

DATABASE

Open Access



# GPCR-BSD: a database of binding sites of human G-protein coupled receptors under diverse states

Fan Liu<sup>1,2</sup>, Han Zhou<sup>1,2,3</sup>, Xiaonong Li<sup>3</sup>, Liangliang Zhou<sup>3</sup>, Chungong Yu<sup>1,4</sup>, Haicang Zhang<sup>1,4</sup>, Dongbo Bu<sup>1,4,5\*</sup> and Xinmiao Liang<sup>1,2,3\*</sup>

\*Correspondence:  
dbu@ict.ac.cn; liangxm@dicp.ac.cn

<sup>1</sup> University of Chinese Academy of Sciences, Beijing 101408, China

<sup>2</sup> Key Laboratory of Phytochemistry and Natural Medicines, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, Liaoning, China

<sup>3</sup> Jiangxi Provincial Key Laboratory for Pharmacodynamic Material Basis of Traditional Chinese Medicine, Ganjiang Chinese Medicine Innovation Center, Nanchang 330000, Jiangxi, China

<sup>4</sup> SKLP, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

<sup>5</sup> Central China Institute of Artificial Intelligence, Zhengzhou 450046, Henan, China

## Abstract

G-protein coupled receptors (GPCRs), the largest family of membrane proteins in human body, involve a great variety of biological processes and thus have become highly valuable drug targets. By binding with ligands (e.g., drugs), GPCRs switch between active and inactive conformational states, thereby performing functions such as signal transmission. The changes in binding pockets under different states are important for a better understanding of drug-target interactions. Therefore it is critical, as well as a practical need, to obtain binding sites in human GPCR structures. We report a database (called GPCR-BSD) that collects 127,990 predicted binding sites of 803 GPCRs under active and inactive states (thus 1,606 structures in total). The binding sites were identified from the predicted GPCR structures by executing three geometric-based pocket prediction methods, fpocket, CavityPlus and GHECOM. The server provides query, visualization, and comparison of the predicted binding sites for both GPCR predicted and experimentally determined structures recorded in PDB. We evaluated the identified pockets of 132 experimentally determined human GPCR structures in terms of pocket residue coverage, pocket center distance and redocking accuracy. The evaluation showed that fpocket and CavityPlus methods performed better and successfully predicted orthosteric binding sites in over 60% of the 132 experimentally determined structures. The GPCR Binding Site database is freely accessible at <https://gpcrbs.bigdata.jcm-sc.cn>. This study not only provides a systematic evaluation of the commonly-used fpocket and CavityPlus methods for the first time but also meets the need for binding site information in GPCR studies.

**Keywords:** G-protein coupled receptor, State-specific structure, Binding site, Key residue, Database

## Background

G-protein coupled receptors (GPCRs) constitute the largest family of membrane proteins in human body and serve as highly prevalent signaling hubs, transmitting numerous extracellular signals and drugs into intracellular pathways [1]. The identified 826 GPCRs can be classified into 5 classes, including classes A (rhodopsin), B



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

(secretin and adhesion), C (glutamate), F (Frizzled) and T (Taste) [2], and approximately 34% of FDA-approved drugs target GPCRs [3].

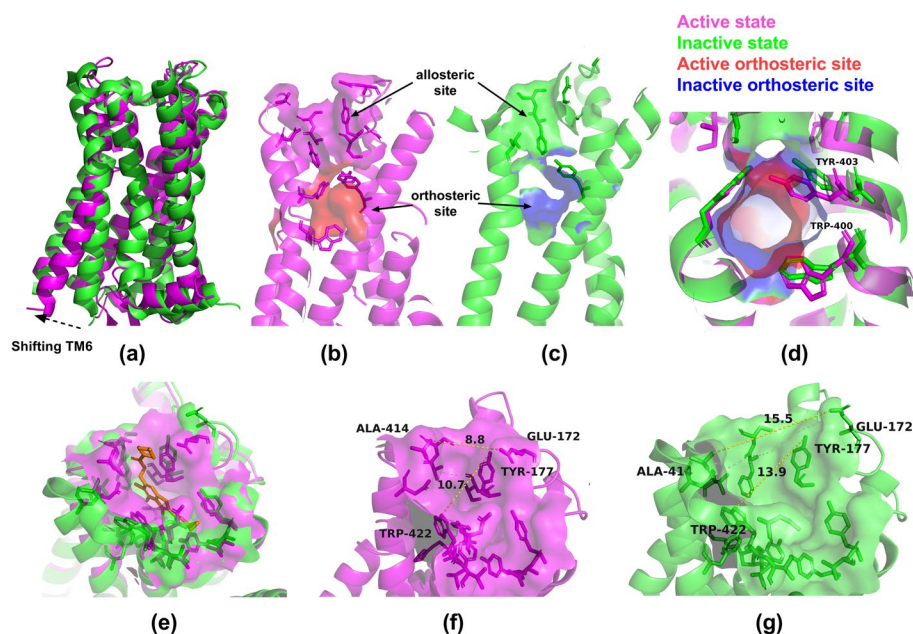
GPCR structures are characterized by seven transmembrane helices (TMHs) that generate moderately high levels of hydrophobicity, and three extracellular and three intracellular loops that are flexible and participate in ligand identification and G-protein attachment [4]. GPCRs exist in two interconvertible conformations: active ( $R^*$ ) and inactive ( $R$ ) states, which are driven by extracellular stimuli, such as the binding of an agonist, and result in the release of coupled G-protein [5–7].

The characteristic structure of GPCRs and the conformation changes with states make it crucial to understand their precise structures. However, it is challenging to obtain experimentally determined crystal structures of GPCRs due to their high hydrophobicity and low thermostability. As of September 2023, only 156 human GPCRs, less than 20% of all 826 GPCRs, have their structures experimentally resolved and archived in the Protein Data Bank (PDB). As a result, most of the current studies on the characteristics of GPCR binding sites were conducted based on their experimental structures [8–11].

In 2021, AlphaFold2 (AF2) achieved remarkable advancements in predicting the structures of human proteins, including the unresolved GPCRs [12, 13]. Since then, a variety of progress has been achieved in the recognition of novel active compounds based on AF2-predicted GPCR structures [14–16]. In 2022, Heo, et al. [17] predicted GPCR structures in both  $R$  and  $R^*$  states using state-specific input templates and developed an AlphaFold-Multistate dataset, which compensates the lack of multi-state conformations in the original AlphaFold2 database.

Gaining insights into the specific ligand binding sites of GPCRs is essential for understanding drug-target interactions and serves as a foundation for structure-based drug design. Taking human muscarinic acetylcholine receptor M2 as an example, it is a representative class A GPCR with an orthosteric binding site in the middle of the seven TMHs and an allosteric binding site near the third extracellular loop (Fig. 1), and the hallmark feature of the activation process is the outward movement and rotation of TMH6 on the cytosolic side, which leads to a significant change of both the orthosteric and the allosteric binding pocket surfaces between the two states. For the orthosteric pocket, repacking of residues W400 and Y403 leads to a change on the pocket surface. For the allosteric site, by comparing the distance between the two pairs of amino acids (W422-Y177 and A414-E172) that make up the surface of the pocket, it can be seen that the pocket in the active state narrows and thus forms good contacts with positive allosteric modulators [18], whereas the inactive state does not.

Recognition of the binding sites of GPCRs is a prerequisite for learning drug-target interactions. For proteins having experimentally resolved structures, if they are bound with ligands, the binding site is usually recognized based on the position of the ligand. For proteins lacking bound ligands in the PDB structures or those with predicted structures, the recognition of binding sites continues to pose a challenge. GPCRs possess unique structural characteristics, making traditional methods for identifying binding sites heavily reliant on aligning subfamilies or homologous members. However, when dealing with GPCRs lacking such loosely related members, a more generalized technique for detecting binding sites becomes necessary.



**Fig. 1** Comparison of muscarinic acetylcholine receptor M2 receptor structure in active (PDB: 4MQT) and inactive (PDB: 5YC8) states: **a** shifting in TM6 backbones; **b** orthosteric and allosteric binding sites in the active state; **c** orthosteric and allosteric binding sites in the inactive state; **d** shifting of the orthosteric binding pocket residues and surfaces **e** shifting of the allosteric binding pocket surfaces, **f** pocket width of the allosteric binding site measured in the active state, **g** pocket width of the allosteric binding site measured in the inactive state

Over the years, several approaches have been developed to predict the binding sites of protein structures. Currently, these methods can be classified into two types: geometric-based methods such as PocketPicker [19], fpocket [20], GHCOM [21–23] and CavityPlus [24, 25], and machine learning-based methods such as P2Rank [26], Deep-Pocket [27] and MaSIF-site [28]. In geometry-based methods, the detection of cavities within the protein surface is typically performed either by generating alpha spheres from Voronoi tessellation or by scanning through the protein surface or vacuum spaces with probes in different radii. Following cavity detection, various scoring functions are applied to each detected cavity based on their geometric and physicochemical properties [20]. Deep learning-based methods use neural networks to extract features from both sequence and structural information of proteins, enabling them to learn the intricate relationship between protein structures and ligand binding sites [27]. However, these methods usually rely on the results of geometric-based pocket detection as inputs and primarily focus on pocket evaluation or scoring.

Existing GPCR and binding pocket related databases each have their limitations. Based on the recent advancements of AlphaFold-MultiState, the GPCRdb database now contains a total of 1,102 PDB structures and incorporates state-specific predicted structures of 432 human GPCRs [29, 30]. The GPCRdb provides binding site analysis tools based on sequence and structure alignments. However, as mentioned above, it is difficult to cope with situations where GPCR subfamily members are small. Wang, et.al developed a CavitySpace [31] binding site library based on the AlphaFold2 prediction and the CAVITY algorithm [32]; however, AlphaFold2 only provides predicted structures in one

state. The GPCRmd [33] database contains 1,814 simulation trajectories and analyzes pockets using the mdpocket tool from the fpocket method. However, this database only includes experimentally resolved structures. Thus, it is necessary to develop a comprehensive binding site analysis for all human GPCR-predicted structures that consider different states.

In this study, we predicted the potential binding sites of 803 human GPCR predict structures in both R and R\* states using the most representative binding pocket prediction tools fpocket, CavityPlus and GHECOM, and presented the GPCR-BSD database. Users can visualize the binding sites, compare binding pocket profiles between different states or between predicted and resolved structures, and download binding site information. We also comprehensively evaluated their performance with a test dataset of 132 human GPCR experimentally determined structures (referred as PDB structures). The evaluation included pocket center distance, key residue coverage and redocking. This work will bridge a critical gap in the study of binding pockets on predicted GPCR structures. The resulting resource will serve as a valuable tool for various applications, including molecular simulations, drug design, and related fields, ultimately advancing our understanding of GPCR-ligand interactions and facilitating the development of therapeutics targeting these important receptors.

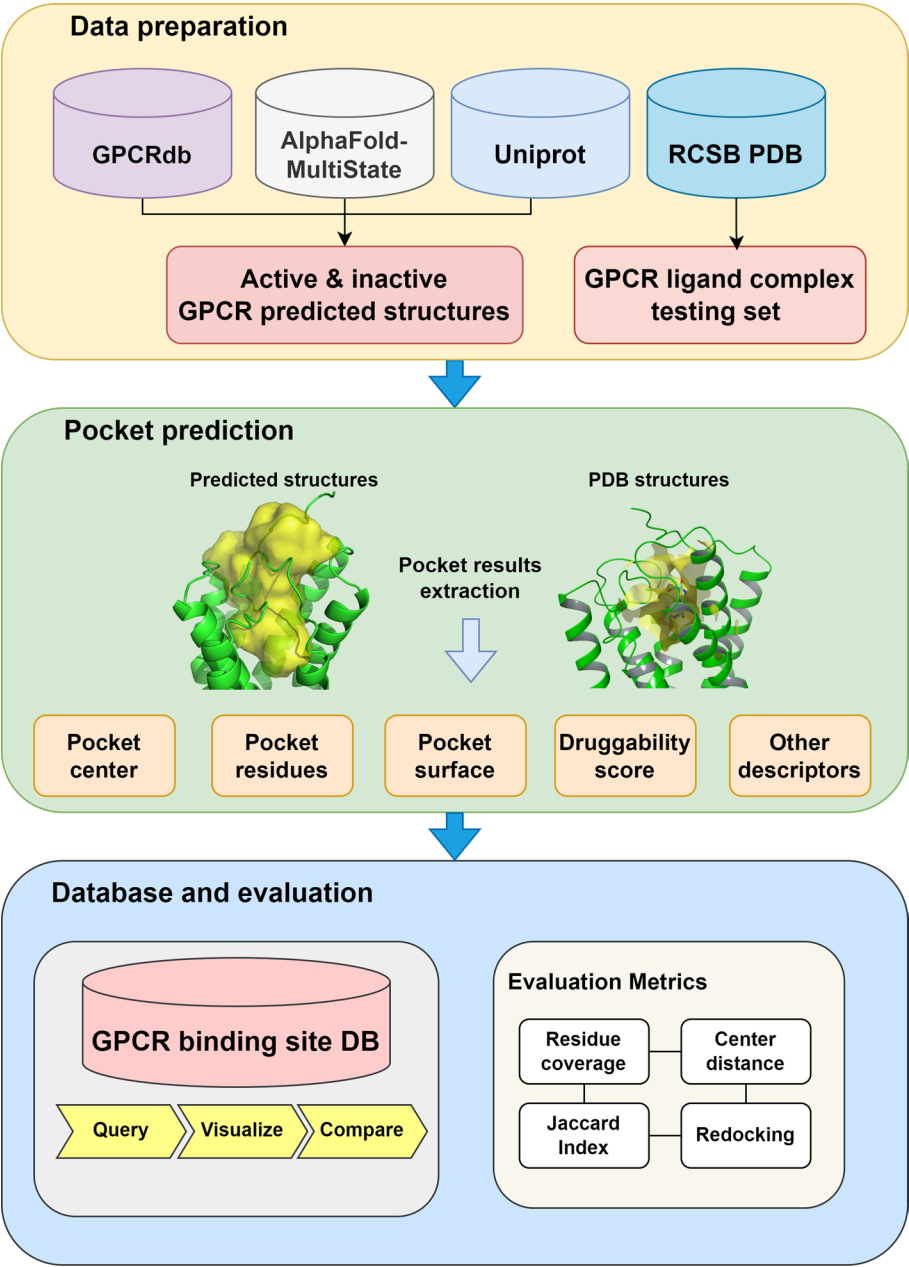
### Construction and contents

Our work included collecting and preparing data, predicting binding pockets, evaluating binding pockets, and building the database (Fig. 2).

#### Datasets

All predicted structures used for pocket prediction and database construction were obtained from public databases and datasets. The active and inactive state-predicted structures of human non-olfactory GPCRs were downloaded from GPCRdb (<https://gpcrdb.org>, accessed on 1 September 2023), which contains 846 active and inactive predicted models of 423 human GPCRs. The predicted structures of human olfactory GPCRs were downloaded from the AlphaFold-Multistate GitHub repository dataset ([https://huhlim.github.io/odorant\\_receptors](https://huhlim.github.io/odorant_receptors), accessed on 1 September 2023), which contains 814 active and inactive predicted models of 407 human GPCRs. We combined the two datasets with a list of 826 human GPCR entries. After data cleaning and deduplication, we selected 1606 predicted structures for 803 GPCRs, which excluded undefined GPCRs annotated by UniProt [34].

To evaluate the effectiveness of predicted binding sites on the predicted structure, a testing set of 132 experimental resolved structures of GPCR-small molecule complexes was collected. The testing set consists of 121 class A, 3 class B, 6 class C, one class F, and one taste receptor (Supplementary Table S1). All crystal structures have small molecule ligands near the center of 7 TMHs' extracellular side, which is the orthosteric site for class A, B, F, and T GPCRs and the allosteric site for class C. To test the prediction accuracy for allosteric binding sites that deviate from the center of the seven TMHS, we also evaluated an additional testing set consisting of 18 class A and 3 class B GPCRs with allosteric ligands binding (Supplementary Table S2). All predicted structures were aligned to the corresponding PDB structures.



**Fig. 2** The flowchart of building the GPCR binding site database and evaluating the pocket prediction results

**Predicting binding pockets**

We applied fpocket, CavityPlus and GHECOM tools to detect potential binding pockets for all predicted GPCR structures. To accommodate most GPCR cases, all programs were run with default parameters. Fpocket and GHECOM successfully processed all the 1606 predicted GPCR structures while CavityPlus failed on 40 pre-dicted structures of 20 GPCRs that have longer sequences and many irregular loops.

### Evaluating binding pocket results

Evaluations were performed based on comparisons between the original and predicted pockets and the redocking of original ligands into predicted pockets.

Considering both contacts and distance interactions, the original binding site residues are defined by all protein residues within 5 Å from any ligand atoms. The center of the original binding site is defined by the center of the ligand. After neglecting all other heteroatoms except the protein and the selected ligand, the original binding site residues and the coordinates of the original pocket centers are extracted from mmCIF files using the Biopython [35, 36] package.

Previous research provided many effective metrics for evaluating the predicted binding pockets [20, 26, 27, 32]. To compare different methods, we adapted three classical metrics: binding site residue coverage, binding site residue Jaccard index, and binding site center distance.

The residue coverage is defined as the ratio between the number of original binding site residues found in the predicted binding site and the number of original binding site residues:

$$\text{coverage} = \frac{\text{predicted binding site residues} \cap \text{original binding site residues}}{\text{original binding site residues}}$$

A successfully predicted binding pocket is defined as successfully predicted more than 70% of the residues around the original binding pocket.

The Jaccard index is defined as the ratio between the number of original binding site residues found in the predicted binding site and the number of the union of the original binding site residues and predicted binding site residues:

$$\text{Jaccard} = \frac{\text{predicted binding site residues} \cap \text{original binding site residues}}{\text{predicted binding site residues} \cup \text{original binding site residues}}$$

A Jaccard index closer to 1 implies that the predicted binding pocket is better overlapped with the original binding pocket. According to previous studies, a Jaccard index higher than 0.3 was set as a valid overlap for two pockets.

The pocket center distance is defined as the distance between the predicted pocket center and the center of the original binding site, which is defined by the center of the ligand. In molecular docking applications, ligand center sampling is typically done within a radius of 10 Å, therefore, we introduced an additional criterion where the distance between the predicted pocket center and the original ligand center should be within 5 Å.

To further compare the difference between the predicted pockets of different methods and analyze the factors affecting pocket prediction results for GPCRs, we additionally calculated two metrics: the pocket volume fold change and the pocket RMSD between the PDB structure and the corresponding predicted structure. The pocket volume fold change describes the comparison of spatial range between the predicted pocket and the original pocket. It is defined as the ratio between predicted pocket volume and original pocket volume. The volume of the predicted pocket is given by the output of the prediction software or server. The volume of the original pocket is calculated by dpocket [20] with a specified ligand. The pocket RMSD is calculated in PyMOL command script



under the following pipeline: First, the original PDB structure and predicted structure are aligned, then the full atom RMSD is calculated with all the original pocket residues defined.

For each prediction, we calculated the coverage of all predicted binding sites and selected the binding site with the highest coverage. Then for each selected binding site prediction result of the test structures, we calculated all the other metrics for further analysis.

To evaluate the performance of predicted pocket positions in real-world docking applications, 121 Class A (Rhodopsin) GPCRs in the testing set were chosen for Schrödinger Induced Fit Docking (IFD) [37–39] experiments using ligands extracted from the original crystal structures. The ligands were prepared using the Ligprep module with no change in ionization, chirality or tautomers. To ensure versatility across various scenarios, the docking protocol employed the Schrödinger IFD's default parameters. For experimentally resolved structures, the criterion of successful docking is the ligand root mean square deviation (RMSD)  $< 2$  Å. Considering the deviation of the predicted structure would affect the RMSD and the center distance between the post-docked ligand poses and the ligand poses in the original crystal structures, we calculated the coverage of pocket residues interacting with the ligand redocking pose in the predicted pocket. The redocking is considered a success when the redocking pose retains more than 50% of the original interacting key residues.

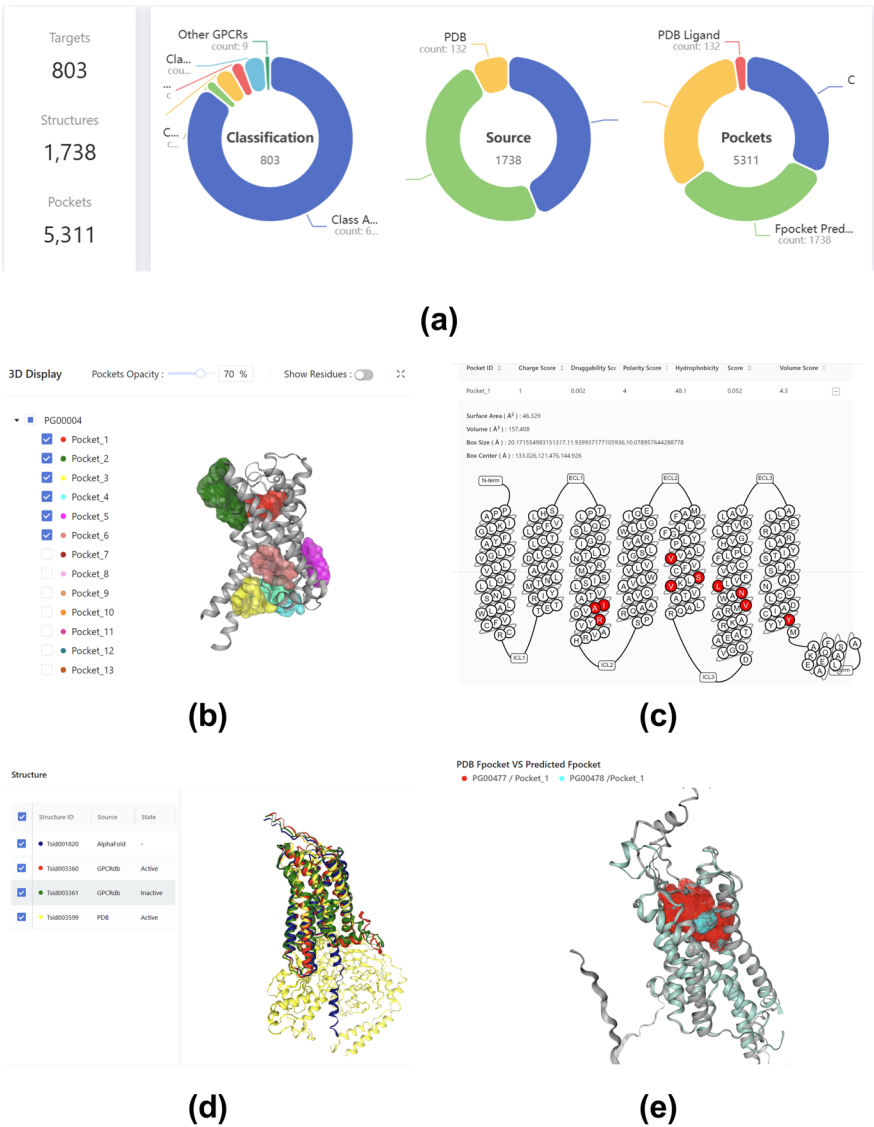
## Utility and discussion

### The GPCR-BSD database

We developed a publicly accessible web server that allows users to conveniently access the pocket information of all human GPCRs (Fig. 3). The server collected 57,904, 21,901 and 43,355 pockets predicted by fpocket, CavityPlus and GHECOM tools respectively for the 1,606 active and inactive GPCR predicted structures, allowing users to query, display, compare, and download data for human GPCRs' predicted pockets. Users can query the database using UniProt Accession numbers, protein names, or gene names and retrieve predicted structures or PDBs of the corresponding proteins in different states as well as their pocket prediction results. The database provides an intuitive interface for visualizing protein structures and binding sites, along with displaying their vacant pocket surfaces (Fig. 3b). Each predicted pocket includes a list of neighboring residues displayed in a snake plot, pocket property parameters provided by the prediction methods, as well as the coordinates of the docking grid center and box parameters (Fig. 2c). Users can rank the pockets by any property parameter or filter pockets by manually inputting the residue list. The database also provides aligned comparisons between predicted and PDB structures in different states (Fig. 3d, e). All pocket prediction results can be downloaded for research uses.

### Evaluating the predicted GPCR binding sites

To compare the changes in the binding pocket of GPCRs between active and inactive states, we compared 30 experimentally resolved structures in the testing set that had both states (Figure S1). The result suggests although the RMSD of binding site atoms



**Fig. 3** The GPCR Binding Site database features: **a** The statistics of all structures and pocket data; **b** 3D visualization of surfaces of predicted binding pockets; **c** snake plot visualization of pocket residues and docking grid parameters; **d** comparison of different structures; **e** comparison of predicted pockets in different structures

between two states is not significant, their interacting residues can be quite different, which illustrates the potential value of the multi-state binding pocket prediction.

To evaluate the prediction performance and adaptability of the three methods on GPCRs, comprehensive assessments were performed based on the testing set with both pocket comparison metrics and redocking.

The results of pocket prediction based on the testing set are shown in Table 1. For experimentally determined structures, under the threshold of pocket residue coverage > 70%, GHECOM successfully predicted 98% of the testing cases and CavityPlus successfully predicted 92% of the testing cases, while fpocket had 71% successful predictions. However, under the criterion of Jaccard index > 0.3, fpocket had 95% successful



**Table 1** Statistics of pocket prediction results by both pocket prediction methods under different criteria

Prediction method	Structure type	Coverage > 70%	Jaccard index > 0.3	Center distance < 5 Å	Both coverage and center distance criteria	Both coverage and Jaccard index criteria
fpocket	Experimentally determined structure	94/132 (71%)	125/132 (95%)	102/132 (77%)	81/132 (61%)	93/132 (70%)
CavityPlus	Experimentally determined structure	121/132 (92%)	60/132 (45%)	84/132 (64%)	84/132 (64%)	60/132 (45%)
GHECOM	Experimentally determined structure	130 (98%)	21/132 (16%)	40/132 (30%)	40/132 (30%)	21/132 (16%)
fpocket	Predicted structure	77/132 (58%)	108/132 (82%)	86/132 (65%)	58/132 (44%)	73/132 (55%)
CavityPlus	Predicted structure	105/132 (80%)	47/132 (36%)	53/132 (40%)	49/132 (37%)	42/132 (32%)
GHECOM	Predicted structure	118/132 (89%)	28/132 (21%)	20/132 (15%)	19/132 (14%)	26/132 (20%)

testing cases which outperformed the other two methods by 50% and 79% respectively. This indicates that CavityPlus and GHECOM tend to predict larger pockets. In terms of the predicted pocket center distance, fpocket outperformed CavityPlus by 7% under the threshold of center distance < 5 Å.

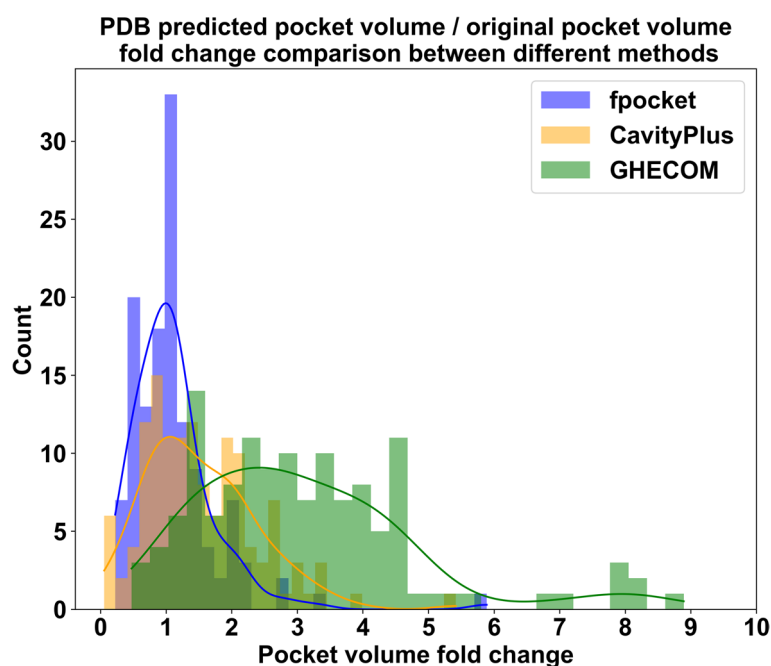
For the experimentally determined structures, both residue coverage and center distance criteria were considered, and the fpocket and CavityPlus methods successfully predicted over 60% of the testing cases, CavityPlus outperforming fpocket by 3%, while GHECOM only succeeded in 30% of the testing cases.

For the predicted structures, since the structural deviation would bias against the predicted pocket center distance and the residue coverage criterion would also bias towards larger pockets, the Jaccard index criterion should be considered alongside the residue coverage criterion. As a result, fpocket successfully predicted 55% of the testing cases and outperformed CavityPlus and GHECOM by 23% and 35% respectively.

These results showed that both fpocket and CavityPlus with default parameters were effective methods in identifying binding sites of experimentally determined structures, while GHECOM prediction results had a lower Jaccard index and higher center distance. The three methods had a better general performance on the experimentally resolved structures than on the predicted structures, and fpocket outperformed the other two methods.

To further investigate the difference in residue coverage and center distance of the prediction results, we compared the predicted versus the real pocket volume fold change in the experimentally determined structures.

As shown in Fig. 4, the pocket volume fold change of fpocket predicted results is more concentrated around 1 than that of CavityPlus, while for pocket volume fold change over 1.5, CavityPlus had more cases than fpocket. The predicted pocket



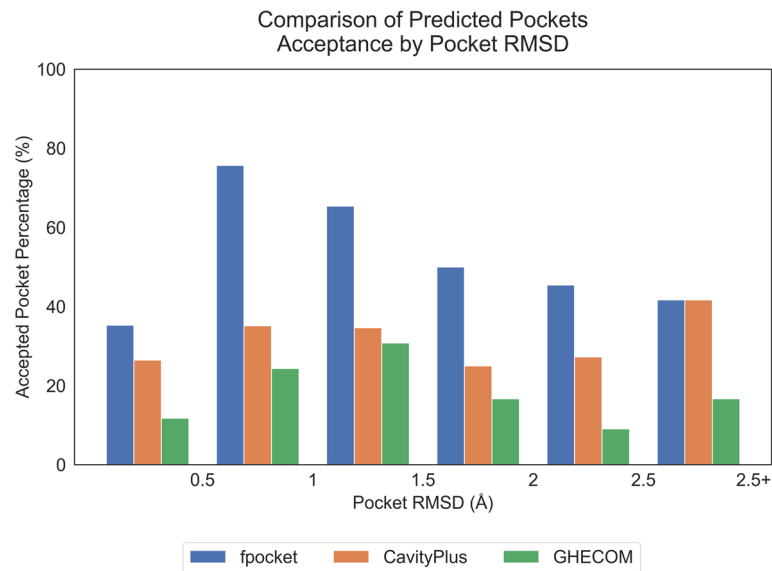
**Fig. 4** Comparison of predicted and actual pocket volume fold change distribution between three methods in experimentally determined structures

volume fold change of GHECOM is generally larger compared with the other two methods. The results are consistent with the findings from the Jaccard index of each predicted pocket, indicating that fpocket tends to predict pockets of smaller volumes, leading to a narrow range of pocket residues, while CavityPlus and GHECOM generate larger pockets with more residues.

To further analyze the factors that affected GPCR pocket prediction results, we compared the percentage of predicted pockets with both residue coverage > 70% and Jaccard index > 0.3 against the pocket RMSD between different states of GPCR.

The pocket RMSD is calculated by all atoms of the original binding pocket residues of experimentally determined structures and the corresponding aligned residues of the predicted structure. Given the distribution of pocket RMSD values of the testing set's pocket RMSD (Figure S4a), the testing set was divided into six groups by different pocket RMSD ranges (Figure S4b).

Figure 5 shows the difference in successful pocket prediction cases between six groups. Generally, fpocket has a better performance than CavityPlus and GHECOM in every group. In the RMSD 0–0.5 Å group, all three methods did not perform well and the acceptance was generally lower than that in other groups. For CavityPlus, the success ratio was below 30%, and for GHECOM the success ratio was below 10%. In other groups (pocket RMSD more than 1 Å), the accepted cases percentage of predicted structures for fpocket went down as the pocket RMSD increased. For CavityPlus, the accepted case percentage reached 40% in the pocket RMSD > 2.5 Å group, and the highest acceptance for GHECOM was in the 1 to 1.5 Å RMSD group.



**Fig. 5** Comparison between pocket RMSD and pocket prediction results of predicted structures

**Table 2** Pocket prediction result comparison for different GPCR states

Prediction method	Structure type	Active	Inactive	Total
fpocket	Experimentally determined structure	51/74 (69%)	42/58 (72%)	93/132 (70%)
CavityPlus	Experimentally determined structure	33/74 (45%)	27/58 (47%)	60/132 (45%)
GHECOM	Experimentally determined structure	12/74 (16%)	9/58 (16%)	21/132 (16%)
fpocket	Predicted structure	47/74 (64%)	26/58 (45%)	73/132 (55%)
CavityPlus	Predicted structure	21/74 (28%)	21/58 (36%)	42/132 (32%)
GHECOM	Predicted structure	15/74 (20%)	11/58 (19%)	26/132 (20%)

The result indicates that the deviation of predicted structures may not be the key factor that affects pocket prediction for these methods, and it depends more on the features of GPCRs.

Table 2 shows the results of acceptance differences between different states. For GPCR PDB structures, both fpocket and CavityPlus methods showed better prediction performance for the inactive state than the active state by 2% or 3% while GHECOM performed equally between the two states. For predicted structures, however, the acceptance of the active state was higher than the inactive state by 19% for fpocket and 1% for GHECOM, while for CavityPlus, the acceptance of active structures went lower than before. Since the inactive predicted structures generally had pocket RMSDs concentrated at a lower range than active ones (Figure S5), the multi-state protein structure prediction method AlphaFold-MultiState fixed the problem that the original AlphaFold2 had, that AlphaFold2 prefers to predict the inactive state GPCRs, and resulted in a better pocket prediction for GPCRs at the active states.

Table 3 shows the redocking result of the selected 121 Class-A GPCRs with the accepting criterion of interacting residue coverage above 50%. Seven testing cases failed in redocking with one of the predicted pockets (Supplementary Table S4). For experimentally determined structures, 66.96% of fpocket predicted pocket redocking cases

**Table 3** Accuracy of redocking original ligand into predicted binding sites

	Experimentally determined structures (ligand RMSD < 2.0)	Experimentally determined structures (key residue coverage > 50%)	Predicted structures (key residue coverage > 50%)
fpocket	77/115 (66.96%)	64/115 (55.65%)	56/115 (48.70%)
CavityPlus	66/115 (57.39%)	58/115 (50.43%)	45/114 (39.47%)
GHECOM	76/115 (66.09%)	60/115 (52.17%)	44/113 (38.94%)

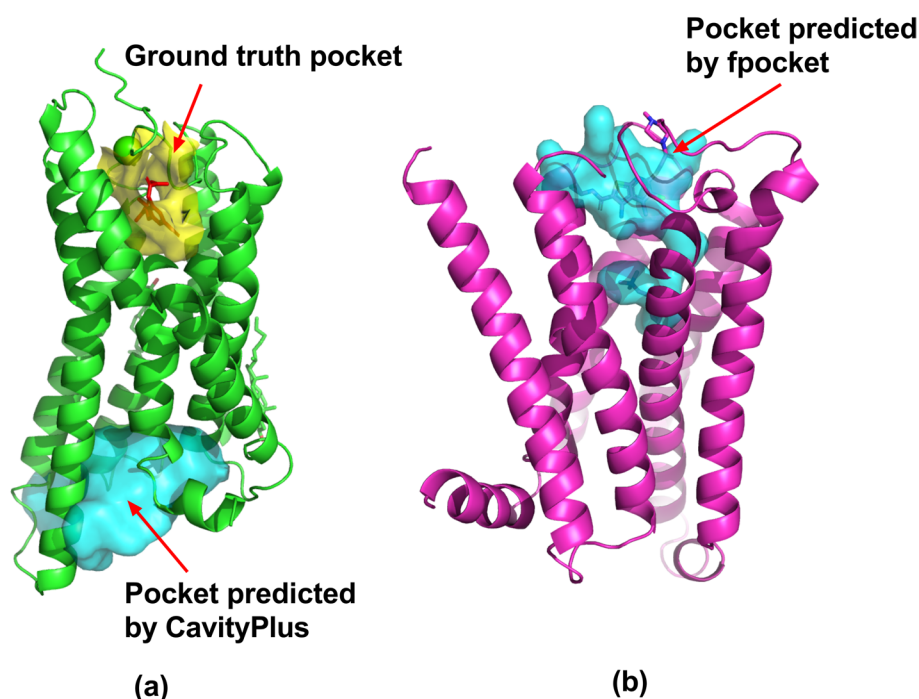
were considered a success under the criterion of ligand RMSD < 2 Å, while CavityPlus and GHECOM had 57.39% and 66.09% of success respectively. For predicted structures, 48.70% of the fpocket predicted pocket redocking cases were successful, and CavityPlus and GHECOM had 39.47% and 38.94% successful cases. The result suggests that on both experimentally determined and predicted structures, it is hard to restore the original poses using predicted pockets by these methods, and although CavityPlus and GHECOM showed better residue coverages compared with the fpocket results, their docking performance was no better.

**Discussion**

During the database construction and evaluation process, considerations emerged in both the pocket prediction and the docking process.

Because of the different algorithms and strategies employed by different pocket prediction tools, the output results can vary significantly. For instance, fpocket tends to predict pockets with smaller binding pocket regions and more concentrated pocket residues, while CavityPlus and GHECOM tend to scan larger pocket spaces. For GPCRs that possess special structural characteristics, this issue is accentuated, as the varying ranges and positions of the predicted binding sites could have a substantial impact on subsequent studies, e.g. molecular docking and other related research. Therefore, caution should be exercised when incorporating predicted binding sites.

Although the primary objective of pocket detection in GPCR structures is to accurately identify the orthosteric binding sites, current pocket recognition tools may not guarantee this goal. For example, in the case of the human GPR35 crystal structure (PDB: 8H8)), CavityPlus failed to identify the orthosteric binding site but identified another intracellular site with a high score, which is likely to be an allosteric binding site. (Fig. 6a). However, in a previous study, CavityPlus managed to detect the orthosteric binding pocket of hGPR35 in the AlphaFold predicted structure and that pocket was confirmed through experimental validation [40]. Therefore, it is necessary to use more than one pocket prediction tool and more than one set of prediction results, and also refer to any reported key residues for accurate binding site prediction. Allosteric binding sites have become increasingly popular as new drug targets in recent years. In GPCRs, allosteric binding sites are typically located outside the center of the TMH, most likely near the intracellular and extracellular loop regions. Due to the high variability of these loop regions, predicting allosteric sites presents a significant challenge. Currently, experimentally resolved GPCR allosteric complexes primarily focused on a limited number of Class A and Class B1 receptors. In our tests, the recognition performance for these sites showed a slight disparity compared to the identification of orthosteric sites



**Fig. 6** Two representative examples of pocket prediction results: **a** CavityPlus prediction (cyan) on PDB 8H8J did not detect the original orthosteric binding site (yellow), **b** fpocket prediction (cyan) on PDB 4MQT found a large pocket region that includes both orthosteric and allosteric ligand

(Supplementary Table S3). For the sub-family of the muscarinic acetylcholine receptors that have an allosteric modulator binding site around the third extracellular loop, such as hCRM2 (PDB: 4MQT), fpocket predicted an elongated binding site that turned out to be a fusion of the orthosteric and allosteric binding sites (Fig. 6b). For similar predicted pocket regions, setting a single docking grid may lead to misalignment of a desired ligand. For some class B1 (secretein) GPCRs, the known negative allosteric modulator binding sites may not be accurately predicted. Hence, in practical docking applications, it is necessary to manually inspect the predicted pockets to ensure consistency with the existing conclusions. For pockets with larger predicted ranges, it is advisable to divide them into multiple sub-regions for further application.

Although AlphaFold provides state-of-the-art structure predictions, confidence between different structural domains may vary, especially for extracellular and intracellular loop regions of GPCRs. Therefore, receptor flexible docking methods such as induced fit docking are recommended to ensure sufficient flexibility of the pocket region to therefore accommodate ligands effectively.

## Conclusion

In this study, we developed a novel GPCR pocket database, which encompasses various GPCR states, and systematically compared three commonly used pocket prediction methods. The state-specific GPCR structures predicted by AlphaFold-MultiState offer new opportunities to analyze the variations in binding sites under different states. This, in turn, facilitates a deeper understanding of GPCR activation

mechanisms and opens up possibilities for further research into the allosteric regulation mechanisms of GPCRs. Our database serves multiple purposes, including molecular docking and virtual screening based on given binding sites, comparative analysis of binding site shapes across different GPCR states, and the identification of potential allosteric sites.

Despite the progress made in the current work, there is still room for improvement in several aspects, which will help to improve the accuracy and confidence of the predictions.

First, more pocket prediction tools should be compared. Currently, we mainly rely on three geometry-based pocket prediction methods, fpocket, CavityPlus and GHECOM, but we plan to extend this aspect in the future. We will consider comparative studies using pocket prediction results from other tools, especially those methods that utilize cutting-edge technologies such as deep learning. That will provide a better understanding of the advantages and limitations of each method, thus broadening our knowledge of GPCR pockets.

Second, more accurate pocket description methods and scoring strategies should be established to better evaluate and compare pocket results generated by different tools, which will more objectively assess the characteristics and properties of pockets and allow us to compare the outputs of different tools. That will help to determine which pockets are biologically significant and how pocket information can be better utilized to guide drug design and biomolecular interaction studies.

Next, parameter optimization strategies should be applied. Although the results of the currently used default parameters are already acceptable, there is still room for improvement. Future work will focus on establishing appropriate parameter optimization strategies to accommodate different types of GPCRs and pocket characteristics. Through careful parameter tuning, we can better adapt to various GPCR structures and properties and improve the consistency and reliability of predictions.

Finally, the prediction of binding pockets should be combined with other in-silico research methods, such as performing blind docking [10] on predicted structures in different states or using molecular dynamics simulation to obtain a broader set of protein–ligand conformations. By utilizing data from multiple perspectives, we will gain a more comprehensive understanding of the changes in binding pockets and the mechanisms of GPCR–ligand interactions.

We continue to strive to improve the accuracy and utility of GPCR pocket prediction in the future to better meet the needs of biomedical research.

#### Abbreviations

GPCR	G-protein coupled receptors
TMH	Transmembrane helix
RMSD	Root mean square deviation
IFD	Induced fit docking
hGPR35	Human G-protein coupled receptor 35
hCRM2	Human muscarinic acetylcholine receptor M2

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12859-024-05962-9>.

Additional file 1.



### Acknowledgements

The web front end was designed using Angular JS. The protein and binding pocket visualization is powered by NGL view. The snake plot illustration script is from GPCRdb open source code. We are sincerely grateful to Gefeng Liu and Xiaoyi Li from our software development team for their invaluable contributions to the development of the server.

### Author contributions

LF: methodology (lead), data curation (lead), software (lead), visualization (lead), writing-original draft (lead), writing-review and editing (equal); ZH: funding, project administration (equal), supervision (equal), writing-review and editing (equal); LXN: project administration (equal) supervision (equal), writing-review and editing (supporting); ZLL: software (equal), visualization (supporting); ZHC: funding, supervision (supporting), YCG: funding, supervision (supporting); BDB: funding, project administration (equal), supervision (equal), writing-review and editing (supporting); LXM: project administration (equal), supervision (supporting).

### Funding

This work was supported by the Innovation Program of Science and Research of DICP, CAS (DICP I202232), Jiangxi Provincial Natural Science Foundation (20224BAB212028), National Natural Science Foundation of China (32271297, 32370657).

### Availability of data and materials

The GPCR Binding Site database is freely accessible at <https://gpcrbs.bigdata.jcm-sc.cn>. The datasets generated and analysed during the current study are available on Zenodo at <https://zenodo.org/records/12744854> (doi: <https://doi.org/10.5281/Zenodo.12744854>).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

Received: 16 July 2024 Accepted: 16 October 2024

Published online: 04 November 2024

### References

1. Foster SR, Hauser AS, Vedel L, Strachan RT, Huang X-P, Gavin AC, Shah SD, Nayak AP, Haugaard-Kedström LM, Penn RB. Discovery of human signaling systems: pairing peptides to G protein-coupled receptors. *Cell*. 2019;179(4):895–908.
2. Yang D, Zhou Q, Labroska V, Qin S, Darbalaei S, Wu Y, Yuliantie E, Xie L, Tao H, Cheng J, et al. G protein-coupled receptors: structure- and function-based drug discovery. *Signal Transduct Target Ther*. 2021;6(1):7.
3. Hauser AS, Attwood MM, Rask-Andersen M, Schioth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov*. 2017;16(12):829–42.
4. Wang J, Hua T, Liu ZJ. Structural features of activated GPCR signaling complexes. *Curr Opin Struct Biol*. 2020;63:82–9.
5. Hauser AS, Kooistra AJ, Munk C, Heydenreich FM, Veprintsev DB, Bouvier M, Babu MM, Gloriam DE. GPCR activation mechanisms across classes and macro/microscales. *Nat Struct Mol Biol*. 2021;28(11):879–88.
6. Mafi A, Kim SK, Goddard WA 3rd. The mechanism for ligand activation of the GPCR-G protein complex. *Proc Natl Acad Sci U S A*. 2022;119(18):e2110085119.
7. Zhou Q, Yang D, Wu M, Guo Y, Guo W, Zhong L, Cai X, Dai A, Jang W, Shakhnovich EI, et al. Common activation mechanism of class A GPCRs. *Elife*. 2019;8:e50279.
8. Divorcy N, Jenkins L, Ganguly A, Butcher AJ, Hudson BD, Schulz S, Tobin AB, Nicklin SA, Milligan G. Agonist-induced phosphorylation of orthologues of the orphan receptor GPR35 functions as an activation sensor. *J Biol Chem*. 2022;298(3):101655.
9. Nicoli A, Dunkel A, Giorgino T, de Graaf C, Di Pizio A. Classification model for the second extracellular loop of class A GPCRs. *J Chem Inf Model*. 2022;62(3):511–22.
10. Hedderich JB, Persechino M, Becker K, Heydenreich FM, Gutermuth T, Bouvier M, Bunemann M, Kolb P. The pocketome of G-protein-coupled receptors reveals previously untargeted allosteric sites. *Nat Commun*. 2022;13(1):2567.
11. Lu S, He X, Yang Z, Chai Z, Zhou S, Wang J, Rehman AU, Ni D, Pu J, Sun J, et al. Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nat Commun*. 2021;12(1):4721.
12. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9.
13. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, Bridgland A, Cowie A, Meyer C, Laydon A, et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021;596(7873):590–6.
14. Liu Q, Yang D, Zhuang Y, Croll TI, Cai X, Dai A, He X, Duan J, Yin W, Ye C, et al. Ligand recognition and G-protein coupling selectivity of cholecystokinin a receptor. *Nat Chem Biol*. 2021;17(12):1238–44.
15. Kim TY, Woo EJ, Yoon TS. Binding mode of brazzein to the taste receptor based on crystal structure and docking simulation. *Biochem Biophys Res Commun*. 2022;592:119–24.

16. He XH, You CZ, Jiang HL, Jiang Y, Xu HE, Cheng X. AlphaFold2 versus experimental structures: evaluation on G protein-coupled receptors. *Acta Pharmacol Sin.* 2023;44(1):1–7.
17. Heo L, Feig M. Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins.* 2022;90(11):1873–85.
18. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hübner H, Pardon E, Valant C, Sexton PM, et al. Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature.* 2013;504(7478):101–6.
19. Weisel M, Proschak E, Schneider G. PocketPicker: analysis of ligand binding-sites with shape descriptors. *Chem Cent J.* 2007;1(1):1–17.
20. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics.* 2009;10(1):1–11.
21. Kawabata T, Go N. Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites. *Proteins.* 2007;68(2):516–29.
22. Kawabata T. Detection of multiscale pockets on protein surfaces using mathematical morphology. *Proteins.* 2010;78(5):1195–211.
23. Kawabata T. Detection of cave pockets in large molecules: spaces into which internal probes can enter, but external probes from outside cannot. *Biophys Physicobiol.* 2019;16:391–406.
24. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J. CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. *Nucleic Acids Res.* 2018;46(W1):W374–9.
25. Wang S, Xie J, Pei J, Lai L: CavityPlus. Update: an integrated platform for comprehensive protein cavity detection and property analyses with user-friendly tools and cavity databases. *J Mol Biol.* 2022;2023:168141.
26. Krivák R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J Cheminf.* 2018. <https://doi.org/10.1186/s13321-018-0285-8>.
27. Aggarwal R, Gupta A, Chelur V, Jawahar CV, Priyakumar UD. DeepPocket: ligand binding site detection and segmentation using 3D convolutional neural networks. *J Chem Inf Model.* 2022;62(21):5069–79.
28. Gainza P, Sverrisson F, Monti F, Rodola E, Boscaini D, Bronstein MM, Correia BE. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat Methods.* 2020;17(2):184–92.
29. Pándy-Szekeres G, Caroli J, Mamyrbekov A, Kermani AA, Keserü GM, Kooistra AJ, Gloriam DE. GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Res.* 2023;51(D1):D395–402.
30. Isberg V, Mordalski S, Munk C, Rataj K, Harpsoe K, Hauser AS, Vroiling B, Bojarski AJ, Vriend G, Gloriam DE. GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 2016;44(D1):D356–364.
31. Wang S, Lin H, Huang Z, He Y, Deng X, Xu Y, Pei J, Lai L. CavitySpace: a database of potential ligand binding sites in the human proteome. *Biomolecules.* 2022;12(7):967.
32. Yuan Y, Pei J, Lai L. Binding site detection and druggability prediction of protein targets for structure-based drug design. *Curr Pharm Des.* 2013;19(12):2326–33.
33. Rodríguez-Espigares I, Torrens-Fontanals M, Tiemann JK, Aranda-García D, Ramírez-Anguila JM, Stepniowski TM, Worp N, Varela-Rial A, Morales-Pastor A, Medel-Lacruz B. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat Methods.* 2020;17(8):777–87.
34. Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, et al. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* 2023;51(D1):D523–31.
35. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009;25(11):1422–3.
36. Hamelryck T, Manderick B. PDB file parser and structure class implemented in Python. *Bioinformatics.* 2003;19(17):2308–10.
37. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem.* 2006;49(2):534–53.
38. Farid R, Day T, Friesner RA, Pearlstein RA. New insights about HERG blockade obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg Med Chem.* 2006;14(9):3160–73.
39. Sherman W, Beard HS, Farid R. Use of an induced fit receptor structure in virtual screening. *Chem Biol Drug Des.* 2005;67(1):83–4.
40. Otkur W, Wang J, Hou T, Liu F, Yang R, Li Y, Xiang K, Pei S, Qi H, Lin H, et al. Aminosaliclates target GPR35, partly contributing to the prevention of DSS-induced colitis. *Eur J Pharmacol.* 2023;949:175719.

## Publisher's Note

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