D3PM: A Comprehensive Database for Protein Motions Ranging from Residue to Domain

Cheng Peng
Shanghai Institute of Materia Medica Chinese Academy of Sciences
Xinben Zhang
Shanghai Institute of Materia Medica Chinese Academy of Sciences
Zhijian Xu
Shanghai Institute of Materia Medica Chinese Academy of Sciences
Zhaoqiang Chen
Shanghai Institute of Materia Medica Chinese Academy of Sciences
Yanqing Yang
Shanghai Institute of Materia Medica Chinese Academy of Sciences
Tingting Cai
Shanghai Institute of Materia Medica Chinese Academy of Sciences
Weiliang Zhu (wlzhu@simm.ac.cn)
Shanghai Institute of Materia Medica https://orcid.org/0000-0001-6699-5299

Database

Keywords: Protein motion, Motion patterns, D3PM database, Amino acids preference

Posted Date: September 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-59550/v1

Version of Record: A version of this preprint was published at BMC Bioinformatics on February 14th, 2022. See the published version at https://doi.org/10.1186/s12859-022-04595-0.

Abstract

Background: Knowledge of protein motions is significant to understand its functions. The currently available databases for protein motions, in general, are focused on overall domain motions, which pay little attention to local residue motions. Albeit with relatively small scale, the local residue motions may play crucial roles in protein functions and its binding with ligand, in particular for those residues within binding pockets.

Results: A comprehensive protein motion database (D3PM) was constructed in this study to facilitate the analysis of protein motions. The D3PM has the motion information ranging from the overall structural changes of macromolecule to the local flip motion of the residues in ligand binding site. Currently, the D3PM has 5,339 entries of overall motions and 2,319 entries of pocket residues' motions. The motion patterns in the database are classified into 4 types of overall structural change and 5 types of pocket residues' shift. Impressively, it was found that less than 15% of the protein pairs have obvious overall conformational adaptations induced by ligand binding, while more than 50% of the protein pairs have significant structural changes in the ligand binding sites, indicating that ligand-induced conformational changes are drastic whereas they are mostly confined around the ligand. By the analysis of pocket residues' preference, we classified amino acids into "pocketphilic" and "pocketphobic" residues, which is helpful to pocket prediction and ligand design.

Conclusion: D3PM is a comprehenssive database about protein motions ranging from residue to domain, which should be useful for exploring diverse protein motions and for understanding protein functions. The database is freely available on www.d3pharma.com/D3PM/index.php.

Background

The conformational diversity of proteins is rooted from their structures and is often a key feature of their functions[1, 2]. A fundamental recognition of how proteins work therefore requires knowledge of their structures and dynamism, which also plays a considerable role in drug discovery and development. For instance, the ensemble docking that treats the protein flexibility is a more popular approach in virtual screening[3, 4]. Such conformational diversity can be studied in various ways. X-ray crystallography and NMR spectroscopy are versatile experimental techniques to provide biomolecular structures[5, 6]. In computational methods, normal mode analysis and molecular dynamics can be used in the refinement of X-ray and NMR structures, as well as the prediction of conformational diversity[7]. With more and more protein structures becoming available, there is an increasing interest to relate such structures to protein motions for studying their functions.

The protein motion has been of particular interest in many areas of the biological sciences. Many studies of protein motions have focused on the domain hinge and shear motions, which has also been a topic of a number of reviews[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 have been complied. Databases of protein domain motions have been also available in recent years[16–19]. In addition, the mobile part in protein collected in some databases has ranged from small loops to entire subunits besides domains. However, in many cases, there are few domain motions for proteins under different conditions such as before and after ligand binding, but significant side-chain motions of pocket residues or catalytic residues[20, 21], which play a crucial role in responding to the access, regiospecificity, stabilization and dissociation of ligand, enzyme catalysis, etc[22, 23]. For example, the most pronounced conformational change simply occurs for F194 in KAI2 protein with a ~ 90° flip of its benzene ring when bound with inhibitor KAR_1[24]. Furthermore, the dynamic residues may impact the conformation of neighboring regions in proteins[25]. Therefore, it is necessary to study those local motions of residues within ligand binding pocket for understanding protein functions.

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 database, but whose entries are redundant for structures of identical protein determined under diverse conditions. Lots of effort has gone into collecting and analyzing these vast amount of data in the PDB, leading to a large number of databases. The MolMovDB[18] is a dominant database containing conformational changes of macromolecule with focus biased towards a collection of data and software pertaining. Other databases, such as the ComSin[27], AH-DB[28], PDBFlex[29], and PSCDB[30] provide protein structures pairs in bound and unbound states to explore protein motions triggered by ligand, among which the AH-DB contains the most entries (> 700,000). However, the motions are sophisticated involving the intrinsic flexibility of protein or experimental conditions (such as temperatures and pH), besides external perturbations like ligand binding[31, 32]. The PCDB[33] and CoDNaS[34] provide a redundant structures' cluster for the same protein under different experimental protocols, but the PCDB has been not available for a while. The CCProf[35] is another conformational diversity database which contains 986,187 protein structure pairs before and after binding and 10 biological features for studying binding target site. However, using these databases to study the local motions of pocket residues is difficult, and there is still a lack of such specific database so far.

In this study we constructed a database that covers all kinds of protein motions ranging from overall structure to local residue, and classified pocket residue motions into 5 types. Considering that the form of structural pairs is more convenient to analyze protein motion features than that of structural cluster, all kinds of protein motions were provided with protein structural pairs in the D3PM. We believe that the D3PM will be helpful to explore diverse protein motions that related to their functions, and promote the drug discovery and development.

Construction And Content

D3PM database construction

All the X-ray structures were downloaded from the PDB with resolution better than 3.0 Å (25th October 2018), and were divided into pairs of identical proteins with 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. Shown in Table S1), which were manually removed in protein-ligand structures. Finally, 7,730,788 pairs of protein structures obtained from the PDB were divided into two parts, according to the overall C_α root mean square deviation (RMSD) threshold of 2.0 Å, which was often used as a threshold for drug discovery[36, 37]. There are a number of redundant pairs of identical protein for the same type of motions. Therefore, a representative pair of identical protein with the most significant motions was selected to construct a non-redundant, contrastive and classified protein motions database.

For the pairs with overall RMSD less than 2.0 Å. Although the overall RMSD of less than 2.0 Å indicates similar structures in pairs of proteins, dramatic motions of a few residues within ligand binding site reminds us the deficiency of the RMSD that hides local motions. To explore how the pocket residues move in responding to ligand binding, we firstly calculated the RMSD matrixes of each pocket residue that around ligand by 5.0 Å for pairs of apo and ligand-bound (holo) structures. It results in 609,433 pairs with at least one residue's RMSD ≥ 2.0 Å. We secondly selected a representative pair with the same motions and the largest residue RMSD, resulting in a final set of 568 cases. To further classify the pocket residues' motions, we analyzed the changes of pocket size on ligand binding using D3Pockets (www.d3pharma.com/D3Pocket/index.php). It is well-known that one type of residue motions can dramatically regulate the binding pocket "on" and "off" states, which is also called 'gatekeeper'[38]. For example, a typical feature inconsistent with its active state of apo NF-κB-inducing kinase is the presence of the guanidinium group of R410 in the adenosine-binding site[39]. However, most of motions simply expand the space of the binding pocket, involving two main causes. A major reason of the expanding is the pocket residues' moving outward. On the other hand, the fusion of more than two small pockets also provides a large enough binding space for ligands. On the contrary, to stabilize bound ligand or take part in the 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 toward the substrate upon ligand binding, forming a putative T-shaped π-stacking interaction[40]. The rest of pocket residues' motions, other than the above four types, have little effect on the space of pocket but form better interactions with ligand. Consequently, the pocket residues' 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'.

Similarly, we also calculated the RMSD matrixes of each pocket residue between pairs of holo structures, including pairs bound with different ligands and 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 ≥ 2.0 Å. We then selected a representative pair with the largest pocket residues' RMSD, resulting in a final set of 967 cases including 620 pairs with the same ligand bound pocket and 347 pairs with different ligand bound pocket.

In fact, the 347 cases could be regarded as pairs of apo and holo pockets. Therefore, they were also classified into five classes (PC, PE, PF, PS and OM). For those pairs bound with the same ligand, there are 1,297,976 pairs with at least one pocket residues' RMSD ≥ 2.0 Å. We then selected a typical pair with the largest pocket residues' RMSD, resulting in a final set of 750 cases.

For the pairs with overall RMSD greater than 2.0 Å. In this set, conformational changes may result from both the inherent flexibility and external perturbations like ligand binding. Consequently, to explore how those motions occur, we classified the pairs with overall RMSD ≥ 2.0 Å into four parts: (a) pairs of apo structures, (b) pairs of apo structure and holo structure (the ligand bound structure), (c) pairs of holo structures with different ligands, (d) pairs bound with the same ligand. For each part, we chose a typical pair of the same motion and protein with less missing residues and larger RMSD, i.e., the one with the least missing residues from top 5 pairs with the largest RMSD. Finally, 5,339 representative pairs were collected. For those pairs of apo structures, inherent flexibility of proteins contributes mostly to their conformational changes, since we have excluded the influence of protein-protein and protein-ligand interactions. Motions induced by the ligand intrigue us when we focus on hit identification by structure-based virtual screening. Set contains apo-holo pairs and pairs of holo structures bound with different ligands is a resource to observe and distinguish the motion induced by ligands. 878 pairs bound with the same ligand have RMSD ≥ 2.0 Å, mostly resulting from the inherently flexibility of protein-ligand complex. For pairs including holo structures, the RMSDs of pocket residues were calculated. 125 apo-holo pairs have obvious pocket residues' motions, which were also classified into five classes (PC, PE, PF, PS and OM).

Finally, the D3PM includes 5,339 entries of overall motions and 2,319 entries of pocket residues motions from 7,730,788 pairs of protein structures as shown in Table 1.

Table 1. Summary of the data available in the D3PM*

The overall protein motions	Number of pairs	Overall RMSD (Å)	
		Mean	Maximum
apo & apo	2173	4.32	41.76
apo & holo	1282	4.34	39.53
holo & holo (different ligands)	1006	4.65	32.02
holo & holo (the same ligand)	878	4.39	26.50
The pocket residues motions		Pocket RMSD (Å)	
		Mean	Maximum
apo & holo	949	1.45	6.42
holo & holo (different ligands)	620	2.60	9.68
holo & holo (the same ligand)	750	1.33	2.50

*: the “apo” referred to as ligand-free protein, and the “holo” referred to as ligand-bound protein.

Linkage of the D3PM and DrugBank databases

The DrugBank is a free available web resource containing comprehensive molecular information about drugs and their targets, which greatly facilitates the drug discovery and development[41]. To make full use of protein motions for drug discovery, drug targets in the DrugBank database are annotated in the protein motion’s list of the D3PM. For example, the drug 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 residues’ motion.

Utility And Discussion

User interface

For easy application, we constructed a web server, which is accessible at www.d3pharma.com/D3PM/index.php, as shown in Fig. 2. The interface to the D3PM has been designed to facilitate both detailed searching of protein motion types and browsing of the whole database. With the website, the users can navigate through the motion types of a specific target protein. Moreover, users can also search the database by PDB ID, Uniprot ID, RMSD, residue’s and ligand’s name etc. Each entry has detail annotation such as PDB ID, Uniprot ID, overall RMSD, pocket RMSD etc.

Comparison of different types of protein motions

The D3PM provides an approach to explore the protein motions caused by either the inherent flexibility of macromolecule or ligand binding or both. In the D3PM, 7,730,788 pairs with the same Uniprot ID are

classified into four types, *viz*, (a) pairs of apo structures, (b) apo-holo pairs, (c) pairs bound with different ligands, (d) pairs bound with the same ligand. To compare the 4 types of protein motions, the pairs included in all the 4 types were selected, resulting in 3,990,497 pairs from 1,970 proteins with the same Uniprot ID. The pairs were regarded as motionless in terms of overall protein structures if their overall RMSD is < 2.0 Å. Consequently, the larger proportion of motionless pairs, the weaker ability to cause overall protein structure motions.

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%). This can be rationalized with the fact that the bound ligands somewhat rigidize protein structures. However, it is noteworthy that there are nearly 5% of the protein-ligand pairs with RMSD ≥ 2.0 Å, which is largely accomplished by the flexible loop structures' motions that are independent of ligand binding such as the active loop of kinases. The proportion of motionless pairs bound with different ligands is 89.4%, with a difference as large as 5.3% compared with the pairs bound with the same ligand. It showed that the protein conformational adaptation induced by ligand binding is somewhat related to the structure of ligands, even though this effect is not very obvious. The apo-holo pairs have the smallest proportion of motionless pairs (85.5%), showing that ligand binding causes the most significant protein conformational adaptation. However, it is important to note that most of the proteins have no obvious overall structural changes on ligand binding, according to all the proportions of motionless pairs of the 4 types are larger than 85%.

To further explore the pocket residues' motions, the RMSD of pocket residues that within 5.0 Å of ligands was calculated (Fig. 3B). Similarly, the bound ligand decreases the flexibility of binding pocket, according to the largest proportion of RMSD < 2.0 Å (90.2%) of the pairs bound with the same ligand. However, structural changes induced by ligands obviously are more significant on the binding pocket, as the smallest proportion of motionless pocket (46.6%) belongs to the pairs bound with different ligands, with a difference as large as 22.8% compared with that of the overall structure. Therefore, more than half of proteins have significant structural changes in binding pockets when bound with different ligands, implying the importance of the flexibility of pocket residues for virtual screening. All the results demonstrated that ligand binding could cause protein conformational changes, especially on the binding pocket, however, stabilize the induced conformations.

The amino acid preference of binding pocket

Interactions with binding pocket residues are indispensable to ligand's binding process, e.g., hydrogen bond, hydrophobic interaction and so on. In order to refresh how amino acids interacting with ligands (pocket residues) differ from that of overall structure, the pocket residues around ligand and whole protein were analyzed. Usually, the definition of pocket residues are the ones with the minimum distance to the ligand of ~5.0 Å[42]. With 178,778 protein-ligand complexes, the residue frequencies of binding pocket were calculated with different definition that around ligand by distances ranging from 2.0 to 6.0 Å. The mean unsigned error (MUE) of the frequencies of the 20 amino acids between the pocket and overall structure was calculated (Fig. S1). The distance of 3.0 Å has the largest difference in residue frequencies

between the binding pocket and overall structure. The larger distance than 3.0 Å, the smaller difference, which indicates that the cutoff value of 3.0 Å could best distinguish the pocket from overall structure. Therefore, using a definition of pocket residues that around ligand by 3.0 and 5.0 Å, we analyzed the frequencies of 20 amino acids for the binding pocket and overall structure, respectively. The Arg, Asp, Ser, Glu, Thr, Lys, Tyr, Asn, His and Cys in binding pocket around ligand by 3 Å significantly overweigh that in overall structure (Fig. 4A), indicating that they are more inclined to interact with the ligand 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 overweigh 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 interaction, could be called “pocketphilic”. Other residues like Leu, Ala, Val, Ile, Pro and Gln have lower frequencies in pocket compared with overall structure with both the cutoff value of 3.0 and 5.0 Å, which could be called “pocketphobic”. The results also provide useful information to improve pocket prediction.

To further explore which residues are easy to shift their positions in pocket in responding to ligand binding, we analyzed the frequencies of residues with RMSD ≥ 2.0 Å using the pocket residues motions database of the D3PM. As shown in Fig. 4B & C, most of the motions are PE type with frequency >56%. The “pocketphilic” resides (Arg, Phe, Tyr, Lys, Glu and Asp) are also easier to move than the “pocketphobic”. In particular, the Arg with the longest side chain is easiest to shift in pockets defined by the cutoff either of 3.0 or 5.0 Å. However, residues with long side chain are not necessarily easy to shift in pockets, such as the motion frequency of Tyr with long side chain is lower than that of Phe. It is also interesting to note that the basic residues (Arg and Lys) are easier to move than the acidic residues (Asp and Glu) in ligand binding pockets.

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

The current strategy of virtual screening based on a selected inhibitor bound conformation as a target protein structure may miss putative ligands, due to the conformational adaptations in the ligand binding site. To evaluate how significant the different conformational adaptations are, the set of 620 pairs of holo proteins with different ligands were used for cross-docking of ligand to a bound receptor structure crystallized in the presence of another ligand. The results of 1240 data (Table S2) showed that the average docking score of ligands docking to its co-crystallized receptor is -9.24, however, the value is obvious smaller when the ligands docking to another ligand bound receptor structure, which is -8.67. In addition, 23% cross-docking cases have the difference of docking score more than 1 kcal/mol, revealing a variety of putative compound may be missed during virtual screening in thousands or even millions of candidate compounds. Therefore, the pocket and its residues flexibility should be treated carefully during virtual screening.

Conclusions

We developed the D3PM to present all kinds of protein motions involving overall structures and residues within binding pockets. In addition, we classified pocket residues' motions into 5 types to explore different function mechanism of ligand binding. Currently, the information provided in the D3PM is in list form. We are currently developing an approach to predict conformational change pathway using a pair of different structures[43]. Therefore, our planned developments of the D3PM will include (i) providing movies of all kinds of protein motions and (ii) regularly updating to process new entries in the PDB database.

Using the D3PM, we firstly compared the ability of different factors to drive protein conformational changes. The results showed that protein motions induced by ligands is significant in the binding pockets according to 53.4% of protein pairs with pocket RMSD ≥ 2.0 Å, but more than 85% of protein pairs have no obvious overall conformational adaptations. The ligands bound rigidize the binding pocket, due to the conformational adaptation of pocket residues to form or stabilize strong protein-ligand interactions. However, there are still nearly 5% of the protein pairs bound with the same ligand with overall RMSD ≥ 2.0 Å. Although the factors including 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 the binding pocket. The results revealed some residues likely to interact with ligand 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 lower frequencies in binding pocket compared with overall structure. The results provided important information for further pocket prediction.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The database generated and analyzed is available at www.d3pharma.com/D3PM/index.php.

Competing interests

The authors declare that they have no competing interests.

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.

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. analysed data for the work; C.P., Z.X. and W.Z. drafted and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Not applicable

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. Femtosecond x-ray crystallography. Rev Mod Phys. 2001;73:17.
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 assessmen. Crit Rev Biochem. 1975;3:141 – 64.
10. Gurd FRN, Rothgeb TM. **Motions in proteins**. Adv Prot Chem. 1979; **33**.
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-D6.
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 Inhibito. Proc Nat Acad Sci. 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. **et al**. Influence of phenylalanine 120 on cytochrome P450 2D6 catalytic selectivity and regiospecificity: crucial role in 7-methoxy-4-(aminomethyl)-coumarin metabolism. Biochem Pharmacol. 2004;68:2263-71., **de** Graaf C.
24. Yongxia G, Zheng Z, James JLC. **et al**. Smoke-derived karrikin perception by the α/β-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. Hrabe 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**.
35. Chang CW, Chou CW, Chang DT. **CCProf**: **exploring conformational change profile of proteins**. **Database** (**Oxford**). 2016; **2016**.
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. 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.
39. **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.
40. 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.
41. 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-D82.
42. 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.
43. Wang J, Shao Q, Xu Z, et al. Exploring transition pathway and free-energy profile of large-scale protein conformational change by combining normal mode analysis and umbrella sampling molecular dynamics. J Phys Chem B. 2014;118:134–43.

Figures

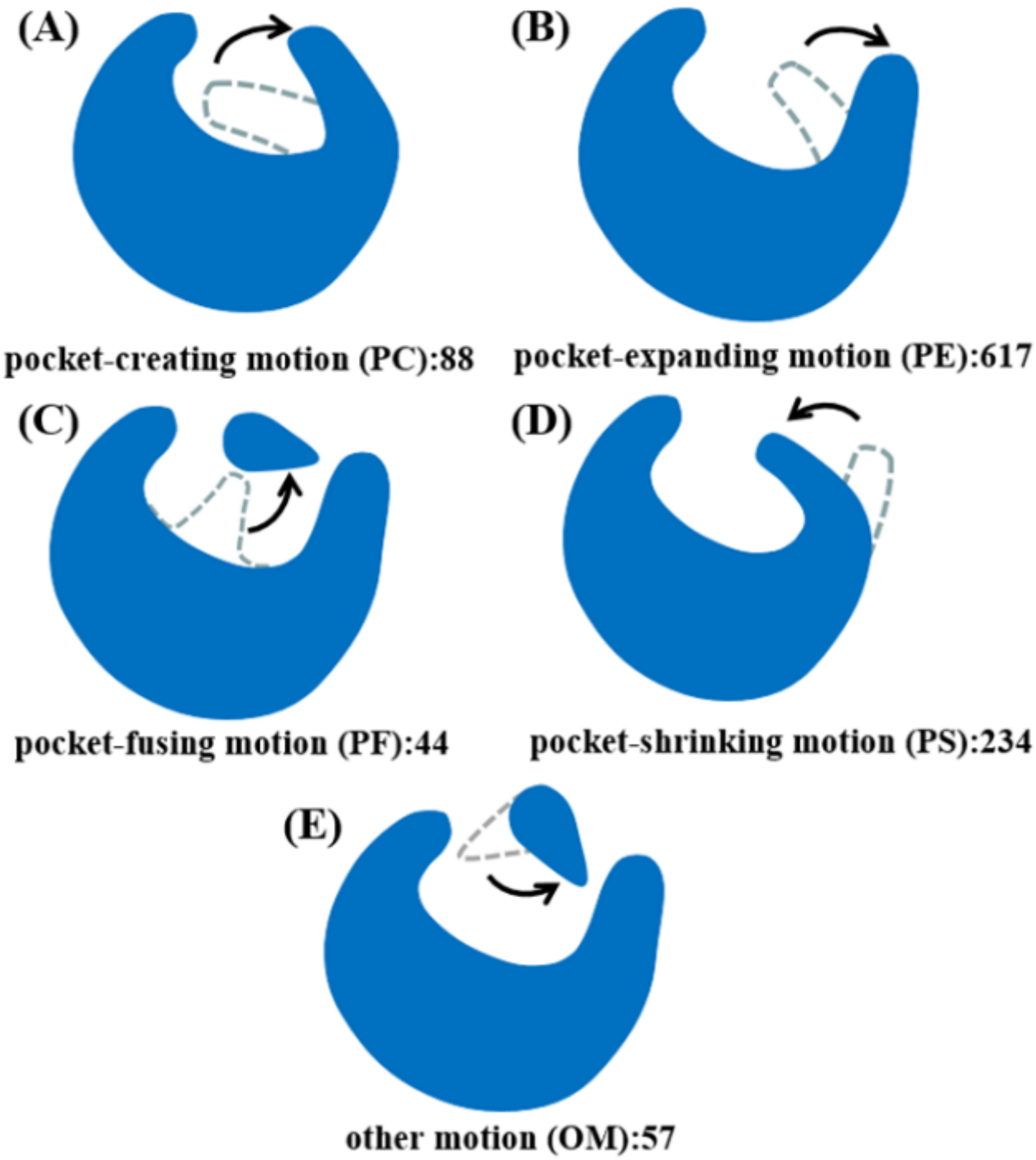

Figure 1

Five classes of pocket residue motions. (A) pocket-creating motion (88), (B) pocket-expanding motion (617), (C) pocket-fusing motion (44), (D) pocket-shrinking motion (234) and (E) other motion (57).

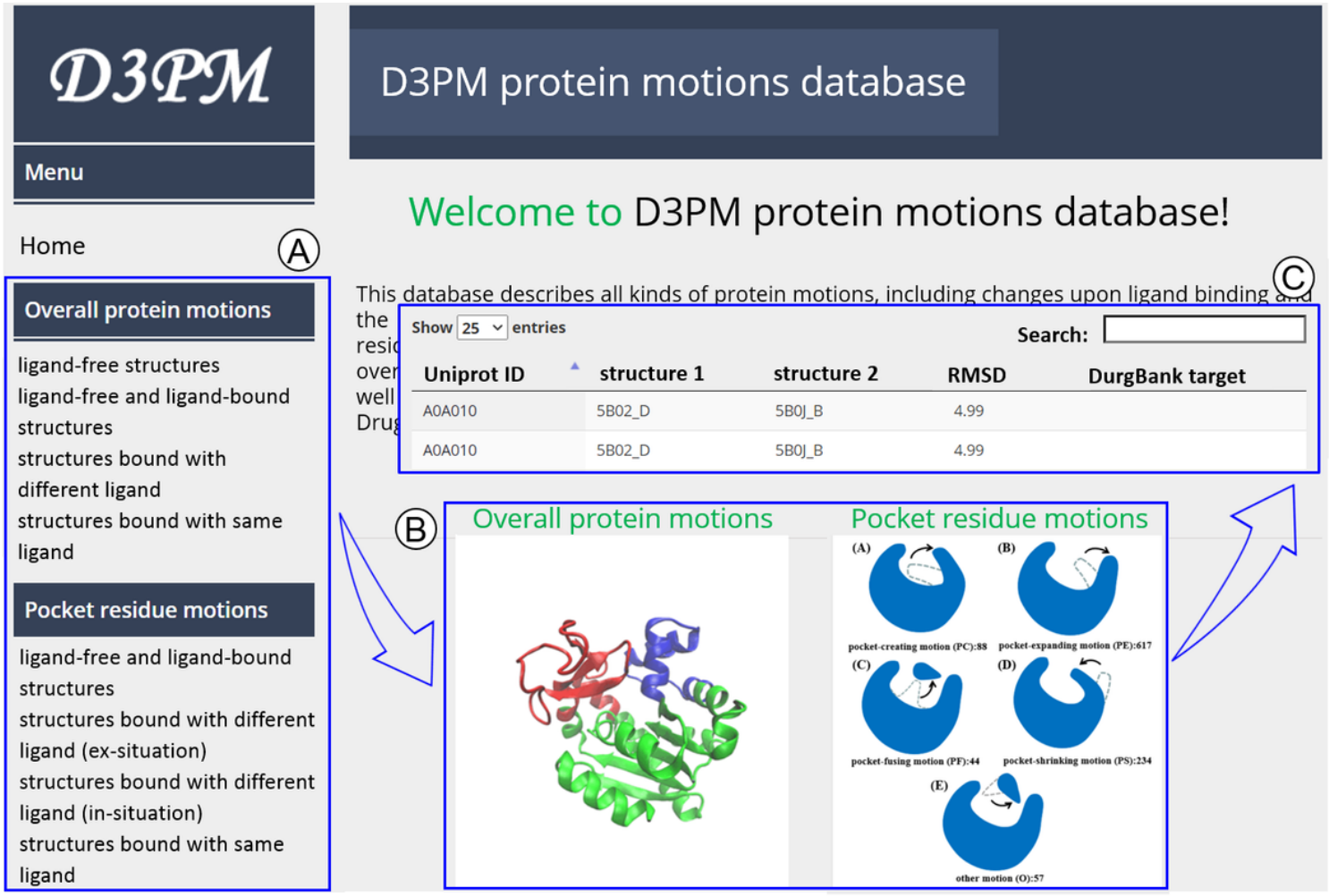

Figure 2

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

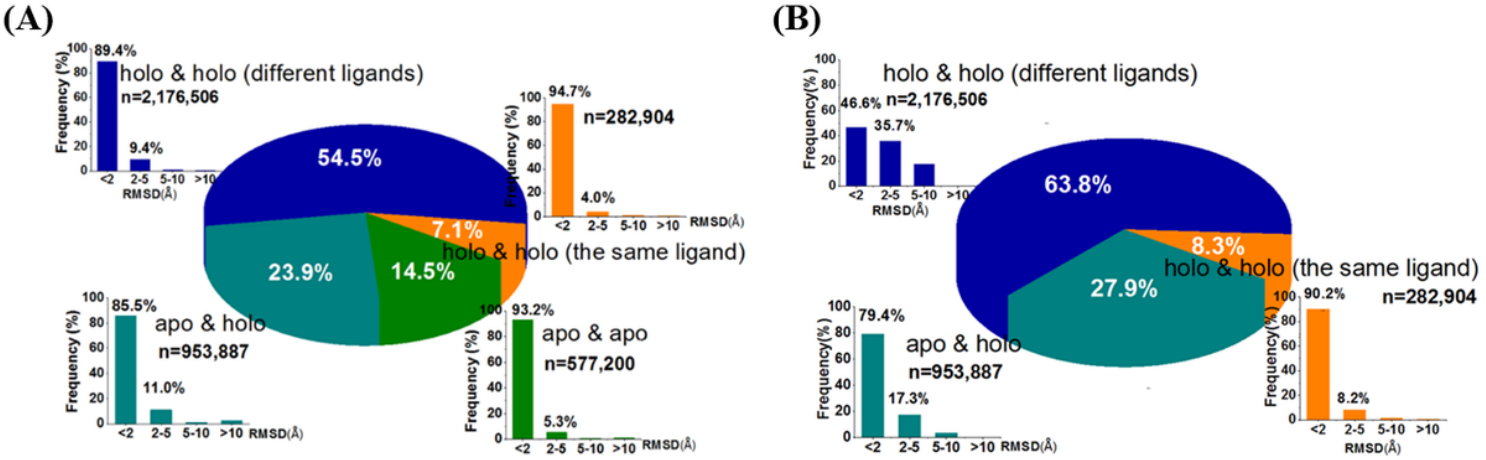

Figure 3

The frequency of four types of protein overall motions (A) and three types of pocket residue motions (B). The “apo” referred to as ligand-free protein, and the “holo” referred to as ligand-bound protein.

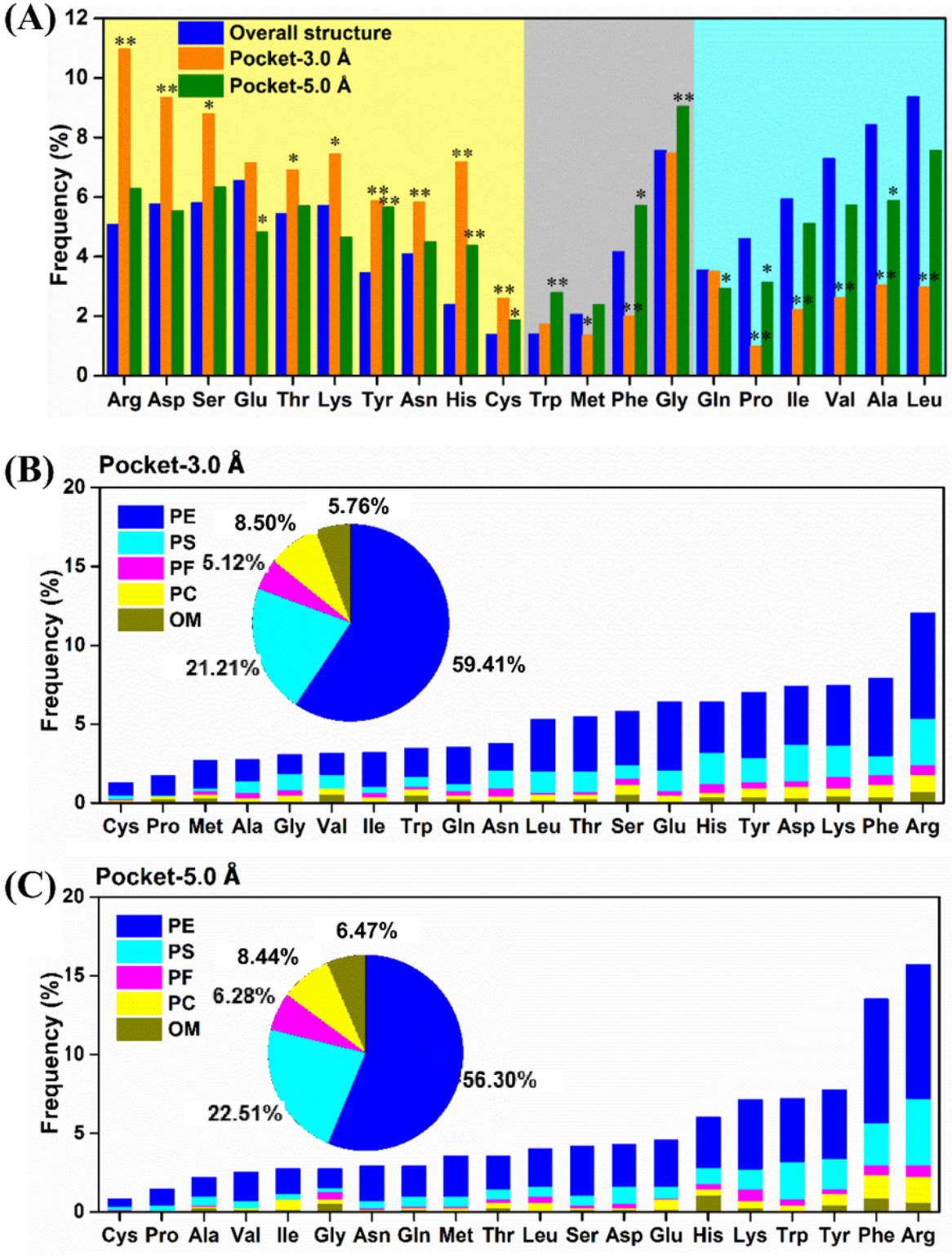

Figure 4

(A) Frequencies of 20 amino acid residues in overall protein structure (blue) or binding sites around ligand by 3.0 Å (orange) and 5.0 Å (green). The residues are grouped in yellow, gray and cyan blocks, according to largest frequency belongs to pocket-3.0 Å, pocket-5.0 Å and overall structure, respectively. (*)

the difference between the overall structure and binding site around ligand by 3.0 or 5.0 Å is statistically significant at the 5% level ($p < 0.05$). (**) the difference between the overall structure and binding site around ligand by 3.0 or 5.0 Å is statistically very significant at the 1% level ($p < 0.01$). Frequencies of 20 amino acid residues easy to move in pocket, defined with the cutoff of 3.0 (B) or 5.0 Å (C)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SID3PM0813.pdf