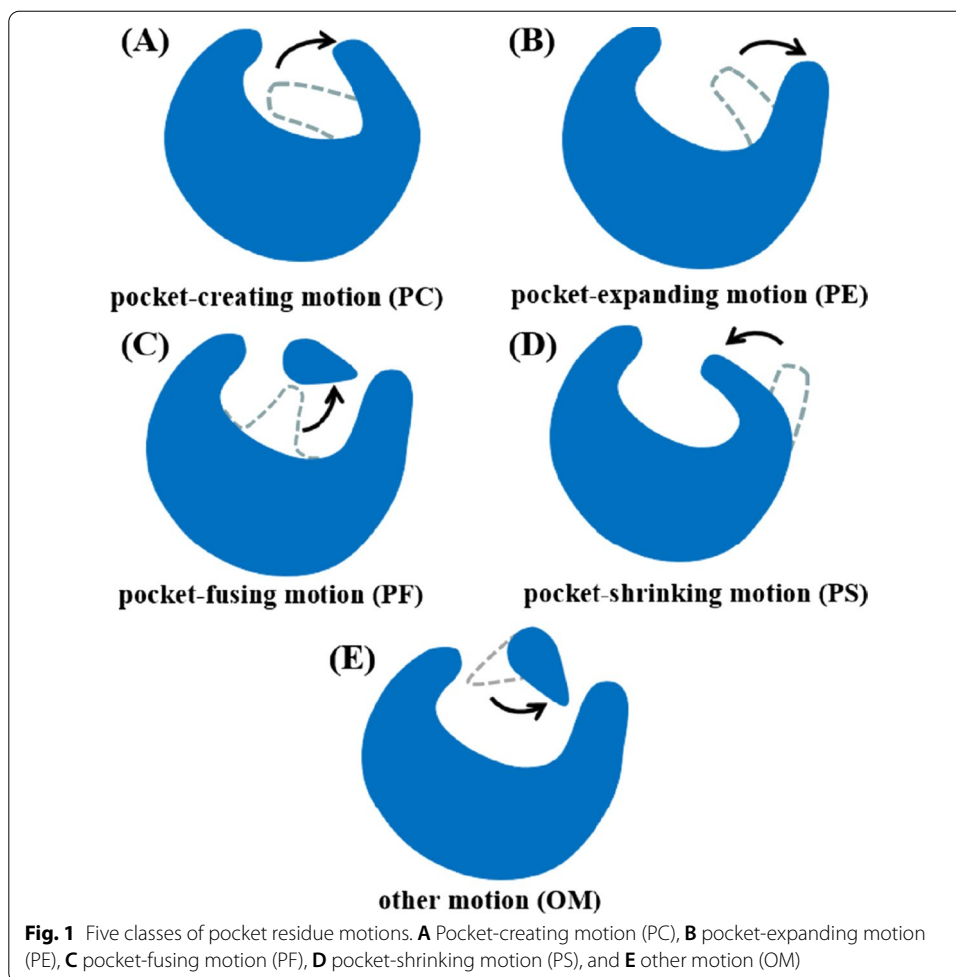
provided with structural pairs in the D3PM. We hope that the D3PM will be helpful to explore diverse protein motions and promote the drug discovery and development.

## Construction and content

### The D3PM database construction

All the X-ray structures with resolution better than 3.0 Å were downloaded from the PDB (25th October 2018 for the initial version, 11th April 2021 for the first update), and were divided into pairs of identical proteins that have the same UniProt ID. The oligomerization state was limited to either monomer or homo-multimer to exclude the influence of protein–protein interactions on structural changes. Many of small molecules bound into proteins are crystallographic additives (PEG, etc. Additional file 1: Table S1), and they were manually removed from the protein–ligand complexes. Finally, protein structural pairs were divided into two subsets with the threshold of 2.0 Å for overall  $C_{\alpha}$  root mean square deviation (RMSD), which was often used as a threshold for drug discovery [36, 37]. There are redundant protein pairs of identical type of motions. Therefore, a typical protein pair with the most significant motion was selected for each type of motion to construct a non-redundant, contrastive, and classified protein motion database.

**For protein pairs with overall RMSD that is smaller than 2.0 Å** Although the overall RMSD that is smaller than 2.0 Å indicates similar structures of a protein pair, distinct motions of a few residues within ligand binding site remind us the deficiency of the overall RMSD that it may hide local motions. To explore how pocket residue moves in responding to ligand binding, we firstly calculated the RMSD matrix of each residue around ligand by 5.0 Å for protein pairs of apo and ligand-bound (holo) structures. As observed by Rebecca et al., the protein dynamics could lead to the opening, closing and adaptation of binding pocket, resulting in the appearance/disappearance of a sub-pocket or an allosteric pocket and the pocket breathing motion [38]. To further analyze the motion of pocket residues upon ligand binding, we calculated the pocket volume using the D3Pockets ([www.d3pharma.com/D3Pocket/index.php](http://www.d3pharma.com/D3Pocket/index.php)). It is well-known that one type of residue motion can dramatically regulate the “on” and “off” states of binding pocket, which is also called ‘gatekeeper’ [39]. For example, the R410 is the gatekeeper of the adenosine-binding site of NIK (NF- $\kappa$ B-inducing kinase) [40]. Noticeably, most of residue motions simply expand the space of binding pocket. A major reason of the expanding is the moving outwards of pocket residues. On the other hand, the fusion of more than two small sub-pockets provides a large space for ligand binding. On the contrary, to stabilize bound ligand or to take part in catalytic process, pocket residues need to approach the ligand, resulting in a shrinking of binding pocket. For example, the F293 in *apo* FOX-4 cephamycinase moves 2.5 Å inwards upon ligand binding, forming a putative T-shaped  $\pi$ -stacking interaction with the substrate [41]. The rest of residue motions, other than the above four types, have little effect on the space of binding pocket but form better interactions with ligands. Consequently, the residue motions could be classified into five classes (Fig. 1): (a) pocket-creating motion (PC), (b) pocket-expanding motion (PE), (c) pocket-fusing motion (PF), (d) pocket-shrinking motion (PS) and (e) other motion (OM). Each class is represented by a code of two characters: for instance,



PC stands for ‘pocket-creating motion.’ Finally, we collected a typical pair of the same type of motions with the largest residue RMSD in the D3PM.

Similarly, we calculated the RMSD matrix of pocket residues for protein pairs of holo structures, including the pairs bound with different ligands and the pairs bound with the same ligand. For the pairs bound with different ligands, there are 2,176,460 pairs with at least one residue’s RMSD that is greater than 2.0 Å. We then selected a typical pair with the largest RMSD, and obtained a final set of 1183 cases, viz., 793 pairs with the same ligand binding pocket and 390 pairs with different ligand binding pockets. The 390 cases could be regarded as protein pairs of apo and holo structures, which could also be classified into the five classes (PC, PE, PF, PS, and OM). For those pairs bound with the same ligand, a final set of 1465 cases was selected from pairs with at least one residue’s RMSD that is greater than 2.0 Å.

**For protein pairs with overall RMSD that is greater than 2.0 Å** In this set, protein conformational change may result from both the inherent flexibility and external perturbations like ligand binding. Consequently, to explore how those motions occur, we classified the protein pairs with overall RMSD that is greater than 2.0 Å into four parts: (a) pairs of apo structures, (b) pairs of apo and holo structures, (c) pairs of holo structures with different ligands, (d) pairs of holo structures with the same ligand. For the pairs of

apo structures, the inherent flexibility of protein contributes mostly to their conformational changes, since we have excluded the influence of protein–protein and protein–ligand interactions. These datasets containing apo-holo pairs and pairs of holo structures bound with different ligands should be a useful resource to evaluate the protein motions induced by ligands. There are 1111 protein pairs bound the same ligand with RMSD that is greater than 2.0 Å, which should result mainly from the inherently flexibility of protein–ligand complex. For the pairs of holo structures, we calculated the RMSD matrix of the pocket residues, and found that 125 apo-holo pairs have obvious pocket residue motions, which could also be classified into the five classes (PC, PE, PF, PS, and OM).

Finally, the D3PM collects 7649 proteins with overall motions and 3513 proteins with pocket residue motions, as shown in Table 1.

### Linkage of the D3PM and DrugBank databases

The DrugBank is a free web resource containing comprehensive drugs information with their targets, which greatly facilitates the drug discovery and development [42]. To make full use of protein motions for drug discovery, druggable targets in the DrugBank database are annotated in the protein motion list in the D3PM. For example, the target carbonic anhydrase 2 (PDB ID: 3HS4) can be found with three kinds of motions in the D3PM, including overall conformational changes caused by its inherent flexibility and ligands binding, and PE type of pocket residue motion.

## Utility and discussion

### User interface

For easy application, we constructed a web server, which is accessible at [www.d3pharma.com/D3PM/index.php](http://www.d3pharma.com/D3PM/index.php) (Fig. 2). The interface to the D3PM was designed to facilitate both detailed searching of protein motions and browsing of the whole database. In the website, users can navigate the protein motion list, and search the database by PDB ID, Uniprot ID, RMSD, residue and ligand name etc. Each entry includes detailed annotations such as PDB ID, Uniprot ID, overall RMSD, pocket RMSD etc. The D3PM has provided entries to download all the data. Three dimensional structures of pocket motions

**Table 1** Summary of the data available in the D3PM (update on 11th April 2021)

The overall protein motions	Number of protein	Overall RMSD (Å)	
		Mean	Maximum
apo & apo	3684	5.31	54.72
apo & holo	1643	4.68	39.53
holo & holo (different ligands)	1211	4.78	63.03
holo & holo (the same ligand)	1111	5.06	63.03
The pocket residues motions	Number of protein	Pocket RMSD (Å)	
		Mean	Maximum
apo & holo	1255	1.91	8.64
holo & holo (different ligands)	793	1.76	8.62
holo & holo (the same ligand)	1465	1.10	9.30

The “apo” referred to as ligand-free protein, and the “holo” referred to as ligand-bound protein. The mean and maximum values were calculated with those representative pairs collected in the D3PM

**D3PM**  
D3PM protein motions database

Menu

Home

**Overall protein motions**

- ligand-free structures
- ligand-free and ligand-bound structures
- structures bound with different ligand
- structures bound with same ligand

**Pocket residue motions**

- ligand-free and ligand-bound structures
- structures bound with different ligand (ex-situation)
- structures bound with different ligand (in-situation)
- structures bound with same ligand

Welcome to D3PM protein motions database!

This database describes all kinds of protein motions, including changes upon ligand binding and the residue over well Drug

Show 25 entries

Uniprot ID	structure 1	structure 2	RMSD	DurgBank target
ADA010	5B02_D	5B0J_B	4.99	
ADA010	5B02_D	5B0J_B	4.99	

**Overall protein motions**

**Pocket residue motions**

- (A) pocket-creating motion (PC):88
- (B) pocket-expanding motion (PE):617
- (C) pocket-forming motion (PF):44
- (D) pocket-shrinking motion (PS):234
- (E) other motion (O):57

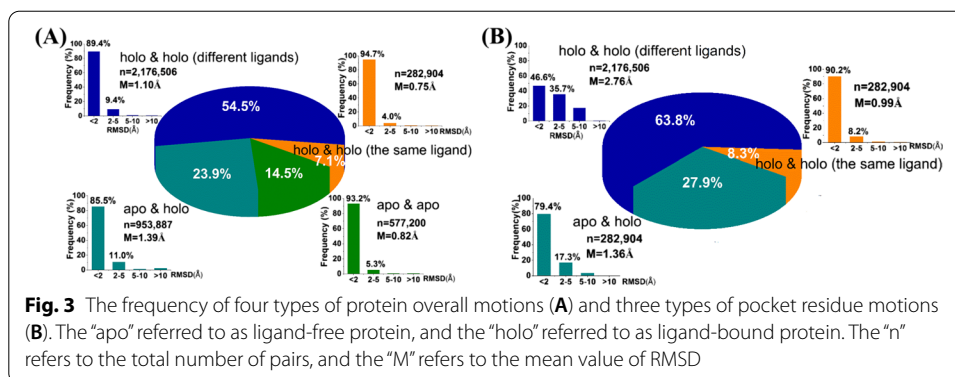
**Fig. 2** The web page of the D3PM database: **A** the overview of types of protein motions included in the D3PM, **B** diagrams for two main types of protein motions, viz. overall protein motions and pocket residue motions, **C** the detailed information of each protein motion

could be shown by JSmol software [43] after clicking the “structure” button. For the structural pairs, the first and second structures are highlighted in red and yellow, respectively. The user has the option to show the aligned structures in cartoon or sticks with the label of pocket residue’s name, and the option to download PDB files containing the aligned structures.

### Comparison of different types of protein motions

The D3PM provides sufficient samples to study protein motions caused by either the inherent flexibility of macromolecule or ligand binding. In the D3PM, 7,730,788 protein pairs are classified into four classes, viz. (a) pairs of apo structures, (b) apo-holo pairs, (c) pairs bound with different ligands, (d) pairs bound with the same ligand. By searching the database, we found 1970 proteins forming 3,990,497 protein pairs, among which each protein possesses all the 4 different motion types.

If a protein pair has overall RMSD that is smaller than 2.0 Å, it was regarded as motionless. In Fig. 3A, the pairs bound with the same ligand have a larger proportion of motionless pairs (94.7%) than that of the pairs of apo structures (93.2%), indicating the weaker ability of the ligand-bound proteins to undergo overall structural motions. The result can be rationalized with the fact that ligand somewhat stabilizes protein structure. However, it is noteworthy that there are nearly 5% of protein pairs bound with the same ligand that have RMSD that is greater than 2.0 Å, which is largely accomplished by flexible loops such as the active loop of kinases. The proportion of motionless for protein pairs bound with different ligands is 89.4%, which is 5.3% less than the pairs bound the same ligand. It indicated that the protein conformational adaptation induced by ligands is somewhat related to the structure of ligand. The apo-holo pairs have the smallest



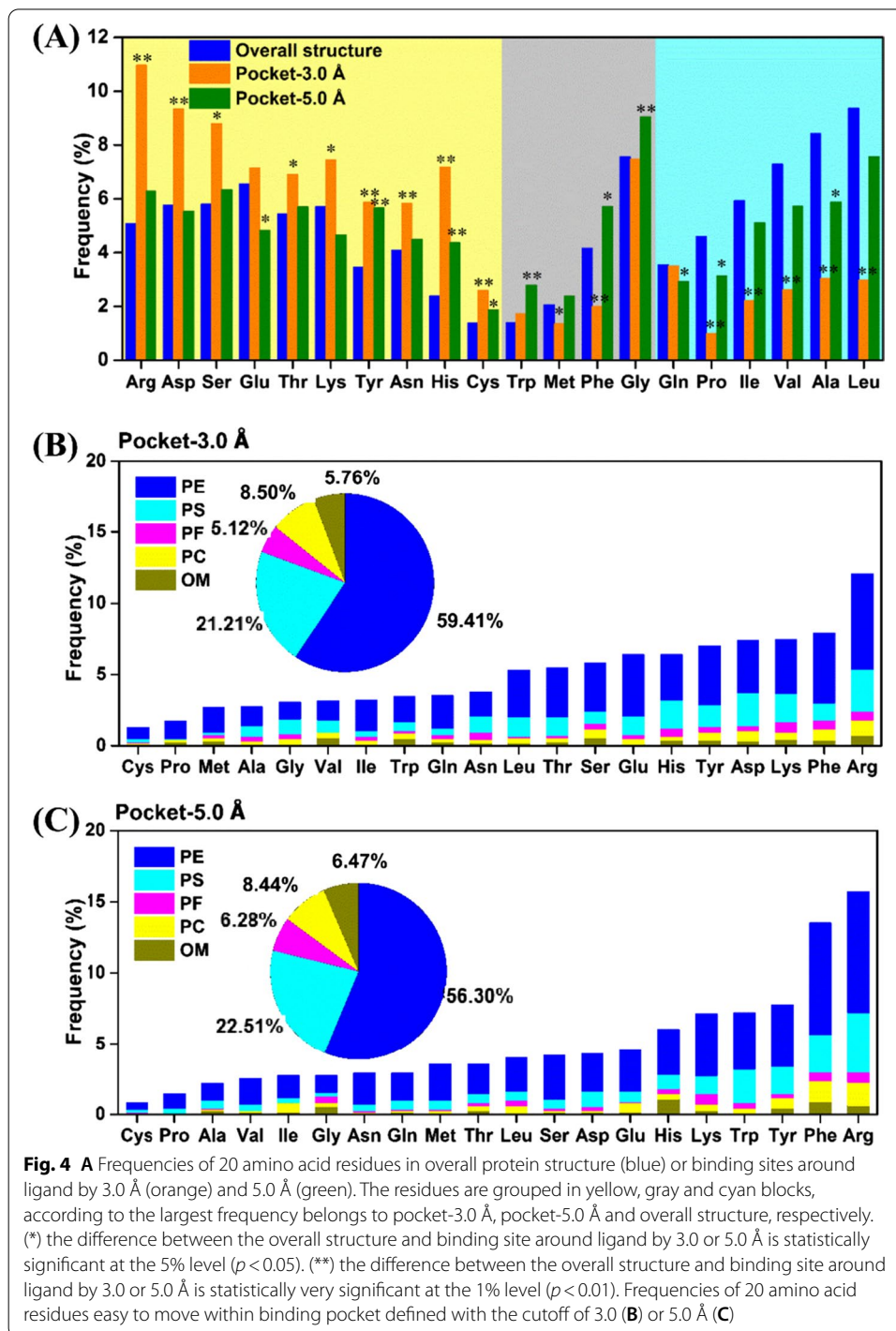
proportion of motionless pairs (85.5%), showing that the ligand binding causes the most significant protein motions. However, it is important to note that most proteins have no obvious overall structural changes upon ligand binding, because the proportions of motionless pairs for all the 4 types of motions are greater than 85%.

To explore pocket residue motions, the RMSD of pocket residues that around ligand by 5.0 Å was calculated (Fig. 3B). Similarly, the bound ligand reduces the flexibility of protein binding pocket, according to the largest proportion of RMSD smaller than 2.0 Å (90.2%) of the pairs bound with the same ligand. However, protein conformational change induced by ligands is much significant on binding pocket, the proportion of motionless pairs with different ligands (46.6%, Fig. 3B) is 42.8% less than that of overall structure (89.4%, Fig. 3A). In other words, more than half of pocket residues have significant structural changes upon ligand binding, implying the importance of the flexibility of pocket residues for virtual screening. In addition, the pairs bound with different ligands have the largest mean value of RMSD (2.76 Å). Therefore, all the results demonstrated that the ligand binding could cause protein conformational changes, especially in binding pocket, however, could also stabilize induced conformations.

### The amino acid preference of binding pocket

Interactions with pocket residues are indispensable to the binding process of ligands, e.g., hydrogen bond, hydrophobic interaction and so on. In order to evaluate how amino acids that bind to ligands (pocket residues) differ from that of overall structure, the residues around ligands and the whole protein were analyzed. Usually, the definition of pocket residues is the ones with a minimum distance to ligand shorter than 5.0 Å [44]. With 178,778 protein–ligand complexes, the residue frequency of binding pocket was calculated by using different distance that around ligand from 2.0 to 6.0 Å. The mean unsigned error (MUE) of the frequencies of the 20 amino acids between binding pocket and overall structure was calculated (Additional file 1: Fig. S1). The distance of 3.0 Å has the largest difference of residue frequencies between binding pocket and overall structure. The larger the distance than 3.0 Å, the smaller the difference, indicating that the cutoff of 3.0 Å could best distinguish binding pocket from overall structure. Therefore, using the distances of pocket residues to ligand of 3.0 and 5.0 Å, we analyzed the frequencies of 20 amino acids for binding pocket and for overall structure, respectively. The Arg, Asp, Ser, Glu, Thr, Lys, Tyr, Asn, His and Cys in binding pocket around ligand by

3.0 Å significantly outweigh that in overall structure (Fig. 4A), indicating that they are more inclined to interact with ligands to form short-range interactions such as hydrogen bond and ionic bond. The frequencies of Gly, Phe, Met and Trp within 5.0 Å of ligands outweigh the corresponding ones in overall structure, indicating they are more inclined to interact with ligands to form long-range interactions. The 14 residues that are likely





to interact with ligands to form short-range or long-range interactions, could be called “pocketphilic”. Other residues like Leu, Ala, Val, Ile, Pro and Gln have lower frequencies in binding pocket compared with overall structure with both the cutoff of 3.0 and 5.0 Å, which could be called “pocketphobic”.

To further explore the frequency of pocket residues that are easy to motion in responding to ligand binding, we analyzed the pocket residue motions in the D3PM. As shown in Fig. 4B & C, most motions are PE type with a frequency that is greater than 56%. The “pocketphilic” residues (Arg, Phe, Tyr, Lys, Glu, and Asp) are easier to move than “pocketphobic” residues. The Arg that has the longest side chain is the easiest to motion. However, it is not necessarily that the longer the side chain of residue is, the easier it is to move. For example, the motion frequency of Tyr is smaller than that of Phe. It is also interesting to note that basic residues (Arg and Lys) are easier to move than acidic residues (Asp and Glu) within ligand binding pockets.

#### **Case study: cross-docking reveals the importance of the pocket residue motions**

The current strategy of virtual screening using a selected inhibitor bound conformation as receptor structure may miss putative ligands, due to protein conformational adaptations in ligand binding site. To evaluate how significant conformational adaptations are for molecular docking, the set of holo proteins bound with different ligands was used for cross-docking of ligand to a bound receptor structure crystallized in the presence of another ligand. The results (Additional file 1: Table S2) showed that the average docking score for ligands docked to its cocrystallized receptor is  $-9.24$  kcal/mol, however, the value is obvious smaller for ligands docked to other holo structures of the same protein, which is  $-8.67$  kcal/mol. In addition, 23% cross-docking cases have a difference of docking score greater than 1 kcal/mol. Taking the difference of 1 kcal/mol as a threshold, the enrichment factor is 1.94. As also shown in receiver operating characteristic (ROC) curves (Additional file 1: Fig. S2), the area under ROC curve (AUC) is 0.70. Therefore, the results showed that the flexibility of pocket residue need to be considered carefully during virtual screening.

#### **Conclusions**

We developed the D3PM database to analyze all kinds of protein motions involving overall structures and binding pocket residues. In addition, we classified pocket residue motions into 5 types for studying different function mechanism of ligand binding. Currently, the information provided in the D3PM is in list form. The D3PM will be regularly updated to reflect new entries in the PDB database.

Using the D3PM, we firstly compared the ability of different factors that are related to protein conformational changes. The results showed that protein motions induced by ligands are significant in binding pocket according to 53.4% of protein pairs have pocket RMSD greater than 2.0 Å, but only less than 15% of protein pairs have obvious overall conformational adaptation. However, there are nearly 5% of protein pairs bound with the same ligand have overall RMSD greater than 2.0 Å. Although factors of both external perturbations like ligand binding and intrinsic flexibility of macromolecule have been studied here, there are still other factors like pH, temperatures and mutation that can impact protein motions, which is valuable for further study.

In addition, we analyzed the preferences of 20 amino acids in binding pocket. The results revealed some residues likely to interact with ligands by forming short-range or long-range interactions, which could be called “pocketphilic”. However, “pocketphobic” residues like Leu, Ala, Val, Glu, Ile, Pro and Gln have smaller frequencies in binding pocket compared with that in overall structure. The results could provide important information for pocket prediction.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-022-04595-0>.

**Additional file 1.** The Additional file 1 contains Figures S1 and S2, Tables S1 and S2. **Figure S1** shows the mean unsigned error of the frequencies of the 20 amino acids between the pocket and overall structure. **Figure S2** shows the ROC curve for the cross docking case. **Table S1** shows the list of crystallographic additives. **Table S2** shows the detailed docking scores of the cross docking case.

## Acknowledgements

Not applicable

## Authors' contributions

Z.X. and W.Z. designed the research; C.P. and X.Z. performed the research; C.P., Z.C., Y.Y. and T.C. analyzed data for the work; C.P., Z.X. and W.Z. drafted and revised the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by the National Key Research and Development Program (2016YFA0502301), the National Natural Science Foundation of China (81872797), and National Science & Technology Major Project “Key New Drug Creation and Manufacturing Program”, China (2018ZX09711002). The funders played no role in the design of the study and collection, analysis, and interpretation of data, and in writing the manuscript.

## Availability of data and materials

The database generated and analyzed is available at [www.d3pharma.com/D3PM/index.php](http://www.d3pharma.com/D3PM/index.php).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>CAS Key Laboratory of Receptor Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, People's Republic of China. <sup>2</sup>University of Chinese Academy of Sciences, No. 19A Yuquan Road, Beijing 100049, People's Republic of China. <sup>3</sup>Open Studio for Druggability Research of Marine Natural Products, Pilot National Laboratory for Marine Science and Technology (Qingdao), 1 Wenhai Road, Aoshanwei, Jimo, Qingdao 266237, People's Republic of China.

Received: 14 August 2020 Accepted: 1 February 2022

Published online: 14 February 2022

## References

1. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol.* 2002;9:646–52.
2. Berendsen HJC. Collective protein dynamics in relation to function. *Curr Opin Struct Biol.* 2000;10:165–9.
3. Venkatraman M, Alan CG, Maxwell DC, et al. Docking: successes and challenges. *Curr Pharm Des.* 2005;11:323–33.
4. Huang SY, Zou X. Advances and challenges in protein-ligand docking. *Int J Mol Sci.* 2010;11:3016–34.
5. Rousse A, Rischel C, Gauthier JC. Colloquium: femtosecond X-ray crystallography. *Rev Mod Phys.* 2001;73:17–31.
6. Bennett WS, Huber R, Engel J. Structural and functional aspects of domain motions in proteins. *Crit Rev Biochem.* 2008;15:291–384.
7. Skjaerven L, Hollup SM, Reuter N. Normal mode analysis for proteins. *J Mol Struct.* 2009;898:42–8.
8. Karplus M, McCammon JA. The internal dynamics of globular protein. *Crit Rev Biochem.* 1981;9:293–349.
9. Careri G, Fasella P. Statistical time events in enzymes: a physical assessment. *Crit Rev Biochem.* 1975;3:141–64.
10. Gurd FRN, Rothgeb M. Motions in proteins. *Adv Prot Chem.* 1979;33:73–165.

11. Cooper A. Conformational fluctuation and change in biological macromolecules. *Sci Prog.* 1980;66:473–97.
12. Konrad H, Aline T, Field MJ. Analysis of domain motions in large proteins. *Proteins.* 1999;34:369–82.
13. Gerstein M, Anderson BF, Norris GE, et al. Domain closure in lactoferrin. *J Mol Biol.* 1993;234:357–72.
14. Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. *Biochem.* 1994;33:6739–49.
15. Nishikawa K, Ooi T, Isogai Y, et al. Representation and computation of the conformations. *J Phys Soc Jpn.* 1972;32:1331–7.
16. Qi G, Lee R, Hayward SA. Comprehensive and non-redundant database of protein domain movements. *Bioinformatics.* 2005;21:2832–8.
17. Echols N. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.* 2003;31:478–82.
18. Piovesan D, Tabaro F, Paladin L, et al. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Res.* 2018;46:D471–6.
19. Flores S, Echols N, Milburn D, et al. The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res.* 2006;34:D296–301.
20. Gelin BR. Sidechain torsional potentials and motion of amino acids in proteins: bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A.* 1975;72:2002–6.
21. Gutteridge A, Thornton J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J Mol Biol.* 2005;346:21–8.
22. Noble MA, Miles CS, Chapman SK, et al. Roles of key active-site residues in flavocytochrome P450 BM3. *Biochem J.* 1999;339:371–9.
23. Keizers PH, Lussenburg BM, de Graaf C, et al. Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regioselectivity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism. *Biochem Pharmacol.* 2004;68:2263–71.
24. Yongxia G, Zheng Z, James JLC, et al. Smoke-derived karrikin perception by the  $\alpha/\beta$ -hydrolase KAI2 from *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2013;110:8284–9.
25. Wang Y, Jardetzky O. Investigation of the neighboring residue effects on protein chemical shifts. *J Am Chem Soc.* 2002;124:14075–84.
26. Bernstein FC, Koetzle TF, Williams GJB, et al. The protein data bank: a computer-based archival file for macromolecular structures. *Eur J Biochem.* 1977;80:319–24.
27. Lobanov MY, Shoemaker BA, Garbuzynskiy SO, et al. ComSin: database of protein structures in bound (complex) and unbound (single) states in relation to their intrinsic disorder. *Nucleic Acids Res.* 2010;38:D283–7.
28. Chang DT, Yao TJ, Fan CY, et al. AH-DB: collecting protein structure pairs before and after binding. *Nucleic Acids Res.* 2012;40:D472–8.
29. Hrabec T, Li Z, Sedova M, et al. PDBFlex: exploring flexibility in protein structures. *Nucleic Acids Res.* 2016;44:D423–8.
30. Amemiya T, Koike R, Kidera A, et al. PSCDB: a database for protein structural change upon ligand binding. *Nucleic Acids Res.* 2012;40:D554–8.
31. Teague SJ. Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov.* 2003;2:527–41.
32. Tokuriki N. Protein dynamism and evolvability. *Science.* 2009;324:203–7.
33. Juritz EI, Alberti SF, Parisi GD. PCDB: a database of protein conformational diversity. *Nucleic Acids Res.* 2011;39:D475–9.
34. Monzon AM, Rohr CO, Fornasari MS, et al. CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. *Database (Oxford).* 2016;2016:baw038.
35. Chang CW, Chou CW, Chang DT. CCPProf: exploring conformational change profile of proteins. *Database (Oxford).* 2016;2016:baw029.
36. Paul N, Rognan D. ConsDock: A new program for the consensus analysis of protein-ligand interactions. *Proteins.* 2002;47:521–33.
37. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31:455–61.
38. Stank A, Kokh DB, Fuller JC, et al. Protein Binding Pocket Dynamics. *Acc Chem Res.* 2016;49:809–15.
39. Emrick MA, Lee T, Starkey PJ, et al. The gatekeeper residue controls autoactivation of ERK2 via a pathway of intramolecular connectivity. *Proc Natl Acad Sci U S A.* 2006;103:18101–6.
40. de Leon-Boenig G, Bowman KK, Feng JA, et al. The crystal structure of the catalytic domain of the NF-kappaB inducing kinase reveals a narrow but flexible active site. *Structure.* 2012;20:1704–14.
41. Lefurgy ST, Malashkevich VN, Aguilan JT, et al. Analysis of the Structure and Function of FOX-4 Cephamycinase. *Antimicrob Agents Chemother.* 2016;60:717–28.
42. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018;46:D1074–82.
43. Herraes A. Biomolecules in the computer—Jmol to the rescue. *Biochem Mol Biol Educ.* 2006;34:255–61.
44. Borrel A, Regad L, Xhaard H, et al. PockDrug: a model for predicting pocket druggability that overcomes pocket estimation uncertainties. *J Chem Inf Model.* 2015;55:882–95.

## Publisher's Note

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