www.nature.com/aps

Acta Pharmacologica Sinica

ARTICLE

CB-Dock: a web server for cavity detection-guided protein–ligand blind docking

Yang Liu[1], Maximilian Grimm[1], Wen-tao Dai[2], Mu-chun Hou[1], Zhi-Xiong Xiao[1] and Yang Cao[1]

As the number of elucidated protein structures is rapidly increasing, the growing data call for methods to efficiently exploit the structural information for biological and pharmaceutical purposes. Given the three-dimensional (3D) structure of a protein and a ligand, predicting their binding sites and affinity are a key task for computer-aided drug discovery. To address this task, a variety of docking tools have been developed. Most of them focus on docking in the preset binding sites given by users. To automatically predict binding modes without information about binding sites, we developed a user-friendly blind docking web server, named CB-Dock, which predicts binding sites of a given protein and calculates the centers and sizes with a novel curvature-based cavity detection approach, and performs docking with a popular docking program, Autodock Vina. This method was carefully optimized and achieved ~70% success rate for the top-ranking poses whose root mean square deviation (RMSD) were within 2 Å from the X-ray pose, which outperformed the state-of-the-art blind docking tools in our benchmark tests. CB-Dock offers an interactive 3D visualization of results, and is freely available at http://cao.labshare.cn/cb-dock/.

Keywords: bioinformatics; computer-aided design; computer-aided drug discovery

Acta Pharmacologica Sinica (2020) 41:138–144; https://doi.org/10.1038/s41401-019-0228-6

INTRODUCTION

Protein–ligand docking has been widely used to predict binding modes and affinities of ligands. Protein–ligand docking is a powerful tool for computer-aided drug discovery (CADD). Currently, there are dozens of commercial and academic tools available for protein–ligand docking [1–12]. Most docking tools require the ligand binding region (the rotation and translation of a ligand in this region) in advance to search for the most energy favorable binding mode. The binding region is usually represented as a cubic box, so its size and center are critical for accurate docking because it defines the boundaries of the conformational sampling space. In many application scenarios, the binding regions are unknown. To identify potential interactions between a given protein and a ligand, docking has to be performed on the entire protein surface to find the most probable binding mode. This process is called blind docking [13–16]. Compared to regular docking, blind docking is less reliable and stable as the docking space is usually too large to sufficiently sample using a limited number of random searches. Nevertheless, blind docking is particularly valuable for discovering unexpected interactions that may occur in unidentified binding modes [17].

Traditionally, blind docking is performed on the entire protein surface. Alternatively, docking on putative binding regions of the given protein usually improves the sampling efficiency and reduces the computational cost of blind docking [18]. Currently, many binding site detection tools have been developed [19–29]. These methods help users find residues that potentially bind with ligands. However, users must cluster residues into groups and estimate the parameters manually and then perform several rounds of protein–ligand docking to obtain the final result. Although this process is feasible, it is not efficient and has not been systematically optimized. To address this problem, several blind docking tools have been developed in recent years that have integrated cavity detection with a focused docking module. For example, popular software SwissDock [30, 31], QuickVina-W [15] and BSP-SLIM [32] provide particularly valuable services for blind docking. In this paper, we described a new blind docking tool, named CB-Dock, which focuses on enhancing the docking accuracy. CB-Dock predicts binding regions of a given protein, calculates the centers and sizes with a curvature-based cavity detection approach, and performs docking with the state-of-the-art docking software Autodock Vina [33]. CB-Dock also ranks the binding modes according to Vina scores and provides an interactive 3D visualization of the binding modes. Our benchmark tests show an ~70% success rate for the top-ranking poses whose root mean squared deviation (RMSD) was within 2 Å from the position in the X-ray crystal structure. It is notably higher than the traditional blind docking method (~40%) and outperformed other popular blind docking tools. The server of CB-Dock is freely available at http://cao.labshare.cn/cb-dock/, together with additional documentation and tutorials.

[1]Center of Growth, Metabolism and Aging, Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China and [2]Shanghai Center for Bioinformation Technology & Shanghai Engineering Research Center of Pharmaceutical Translation, Shanghai Industrial Technology Institute, Shanghai 201203, China
Correspondence: Yang Cao (cao@scu.edu.cn)
These authors contributed equally: Yang Liu, Maximilian Grimm.

Received: 5 December 2018 Accepted: 14 March 2019
Published online: 1 July 2019

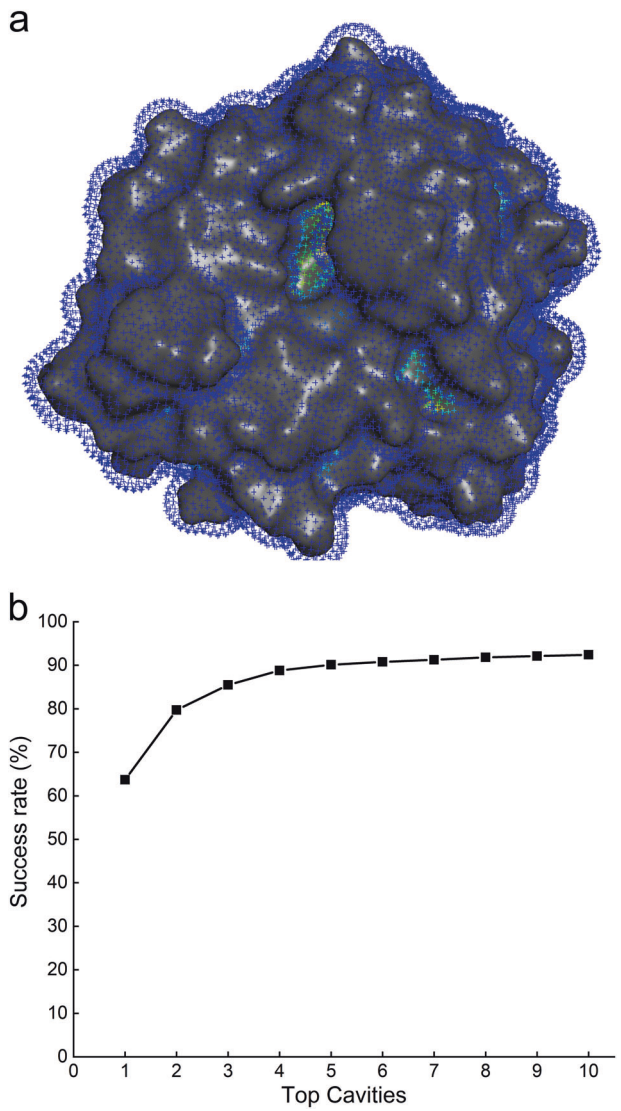

Fig. 1 The cavities detected by CB-Dock. **a** The putative cavities of aminopeptidase (PDB ID: 1TXR) were highlighted in green. **b** Success rate of detected cavities from top 10 to 1 in the PDBbind Set. (Top 5 = 90.1%)

MATERIALS AND METHODS

Benchmark dataset

PDBbind Set. A total of 1684 protein–ligand structures were selected from PDBbind (v2018) [8, 34]. The molecular weights of the proteins were limited to 150~500 g/mol, and the numbers of rotatable bonds were within 10. In addition, the proteins that share 60% or more similarity to the Astex Diverse Set or MTiAutoDock data were eliminated. The structures can be downloaded from our website http://cao.labshare.cn/cb-dock/.

Astex Diverse Set. The Astex Diverse Set contains 85 protein–ligand complexes [35], which were downloaded from the Protein Data Bank [36]. The redundant identical chains, water molecules, and heteroatoms were discarded.

MTiAutoDock Set. The test data are from the benchmark set of MTiOpenScreen [37]. The data contains 27 crystal structures that cover important drug targets, including enzymes, GPCRs, nuclear receptors, and PPIs.

Apo Structure Set. The above Astex Diverse Set is composed of protein–ligand complex (holo) structures. To test the docking in the unbound state (apo) of proteins, we collected 19 apo protein structures [18] available in the Astex Diverse Set. Each apo structure corresponds to a holo structure in the Astex Diverse Set. The sequence identity and coverage of each pair are greater than 95%. To compare the accuracy of the docking results, we superimposed each apo structure onto its corresponding holo structure.

The traditional blind docking and redocking protocols

The parameters of traditional blind docking were customized as described by the protocol from Di Muzio et al. [38]. The docking center is the spatial geometric center of all the heavy atoms of the protein. To obtain the sizes of the docking box, distances between the center and each atom along the three axes (*x, y*, and *z*) were calculated. Then, the maximum value of the distance along each dimension is doubled and adds an additional 5 Å as the size of the docking box [38].

Redocking was performed with known binding sites. The docking parameters were customized by following the method from Wei and Michal [39]. In general, the search box size is equal to 2.857 times the radius of gyration of the ligand, which consistently obtains the highest prediction accuracy when using AutoDock Vina [39].

RESULTS

Detecting cavities on proteins

Most small-molecule binding occurs in protein pockets or cavities because high affinity can only be gained by sufficiently large interaction interfaces [40]. CB-Dock searches for concave surfaces to detect cavities. Briefly, CB-Dock generates a set of points to represent the solvent-accessible surface and calculates the curvature factor of each point using the method from our previous work [41, 42]. These points at the concave surface (curvature factor >8) are clustered by a density-peak-based clustering algorithm [43]. Thus, we obtained several clusters of points that represent cavities on the protein surface. We present the example of aminopeptidase (PDB ID: 1TXR), whose cavities are highlighted in Fig. 1a. The cavities were ranked according to their sizes. We compared our method (called CurPocket) with state-of-the-art protein–ligand binding site prediction methods using the benchmark set of COACH [23], which is one of the best prediction methods. The results showed that our method is comparable to that of COACH in terms of Matthews correlation coefficient, precision, and recall (see Supplementary Table S1). Unlike traditional binding site prediction methods, our method detected the real binding cavities as much as possible to offer options for blind docking. To investigate its performance in detecting real binding cavities, we submitted 1684 structures from PDBbind to CurPocket (see the Materials and methods section) and examined their success rates by comparing the top 10 cavities with the real binding cavities from the crystal structures. Test results showed that the predicted success rates [44] of the top 1 to 10 cavities increased from 63.7% to 92.4%, respectively (Fig. 1b). From the top 10 to top 5, the success rate dropped only 2%. To balance the computational expense and cavity detection accuracy, we selected the top 5 cavities as candidates for focused docking.

Calculating centers and sizes of docking boxes

For a putative cavity, CB-Dock needs to customize a docking box for the following computation. A good docking box should enclose the native binding pose and exclude as many as possible irrelevant poses. The center and size of the docking box are the key parameters in this process. The center of the ligand from the crystal structure is the best choice for the docking box; however, we can base these parameters only on the putative cavity and

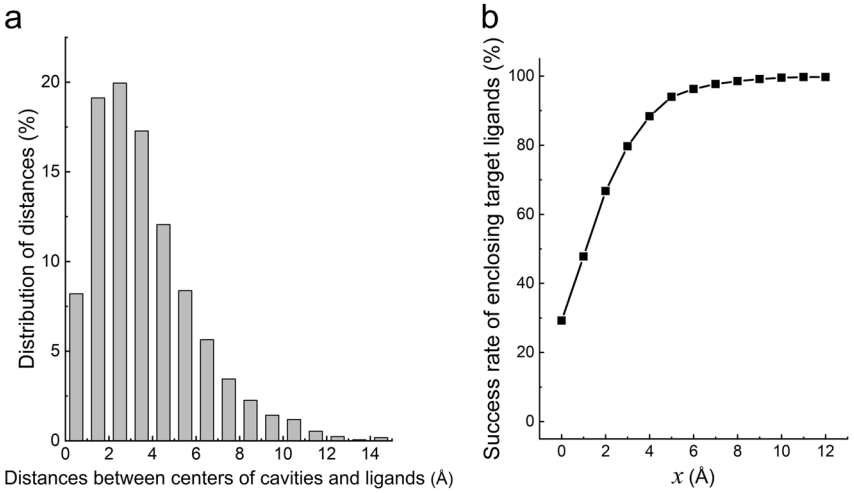

Fig. 2 The parameters of docking box were optimized by the statistical analysis of known binding cavities. **a** Distribution of distances between centers of ligands in crystal structures and centers of putative cavities which are closest to the ligand. **b** The proportion of the test cases whose docking boxes successfully enclose the target ligands for the given extra edge length *x* of the box

unbound ligands to estimate the center and size. Hence, we first selected the center of the putative cavity, i.e., the center of points at the concave surface, as the docking center. To quantify its deviation from the best center, we calculated distances between the two centers using the PDBbind data set (see the Materials and methods section). The distances between centers of real and putative target cavities were distributed from 1 to 10 Å (Fig. 2a). For most of the data (76.6%), the distances were within 5 Å and up to 97.7% when distances were within 10 Å. The result indicated that for the majority of the data, the center of the cavity was close to the ideal center. Second, we needed to determine the lengths of the docking box in each dimension, which was related to the size of the cavity, the size of the ligand and the deviation of the putative center from the ideal center. After systematical examination of the outcome from docking, we finally calculated the *i* axis length L_i of the docking box by a constant *x* plus the maximum of the length C_i of the putative cavity or gyration radius *R* of the given ligand as follows:

$$L_i = x + \max(R, C_i)$$

The constant *x* is used to compensate for the deviation of the putative center and to ensure that the ligand is enclosed in the docking box. To determine *x*, we tested the above protein–ligand structures to investigate the proportion of docking boxes that enclosed the ligands by gradually increasing *x* from 0 to 12 Å (Fig. 2b). The results showed that the proportion grows rapidly when *x* increases from 0 to 5 Å. When *x* is 10 Å, all the ligands are enclosed in the docking box. Thus, we choose $x = 10$ Å in our program. Detailed analysis shows that the sizes of the docking box by the above formula were mostly less than 30 Å, which was within the recommended upper limit (http://vina.scripps.edu/manual.html#faq).

The guidance of cavity detection improved blind docking

To assess the performance of CB-Dock, we compared it with traditional blind docking using a protein–ligand complex from Astex Diverse Set [35]. The docking parameters of traditional blind docking are described in section 'The traditional blind docking and redocking protocols'. In addition, to determine the upper limit of this blind docking, we also tested redocking the centers and sizes of docking boxes that were obtained from crystal structures [39]. We measured the accuracy by RMSD between the predicted binding mode with the lowest docking score and the native mode in the crystal structures. The performances of these methods were quantified by the percentage of correct predictions (RMSD < 2 Å) (Fig. 3a). The results show that for traditional blind docking, redocking, and CB-Dock, the prediction accuracies were 38.8%, 76.5%, and 69.4%, respectively. As we expected, CB-Dock had significant improvements (~30% higher) over traditional blind docking, and the overall accuracy was much closer to redocking and the upper limit of docking using Autodock Vina. Particularly, when the prediction was correct, CB-Dock and redocking had nearly identical RMSD values (Fig. 3b). This result implies that the cavity detection and docking parameters of CB-Dock work rather well. As AutoDock Vina is based on a random algorithm, whose results may be different from the repeat runs, we repeated the test for 3 rounds to investigate the stability of the three methods. The results showed that the RMSD variations of CB-Dock and redocking were less than 5%, while it was up to 10% for traditional blind docking. We argued that CB-Dock appropriately decreased the sampling space and thereby reduced the randomness of the results. In all, cavity detection is a powerful approach to improve blind docking.

Comparison of CB-Dock with existing blind docking tools

To gain an overall performance of CB-Dock, we further compared it with four state-of-the-art docking tools, including DockingApp [38], MTiAutoDock [37], rDock [45], and SwissDock [30, 31]. Though the tools provide multiple usages, we focused on their performance of blind docking. DockingApp searches for binding sites over the whole protein surface by AutoDock Vina [33]. MTiAutoDock uses the same strategy but is powered by AutoDock 4.2.6 [5]. rDock and SwissDock perform docking in the vicinity of predicted cavities. Unlike curvature-based cavity detection in CB-Dock, rDock uses a two-probe sphere method [45], and SwissDock employs a variant of the grid-based LIGSITE algorithm [46] to identify cavities. In general, DockingApp and MTiAutoDock follow the traditional strategy, while rDock, SwissDock, and CB-Dock only allow docking in the putative binding regions. We conducted the benchmarks on the Astex Diverse Set and MTiAutoDock data (see the Materials and methods section). In the first dataset, DockingApp, MTiAutoDock, rDock, SwissDock (accurate mode) and CB-Dock achieved 42.4%, 42.4%, 41.2%, 53.0%, and 69.4% success rates of top-ranking poses within the RMSD of 2 Å from crystal structures, respectively (Fig. 4a). In the second set, the five tools achieved 33.3%, 51.9%, 33.3%, 70.4% and 74.1% success rates, respectively (Fig. 4b). Both benchmarks illustrated that, in terms of success rates for top-ranking poses, CB-Dock outperformed other blind docking tools. As blind docking strongly depends on the

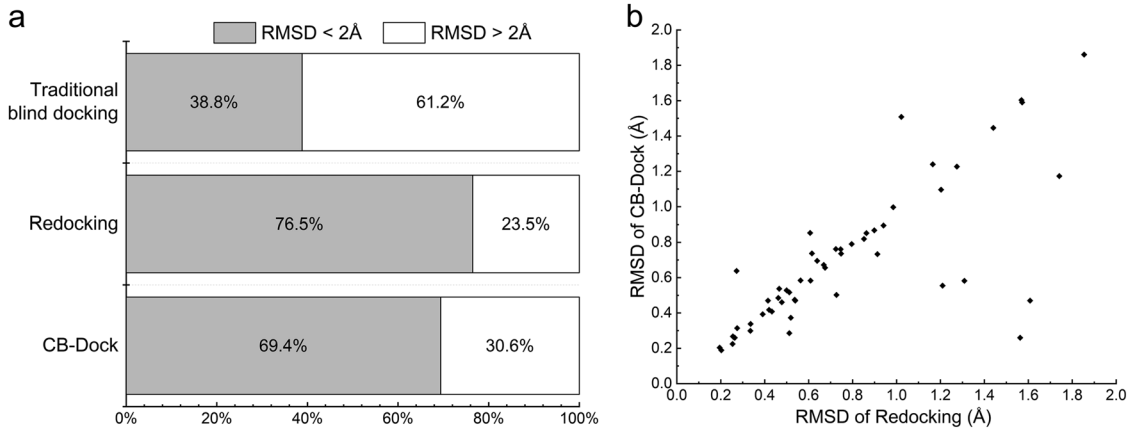

Fig. 3 The performance of traditional blind docking, redocking and CB-Dock on Astex Diverse Set. **a** The percentage of top-ranked poses with an RMSD below 2 Å of the three methods. **b** RMSD of CB-Dock versus redocking when RMSDs < 2 Å

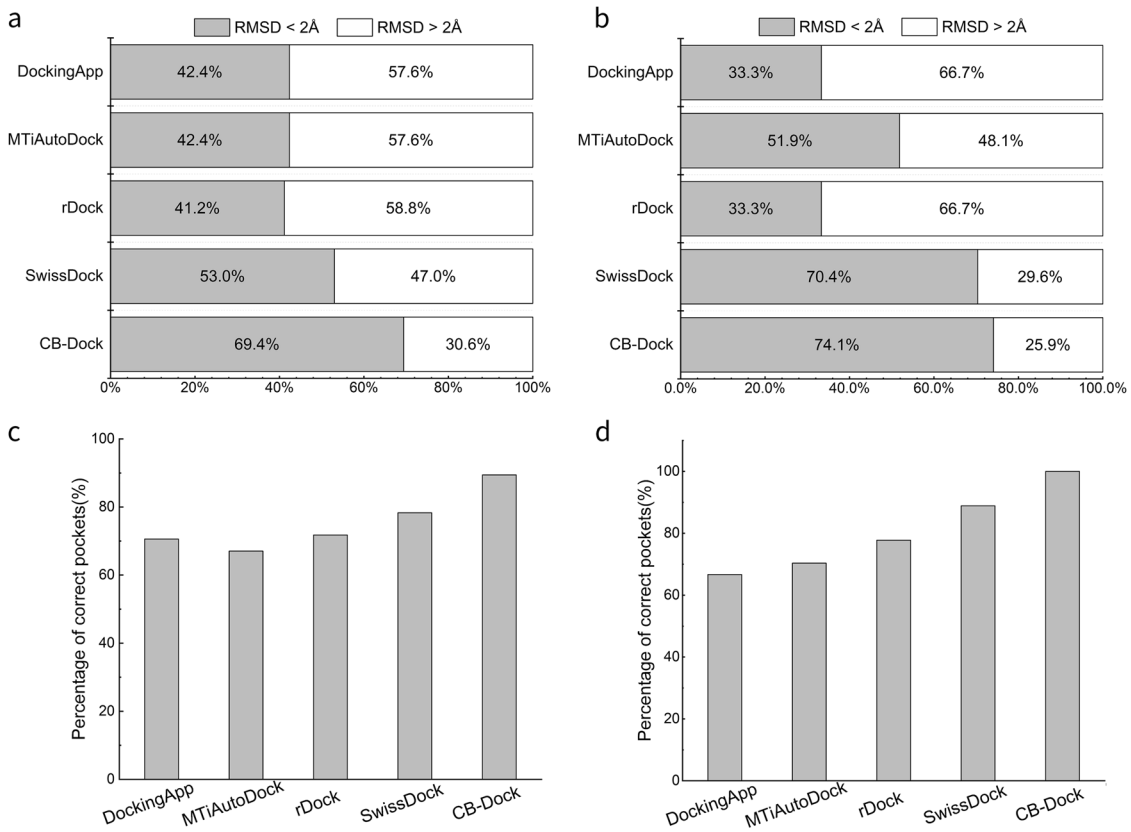

Fig. 4 The performance of DockingApp, MTiAutoDock, rDock, SwissDock (accurate mode), and CB-Dock on Astex Diverse Set and MTiAutoDock Set. **a** The success rates of the top-ranking binding modes in Astex Diverse Set. **b** The success rates of the top-ranking binding modes in MTiAutoDock Set. **c** The percentages of correct binding pockets for top-ranking poses in Astex Diverse Set. **d** The percentages of correct binding pockets for top-ranking poses in MTiAutoDock Set

accuracy of predicted binding sites, we further compared the average percentage of correctly predicted binding sites [44]. The results showed that the accuracies were 70.6%, 67.1%, 71.8%, 78.3%, and 88.2% for DockingApp, MTiAutoDock, rDock, SwissDock (accurate mode) and CB-Dock, respectively, on the Astex Diverse Set data (Fig. 4c) and were 70.4%, 70.4%, 77.8%, 88.9%, and 100%, respectively, on the MTiAutoDock data (Fig. 4d). These results exhibited good correlations with the above success rates of predicting the binding sites and indicated the significance of the binding site prediction in CB-Dock.

The above tests benchmarked blind docking on the ligand-bound states (holo) of receptors from the protein–ligand complex structures. The blind docking in unbound (apo) structures is much more challenging as the conformational changes of proteins are difficult to predict. We performed blind docking using the 19 apo crystal structures available in Astex Diverse Set (see Apo Structure Set in the Materials and methods section). The results showed that the average percentages of correctly predicted binding sites [44] of the top-one predictions are 47.4%, 36.8%, 47.4%, 31.6%, and 68.4% for DockingApp, MTiAutoDock, rDock, SwissDock (accurate mode),

Table 1. The RMSDs of five blind docking tools benchmarked in Apo Structure Set

Target protein	DockingApp (Å)	MTiAutoDock (Å)	rDock (Å)	SwissDock (Å)	CB-Dock (Å)
1hq2	**3.019**	12.796	**3.018**	**3.601**	**3.037**
1ke5	7.543	21.849	16.685	37.472	7.535
1l2s	15.528	**3.874**	7.361	31.838	15.497
1l7f	26.68	19.626	34.651	17.328	**0.865**
1n1m	34.526	18.808	24.157	21.524	23.342
1n2v	22.755	**3.903**	**3.836**	27.137	**4.517**
1oq5	17.044	12.092	14.372	5.265	**1.817**
1oyt	**0.339**	**0.538**	**4.766**	**3.783**	**0.335**
1q41	**2.219**	**1.382**	29.771	**1.483**	**2.145**
1s3v	**3.912**	**4.098**	**1.802**	**3.243**	**4.622**
1t40	**4.750**	30.219	28.091	30.17	**4.726**
1t46	17.22	32.354	18.139	17.82	17.206
1v0p	37.477	**4.673**	36.866	23.273	**4.212**
1v48	7.971	15.968	13.423	13.615	7.908
1w1p	9.203	10.918	**1.721**	27.866	13.354
1yvf	16.433	60.384	19.995	13.263	14.687
1ywr	**4.428**	25.78	**4.284**	24.156	**4.381**
2br1	**4.383**	5.081	**3.749**	**1.589**	**4.381**
2bsm	16.412	16.78	**3.744**	20.386	**0.784**
Average	13.255	15.849	14.233	17.095	7.124
RMSD < 5 Å	36.8%	31.6%	42.1%	26.3%	63.2%

The RMSD values < 5 Å are highlighted in bold

and CB-Dock, respectively. The RMSDs exhibited a similar trend. The success rates of top-ranking sites within the RMSD of 5 Å are 36.8%, 31.6%, 42.1%, 26.3%, and 63.2%, respectively (see Table 1). CB-Dock achieved the highest accuracy in the Apo Structure Set. However, the success rate was notably lower than that on the holo structure set. Analysis showed that the conformational differences between apo and holo structures may result in two types of inaccurate docking. One type is that CB-Dock identifies accurate cavities for docking; however, the detailed conformation of cavities was different between apo and holo structures. If the differences were critical for binding, docking may not be accurate because CB-Dock does not model the conformational changes between apo and holo structures. An example of this type is the PDB structure 1L2S (see Fig. S1a and S1b). The side chain of Ser64 at the protein–ligand interface was turned 44.5° from the apo structure (PDB ID: 2BLS) to the holo structure (PDB ID: 1L2S) to avoid atomic clashes. This difference misled docking on the apo structure. The other type of inaccurate docking was that the top five cavities of the apo structure do not include the real binding cavity. An example of this type is the PDB structure 1YVF (see Fig. S1c and S1d). The real binding cavity was ranked in the top five cavities on the holo structure (PDB ID: 1YVF), while it was too small to rank in the top five cavities on apo structure (PDB ID: 2GIR). Hence, the docking has a very large RMSD.

Computational speed is another critical feature of docking in high-throughput virtual screening. Because only DockingApp, rDock, and CB-Dock provided a stand-alone version, the time consumption was analyzed for the three blind docking tools. The results showed that the average running times of DockingApp, rDock, and CB-Dock on Astex Diverse Set were 44.4, 75.8, and 62.7 s per blind docking, respectively, on an AMD Ryzen1700 processor (see Table S2). Detailed data showed that the running time of CB-Dock and DockApp did not show any correlation with the size of the protein (number of residues) but was slightly related to the flexibility of ligand (quantified by the number of rotatable bonds) (See Fig. S2). In contrast, the time consumption of rDock had a strong relationship with the size of the protein but not the flexibility of the ligand (see Fig. S2). Although the precise time consumption of MTiAutoDock and SwissDock was not available, based on our tests, their online usages took over 10 min on average to return a docking result. Taken together, we argue that CB-Dock serves as a relatively rapid blind docking tool. In particular, the protein-size-independent feature of CB-Dock is suitable for docking-based inverse virtual screening.

CB-Dock web server

To facilitate the use of CB-Dock, we constructed a web server at http://cao.labshare.cn/cb-dock/, which only requires the input of a protein file to be in the PDB format and a ligand file in the MOL2, MOL, or SDF. After submission, CB-Dock checks the input files and converts them to pdbqt formatted files using OpenBabel [47] and MGLTools [5]. Next, CB-Dock predicts cavities of the protein and calculates the centers and sizes of the top N ($n = 5$ by default) cavities. Each center and size, as well as the pdbqt files, are submitted to AutoDock Vina for docking. The final results are displayed after the computation of N rounds. Users can browse binding scores, cavity sizes, and docking parameters of the predicted binding modes in a table. Moreover, users can inspect the 3D structures of any binding modes on the web page by clicking the structures in the related table. The interactive 3D structures are drawn by NGL Viewer [48], which is supported by most modern browsers. Users are able to display atom-specific information, rotate and translate molecules, select models and colors. For more details, users could refer to the manual on the CB-Dock homepage.

Here, we present a case study of the software CB-Dock (Fig. 5). Nultin3a, a potential anti-cancer drug, is able to bind with the E3 ubiquitin-protein ligase MDM2 and inhibit the MDM2–P53 interaction. The MDM2 protein structure was downloaded (PDB ID: 4HG7) from PDB. The Nutlin-3a mol2 file was generated by the PRODRG software [49]. The two files were uploaded and

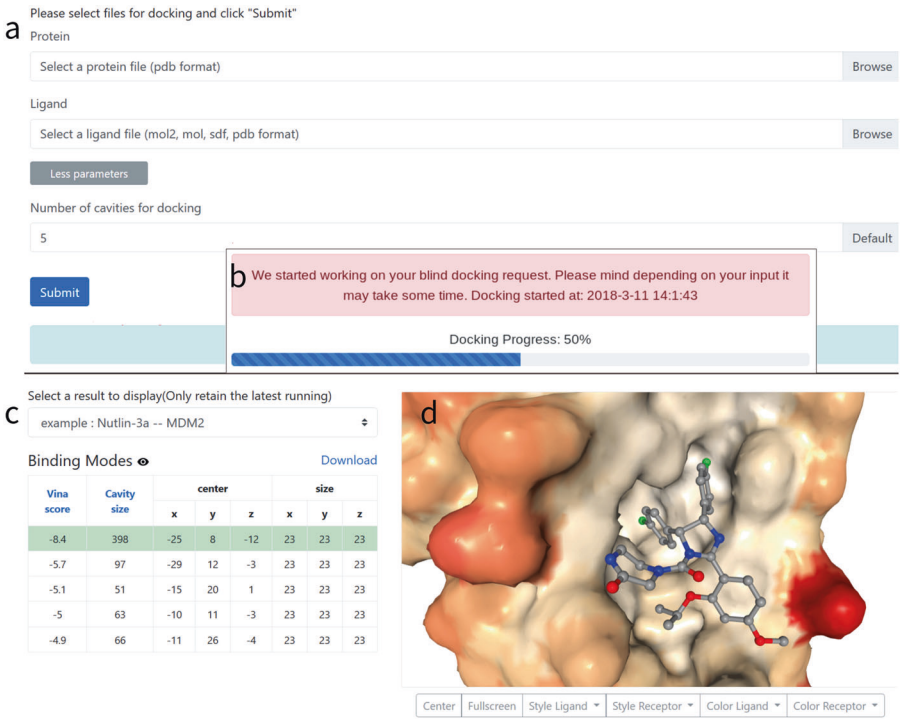

Fig. 5 The web interface of CB-Dock. **a** The interface for uploading protein and ligand files. **b** The progress bar in job running. **c** The Vina scores and cavity information of results. **d** The interactive 3D viewer illustrating selected binding modes

submitted to the CB-Dock server by clicking the button "Submit". While processing docking, a progress bar appeared to indicate the status of the job. When the processing was complete (after approximately 2 min), the web page was updated with the results. The table listed Vina scores, cavity sizes, docking centers, and sizes of predicted cavities. Once a ligand in the table is selected, the structure in the interactive 3D graphics is visualized. In our example, the top binding mode with a Vina score of −8.4 also had the largest binding cavity. The binding mode was almost identical to the mode of ligand in the crystal structure (RMSD = 0.484 Å).

DISCUSSION

Discovering protein–ligand binding sites and conformations are particularly important in drug discovery. Blind docking is a powerful method for obtaining that information. Blind docking is also one of the key components in high-throughput screening and inverse docking [50–53]. Therefore, it is of great value to develop accurate blind docking tools. Thanks to the well-established AutoDock Vina docking software, we focused on developing methods of cavity detection and docking parameter optimization, which are critical for blind docking. CB-Dock is the first cavity detection-guided blind docking tool designed with AutoDock Vina among many popular Vina-based tools (http://vina.scripps.edu/manual.html#faq). The benchmark tests show that CB-Dock outperforms other state-of-the-art blind docking tools in terms of predicting binding sites and binding conformations. This performance is attributed to the curvature-based cavity detection that precisely narrows down the docking space as well as the optimized parameters for AutoDock Vina. Some shortcomings of CB-Dock were also observed in the test. First, compared to regular docking, CB-Dock was more time expensive because the docking was performed iteratively in five cavities. To reduce time consumption, cavity detection should be further improved in the future. Second, if the size of cavities was notably greater than that of the ligand, the accuracy of docking tends to decrease. A typical example is the huge cavity detected on nitric-oxide synthase (PDB ID: 1MMV), in which the predicted docking position is at the opposite side of the cavity (see Fig. S3). This result is mainly related to the accuracy of the scoring function, which is supposed to distinguish the global minimum from local minimums. Using an additional scoring function to rerank binding positions could be a solution to this problem. Third, CB-Dock needs to improve the accuracy of docking in apo structures. Compared to holo structures, apo structures show conformational rearrangement in ligand binding sites, which has not been captured in current CB-Dock software. In the following developments of CB-Dock, the protein conformation sampling method will be incorporated in CB-Dock to enhance docking in apo structures.

Apart from blind docking capabilities, user-friendly interfaces are also very important for docking tools. CB-Dock offers a convenient web service that allows even nonexpert users to perform protein–ligand docking and visualize results in 3D. We believe that CB-Dock can contribute to the characterization of newly determined protein structures and suggest novel therapeutic targets for biological and pharmaceutical studies.

ACKNOWLEDGEMENTS

The authors thank Professor Jian-yi Yang of Nankai University for helping in running COACH-D and Dr. Holger Stitz of Johannes Kepler University Linz for his invaluable editing of the manuscript. We also thank Professor Yang Zhang and Cheng-xin Zhang of the University of Michigan, Professor Xiang-jun Du of Sun Yat-sen University and Dr. Zhi-chao Miao of Cambridge University for invaluable discussions. This work was supported by the National Natural Science Foundation of China (Grant numbers 31401130, 81830108, and 81672736), the National Key R&D Program of China (2018YFC0910500), the Shanghai Sailing Program (16YF1408600), the funding for prevention and control technology of African swine fever (2018NZ0151) and the Shanghai Industrial Technology Institute (17CXXF008).

AUTHOR CONTRIBUTION

YL designed and optimized the CB-Dock tool and wrote the manuscript. MG built the CB-Dock web server. WTD benchmarked the program. MCH tested the server. ZXX guided the experiments. YC designed the project and wrote the manuscript.

ADDITIONAL INFORMATION

The online version of this article (https://doi.org/10.1038/s41401-019-0228-6) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev. 2017;9:91–102.
2. Yuriev E, Holien J, Ramsland PA. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. J Mol Recognit. 2015;28:581–604.
3. Meiler J, Baker D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins. 2006;65:538–48.
4. Marialke J, Tietze S, Apostolakis J. Similarity based docking. J Chem Inf Model. 2008;48:186–96.
5. Morris G, Huey R. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem. 2010;30:2785–91.
6. Bolia A, Ozkan SB. Adaptive BP-Dock: an induced fit docking approach for full receptor flexibility. J Chem Inf Model. 2016;56:734–46.
7. Allen WJ, Balius TE, Mukherjee S, Brozell SR, Moustakas DT, Lang PT, et al. DOCK 6: impact of new features and current docking performance. J Comput Chem. 2015;36:1132–56.
8. Liu Z, Su M, Han L, Liu J, Yang Q, Li Y, et al. Forging the basis for developing protein-ligand interaction scoring functions. Acc Chem Res. 2017;50:302–9.
9. Lam PCH, Abagyan R, Totrov M. Ligand-biased ensemble receptor docking (LigBEnD): a hybrid ligand/receptor structure-based approach. J Comput Aided Mol Des. 2018;32:187–98.
10. Padhorny D, Hall DR, Mirzaei H, Mamonov AB, Moghadasi M, Alekseenko A, et al. Protein–ligand docking using FFT based sampling: D3R case study. J Comput Aided Mol Des. 2018;32:225–30.
11. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 1997;267:0–748.
12. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003;52:609–23.
13. Hetényi C, Van Der Spoel D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. FEBS Lett. 2006;580:0–1450.
14. Hetényi C, van der Spoel D. Efficient docking of peptides to proteins without prior knowledge of the binding site. Protein Sci. 2002;11:1729–37.
15. Hassan NM, Alhossary AA, Mu Y, Kwoh CK. Protein-ligand blind docking using QuickVina-W with inter-process spatio-temporal integration. Sci Rep 2017;7:15451.
16. Sánchez-Linares I, Pérez-Sánchez H, Cecilia JM, García JM. High-throughput parallel blind virtual screening using BINDSURF. BMC Bioinformatics 2012;13(Suppl 14):S13.
17. Iorga B, Herlem D, Barré E, Guillou C. Acetylcholine nicotinic receptors: finding the putative binding site of allosteric modulators using the "blind docking" approach. J Mol Model. 2006;12:366–72.
18. Ghersi D, Sanchez R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. Proteins. 2009;74:417–24.
19. Dai W, Wu A, Ma L, Li YX, Jiang T, Li YY. A novel index of protein-protein interface propensity improves interface residue recognition. BMC Syst Biol. 2016;10:381–92.
20. Shin WH, Seok C. GalaxyDock: Protein-ligand docking with flexible protein side-chains. J Chem Inf Model. 2012;52:3225–32.
21. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol. 2009. https://doi.org/10.1371/journal.pcbi.1000585.
22. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, et al. 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:W374–W379.
23. Yang J, Roy A, Zhang Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics. 2013;29:2588–95.
24. Levitt DG, Banaszak LJ. POCKET: A computer graphies method for identifying and displaying protein cavities and their surrounding amino acids. J Mol Graph. 1992;10:229.
25. Laskowski RA. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. J Mol Graph. 1995;13:323–30.
26. Brylinski M, Skolnick J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc Natl Acad Sci USA. 2008;105:129–34.
27. Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model. 2003;21:289–307.
28. Brylinski M, Feinstein WP. EFindSite: improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands. J Comput Aided Mol Des. 2013;27:551–67.
29. Wu Qi, Peng Zhenling, Yang Zhang JY. COACH-D: improved protein–ligand binding sites prediction with refined ligand-binding poses through molecular docking. Nucleic Acids Res. 2018;46:313–38.
30. Grosdidier A, Zoete V, Michielin O. Blind docking of 260 protein-ligand complexes with eadock 2.0. J Comput Chem. 2010;30:2021–30.
31. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res. 2011;39:270–7.
32. Lee HS, Zhang Y. BSP-SLIM: a blind low-resolution ligand-protein docking approach using predicted protein structures. Proteins. 2012;80:93–110.
33. Trott O, Olson AJ. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2009;31:455–61.
34. Liu Z, Li Y, Han L, Li J, Liu J, Zhao Z, et al. PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics. 2015;31:405–12.
35. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, et al. Diverse, high-quality test set for the validation of protein-ligand docking performance. J Med Chem. 2007;50:726–41.
36. Burley SK, Berman HM, Christie C, Duarte JM, Feng Z, Westbrook J, et al. RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. Protein Sci. 2018;27:316–30.
37. Labbé CM, Rey J, Lagorce D, Vavruša M, Becot J, Sperandio O, et al. MTiOpenScreen: A web server for structure-based virtual screening. Nucleic Acids Res. 2015;43:448–54.
38. Di Muzio E, Toti D, Polticelli F. DockingApp: a user friendly interface for facilitated docking simulations with AutoDock Vina. J Comput Aided Mol Des. 2017;31:213–8.
39. Feinstein WP, Brylinski M. Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets. J Cheminform. 2015;7:1–10.
40. Sotriffer C, Klebe G. Identification and mapping of small-molecule binding sites in proteins: Computational tools for structure-based drug design. Farmaco. 2002;3:243–51.
41. Cao Y, Li L. Improved protein-ligand binding affinity prediction by using a curvature-dependent surface-area model. Bioinformatics. 2014;30:1674–80.
42. Cao Yang, Wentao Dai ZM. Evaluation of protein–ligand docking by cyscore. Comput Drug Discov Des. 2018;1762:223–32.
43. Rodriguez A, Laio A, Xu R, Wunsch D, Frey BJ, Dueck D. et al.Machine learning. Clustering by fast search and find of density peaks. Science. 2014;344:1492–6.
44. Schmidt T, Haas J, Gallo Cassarino T, Schwede T. Assessment of ligand-binding residue predictions in CASP9. Proteins. 2011;79:126–36.
45. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, et al. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. PLoS Comput Biol. 2014;10:e1003571 https://doi.org/10.1371/journal.pcbi.1003571.
46. Hendlich M, Rippmann F, Barnickel G. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model. 1997;15:359–63.
47. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An Open chemical toolbox. J Cheminform. 2011;3:33.
48. Rose AS, Bradley AR, Valasatava Y, Jose M, Prli A, Rose PW. NGL Viewer : Web-based molecular graphics for large complexes. Bioinformatics. 2018;34:3755–8.
49. Schüttelkopf AW, Van Aalten DMF. PRODRG: A tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr Sect D Biol Crystallogr. 2004;60:1355–63.
50. Sánchez-Linares I, Pérez-Sánchez H, Cecilia JM, García JM. High-Throughput parallel blind Virtual Screening using BINDSURF. BMC Bioinformatics. 2012;13:S13 https://doi.org/10.1186/1471-2105-13-S14-S13.
51. Pérot S, Sperandio O, Miteva MA, Camproux AC, Villoutreix BO. Druggable pockets and binding site centric chemical space: A paradigm shift in drug discovery. Drug Discov Today. 2010;15:656–67.
52. Schwardt O, Cutting B, Kolb H, Ernst B. Drug discovery today. Front Med Chem. 2005;3:1–9.
53. Kharkar PS, Warrier S, Gaud RS. Reverse docking: A powerful tool for drug repositioning and drug rescue. Future Med Chem. 2014;6:333–42.