Published online 26 May 2018 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* ***W423–W431***
doi: 10.1093/nar/gky398

HSYMDOCK: a docking web server for predicting the structure of protein homo-oligomers with Cn or Dn symmetry

Yumeng Yan, Huanyu Tao and Sheng-You Huang*

Institute of Biophysics, School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

Received January 26, 2018; Revised April 07, 2018; Editorial Decision April 21, 2018; Accepted May 03, 2018

ABSTRACT

A major subclass of protein–protein interactions is formed by homo-oligomers with certain symmetry. Therefore, computational modeling of the symmetric protein complexes is important for understanding the molecular mechanism of related biological processes. Although several symmetric docking algorithms have been developed for Cn symmetry, few docking servers have been proposed for Dn symmetry. Here, we present HSYMDOCK, a web server of our hierarchical symmetric docking algorithm that supports both Cn and Dn symmetry. The HSYMDOCK server was extensively evaluated on three benchmarks of symmetric protein complexes, including the 20 CASP11–CAPRI30 homo-oligomer targets, the symmetric docking benchmark of 213 Cn targets and 35 Dn targets, and a nonredundant test set of 55 transmembrane proteins. It was shown that HSYMDOCK obtained a significantly better performance than other similar docking algorithms. The server supports both sequence and structure inputs for the monomer/subunit. Users have an option to provide the symmetry type of the complex, or the server can predict the symmetry type automatically. The docking process is fast and on average consumes 10∼20 min for a docking job. The HSYMDOCK web server is available at http://huanglab.phys.hust.edu.cn/hsymdock/.

INTRODUCTION

Protein–protein interactions play critical roles in many biological processes such as signal transduction, cell regulation, DNA replication and repair, and RNA transcription. Therefore, determining the complex structure of the interactions is crucial for understanding the molecular mechanism and developing drugs targeting these interactions (1). Due to the high cost and technical difficulties in experimental methods, protein–protein docking has played an important role in filling the gap between the large number of individual protein structures and the limited number of complex structures in the Protein Data Bank (PDB)(2). Given two individual proteins, protein–protein docking samples possible binding modes of one protein within a binding site if the binding site is available or around the whole surface of the other protein if the binding site is unknown. An energy scoring function is then used to rank the generated binding modes, and the top-scored modes are predicted as the complex structures. For years, a variety of algorithms have been developed for protein–protein docking and achieved many successes in various systems (1,3).

To make protein–protein docking user-friendly, a number of protein–protein docking servers such as ClusPro (4), HADDOCK (5), RosettaDock server (6), GRAMM-X (7), 3D-Garden (8), HEX server (9), SwarmDock (10), ZDOCK server (11), PatchDock (12), ATTRACT (13), pyDockSAXS (14), InterEvDock (15) and HDOCK server (16) have been developed and made available for public access. With these servers, users can easily run local or global docking to obtain predicted complex structures with individual proteins by *ab initio* docking or template-based docking, where experimental binding information and co-evolutionary data from structures or sequences may be used.

A major subclass of protein–protein interactions is formed by symmetric homo-oligomers (17–19). For example, as of 30 December 2017, more than one third (i.e. 52367) of the total entries in the PDB have certain types of symmetry. It has been thought that the symmetry of homo-oligomeric protein assemblies is associated with many potential benefits like greater stability, reduced aggregation, and robustness against errors in synthesis (17,18). Many transmembrane proteins like ion channels are formed by symmetric homo-oligomer assemblies. The interfaces between symmetric homo-oligomers are often the targeting sites for the regulation of biological processes (20). Therefore, determination of symmetric homo-oligomer structures

*To whom correspondence should be addressed. Tel: +86 27 87543881; Fax: +86 27 87556576; Email: huangsy@hust.edu.cn

is essential for understanding the biological functions of symmetric proteins at the molecular level. Theoretically, one may use a general protein–protein docking algorithm to generate the structure of symmetric homo-oligomers. However, due to the constraints introduced from symmetry, general docking algorithm is often not efficient in predicting the structure of symmetric homo-oligomers. On the one hand, without the symmetry constraints, general docking algorithms need to search much more space; On the other hand, general docking algorithm may not be able to construct complex structures with strict symmetry. Therefore, specialized docking algorithms are necessary for the structure prediction of symmetric protein complexes.

There are two major groups of symmetry in the PDB. One is cyclic (Cn symmetry), for which oligomeric structure can be described by a rotation around a single rotation axis of one subunit. The other is the dihedral group (Dn symmetry), which combines one rotational Cn axis with a perpendicular axis of 2-fold symmetry (18). Several algorithms have been developed for symmetric protein docking. Wolfson and colleagues developed a geometry-based docking algorithm for the prediction of cyclically symmetric complexes, which is referred to as SymmDock (21). SymmDock uses local feature matching to produce the candidate set of transformations by restricting the search to symmetric cyclic transformations. The Weng group used an FFT-based algorithm for protein–protein docking with Cn symmetry by restricting the search to the cyclic symmetry (M-ZDOCK) (22). With more and more symmetric structures available in the PDB, several web servers that use template-based methods like ROBETTA (23), SWISS-MODEL (24), and GalaxyGemini (25) have been proposed to predict the homo-oligomer structure. Very recently, the Seok group has proposed an approach for homo-oligomer structure prediction from a monomer sequence or structure by template-based modeling or *ab initio* docking (GalaxyHomomer) (26). However, few docking servers support *ab initio* docking of homo-oligomers with Dn symmetry, except SAM (27). Given that a significant portion (about one sixth) of symmetric complexes belongs to Dn symmetry, an algorithm that can handle the homo-oligomer docking with Dn symmetry is necessary. Meeting the needs, we have developed a web server for the symmetric docking of homo-oligomers with both Cn and Dn symmetry by implementing our symmetric docking algorithm for the Critical Assessment of Prediction of Interactions (CAPRI) experiments (28), which is referred to as HSYMDOCK. Compared to similar docking servers, our HSYMDOCK server accepts not only structure but also sequence as input for the subunit. The docking process is fully automatic and the results are interactively provided to users through a user-friendly web page.

MATERIALS AND METHODS

The docking protocol of HSYMDOCK

HSYMDOCK is supported by our hierarchical symmetric docking algorithm (29) and iterative scoring function for protein–protein interactions (30). Several third-party programs and a set of tools developed by our group are used to streamline the docking protocol. The workflow of HSYMDOCK is shown in Figure 1.

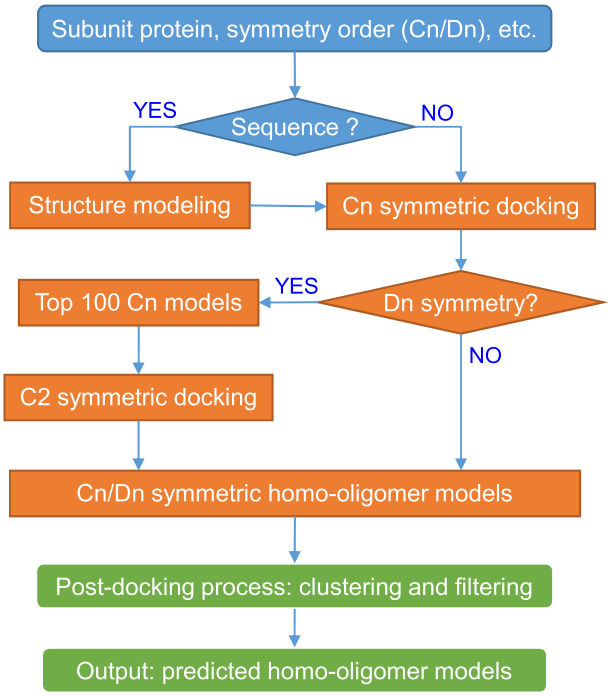

Figure 1. Flowchart of the HSYMDOCK web server including three stages: (1) data input, (2) symmetric docking and (3) output, which are shown in blue, orange and green, respectively.

HSYMDOCK starts from a monomer or subunit and a symmetry type. The server accepts both sequence and structure for the subunit. Both Cn and Dn symmetries are supported. Users are also given an option to provide information about the binding site.

Then, the server will check the input for the subunit. If the input is a structure, the pipeline will go to the next docking step. If the input for the subunit is a sequence, its structure will be constructed by homology modeling as follows. A sequence similarity search is conducted against the PDB database to find the homologous templates for the target protein. Here, the HHSuite package is used for the sequence search (31), due to its efficient detection of remote homologs. If multiple templates are detected, the one with the highest sequence coverage, the highest sequence similarity, and the highest resolution will be selected. With the selected template, the 3D structure of the subunit is built using MODELLER (32), in which the sequence alignment is conducted using ClustalW (33,34).

With the user-provided symmetry type (Cn or Dn) and the structure modeled by our server or uploaded by users, the workflow enters the next step, i.e. symmetric protein docking. Here, a hierarchical FFT-based docking program developed by our group, is used to globally sample putative binding modes with Cn symmetry (29). If the job is to predict the structure of a Cn symmetric homo-oligomer, only one round of Cn symmetric docking step will be performed.

If the job is to predict a homo-oligomer with Dn symmetry, an additional C2 symmetric docking with previously predicted Cn complexes will be conducted to construct the Dn homo-oligomers around an additional perpendicular axis, in which the top 100 previously predicted Cn complexes are used. The docking process can also incorporate the binding site information if users have provided such information at the time of submission. The final top 100 predictions are provided to users through a result web page, in which users can interactively view the top 10 binding models through a Jmol web interface (35).

Docking and scoring method

We have developed an FFT-based symmetric protein docking algorithm to globally sample putative homo-oligomers with Cyclic symmetry by restricting the translational search on the x–y plane perpendicular to the z-axis of Cn symmetry for each rotation. An improved pairwise scoring function is used for initial shape matching during the global search (36). The key point of our new scoring function is that the score for a ligand grid will take into account the contributions not only from its nearest neighboring receptor grids but also from other receptor grids by a form of $\sim e^{-1/r^2}$, where r is the distance between the ligand grid and the receptor grids. An angle interval of 10° is used for rotational sampling, and a spacing of 1.2 Å is adopted for the FFT-based translational search. For each rotation, the four translations, which correspond to the best shape complementarity scores in the four quadrants of the x–y plane, are selected. For Dn symmetric docking, an additional C2 symmetric docking is performed with the previously constructed complexes of Cn symmetry by using an additional perpendicular axis of symmetry. All the binding modes are evaluated by ITScorePP (30) and ranked by their binding scores. The final ranked binding modes are clustered with an RMSD cutoff of 5 Å (3), where the RMSD is calculated using backbone atoms. If two binding modes have a ligand RMSD of ≤5 Å, the one with the better score is kept. By default, the top 100 homo-oligomeric models are output to users through a result web page.

Input

HSYMDOCK requires users to provide a monomer or subunit for the prediction of homo-oligomeric complexes. For the subunit, both structure and sequence inputs are supported. The server accepts four types of inputs for the subunit protein, two of which are for structure and the other two are for sequence.

- Upload your pdb file in PDB format.
- Provide your structure in PDB ID:ChainID (e.g. 5A6X:A).
- Copy and paste your **protein** sequence in FASTA format.
- Upload your **protein** sequence file in FASTA format

Only one type of input is needed. For sequence input, the maximum number of amino acids is set to 2000 and modified amino acids are not supported. For structure input, the maximum number of atoms is set to 20 000 and modified amino acids are treated as the most basic types. Users can upload their own pdb files or provide the PDB: chain ID(s). Since our structural modeling protocol is currently designed to model single-chain proteins from sequences, users are recommended to upload their own structures if the subunit protein contains multiple chains.

In addition to the subunit protein, users have an option to provide a symmetry type, or the server will predict the symmetry type automatically. The server supports both Cyclic (Cn) and Dihedral (Dn) symmetry. By default, the top 100 symmetric homo-oligomer complexes are output, while users can change the number within the range from 1 to 500 when submitting their job.

In addition, users also have an option to provide information about the binding site, i.e. the residues at the binding interface between symmetric subunits of the complex. The binding site information, if provided, will be used in the post-docking clustering stage as a filter after the FFT-based search. A few residues for the binding site are sufficient for constraining correct binding modes. Users may also give a name to their job and provide an email address for notification of job completion.

Benchmarks

CASP–CAPRI targets. In collaboration with recent CASP sessions (37), CAPRI had two rounds for the prediction of homo-oligomer complexes. There are a total of 30 homomer targets with Cn or Dn symmetry in these CAPRI challenges, of which 23 targets are from round 30 (i.e. CASP11–CAPRI30 challenges) and 7 targets are from round 37 (i.e. CASP12-CAPRI37 challenges). These targets formed a valuable benchmark for blindly testing the performance of a docking algorithm and/or scoring function.

As biological information about binding may be used during the blind prediction of CASP–CAPRI challenges, the performance in CAPRI experiments may not reflect the predictive capability of the docking algorithm alone. Therefore, we have tested our HSYMDOCK on the 20 homo-oligomer proteins among the targets of CASP11–CAPRI30 by using the monomer models from the CASP11 I-TASSER server predictions (Supplementary Table S1) (37), in which no biological information about binding was used.

Symmetric docking benchmark. In addition to the CASP–CAPRI homo-multimer targets, we have also constructed a non-redundant benchmark for symmetric protein–protein docking based on the experimentally determined structures in the PDB, the details of which were described elsewhere (http://huanglab.phys.hust.edu.cn/SDBenchmark/). Briefly, all the homo-oligomeric protein complexes with certain symmetry (e.g. Cn, Dn, H, I, T, etc.) were collected from the crystal structures with resolution better than 2.5 Å in the PDB. Only the biological unit was considered when determining its symmetry type. These biologically symmetric homo-oligomer complexes were then clustered according to their SCOP (version 1.75) family IDs (38). For the symmetric complexes belonging to the same family, the one with

the best resolution was selected as the representative of the family, corresponding to a bound case of our benchmark, in which each subunit is called the bound structure of the complex. Namely, the symmetric complex for a bound case is the cocrystallized one of bound structures. For the bound structure in each bound case, the unbound structure was identified by searching against the PDB database for the asymmetric structures using the FASTA program (39). If an asymmetric structure had >95% sequence identity with the bound structure and covered >95% of the sequence alignment, the asymmetric structure was regarded as a candidate of the unbound structure. Namely, the unbound structure is either in free form or bounds with another protein. If there were multiple unbound structures for a subunit protein, the one with the high resolution was selected as the representative. This yielded a total of 213 homo-oligomeric protein complexes with Cn symmetry and 35 homo-oligomer complexes with Dn symmetry (Supplementary Table S2). There targets were used to further validate the symmetric docking algorithm of HSYMDOCK.

Transmembrane proteins. Among the class of symmetric protein complexes, an important part that acting as homo-oligomers are transmembrane (TM) proteins like ion channels. Due to the effect of membranes, this category of proteins may present different challenges in predictions. Therefore, to test the predictive power of HSYMDOCK on such proteins, we have also constructed a test set of symmetric membrane proteins. Specifically, we filtered all the PDB entries in the PDBTM database (version: 16 June 2017) (40) with the following criteria. First, the proteins are crystal structures with resolution of ≤2.0 Å. Second, each subunit of the protein consists of only one protein chain. Then, all the proteins were clustered with a sequence identity cutoff of 70%, and the one with the best resolution in each cluster was selected as the representative. This yielded a final set of 55 symmetric transmembrane proteins (Supplementary Table S3). For each TM protein, the monomer structure was modeled by our in-house homology modeling protocol of HSYMDOCK, in which the templates with higher than 40% sequence identity were excluded.

Evaluation criteria

The quality of a predicted binding mode is measured by three parameters used in CAPRI: the ligand RMSD (L_{RMSD}) after superimposition of receptor proteins, the interface RMSD after superimposition of interface residues between the predicted and native structures (I_{RMSD}), and the percentage of native contacts (f_{nat}) (28). Here, the RMSD is calculated based on the backbone atoms, and the interface is defined as those residues that are within 10 Å from the partner molecules in the native structure. According to the CAPRI criteria, the accuracy of a prediction can be divided into four categories: 'High', 'Medium', 'Acceptable' and 'Incorrect' (28). This is the default criteria, unless otherwise specified. A prediction with an at least acceptable accuracy is defined as a hit. For evaluation on a benchmark, the docking performance is measured by the success rate, i.e. the percentage of the cases with at least one hit within a certain number of predictions.

RESULTS

HSYMDOCK server

The HSYMDOCK server is hosted on a Linux server of two compute nodes, each of which includes two Intel(R) Xeon E5-2690 v4 2.60GHz CPUs with 28 cores and 256 GB of Memory. A maximum of 50 jobs can be running at the same time while hundreds of jobs can be queued in the background. The docking process is fast and the average running time for a docking calculation is 10–20 min. The web server is built on Apache HTTP, HTML, PHP and the JSmol web applet for the docking pipeline and binding model visualization. The web service does not require registration and can be freely accessed.

After users submit input data, the HSYMDOCK server will put the docking job in queue immediately. At the same time, the web interface will be redirected to a web page showing the job ID and running status. The job status including 'QUEUED', 'RUNNING' and 'RESULTS' will be updated every 10 s on the status page. The URL to the docking results is something like http://huanglab.phys.hust.edu.cn/hsymdock/data/jobid, where 'jobid' is a unique job ID assigned by the server. Users can bookmark the job status page for access to the docking results at a later time. Users will also be notified by email when the job is finished if a valid email address is provided at the time of job submission.

Output

When a job is done, the docking results will be provided to users on a result page, as shown in Figure 2. The docking results include two types of files for download:

- The individual symmetric homo-oligomer models for the top 20 predictions .
- The compressed packages for the top 10 predictions, the top 100 predictions, and all the docking results.

Since the top 10 homo-oligomer models are normally regarded as the most important predictions, the result page provides an interactive view of the top 10 models using the Jmol software (35). Users can choose to view any of the top 10 models or all together by different representations and styles.

The result page also gives a summary of the rankings and docking scores for the top 10 homo-oligomer models. However, it should be noted that the docking scores here do not reflect the true binding affinities, but a relative ranking of the homo-oligomer models. It is also recommended that users download their docking results as soon as their job is done, as the job results will only be stored on our server for two weeks.

In addition, if only a sequence is provided as the input for a protein, the result page will also show the information of homology modeling, including the used template, model quality, sequence alignment and sequence identity between the template and the input sequence. The model quality is classified as 'High' if the sequence identity is >50%, 'Medium' if the sequence identity is between 30% and 50%, and 'Low' if the sequence identity is <30%.

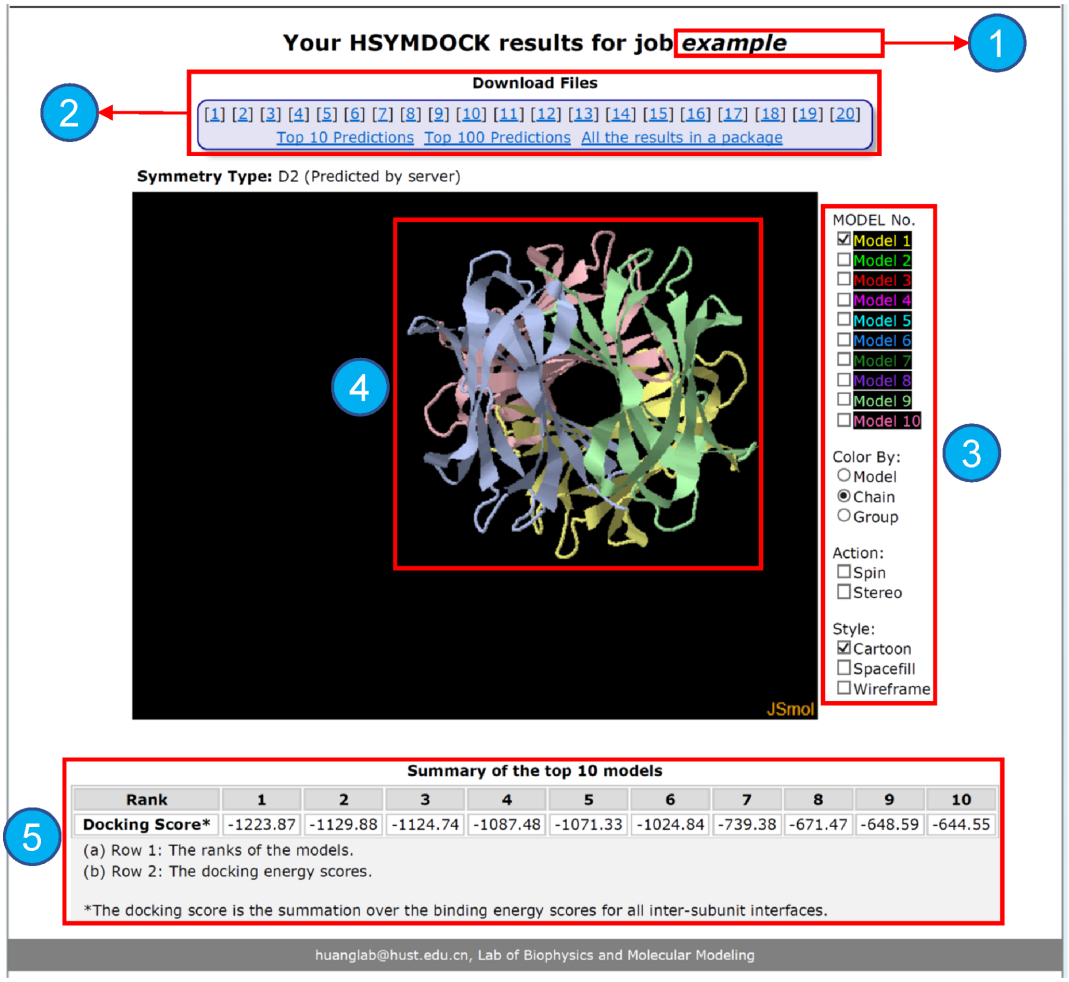

Figure 2. HSYMDOCK server output page. At the top of the page is the user-provided job name or a unique job ID (1), and under it are the files for download (2). Optional buttons on the right can control Jmol to choose which model to be viewed and how to view (3) on the left (4). The docking scores of the top 10 models are shown on the bottom (5).

Docking performance

CASP–CAPRI targets. We have participated in recent CASP–CAPRI experiments for the prediction of all 30 homo-oligomeric protein complexes in rounds 30 and 37. Of the 23 homo-oligomer targets from CASP11–CAPRI30, our docking approach, similar to that used in HSYMDOCK, obtained the structures with an acceptable accuracy or better for 14 targets, though it should be noted that biological information like templates were used in the CAPRI experiments. The results ranked our approach as the top predictor based on the number of targets for which at least one acceptable solution was reached (29,37). Of the seven homo-oligomer proteins in CASP12–CAPRI30, our group achieved an at least acceptable prediction for five targets, compared to six targets for the best-performed group (41).

We have further tested HSYMDOCK on the 20 targets from CASP11–CAPRI30 by using modeled monomer structures. Figure 3 shows the number of targets with at lease one hit by HSYMDOCK when the top 1 and top 5 predictions were considered. The accuracies and ranks of first hits for all the cases are listed in Supplementary Table

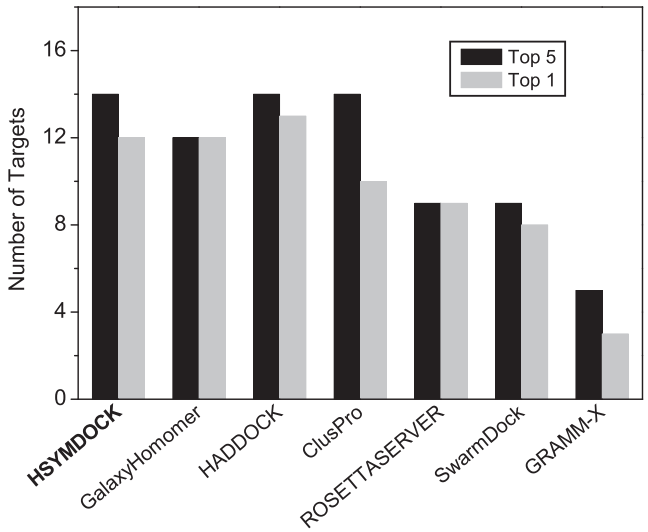

Figure 3. The numbers of targets with an at least acceptable mode predicted by HSYMDOCK and other six similar docking servers on the test set of 20 homo-oligomer targets from CASP11–CAPRI30. The results for the methods other than HSYMDOCK were taken from the literature (26).

S1. For comparison, Figure 3 also shows the results by six other state-of-the-art docking severs. It can be seen from the figure that when the top 1 prediction was considered, HSYMDOCK and GalaxyHomomer obtained successful predictions for 12 targets, compared to 13 targets for HADDOCK and 10 targets or fewer for other servers. When the top five predictions were considered, HSYMDOCK, HADDOCK, and ClusPro perform the same (14 successful targets), and better than other methods. The results indicate the robustness of HSYMDOCK in *ab initio* docking.

Symmetric docking benchmark. We further tested our HSYMDOCK algorithm on the symmetric protein docking benchmark of 213 Cn targets and 35 Dn targets. For Cn symmetry, the quality of a docking model was measured according to the CAPRI criterion (28) For Dn symmetry, the quality of a binding mode was evaluated by the average ligand RMSD of the other subunits after the first subunit was superimposed onto the native structure. The performance was characterized by the success rate, i.e. the fraction of the targets with at least one hit in the test set when a certain number of top predictions are considered. Here, a hit or successful prediction is defined as a model with an acceptable accuracy or better for Cn targets according to the CAPRI criteria (28) or with an average ligand RMSD of <10 Å for Dn targets (3).

Figure 4 shows the success rate of HSYMDOCK as a function of the number of top predictions in binding mode prediction for the 213 cases with Cn symmetry. The accuracies and ranks of first hits for all the cases are listed in Supplementary Table S2. For comparison, the figure also gives the corresponding results for three other state-of-the-art Cn symmetric docking algorithms, M-ZDOCK (22), SymmDock (21) and SAM (27), on this benchmark, in which the same clustering criteria for the binding modes have been applied.

It can be seen from Figure 4 that HSYMDOCK obtained a significantly better performance than the other three docking methods for bound docking and achieved a success rate of 76.1% within top 10 predictions, compared to 66.2% for M-ZDOCK, 61.0% for SAM and 42.3% for SymmDock, respectively (Figure 4A). Similar trend can also be observed in the results of unbound docking, the performance differences are not as much as those in bound docking due to the effect of conformational changes. Namely, HSYMDOCK performed significantly better than the other three methods for unbound docking and obtained a success rate of 48.8% for top 10 predictions, compared to 44.1% for M-ZDOCK, 40.4% for SAM and 37.1% for SymmDock, respectively (Figure 4B).

When tested on the targets with Dn symmetry, HSYMDOCK was also efficient and had a success rate of 85.7% for bound docking and 54.3% for unbound docking when the top 10 predictions were considered, compared to 51.4% and 37.1% for SAM, respectively (Figure 4C and D). The higher success rate on the Dn targets than that on the Cn targets can be understood because the additional two-fold symmetry of Dn considerably reduces the search space during docking and thus increase the probability of finding a correct model.

Transmembrane proteins. Figure 5 shows the the success rate of HSYMDOCK on the nonredundant benchmark of 55 transmembrane (TM) proteins when the top 1 and top 5 predictions were considered. For comparison, the figure also lists the results of other two methods, GalaxyHomomer and HH+MODELLER, on a similar benchmark of 47 transmembrane proteins developed by the Seok group (26). It can be seen from the figure that HSYMDOCK achieved a significantly better performance and obtained a success rate of 54.5% and 65.5% when the top 1 and top 5 predictions were considered, compared to 40.4% and 40.4% for GalaxyHomomer and 29.8% and 38.3% for HH+MODELLER, respectively. The high success rate of HSYMDOCK suggests the strong capability in predicting the complex structures of TM proteins.

Discussions. Despite the good performances on a variety of benchmarks of symmetric proteins, HSYMDOCK also missed correct predictions for a significantly number of targets, especially for unbound docking (see, e.g. Figure 4). Therefore, we have manually examined those cases for which HSYMDOCK failed to give any correct binding modes even the top 1000 predictions were considered. It was found that three factors may contribute to the failure in bound and also unbound cases. First, the binding interfaces between subunits in some complexes are small and therefore the native complex can not be ranked favorably during docking. Second, some complexes are formed by intertwined monomers and therefore post a challenge in sampling. In addition, some proteins are not strictly symmetric and thus difficult to predict. For unbound cases, there are two major reasons for the failure. One is the conformational changes in the unbound structures. The other is due to some missing parts or extra segments in the unbound structures, compared to the corresponding bound structures, which will significantly change the binding interface.

Comparing Figures 4 and 5, one can find that HSYMDOCK performs significantly better on the transmembrane proteins than on the general symmetric complexes. This may be understood as follows. The monomers of transmembrane proteins are normally in linear shape, especially for α-helical transmembrane proteins. They are often much longer in one direction (say Z direction) than in other directions (say X and Y directions). Therefore, the binding interface in head-to-head style is much smaller than the correct interface in side-by-side way. This effectively reduce the number of favorable binding modes because of the small binding interfaces of head-by-head binding modes, and thus increase the probability of ranking correct binding modes within top predictions. In contrast, the general symmetric complexes are formed by globule proteins. The monomers of globule proteins are relatively more symmetric than transmembrane proteins and therefore can form favorable binding interfaces in all directions, causing a challenge in ranking.

Examples of the docking models

Figure 6 shows two examples of the docking models predicted by our HSYMDOCK web server. One is a homo-oligomeric protein complex with Cn symmetry, which is the

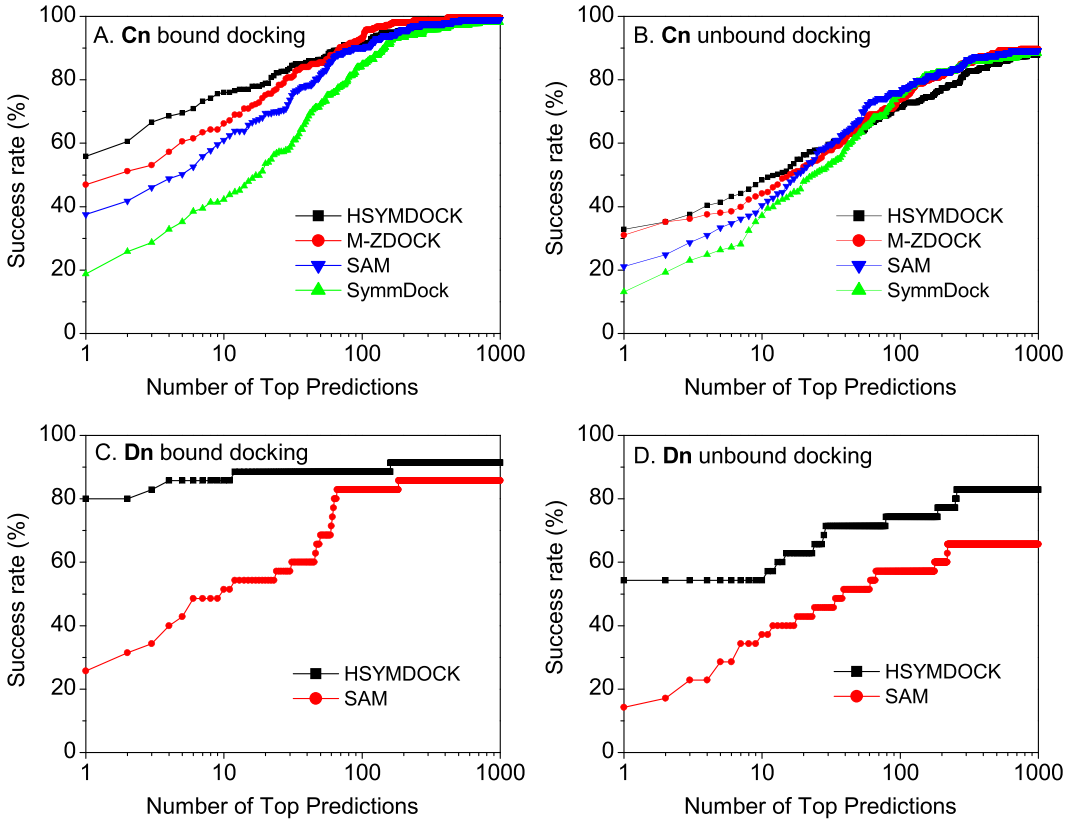

Figure 4. The success rates as a function of the number of top predictions in binding mode predictions for HSYMDOCK and three similar methods, M-ZDOCK, SAM, and SymmDock, on a non-redundant docking benchmark of 213 Cn targets and 35 Dn targets.

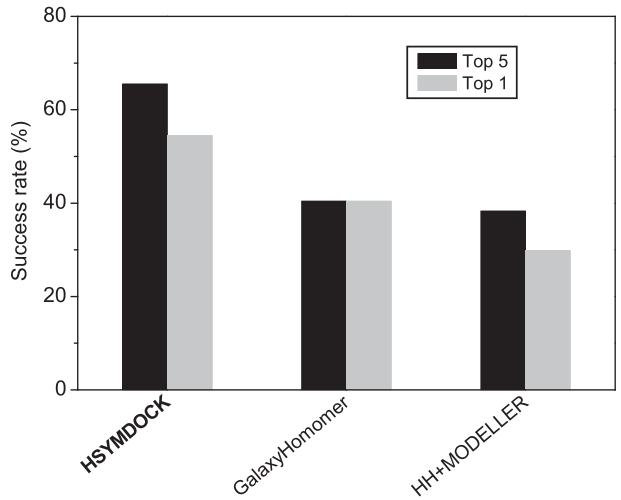

Figure 5. The success rates by HSYMDOCK, GalaxyHomomer and HH+MODELLER on a nonredundant benchmark of 55 transmembrane proteins when the top 1 and top 5 predictions were considered, respectively. The results for the methods other than HSYMDOCK were taken from the literature (26).

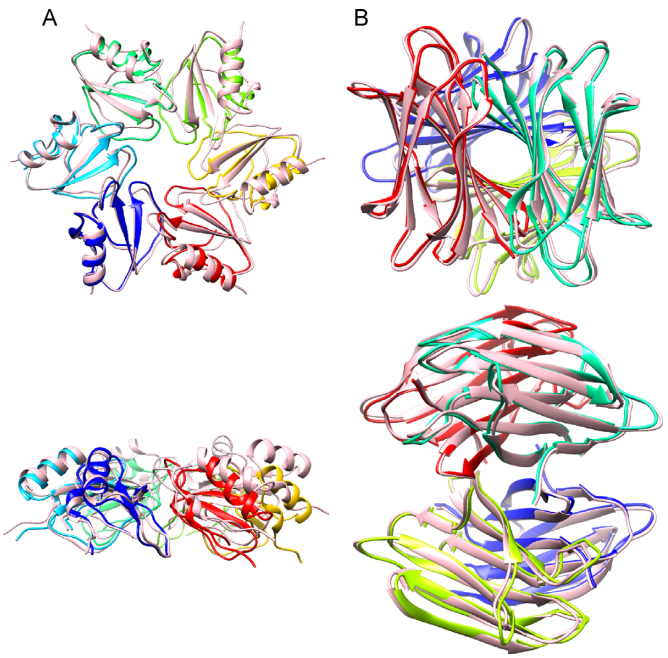

Figure 6. Comparison between the crystal structure and HSYMDOCK server prediction for two symmetric holo-oligomer examples, where the crystal structure is colored in pink and the predicted structure is colored by chain: (**A**) C6 symmetric target (PDB code: 2CI2); (**B**) D2 symmetric target (PDB code: 5A6X). The upper and lower rows are for the top and side views of the complexes, respectively.

serine proteinase inhibitor CI-2 from barley seeds (PDB code: 2CI2) (42). The other is a homo-oligomeric protein complex with Dn symmetry for the HSYMDOCK example page, which is the LecB lectin from Pseudomonas aerug-

inosa strain PA14 (PDB code: 5A6X) (43). Only the sequence of the subunit was provided as input and no binding site information was given when submitting the docking jobs. Among the top 10 constructed models, for the Cn target, the first model gives a medium accuracy with an interface RMSD of 1.07 Å from the native structure (Figure 6A), while for the Dn target, the first model has an high accuracy with an interface RMSD of 0.46 Å (Figure 6B).

CONCLUSION

We have developed HSYMDOCK, a user-friendly web server for predicting the structures of homo-oligomer complexes with Cn or Dn symmetry. The docking server accepts both sequence and structure as input for the subunit. The server efficiently integrates multiple components including sequence search, template selection, model building, and global symmetric docking. Our symmetric docking algorithm, similar to that in HSYMDOCK, obtained correct predictions for 14 cases of 23 homo-oligomer targets in recent CASP11–CAPRI30 challenges. HSYMDOCK was extensively tested on three benchmarks of diverse symmetric protein complexes and showed a significantly better performance than other similar docking algorithms or servers. Without using biological information, HSYMDOK yielded at least acceptable modes for 14 of 20 homo-oligomer targets from CASP11–CAPRI30 within top 5 predictions. When tested on a nonredundant symmetric protein docking benchmark of 213 Cn targets, HSYMDOCK obtained a success rate of 76.1% and 48.8% for bound and unbound docking when the top 10 predictions were considered. HSYMDOCK is also efficient on the targets with Dn symmetry, and obtained a success rate 85.7% for bound docking and 54.3% for unbound docking within top 10 predictions. In addition, HSYMDOCK also showed a strong predictive power on membrane proteins, and obtained a success rate of 65.5% within top 5 predictions on a benchmark of 55 symmetric transmembrane proteins. These results validated our HSYMDOCK as an efficient symmetric docking web server.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [31670724]; National Key Research and Development Program of China [2016YFC1305800, 2016YFC1305805; Huazhong University of Science and Technology. Funding for open access charge: NSFC [31670724].
Conflict of interest statement. None declared.

REFERENCES

1. Wodak,S.J. and Janin,J. (1978) Computer analysis of protein–protein interaction. *J. Mol. Biol.*, **124**, 323–342.
2. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
3. Huang,S.-Y. (2014) Search strategies and evaluationin protein–protein docking: principles, advances and challenges. *Drug Discov. Today*, **19**, 1081–1096.
4. Comeau,S.R., Gatchell,D.W., Vajda,S. and Camacho,C.J. (2004) ClusPro: a fully automated algorithm for protein–protein docking. *Nucleic Acids Res.*, **32**, W96–W99.
5. de Vries,S.J., van Dijk,M. and Bonvin,A.M. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.*, **5**, 883–897.
6. Lyskov,S. and Gray,J.J. (2008) The RosettaDock server for local protein–protein docking. *Nucleic Acids Res.*, **36**, W233–W238.
7. Tovchigrechko,A. and Vakser,I.A. (2006) GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res.*, **34**, W310–W314.
8. Lesk,V.I. and Sternberg,M.J. (2008) 3D-Garden: a system for modelling protein–protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. *Bioinformatics*, **24**, 1137–1144.
9. Macindoe,G., Mavridis,L., Venkatraman,V., Devignes,M.D. and Ritchie,D.W. (2010) HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.*, **38**, W445–W449.
10. Torchala,M., Moal,I.H., Chaleil,R.A., Fernandez-Recio,J. and Bates,P.A. (2013) SwarmDock: a server for flexible protein–protein docking. *Bioinformatics*, **29**, 807–809.
11. Pierce,B.G., Wiehe,K., Hwang,H., Kim,B.H., Vreven,T. and Weng,Z. (2014) ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, **30**, 1771–1773.
12. Schneidman-Duhovny,D., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.
13. de,Vries S.J., Schindler,C.E., Chauvot de Beauchene,I. and Zacharias,M. (2015) A web interface for easy flexible protein–protein docking with ATTRACT. *Biophys J.*, **108**, 462–465.
14. Jimenez-Garcia,B., Pons,C., Svergun,D.I., Bernado,P. and Fernandez-Recio,J. (2015) pyDockSAXS: protein–protein complex structure by SAXS and computational docking. *Nucleic Acids Res.*, **43**, W356–W361.
15. Yu,J., Vavrusa,M., Andreani,J., Rey,J., Tuffery,P. and Guerois,R. (2016) InterEvDock: a docking server to predict the structure of protein¨Cprotein interactions using evolutionary information. *Nucleic Acids Res.*, **44**, W542–W549.
16. Yan,Y., Zhang,D., Zhou,P., Li,B. and Huang,S.-Y. (2017) HDOCK: a web server for protein–protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.*, **45**, W365–W373.
17. Goodsell,D.S. and Olson,A.J. (2000) Structural symmetry and protein function. *Ann. Rev. Biophys. Biomol. Struct.*, **29**, 105–153.
18. Andre,I., Strauss,C.E., Kaplan,D.B., Bradley,P. and Baker,D. (2008) Emergence of symmetry in homooligomeric biological assemblies. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 16148–16152.
19. Poupon,A. and Janin,J. (2010) Analysis and prediction of protein quaternary structure. *Methods Mol. Biol.*, **609**, 349–364.
20. Petsalaki,E. and Russell,R.B. (2008) Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.*, **19**, 344–350.
21. Schneidman-Duhovny,D., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.
22. Pierce,B., Tong,W. and Weng,Z. (2005) M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. *Bioinformatics.*, **21**, 1472–1478.
23. DiMaio,F., Leaver-Fay,A., Bradley,P., Baker,D. and Andre,I. (2011) Modeling symmetric macromolecular structures in Rosetta3. *PLoS One*, **6**, e20450.
24. Biasini,M., Bienert,S., Waterhouse,A., Arnold,K., Studer,G., Schmidt,T., Kiefer,F., Gallo,Cassarino T., Bertoni,M., Bordoli,L. *et al.* (2014) SWISS-MODEL: modelling protein tertiary andquaternary structure using evolutionary information. *Nucleic Acids Res.*, **42**, W252–W258.
25. Lee,H., Park,H., Ko,J. and Seok,C. (2013) GalaxyGemini: a web server for protein homo-oligomer structure prediction based on similarity. *Bioinformatics*, **29**, 1078–1080.
26. Baek,M., Park,T., Heo,L., Park,C. and Seok,C. (2007) GalaxyHomomer: a web server for protein homo-oligomer structure

prediction from a monomer sequence or structure. *Nucleic Acids Res.*, **45**, W320–W324.
27. Ritchie,D.W. and Grudinin,S. (2016) Spherical Polar Fourier Assembly of Protein Complexes with Arbitrary Point Group Symmetry. *J. Appl. Crystallogr*, **49**, 158–167.
28. Janin,J., Henrick,K., Moult,J., Ten Eyck,L., Sternberg,M.J.E., Vajda,S., Vasker,I. and Wodak,S.J. (2003) CAPRI: a critical assessment of predicted interactions. *Proteins*, **52**, 2–9.
29. Yan,Y., Wen,Z., Wang,X. and Huang,S.Y. (2017) Addressing recent docking challenges: a hybrid strategy to integrate template-based and free protein–protein docking. *Proteins*, **85**, 497–512.
30. Huang,S.-Y. and Zou,X. (2008) An iterative knowledge-based scoring function for protein-rotein recognition. *Proteins*, **72**, 557–579.
31. Remmert,M., Biegert,A., Hauser,A. and Soing,J. (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods*, **9**, 173–175.
32. Marti-Renom,M.A., Stuart,A., Fiser,A., Sanchez,R., Melo,F. and Sali,A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291–325.
33. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
34. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
35. Hanson,R.M., Prilusky,J., Renjian,Z., Nakane,T. and Sussman,J.L. (2013) JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Isreal J. Chem.*, **53**, 207–216.
36. Yan,Y. and Huang,S.-Y. (2017) A new pairwise shape-based scoring function to consider long-range interactions for protein–protein docking. *Biophys. J.*, **112**, 470a.
37. Lensink,M.F., Velankar,S., Kryshtafovych,A., Huang,S.-Y., Schneidman-Duhovny,D., Sali,A., Segura,J., Fernandez-Fuentes,N., Viswanath,S., Elber,R. *et al.* (2016) Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP–CAPRI experiment. *Proteins*, **84**, 323–348.
38. Lo Conte,L., Ailey,B., Hubbard,T.J., Brenner,S.E., Murzin,A.G. and Chothia,C. (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res.*, **28**, 257–259.
39. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2444–2448.
40. Kozma,D., Simon,I. and Tusnady,G.E. (2013) PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, **41**, D524–D529.
41. Lensink,M.F., Velankar,S., Baek,M., Heo,L., Seok,C. and Wodak,S.J. (2018) The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins*, **86**, 257–273.
42. McPhalen,C.A. and James,M.N. (1987) Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry*, **26**, 261–269.
43. Sommer,R., Wagner,S., Varrot,A., Nycholat,C.M., Khaledi,A., Haussler,S., Paulson,J.C., Imberty,A. and Titz,A. (2016) The virulence factor LecB varies in clinical isolates: consequences for ligand binding and drug discovery. *Chem. Sci.*, **7**, 4990–5001.