Published online 17 May 2017 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* ***W365–W373***
doi: 10.1093/nar/gkx407

HDOCK: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy

Yumeng Yan, Di Zhang, Pei Zhou, Botong Li and Sheng-You Huang*

School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P. R. China

Received January 25, 2017; Revised April 16, 2017; Editorial Decision April 28, 2017; Accepted April 29, 2017

ABSTRACT

Protein–protein and protein–DNA/RNA interactions play a fundamental role in a variety of biological processes. Determining the complex structures of these interactions is valuable, in which molecular docking has played an important role. To automatically make use of the binding information from the PDB in docking, here we have presented HDOCK, a novel web server of our hybrid docking algorithm of template-based modeling and free docking, in which cases with misleading templates can be rescued by the free docking protocol. The server supports protein–protein and protein–DNA/RNA docking and accepts both sequence and structure inputs for proteins. The docking process is fast and consumes about 10–20 min for a docking run. Tested on the cases with weakly homologous complexes of <30% sequence identity from five docking benchmarks, the HDOCK pipeline tied with template-based modeling on the protein–protein and protein–DNA benchmarks and performed better than template-based modeling on the three protein–RNA benchmarks when the top 10 predictions were considered. The performance of HDOCK became better when more predictions were considered. Combining the results of HDOCK and template-based modeling by ranking first of the template-based model further improved the predictive power of the server. The HDOCK web server is available at http://hdock.phys.hust.edu.cn/.

INTRODUCTION

Proteins and nucleic acids are the two most important types of biological macromolecules in the cell. Their interactions are crucial for many biological processes such as signal transduction, cell regulation, protein synthesis, DNA replication and repair, RNA transcription, etc. Therefore, determination of their complex structures is valuable to understand the biological process at the atomic level and thus develop therapeutic interventions or drugs targeting these interactions. Given the high cost and technical difficulties in experimental methods, molecular docking, which computationally predicts the complex from individual structures, has been playing an important role in the determination of complex structures (1,2).

Docking is a process of sampling and scoring (2,3). Given two individual structures, docking tries to sample all possible binding modes of one structure related to the other. A scoring function is then used to rank the sampled binding modes during and/or after sampling. Due to lack of information about binding sites, which is often the case for many protein–protein and protein–DNA/RNA interactions, *ab initio* global docking is normally needed to sample putative binding modes in six degrees (three rotations plus three translations) of freedom (2). With the fast development in structural proteomics, more and more experimental complex structures are becoming available. The interface information from the complex structures has greatly promoted the development of macromolecular docking in both algorithm and accuracy, as observed in the community-wide Critical Assessment of Prediction of Interactions (CAPRI) (4). Accordingly, protein–protein docking has significantly evolved from initially *ab initio* docking (5–7) to interface-guided docking (8,9) in the past decade. Nevertheless, incorporation of homologous complex information into traditional *ab initio* docking is still challenging, especially for non-expert users (2).

For years, a number of docking algorithms and their web servers such as ClusPro (10), HADDOCK (11), RosettaDock server (12), GRAMM-X (13), 3D-Garden (14), HEX server (15), SwarmDock (16), ZDOCK server (17), PatchDock (18), ATTRACT (19), pyDockSAXS (20), InterEvDock (9) and NPDock (21), have been developed and made available for public access. However, these servers except NPDock were all originally developed for protein–protein docking, although they were then adapted to accept nucleic acids. In addition, all current web servers only accept structures as input. However, many proteins and nucleic acids do not have an available structure. It is a challenge for non-expert users to model structures from sequences for docking, as the variation in starting structures can make a significant difference to docking results (22). This will become more complicated when interface infor-

*To whom correspondence should be addressed. Tel: +86 27 8754 3881; Fax: +86 027 8755 6576; Email: huangsy@hust.edu.cn

mation about binding needs to be incorporated, as shown in template-based docking for an example (23).

To efficiently utilize available homologous complexes in the protein data bank (PDB) (24), we have developed a hybrid docking strategy to automatically incorporate the binding interface information into traditional global docking. With the interface information, our algorithm has ranked within the top-performed groups in recent CAPRI sessions (23,25,26), although part of our success was attributed to our iterative knowledge-based scoring function (27,28). Given the good performance of our docking algorithm, here we have presented the HDOCK server, a general framework for protein–protein and protein–DNA/RNA docking that is similar to our hybrid docking pipeline used for CAPRI. Compared to current docking servers, our HDOCK server accepts not only structures but also sequences as input for proteins and can automatically integrate the binding information from the PDB (24). HDOCK uses intrinsic scoring functions for both protein–protein and protein–DNA/RNA docking. The docking process is fully automated and the results are presented interactively to users by a web page and through an email notification if a valid email is provided.

MATERIALS AND METHODS

Workflow of the HDOCK server

HDOCK is an integrated package of multiple components including several third-party programs, our docking algorithm and scoring functions, and a set of tools developed in our group. The workflow implemented in the HDOCK server is shown in Figure 1.

The first step of the workflow is data input that accepts both sequences and structures for proteins. Only structure input is currently supported for DNAs/RNAs, as it is still a challenge to model DNA/RNA structures from sequences. There is also an option for users to provide binding site information.

The second step of the workflow is sequence similarity search. Given the sequences from input or converted from structures, a sequence similarity search is conducted against the PDB sequence database to find the homologous sequences for both receptor and ligand molecules. For proteins, the HHSuite package is used for sequence search (29), as it is well known for its efficient detection of remote homologs. For DNA/RNA, the FASTA (version 3.6) program is used, as FASTA is a robust and easy-to-use program for both protein and DNA/RNA sequence search (30). Thus, this step will yield two sets of homologous templates (one for receptor and the other for ligand), respectively.

Then, it goes to the third step by comparing two sets of templates to see if they have common records with the same PDB codes. If there are such PDB codes, a common template will be selected for both receptor and ligand. If there is no overlap between two sets of homologous templates, the best templates will be selected for the receptor protein and/or the ligand protein from two sets of templates, respectively. If multiple templates are available, the one with the highest sequence coverage, the highest sequence similarity and the highest resolution will be selected. Priority is also given to the template from a complex over an apo structure during the template selection if the differences between two templates are within 10% in sequence coverage, similarity and resolution. With the selected templates, models are built using MODELLER (31), in which the sequence alignment is conducted using ClustalW (32,33).

With the structures modeled by the server or uploaded by users, the workflow now enters the last step, traditional global docking. Here, HDOCKlite, a hierarchical FFT-based docking program developed in our group, is used to globally sample putative binding orientations (25). The docking process will also incorporate the binding site information if users have provided such information at the time of submission. The docking models by HDOCKlite and template-based model by MODELLER are interactively provided to users for download through a web page and an email notification if an email address has been provided. On the result page, users can also view the top 10 binding models through a Jmol web interface (34).

For the sake of computational efficiency and least interruption of service, a local copy of the PDB database is maintained on the HDOCK web server and updated monthly. It should be noted that although users can submit either sequences or structures as input for proteins, there are some differences for sequence and structure inputs during the HDOCK pipeline. With structure inputs, the HDOCK pipeline will perform a template-based docking if a complex template is found through sequence similarity search; otherwise, the server will do free docking with the input structures. With sequence inputs, the server will search their homologous templates and then build the structures from monomer or complex templates for docking. If no template is found, docking will not be conducted. Nevertheless, sequence inputs should be good enough for most of real applications given the capability of HHSuite and FASTA for remote homology detection.

Docking and scoring methods

An FFT-based global docking program developed in our group, referred to as HDOCKlite, is used to globally sample putative binding modes in the HDOCK server, in which an improved shape-based pairwise scoring function has been used (35). The key point of our new scoring function is that during sampling 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 15° is used for rotational sampling, and a spacing of 1.2 Å is adopted for FFT-based translational search. For each rotation, the top 10 translations with best shape complementarities from the FFT search are optimized by our iterative knowledge-based scoring functions (27,28). The same scoring function is used for both protein–DNA and protein–RNA interactions as DNA is similar to RNA in terms of residue and atom types. One binding mode, that corresponds to the best-scored translation, is kept for each rotation. Given the angle interval of 15°, there are around 4392 evenly distributed rotations in the Euler space. Thus, we have a total of 4392 sampled binding modes for a docking run. The ranked binding modes are clustered with an rmsd cutoff of 5 Å as used in

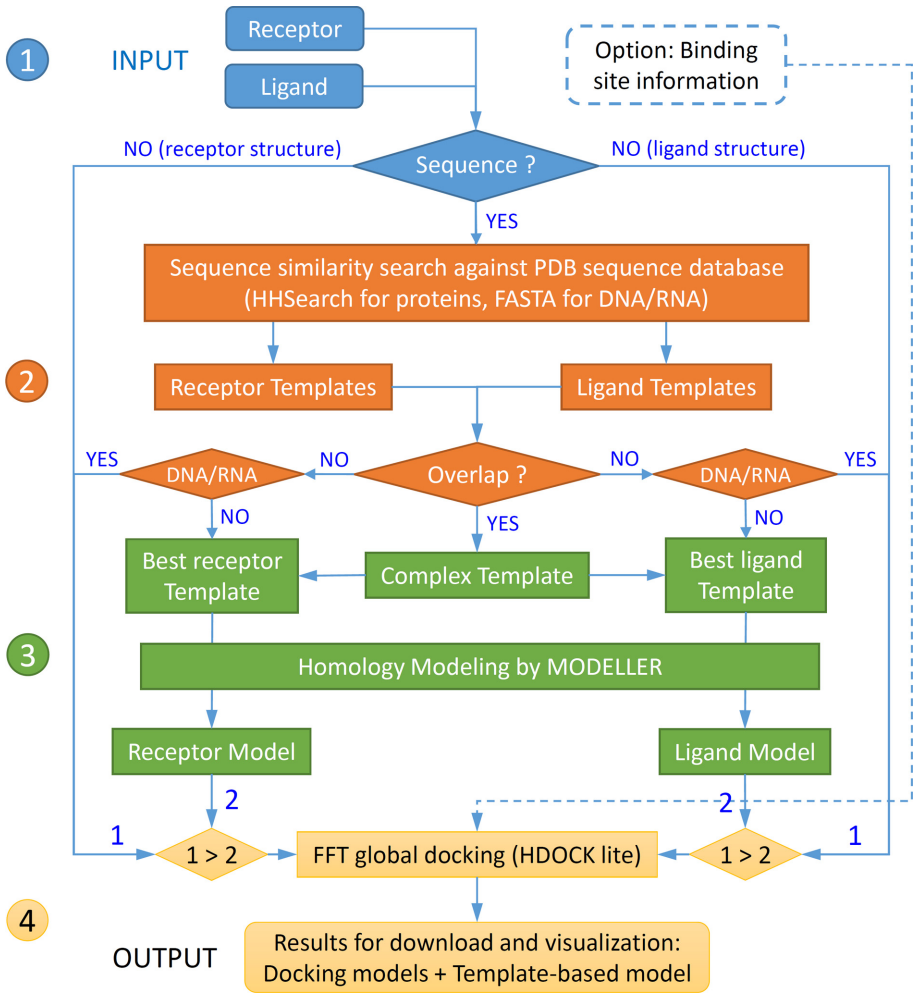

Figure 1. The workflow of the HDOCK web server that is divided into four stages: (**1**) data input, (**2**) sequence similarity search, (**3**) structure modeling and (**4**) FFT-based global docking in which priority is given to user-input structures.

other docking studies (2), where the RMSD is calculated using backbone atoms. If two binding modes have a ligand rmsd of $\leq$5 Å, the one with the better score is kept. By default, the top 100 binding models are pre-generated for users, though users can download the docking output file including all the 4392 binding solutions and generate their own binding models with the program provided on the server help page.

Input

To best facilitate the use of the HDOCK server by normal users, especially for non-expert users, the server is designed to accept both sequence and structure inputs for proteins. Figure 2 shows the web interface of the HDOCK server. For each molecule, the server accepts four types of inputs for proteins, two for structures and two for sequences, as follows:

i) Upload your pdb file in PDB format.
ii) Provide your pdb file in PDB ID:ChainID (e.g. 1CGI:E).
iii) Copy and paste your **protein** sequence below in FASTA format.
iv) Upload your **protein** sequence file in FASTA format.

Only one type of input is needed for each molecule. Currently, the server only supports structure inputs for DNAs/RNAs, as automatic modeling of DNA/RNA structures from sequences is still challenging. For structure input, users can upload their own pdb files or provide the PDB: chain ID(s). Since the server is now designed to model single-chain protein structures from sequences, users are recommended to upload their own structures for better accuracy if their protein contains multiple chains.

In addition, users also have an option to provide binding site information in two forms. One is the residue information of the binding site on one molecule. The other is the residue distance constraint at the binding interface between two molecules. The binding site information if provided will be used during the docking process as well as the post-docking clustering stage as a filter. A few residue information about the binding site is good enough to constrain correct binding modes.

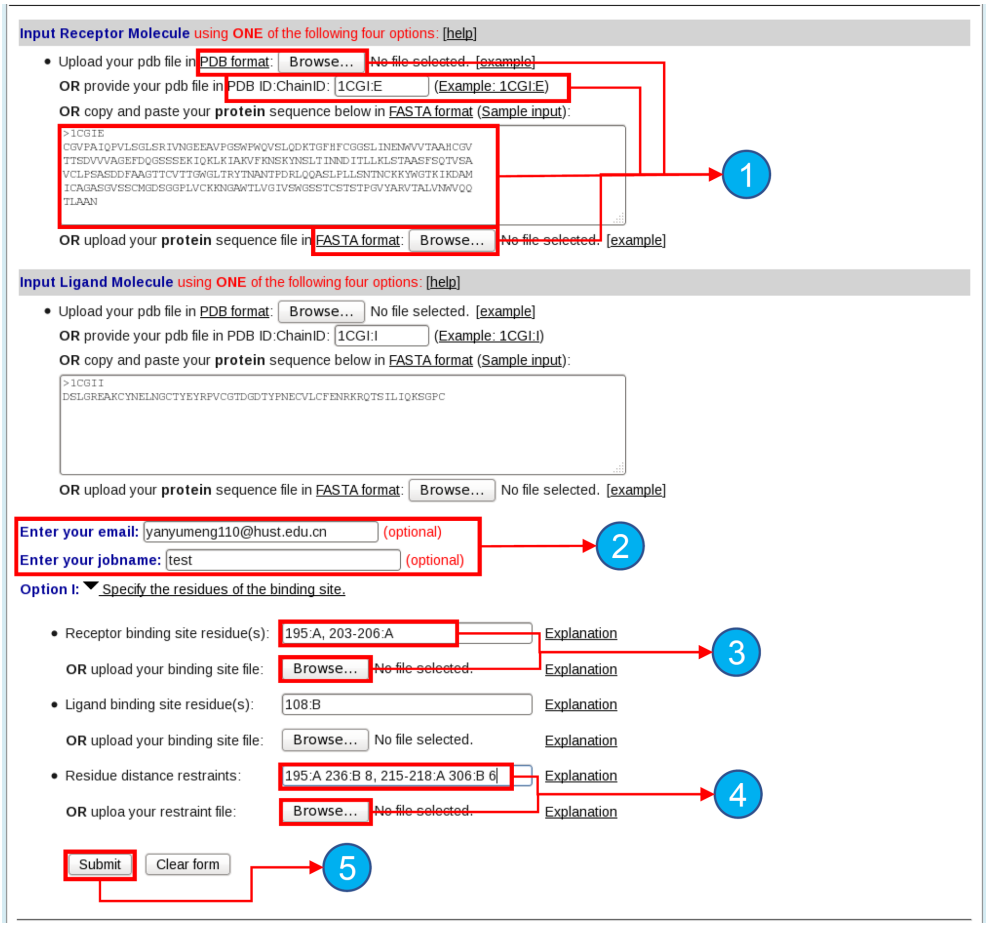

Figure 2. HDOCK server job submission, that includes: PDB file/sequence input fields (**1**), optional user email address and job name (**2**), optional binding site residues (**3**) and distance restraints (**4**) and the 'submit' button (**5**).

Test sets

Although a docking algorithm similar to that in HDOCK has been well tested in CAPRI experiments (23), more evaluations were also done here on public benchmarks to show the robustness and efficacy of our HDOCK algorithm as a server.

If a highly homologous complex template can be found in the PDB, free docking is not needed as a reliable complex structures may be directly constructed by just using template-based modeling techniques. However, in real applications, only weakly homologous complexes might be available, as shown in CAPRI experiments. These weakly homologous templates may not necessarily have similar binding interfaces. Therefore, template-based modeling for such cases may result in misleading predictions. This is just the issue that our HDOCK algorithm is trying to address. As such, we have selected those cases in the twilight zone that have weakly homologous complexes of <30% sequence identity during the preparation of datasets.

The test set for protein–protein docking was from the protein–protein docking benchmark 4.0 developed by the Weng group (36). We have selected those cases that have weakly homologous templates and whose proteins are single-chain structures, resulting in a total of 54 protein–protein complexes (Supplementary Table S1).

The dataset for protein–DNA docking was taken from the protein–DNA docking benchmark created by the Bonvin group (37). Similarly, we have selected those cases with single-chain proteins and weakly homologous templates. This yielded a total of 23 test cases (Supplementary Table S2).

The test sets for protein–RNA docking included three published benchmarks. They were the protein–RNA docking benchmark v1.1 from the Fernandez-Recio group (38), the protein–RNA docking benchmark version 2 from the Bahadur group (39), and the protein–RNA docking benchmark 1.0 by the Zou group (40), respectively. We have used the similar way to select the test cases, and obtained a total of 33 cases for the Fernandez benchmark (Supplementary Table S3), 33 cases for the Bahadur benchmark (Supplementary Table S4) and 25 cases for the Zou benchmark (Supplementary Table S5).

RESULTS

HDOCK server

The HDOCK server is hosted on a Linux server of two Intel(R) Xeon E5-2690 v4 2.60GHz CPUs with 28 cores and

256 GB of Memory. A maximum of 20 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 about 10–20 min. The web server is based 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 is freely available.

After users submit input data, the HDOCK server will create a docking job and put it 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 job status also shows how soon the job will be finished if the job is running. The URL to the docking results is something like http://hdock.phys.hust.edu.cn/date/jobid, where 'jobid' is a unique job id shown on the status web page. Users can bookmark the job status page for access to the docking results at a later time. Users will also be notified by an email once the job is finished if a valid email address has been provided at the time of submission.

Output

Once a job is finished, the job status page will be automatically redirected to the result page, as shown in Figure 3. The docking results include three basic files:

i) Receptor PDB file uploaded by users or constructed by the server from the FASTA sequence provided by users.
ii) Ligand PDB file uploaded by users or constructed by the server from the FASTA sequence provided by users.
iii) HDOCK Output that include 4392 ligand binding modes represented by their transformations.

In addition, the result page also shows the template information for receptor and ligand on the top and a docking summary of the top 10 models on the bottom. The server pre-generates the top 100 binding models for each job. In addition, the template-based model if available, ranking #0, is also provided for viewing and download. Users can download any of the top 20 binding models individually, or choose to download all the top 10 predictions or the top 100 predictions as a package. Users may also download all the results in a single package that includes the Receptor PDB file, Ligand PDB file, HDOCK Output and the top 100 predictions.

Users who want to get more than 100 binding models may download our binding model generation program and follow the instruction on the help page to construct a specified number of binding models.

As the top 10 binding models are normally deemed as the most important models, the result page also provides the interactive view of the top 10 models using the Jmol software (34). Users can choose to view any of the top 10 models or all together by different representations and styles.

Performance of the HDOCK server

Performance for CAPRI. A docking approach similar to that in HDOCK has been used for CAPRI experiments as a predictor and ranked one of the top-performed algorithms in recent CAPRI sessions (25,26). In the recent CAPRI-CASP11 experiment for symmetric oligomer modeling, a template-based docking pipeline similar to that in HDOCK obtained correct complex structures of acceptable accuracy or better for 16 out of the total 25 targets, ranking #1 as a predictor based on the number of targets for which at least one acceptable solution was reached (23,41). When evaluated on the 20 targets of the CAPRI-CASP11 challenge for which we have identified a weakly template, HDOCK obtained a success rate of 75%, compared to 50% for template-based modeling (Supplementary Table S6) (23). Our docking algorithm also performed well on general protein–protein docking and ranked #6 in latest CAPRI challenges (42).

Performance on five public benchmarks. During the evaluations for HDOCK sever, those homologous complexes of $\geq$30% sequence identity with the test cases were excluded. For protein–protein docking, the sequences of unbound structures were used as input for both receptor and ligand. For protein–nucleic acid docking, sequences were input for proteins and the unbound structure if available or the bound structure from the benchmark was provided as input for nucleic acid. For comparison, we have also performed free unbound docking with HDOCKlite by using the unbound structures when provided or the bound structures in the benchmark. During the assessment, a success was defined as that the predicted binding mode has an acceptable accuracy or better according to the CAPRI criteria (4). The success rate was defined as the number of cases with at least one correct model divided by the total number of cases in the test set when a specific number of predictions were considered.

Figure 4A–E showed the success rates of HDOCK server as a function of the number of top predictions in binding mode prediction for those cases with weakly homologous complexes on five benchmarks, respectively. The success rates for template-based modeling and traditional free unbound docking were also shown in the figures. The detailed docking results and used weakly complex templates for the five benchmarks were listed in Supplementary Tables S1–5.

For protein–protein docking (Figure 4A), although template-based modeling obtained a better performance with a success rate of 38.9% when the top prediction was considered, compared to 24.1% for HDOCK, the success rate for HDOCK increased with more predictions considered. When the top 10 predictions were considered, the HDOCK server tied with template-based modeling at the success rate of 38.9%. After top 10 predictions, the success rate of HDOCK server became higher than pure template-based modeling and obtained the success rates of 59.3 and 72.2% for top 100 and 1000 predictions, respectively. Compared to template-based modeling and HDOCK server, free unbound docking had the worst performance and obtained the success rates of 11.1 and 29.6% for top 1 and 10 predictions, respectively (Figure 4A).

An example of the protein–protein binding models predicted by HDOCK server and template-based modeling for target 1OPH is shown in Figure 5. As shown in the fig-

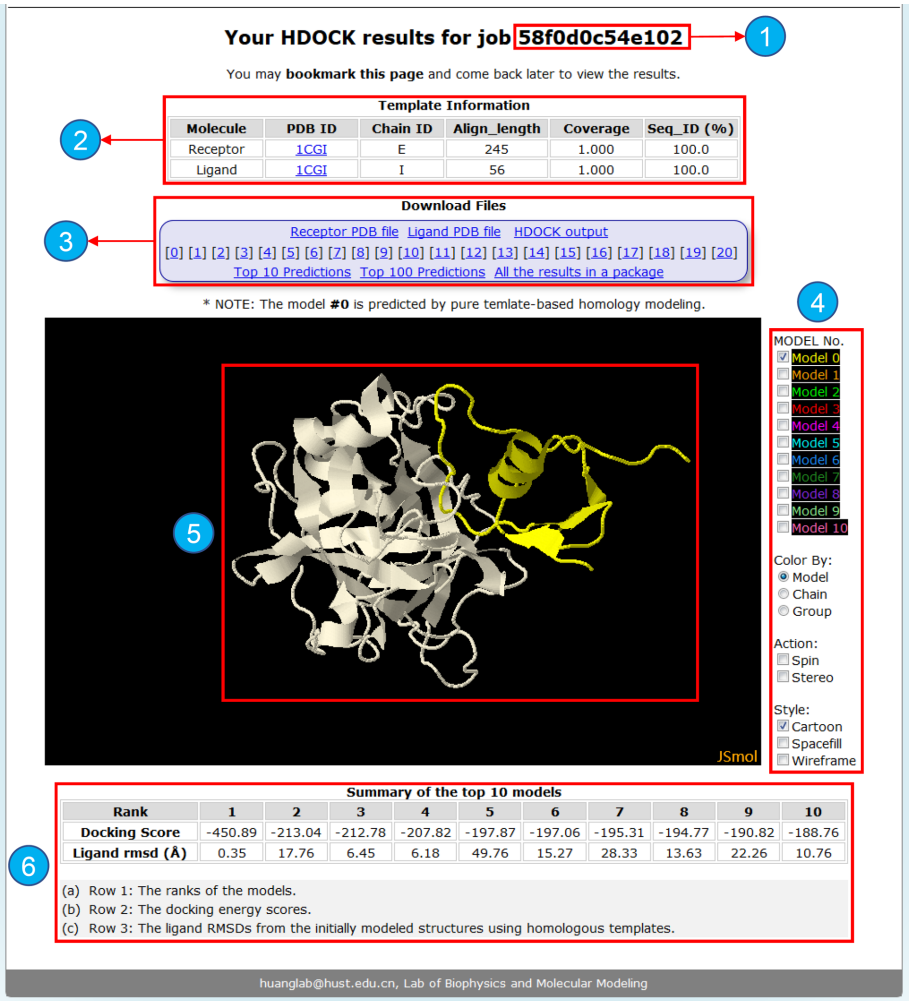

Figure 3. HDOCK server output page. At the top of the page is the unique job ID (**1**), and under it are the template information for receptor and ligand molecules (**2**) and the files for download (**3**). Optional buttons on the right can control Jmol to choose which model to view and how to view (**4**) on the left (**5**). A docking summary of the top 10 models is shown on the bottom (**6**).

ure, due to the different binding interfaces between the target 1OPH and the template 3H5C, template-based modeling generated an incorrect prediction that is far away from the native structure. However, with the docking step, the HDOCK server was able to obtain a binding mode of acceptable accuracy at rank #5, suggesting a rescuing capability in obtaining correct models for cases with misleading templates.

For protein–DNA docking (Figure 4B), similar trends to the performances for protein–protein docking were also observed. When the top prediction was considered, template-based modeling performed the best with a success rate of 26.1%, compared to 17.4% for HDOCK server and 4.3% for free unbound docking (Figure 4B). When the top five predictions were considered, the HDOCK server reached the success rate of template-based modeling, compared to 8.7% for free unbound docking. When the top 100 predictions were considered, the HDOCK server had the highest success rate of 47.8%, compared to 26.1% for template-based modeling and 34.8% for free unbound docking.

For protein–RNA docking (Figure 4C–E), the performances of three approaches depended on specific benchmarks. For example, when the top prediction was considered, the HDOCK server performed better than template-based modeling on the Fernandez and Zou benchmarks, but slightly worse than template-based modeling on the Bahadur benchmark. Specifically, on the Fernandez benchmark, the HDOCK server obtained the success rates of 33.3 and 54.5% when the top 1 and 10 predictions were considered, compared to 21.2 and 42.4% for free unbound docking and 24.2% for template-based modeling (Figure 4C). On the Bahadur benchmark, the HDOCK server obtained the success rates of 33.3 and 51.5% when the top 1 and 10 predictions were considered, compared to 18.2 and 45.5% for free unbound docking and 36.4% for template-based modeling (Figure 4D). On the Zou benchmark, the HDOCK server obtained the success rates of 52 and 64% when the top 1 and 10 predictions were considered, compared to 52 and 64% for free unbound docking and 40% for template-based modeling (Figure 4E).

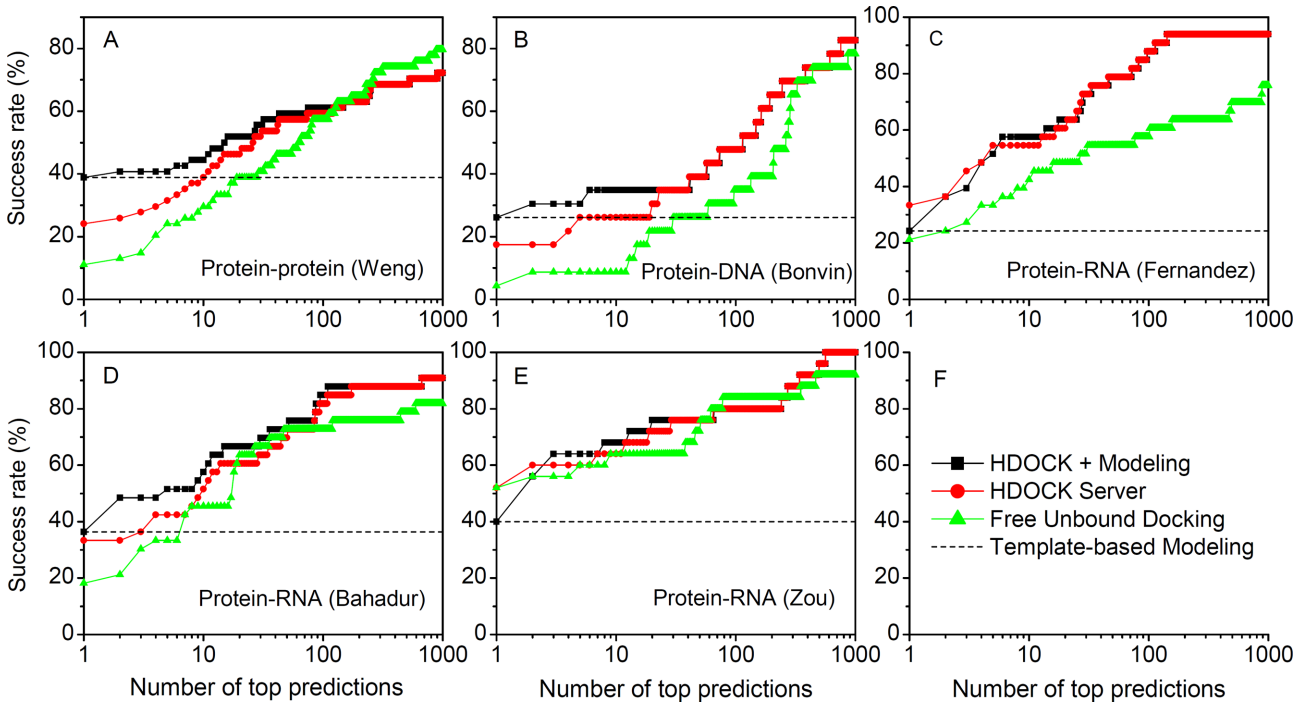

Figure 4. The success rates of HDOCK server and template-based modeling as well as their combination for those cases with weakly homologous complexes from the protein–protein docking benchmark by the Weng group (**A**), the protein–DNA docking benchmark by the Bonvin group (**B**), the protein–RNA docking benchmark v1.1 by the Fernandez group (**C**), the protein–RNA docking benchmark version 2 by the Bahadur group (**D**), and the protein–RNA docking benchmark 1.0 by the Zou group (**E**). For reference, the success rates of free docking with unbound structures from the benchmark were also shown in the figure. The figure legend in panel (**F**) applies to all other panels.

Figure 5. Comparison of the crystal structure (blue), HDOCK server prediction (yellow) and template-based model (green) for the protein–protein complex 1OPH after the receptor structures (red) were superimposed, where the server prediction, ranking #5, has an acceptable accuracy according to the CAPRI criteria (see Supplementary Table S1).

The overall better performance of the HDOCK pipeline than template-based modeling and traditional free unbound docking for top 10 predictions on the five benchmarks demonstrated the functionality and efficacy of the HDOCK server in utilizing the binding information from the PDB for protein–protein and protein–DNA/RNA docking. From Figure 4, one can also see that both HDOCK server and template-based modeling performed the best on some of the five test sets when the top prediction was considered. Therefore, we have combined the results from template-based modeling and HDOCK by ranking first of the template-based model. As shown in the figure, the HDOCK + modeling combination performed the best among the four approaches when the top 10 predictions were considered.

CONCLUSIONS AND FUTURE DEVELOPMENT

The HDOCK server provides a user-friendly web access to our robust hybrid algorithm of template-based modeling and free docking for protein–protein and protein–DNA/RNA complexes. The docking server accepts both sequences and structures as input for proteins. It efficiently integrates various components including sequence search, template selection, model building and global docking. With homologous complexes excluded, overall the HDOCK server obtained a significantly better performance in binding mode prediction than template-based modeling and traditional free unbound docking for protein–protein and protein–DNA/RNA docking on five public benchmarks when the top 10 predictions were considered, demonstrating the efficacy of HDOCK in incorporating the binding interface information from the PDB. The predictive power of the HDOCK server can be further improved by ranking first of the template-based model.

Despite the good performances of HDOCK with weakly homologous complex templates in CAPRI and on the five benchmarks, real docking applications like CAPRI experiments is still challenging, as high-quality homologous complex templates are often not available for new targets. However, with more and more complex structures determined, template-based hybrid docking is expected to play an increasing role in modern docking development. In addition, evaluations are also being done to assess the reliability of automatic modeling of DNA/RNA structures from sequences, aiming to include sequence input option for DNA/RNA in the future development of the HDOCK server.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China (NSFC) [31670724]; National Key Research and Development Program of China [2016YFC1305800, 2016YFC1305805]; Huazhong University of Science and Technology Startup Grant [2006012126, 3004012104]. 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. Huang,S.-Y. (2014) Search strategies and evaluationin protein-protein docking: principles, advances and challenges. *Drug Discov. Today*, **19**, 1081–1096.
3. Vajda,S., Hall,D.R. and Kozakov,D. (2013) Sampling and scoring: a marriage made in heaven. *Proteins*, **81**, 1874–1884.
4. 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.
5. Katchalski-Katzir,E., Shariv,I., Eisenstein,M., Friesem,A.A., Aflalo,C. and Vakser,I.A. (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 2195–2199.
6. Gabb,H.A., Jackson,R.M. and Sternberg,M.J. (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.*, **272**, 106–120.
7. Vakser,I.A. (1997) Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins*, **29**(Suppl. 1), 226–230.
8. de Vries,S.J., van Dijk,A.D., Krzeminski,M., van Dijk,M., Thureau,A., Hsu,V., Wassenaar,T. and Bonvin,A.M. (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins*, **69**, 726–733.
9. Yu,J., Vavrusa,M., Andreani,J., Rey,J., Tuffery,P. and Guerois,R. (2016) InterEvDock: a docking server to predict the structure of protein-protein interactions using evolutionary information. *Nucleic Acids Res.*, **44**, W542–W549.
10. 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.
11. 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.
12. Lyskov,S. and Gray,J.J. (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.*, **36**, W233–W238.
13. Tovchigrechko,A. and Vakser,I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.*, **34**, W310–W314.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. Tuszynska,I., Magnus,M., Jonak,K., Dawson,W. and Bujnicki,J.M. (2015) NPDock: a web server for protein-nucleic acid docking. *Nucleic Acids Res.*, **43**, W425–W430.
22. Huang,S.-Y. (2015) Exploring the potential of global protein-protein docking: an overview and critical assessment of current programs for automatic ab initio docking. *Drug Discov. Today*, **20**, 969–977.
23. 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.
24. 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.
25. Huang,S.-Y. and Zou,X. (2010) MDockPP: a hierarchical approach for protein-protein docking and its application to CAPRI rounds 15-19. *Proteins*, **78**, 3096–3103.
26. Huang,S.-Y., Yan,C., Grinter,S.Z., Jiang,L. and Zou,X. (2013) Inclusion of the orientational entropic effect and low-resolution experimental information for protein-protein docking in CAPRI. *Proteins*, **81**, 2183–2191.
27. Huang,S.-Y. and Zou,X. (2008) An iterative knowledge-based scoring function for protein-protein recognition. *Proteins*, **72**, 557–579.
28. Huang,S.-Y. and Zou,X. (2014) A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.*, **42**, e55.
29. 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.
30. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 2444–2448.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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**(Suppl. 1), 470a.
36. Hwang,H., Vreven,T., Janin,J. and Weng,Z. (2010) Protein-protein docking benchmark version 4.0. *Proteins*, **78**, 3111–3114.

37. van Dijk,M. and Bonvin,A.M. (2008) A protein-DNA docking benchmark. *Nucleic Acids Res.*, **36**, e88.
38. Perez-Cano,L., Jimenez-Garcia,B. and Fernandez-Recio,J. (2012) A protein-RNA docking benchmark (II): extended set from experimental and homology modeling data. *Proteins*, **80**, 1872–1882.
39. Nithin,C., Mukherjee,S. and Bahadur,R.P. (2017) A non-redundant protein-RNA docking benchmark version 2.0. *Proteins*, **85**, 256–267.
40. Huang,S.-Y. and Zou,X. (2013) A nonredundant structure dataset for benchmarking protein-RNA computational docking. *J. Comput. Chem.*, **34**, 311–318.
41. Lensink,M.F., Velankar,S., Kryshtafovych,A., Huang,S.-Y., Schneidman-Duhovny,D., Sali,A. *et al.* (2016) Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment. *Proteins*, **84**, 323–348.
42. Lensink,M.F., Velankar,S. and Wodak,S.J. (2017) Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*, **85**, 359–377.