nature protocols

PROTOCOL

https://doi.org/10.1038/s41596-020-0312-x

Check for updates

The HDOCK server for integrated protein–protein docking

Yumeng Yan[1,2], Huanyu Tao[1,2], Jiahua He[1] and Sheng-You Huang[1]✉

The HDOCK server (http://hdock.phys.hust.edu.cn/) is a highly integrated suite of homology search, template-based modeling, structure prediction, macromolecular docking, biological information incorporation and job management for robust and fast protein–protein docking. With input information for receptor and ligand molecules (either amino acid sequences or Protein Data Bank structures), the server automatically predicts their interaction through a hybrid algorithm of template-based and template-free docking. The HDOCK server distinguishes itself from similar docking servers in its ability to support amino acid sequences as input and a hybrid docking strategy in which experimental information about the protein–protein binding site and small-angle X-ray scattering can be incorporated during the docking and post-docking processes. Moreover, HDOCK also supports protein–RNA/DNA docking with an intrinsic scoring function. The server delivers both template- and docking-based binding models of two molecules and allows for download and interactive visualization. The HDOCK server is user friendly and has processed >30,000 docking jobs since its official release in 2017. The server can normally complete a docking job within 30 min.

Introduction

Protein–protein interactions are involved in various biological processes in cells such as signal transduction, cell regulation, protein synthesis, DNA replication and repair and RNA transcription. Knowledge of the complex structure of the proteins involved is crucial to understanding their interaction mechanism and thus for the development of therapeutic interventions that depend on drugs targeting the interactions. For years, many protein–protein complex structures have been experimentally determined and deposited in the Protein Data Bank (PDB)[1]. However, due to the high cost and technical difficulties in experimental methods, the number of structures of protein complexes in the PDB is still very limited compared with the number of structures of individual protein molecules. As such, molecular docking, which computationally predicts the interaction between two molecules, has played an important role in obtaining structural information of protein dimers and larger complexes[2–4]. Using the individual structures of two molecules of interest, molecular docking samples putative binding modes through a search algorithm and evaluates them with an energy scoring function. Then, all the putative binding modes are ranked according to their binding energy scores, in which the top-scored models are selected as the predicted complex structures[5].

Since the pioneer work by Wodak and Janin[6], molecular docking has advanced considerably, evolving from early ab initio template-free docking (which uses the structures of two molecules only to predict the structure of the complex)[7–9] to information-guided docking (which uses experimental binding information to constrain the sampling of putative binding modes during the docking process)[10–13] in the past decade[5]. Correspondingly, new challenges have been presented during the development process of molecular docking approaches, as seen by the community-wide experiment, Critical Assessment of PRediction of Interactions (CAPRI: https://www.ebi.ac.uk/pdbe/complex-pred/capri/)[14,15]. First, with the rapid development of structural genomics[16], more and more protein–protein complex structures are being experimentally determined. As such, more information about the binding interfaces of involved proteins is becoming available in the PDB[1]. In addition, information about the residue contacts between proteins may also be derived through an evolutionary analysis in sequences[17] or deep learning[18]. However, how to efficiently incorporate such binding interface information into docking is challenging, especially for non-expert biologists. Second, molecular docking starts from the 3D structures of molecules[5]. However, most of the proteins only

[1]School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, P. R. China. [2]These authors contributed equally: Yumeng Yan, Huanyu Tao. ✉e-mail: huangsy@hust.edu.cn

Box 1 | Protein–RNA/DNA docking

As one important function of the HDOCK sever, protein–RNA/DNA docking is also supported. However, predicting accurate structures of RNAs and modeling RNA flexibility upon binding are still challenging in current protein–RNA/DNA docking. The server is being developed further to address these two challenges.

Server Input

To perform protein–RNA/DNA docking with the HDOCK server, users should submit the 3D structure of RNA/DNA whenever possible. RNA/DNA can be input as either receptor or ligand. The server will automatically detect the molecular types of input molecules and start the corresponding docking procedure.

! CAUTION Nucleotide sequence input for RNA/DNA is still under development in the HDOCK server because RNAs/DNAs are much less conserved than proteins in their 3D structures for the same homology. Therefore, structure input of RNA/DNA is strongly recommended for better performance of protein–RNA/DNA docking.

Procedure workflow

The approach for protein–RNA/DNA docking is similar to those for protein–protein docking except two RNA/DNA-relevant ones. First, the FASTA program[100] is used as the sequence similarity search tool to identify possible homologous templates of RNA/DNA for template-based modeling. Second, the intrinsic scoring function for protein-RNA interactions[37] is integrated in template-free docking.

RNA/DNA starting structures

Users may obtain an experimental RNA/DNA structure from the PDB or build a 3D RNA/DNA model using the nucleotide sequence through a third-party RNA/DNA structure prediction approach. Examples of RNA/DNA structure prediction methods include ModeRNA for RNA comparative modeling, 3DNA[73] for DNA/RNA duplex or single strand structural building and 3dRNA[74] for template-free RNA structure prediction. The HDOCK server is also being developed to automatically model 3D structures from sequences for RNA/DNA by integrating such state-of-the-art RNA/DNA structure prediction tools.

! CAUTION The RNA/DNA 3D structure prediction is still in the developing stage[101,102]. Compared to proteins, RNAs are much less conserved than proteins in the 3D structure for homologs with similar sequences. Therefore, good quality models may be generated only through comparative modeling for the targets with high homologous templates[103]. For template-free RNA structure prediction methods like 3dRNA, they normally rely on known RNA 3D structures to construct their fragment template library[14,75]. However, the number of RNA structures is still very limited in the PDB[1,14], which makes the RNA 3D structure prediction extremely challenging. Therefore, for the RNA molecules without high homologous templates and for those non-canonical DNA structures, they may result in low quality or even wrong 3D models. In such cases, users should not solely rely on the predicted protein–RNA/DNA complex structures, and would better validate the predicted complex structure in combination with their experimental data.

have sequences and do not have a 3D structure available[1]. As many users of molecular docking are non-expert biologists who are not familiar with structure prediction, it was necessary to develop a protein–protein docking protocol that also supports amino acid sequence input.

To address these challenges, we have developed a highly integrated HDOCK server for protein–protein docking (http://hdock.phys.hust.edu.cn/) through a hybrid strategy of template-based modeling and ab initio template-free docking. The server automatically incorporates the binding interface information from the PDB and/or user-input biological information like residue restraints and molecular size/shape information obtained from small-angle X-ray scattering (SAXS), supports both amino acid sequence and structure inputs and uses an intrinsic scoring function for protein–protein interactions[19]. In addition, the server is also being developed to support protein–RNA/DNA docking, a relatively new area of molecular docking (see Box 1 for details). Our server is user friendly and has processed >30,000 docking jobs from around the world since its first release in 2017. Recently, HDOCK was ranked as the number one docking server for multimeric protein structure prediction in the community-wide critical assessment of structure prediction 13 (CASP13)-CAPRI experiment in 2018[20]. Here, we present a detailed protocol of our HDOCK server with instructions for server input, job monitoring, result interpretation and troubleshooting, facilitating non-expert biologists to use the docking service. We anticipate that the protocol will also be beneficial for the developers of similar bioinformatics tools.

Development of the HDOCK server

The development of HDOCK has significantly benefited from the CAPRI[15], the community-wide blind evaluation of protein–protein interaction modeling approaches, and has continuously been motivated by the needs of experimental biologists. Since it was launched in 2001, the CAPRI has served as a valuable platform for promoting the development and improvement of new and existing docking algorithms. We started to develop our HDOCK server with our free protein–protein docking

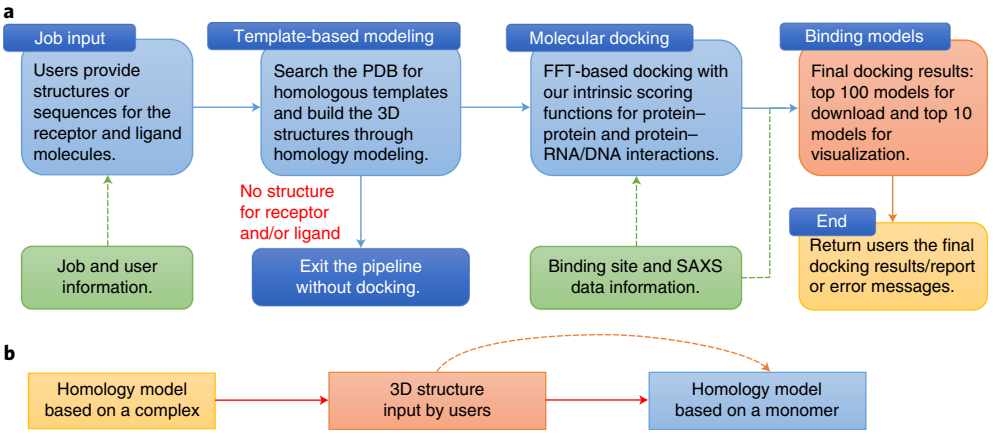

Fig. 1 | Procedure of the HDOCK server. a, Workflow used by the HDOCK server, including four stages: job input, template-based modeling and structure prediction, molecular docking and final results delivery. The steps in the green boxes are optional. **b**, The priority order of the 3D structures used by the HDOCK server for docking, where the orders for the protein in the 'hybrid' and 'template-free' docking modes are indicated by the red solid and orange dashed arrows, respectively.

program, HDOCKlite (http://huanglab.phys.hust.edu.cn/software/hdocklite/)[21,22]. The docking program first samples the putative binding modes between two proteins through a fast Fourier transform (FFT)–based global search method[7] and then evaluates the sampled binding modes with an improved iterative knowledge-based scoring function for protein–protein interactions[21]. Such FFT-based approaches and hierarchical docking algorithms have demonstrated their advantages and good performance in the CAPRI experiments[23,24].

Until 2016, many steps of our docking pipeline still needed to be manually done, such as building structures from sequences, incorporating binding interface information and running docking commands, although our docking protocol achieved an excellent performance in the multimeric protein structure prediction in the community-wide CASP11-CAPRI and CASP12-CAPRI experiments[21,25,26]. Then, considering that many experimental biologists do not have the expertise and computer resources to manually conduct such docking computations step by step, we streamlined the whole process of our docking protocol and published the first automatic version of HDOCK[27], a web server for protein–protein and protein–RNA/DNA docking based on a hybrid strategy, in 2017. In this version, both structures and amino acid sequences are supported for proteins, but only structures are accepted for RNA/DNA, given that automatic structure prediction of RNAs/DNAs is much more challenging than that of proteins. Compared to proteins, RNAs and non-canonical DNAs are much less conserved in the sequence–structure relationship. As such, predicting the structures from sequences is much more difficult for RNA/DNA than for proteins, and RNAs/DNAs are also much more flexible than proteins, which make RNA/DNA-docking extremely difficult. Currently, template-free structure prediction of RNAs is still not reliable, and good-quality models can be generated based only on high-homologous templates for RNA/DNA molecules (see Box 1 for details).

Since its release in 2017, the HDOCK server has been well received by the research community[28–35] and completed >30,000 jobs from around the world. Meanwhile, major updates have been done with our HDOCK server, including the ability to incorporate information from experimental SAXS data and improvement of homology search, template-based modeling, and protein structure prediction. Recently, HDOCK was ranked as the number one docking web server for multimeric protein structure prediction in the community-wide CASP13-CAPRI challenge in 2018[20].

Overview of the procedure

The workflow of our HDOCK server is shown in Fig. 1a. There are two working modes in the server: one is the default hybrid docking mode, and the other is the template-free docking mode. First, users are required to provide the input of two individual proteins (Steps 2 and 3), one for receptor and the other for ligand, in which both amino acid sequences and PDB structures are supported. Then, the server will do the template-based modeling of the receptor and ligand molecules by searching the PDB for putative homologous templates based on the sequences of proteins. If a homologous complex template is found, new structures will be built for the receptor and/or ligand based on the found

template whether the input is a sequence or structure; otherwise, the 3D structure will be built from a monomer template if the input is a sequence. If no template is found for an input sequence, the docking pipeline will exit without predictions (Fig. 1a). Therefore, there could be three possible types of individual protein structures used by the HDOCK server (Fig. 1b): (i) homology models based on a complex template, (ii) PDB structures input by users, and (iii) homology models based on a monomer template, of which the first available one will be selected according to the order of their priorities. For 'hybrid' and 'template-free' docking modes, the priority orders are 1-2-3 and 2-3, respectively. Next, with the structures of receptor and ligand molecules, the HDOCK server will perform global docking to sample putative binding modes through an FFT-based search method and then evaluate them with our intrinsic scoring function for protein–protein interactions. Biological information, such as experimental data on the protein–protein binding site or SAXS profile, can be incorporated during the docking (Step 7) and/or post-docking processes (Steps 6 and 7). Finally, the top 100 predicted complex structures are provided to users for download, of which the top 10 models can be visualized through an interactive NGL viewer[36] on the result web page (Steps 13 and 14).

Applications of the HDOCK server

Studying protein–protein binding

One basic function of the HDOCK server is to predict the interaction interface of two interacting proteins. One important part of the HDOCK server is that it supports amino acid sequence inputs and can automatically incorporate experimental information that is available about the interaction between the two proteins. Therefore, HDOCK can be used to study the molecular mechanism of protein–protein interactions through the predicted complex structure of involved proteins, especially when some biological information about binding is available. For example, Dudenhoeffer et al.[31] have used the HDOCK server to model the complex structure between the acidic repeat 2 domain of N-utilization substance factors A (NusA-AR2) and the inositol monophosphatase SuhB. The chemical shift perturbations of the [^{1}H,^{15}N]-heteronuclear single quantum coherence titration of ^{15}N-NusA-AR1-AR2 with SuhB could be used to determine the key residues of NusA-AR2 that interact with SuhB. In this example, the NMR experimental information could be translated into the residue restraints as input of the HDOCK server. With the protein structures and residue restraints as input, an appropriate model was generated by the HDOCK server, which shows that NusA-AR2 binds to the positively charged area of SuhB via negatively charged helices α3 and α5[31].

Studying protein–RNA/DNA interactions

Protein–RNA/DNA docking is still being developed in the HDOCK server due to the extreme difficulty in the structure prediction of RNAs/DNAs and high flexibility of RNA/DNA upon binding (see Box 1 for details). Therefore, users should not solely rely on the predicted protein–RNA/DNA complex structures. HDOCK integrates an intrinsic scoring function for protein–RNA docking[37] and can also be used to study the molecular mechanism of protein–RNA/DNA interactions based on the predicted complex structure. For example, to investigate the interaction mechanism between virulence-associated proteins C (VapC) toxins and their substrates, Deep et al.[33] performed molecular docking to construct the complex of VapC11 and tRNA-LeuCAG using the HDOCK server, using the individual structures as input. They used the criterion that the cleavage site in the anticodon loop of tRNA should be close to the active site in VapC11 to manually select the appropriate model. According to the predicted model, they found that tRNA might interact with the surface-exposed arginine residues of VapC11 such as Arg14, Arg25 and Arg94. The subsequent mutational experiments validated that Arg94 is crucial for binding tRNA substrate[33].

Determining the complex structures

In addition to studying the interaction between two molecules by combining experimental information, HDOCK may be used to help experimental biologists determine protein complex formation of two interacting molecules from scratch. For example, Kostareva et al.[34] used the HDOCK server to predict the interaction interface between the V_HH domain and IL-17A dimer, using their individual structures as input. Structural comparison between the predicted model and the experimental structure of IL-17A bound to the CAT 2200 antibody Fab fragment (PDB ID: 2VXS) reveals that the V_HH/IL-17A complex shows more extensive contact areas than the complex with CAT 2200 antibody, which explains the high affinity of the V_HH domain for IL-17A[34].

Table 1 | Comparison between HDOCK and other docking web servers

Web server name	Ab initio docking	Amino acid sequence input	PDB structure input	Template-based modeling	Protein–protein docking	Protein–RNA/DNA docking[a]	SAXS data
HDOCK	√	√	√	√	√	√	√
ClusPro	√	×	√	×	√	×	√
HADDOCK	√	×	√	×	√	×	√
GRAMM-X	√	×	√	×	√	×	×
HEX	√	×	√	×	√	×	×
NPDOCK	√	×	√	×	×	√	×
pyDockWEB	√	×	√	×	√	×	×
pyDockSAXS	√	×	√	×	√	×	√
ZDOCK	√	×	√	×	√	×	×
PatchDock	√	×	√	×	√	×	×
SWARMDOCK	√	×	√	×	√	×	×

[a]Indicates whether the method includes an intrinsic scoring function for protein–RNA/DNA interactions.

Comparison with other methods

A number of docking servers, such as ClusPro[38,39], HADDOCK[10–12], GRAMM-X[40], 3D-Garden[41], HEX[42], SWARMDOCK[43], ZDOCK[44], PatchDock[45], RosettaDock[46], ATTRACT[47], pyDockWEB[48,49], pyDockSAXS[50], InterEvDock[13] and NPDock[51], have been developed in the past decade. However, HDOCK distinguishes itself from similar docking servers in several major aspects (Table 1). First, unlike in other docking servers where the binding interface information is used as a restraint during sampling, HDOCK implicitly integrates the binding interface information in the PDB into the structures of receptor and ligand through an efficient hybrid strategy, which significantly improves the accuracy and efficiency of incorporating such biological information[21]. Second, HDOCK supports both structure and amino acid sequence inputs for proteins by integrating state-of-the-art protein structure prediction approaches (HH-suite[52], ClustalW[53] and MODELLER[54]), which is especially useful for experimental scientists who are not familiar with structure prediction. Third, HDOCK uses a new pairwise scoring function to consider long-range shape complementarity in FFT-based docking[22,55], which significantly increases the sampling enrichment of near-native binding modes. Fourth, HDOCK integrates intrinsic statistical mechanics-based iterative scoring functions for protein–protein interactions[19] and protein–RNA interactions[37] during the docking process. Fifth, HDOCK supports options for many types of biological information during the docking/post-docking process, including SAXS data, homologous templates and binding site information.

Moreover, HDOCK is one of the fastest docking servers and can normally finish a docking job within 30 min. Despite its fast speed, the HDOCK server is also accurate due to its efficient integration of biological information. We have participated in the latest community-wide CASP13-CAPRI Assembly prediction challenge in 2018, where only sequences are provided for the monomer, and participants are required to submit their predicted multimeric structures before a deadline. As shown in Table 2, among the eight participating servers including HDOCK[27], SWARMDOCK[43], ClusPro[38], LZERD[56], MDOCKPP[57], HADDOCK[11], GALAXYPPDOCK[58] and HAWKDOCK[59] that submitted predictions to CAPRI, our HDOCK protocol was ranked as the number one docking server for multimeric protein structure prediction in the CASP13-CAPRI experiment according to the CAPRI official evaluation[20]. Specifically, our HDOCK server predicted the most targets with acceptable accuracy or better for all three difficulty categories of targets (Table 2), and gave success rates of 100%, 27.3% and 60.0% within the top 10 predictions for the 9 'Easy', 11 'Difficult' and 'All' 20 targets, respectively (Fig. 2a–c). In addition, using the HDOCK server, our group also achieved the best performance among the 38 CAPRI+CASP groups for those cryo-electron microscopy (CryoEM) targets according to the CASP official evaluation[60] (Fig. 2d).

Despite the advantages and superior performance of HDOCK, there are some other protein–protein docking approaches that distinguish themselves from HDOCK in handling flexibility, and these approaches may be selected by users if flexibility is a concern. For example, HADDOCK is well-known for its efficient incorporation of interaction restraints by allowing for full structural

Table 2 | Performances of the participating web servers in the CASP13-CAPRI challenge

Rank	Server name	Easy targets		Difficult targets		All targets	
		Number[a]	Success[b]	Number[a]	Success[b]	Number[a]	Success[b]
1	HDOCK	9	9/5***/4**	11	3/2**	20	12/5***/6**
2	SWARMDOCK	9	8/4***/4**	11	1***	20	9/5***/4**
3	ClusPro	9	9**	11	3/1**	20	12/10**
4	LZERD	9	7/3***/4**	11	2**	20	9/3***/6**
5	MDOCKPP	9	8/2***/4**	11	3	20	11/2***/4**
6	HADDOCK	9	9/3***/3**	10	0	19	9/3***/3**
7	GALAXYPPDOCK	6	5/2***/2**	11	3/1***	17	8/3***/2**
8	HAWKDOCK	2	2/1***	5	0	7	2/1***

Shown are the numbers of successfully predicted targets within the top 10 predictions for 9 'Easy', 11 'Difficult' and 'All' 20 targets, where the ranking is based on the results of all targets according to the CAPRI official evaluations. [a]The total number of targets tried by the corresponding server. [b]The success is represented by the numbers of targets with an acceptable or better/high (***)/medium (**) accuracy.

flexibility of both side chains and backbone[11]. ATTRACT[47] and SWARMDOCK[43] use normal modes to consider large conformational changes of backbone. ClusPro is a cluster-based docking approach that may consider some entropic effects[38]. RosettaDock is developed for local flexible protein–protein docking through side-chain repacking[46]. In addition, some approaches focus on rescoring. For example, InterEvDock integrates the InterEvScore potential to account for co-evolutionary information in the docking process[13]. pyDOCKWEB is a global docking server based on the FFT-based docking program FTDock[7] and a new version of the pyDock scoring algorithm[47]. HAWKDOCK is a post-docking protocol to rescore and cluster the binding decoys generated by the docking program ATTRACT[59].

Limitations

Molecular flexibility

The docking engine in the HDOCK server performs rigid-body docking by mapping the receptor and ligand molecules onto grids. Therefore, flexibility within the proteins may have a significant impact on the docking results, for instance, if the receptor and/or ligand molecules experience large conformational changes upon binding, though some degree of flexibility can be implicitly considered through grids. This is an inherent issue for HDOCK or any other rigid-body docking approach[5]. However, this limitation may be substantially reduced if a homologous complex template is available for the receptor and ligand to be docked, as the template-based docking in HDOCK can incorporate larger conformational changes compared to template-free docking. The impact of flexibility may also be indirectly addressed if users provide a few residue restraints about binding, which can significantly help the HDOCK protocol select the correct binding modes.

Accuracy of protein homology models

One highlight in the HDOCK server is that it supports amino acid sequence input for proteins. However, it should be noted that the accuracy of 3D structures generated by homology modeling is strongly dependent on the sequence identity between target and template[36], although our docking strategy and scoring function can consider some degree of structural uncertainties. Generally, models tend to be reliable with an overall root mean square deviation (RMSD) of <1 Å between the model and the experimental structure for sequence identity of >50%[61]. Models are also acceptable with some errors in loops in the 30–50% identity range; however, below 30% sequence identity, serious errors can occur even in the basic folds[61]. Normally, our HDOCK server is able to build reliable 3D structures from sequences for easy and medium-difficult targets[62]. However, for those difficult targets, the HDOCK server may not be able to find a template to construct a structure and thus will give no predictions on the result page. In those cases, users are recommended to use a specialized structure prediction method like I-TASSER[63–65] for structural modeling and then submit the modeled 3D structures to the HDOCK server for docking. In addition, HDOCK is currently designed to model single-chain protein structures from sequences. Therefore, users are recommended to upload their own 3D structures if their proteins contain multiple chains.

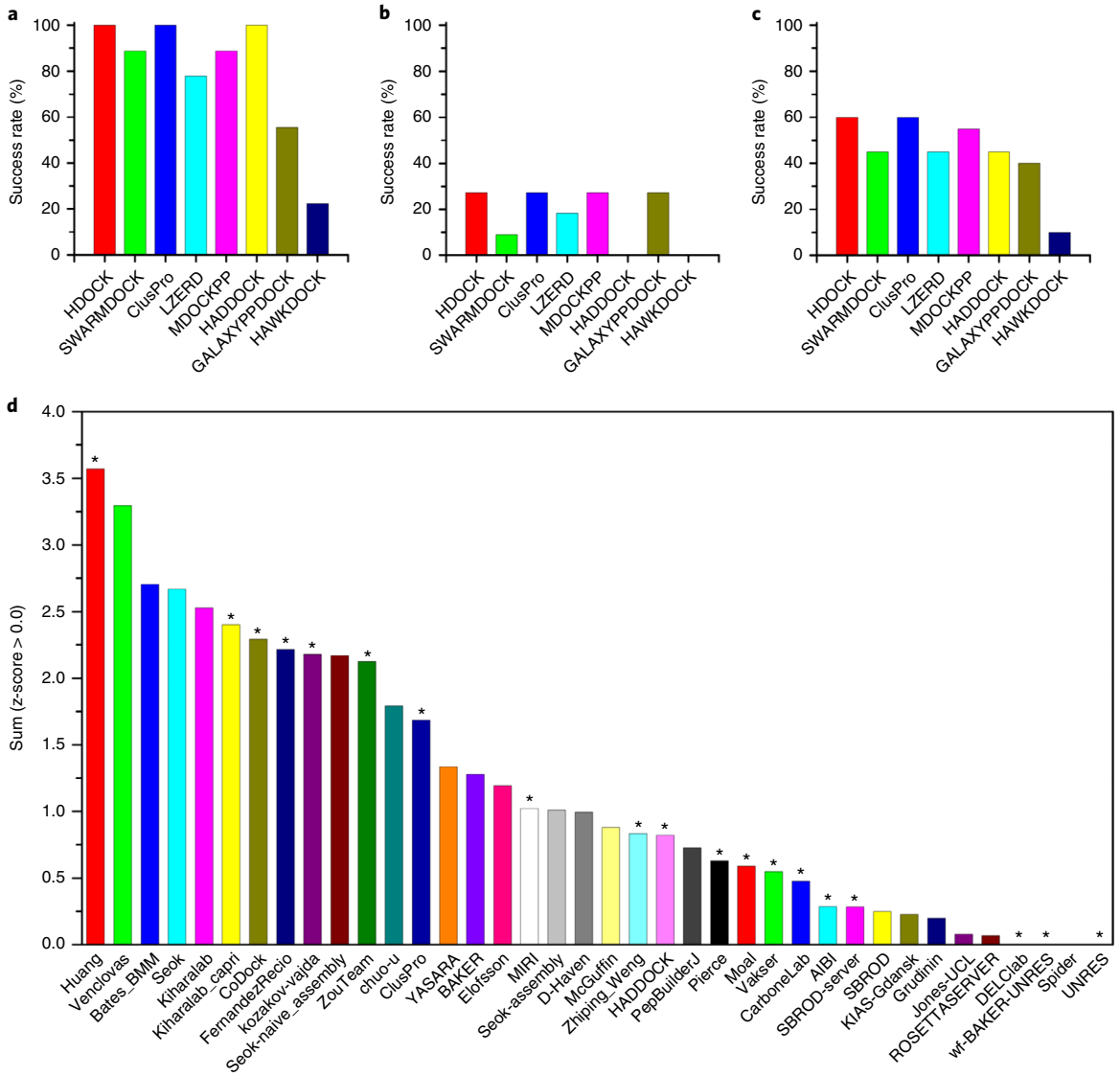

Fig. 2 | Performance of the HDOCK approach and other docking servers/methods on different categories of targets in the CASP13-CAPRI assembly challenge in 2018. a–**c**, Success rates on the 9 'Easy' targets (**a**), 11 'Difficult' targets (**b**) and 'All' 20 targets (**c**) for the top 10 predictions according to the CAPRI criteria. **d**, Summed z-scores of 38 CASP+CAPRI groups on those 'CryoEM' CAPRI targets for Model 1, where those indicated by a '*' stand for CAPRI groups and 'Huang' is our group. Data are adapted from the CASP13-CAPRI assessment report[20] and the CASP13 ranking results for multimeric targets published on the CASP website (http://predictioncenter.org/casp13/zscores_multimer.cgi), respectively.

Experimental design

Protein structure prediction

One important feature of the HDOCK server is its ability to support amino acid sequence inputs by integrating state-of-the-art structure prediction methods. Given an amino acid sequence, the server can automatically construct the protein's 3D structure through homology modeling. The HHSuite package[52] is used to search the PDB for the homologous templates of the protein. If there is more than one template, the server will select the best one according to the criteria of sequence coverage, similarity and resolution, or the one from a protein complex over a monomer if two templates are <10% different in sequence coverage, similarity and resolution[21]. Then, the 3D model is built based on the selected template using MODELLER[54], where the sequence alignment between the target and the template is performed using ClustalW[53]. If no reasonable template is found for an input protein sequence, the HDOCK server will exit the docking pipeline without predictions. After a 3D structure is modeled, the structure will be further refined through a 5,000-step AMBER[66] MD minimization to remove severe atomic clashes.

Template-based docking

Another notable feature that distinguishes HDOCK from similar servers is that it efficiently integrates template-based modeling into docking. Specifically, given a receptor and a ligand, a sequence similarity search is performed against the PDB to identify possible complex templates through the BLAST program[67]. If both the receptor and the ligand are homologous to the corresponding partners of a complex in the PDB, the complex is regarded as a template candidate. If multiple templates are available, the one with the highest sequence coverage, the highest sequence similarity and the highest resolution is chosen. Next, with the selected complex template, the 3D structures of receptor and ligand are built based on the corresponding homologous partners in the complex template using MODELLER[54]. Then, the corresponding template-based complex of receptor and ligand can be constructed by superimposing the modeled receptor and ligand structures onto the complex template, where a 5,000-step AMBER MD minimization[66] is also conducted to remove severe atomic clashes at the interface. Theoretically, the docking pipeline may stop here if the template-based complex is reliable. However, homologous proteins may not necessarily have a similar binding interface when forming a complex. They may either have completely different binding interfaces from the true complex or different binding orientations even if they share a common binding site. In such cases, pure template-based modeling may lead to wrong predictions or no prediction due to lack of homologous complex templates, as demonstrated in our previous study[21]. Therefore, to increase the robustness of our docking protocol, an additional free docking is also performed using the receptor and ligand partners from the homology-modeled complex.

Template-free or ab initio docking

In HDOCK, docking of two molecules is conducted using an FFT-based hierarchical approach[7]. First, possible binding modes are globally sampled through an FFT-based global search strategy with an improved shape complementarity scoring method[22,55], where proteins are represented by a reduced model. Specifically, one molecule (e.g., the receptor) is fixed, and the second molecule (e.g., the ligand) is evenly rotated in 3D Euler (i.e., rotational) space. For each rotation of the ligand, both the receptor and ligand molecules are mapped onto grids by considering the long-range interactions of atoms[22,55]. Then, the putative binding modes between receptor and ligand are exhaustively sampled in 3D translational space through an FFT-based search approach. During the docking calculation, evenly distributed orientations are sampled in rotational Euler space with an angle interval of 15° and translational space with a grid spacing of 1.2 Å. For every rotation of the ligand in 3D Euler space, the top 10 orientations with the best shape complementarity scores in 3D translational space are evaluated by a new version of our scoring function ITScorePP, and the top-scored one is kept. For the angle interval of 15°, we have 4,392 evenly distributed rotations in 3D Euler space, which results in a total of 4,392 final binding modes for a global docking in 6D space. Finally, the 4,392 binding modes are ranked according to their binding energy scores and clustered. If the ligand RMSD of two binding modes is below a cutoff of 5 Å, the one with the better score will be selected, where the RMSD is calculated based on the backbone atoms of the ligand.

Symmetric multimer docking

We recently included an option into HDOCK for symmetric multimer docking in the 'Advanced Options' section. However, users can always use the general HDOCK protocol for homo-multimer docking with the same receptor and ligand molecules[27] or our specialized HSYMDOCK server for symmetric multimer docking[68]. For the symmetric multimer docking mode in HDOCK, only the receptor molecule is needed, and the ligand molecule is ignored. Both cyclic (Cn) and dihedral (Dn) symmetries are supported. The binding mode sampling algorithm and scoring functions for symmetric multimer docking are similar to those for general template-free docking, as described above, except that the symmetry is used as a restriction to reduce the search space in the symmetric docking[69]. Specifically, for Cn symmetric docking, given a monomer, we obtain the 'Receptor' orientation by rotating the monomer by a set of angles in Euler space and the 'Ligand' by rotating the 'Receptor' by 360°/n about the z-axis. Then, the binding modes between the 'Receptor' and 'Ligand' can be exhaustively sampled in the x-y plane using the same FFT-based search algorithm. Repeating the above process will result in a global Cn symmetric docking for the monomer. For Dn symmetric multimer docking, it can be realized by a Cn plus C2 docking[68]. Namely, the Cn symmetric complex structures are first generated through a Cn symmetric multimer docking, and then the Dn symmetric binding modes can be obtained through an additional C2 symmetric docking around a perpendicular axis by using the previously predicted Cn complexes. During symmetric

docking, an angle interval of 10° is used for rotational sampling, and a grid spacing of 1.2 Å is adopted for FFT-based translational search[69].

Incorporating residue restraints

The residue restraints for binding, if provided by users, are incorporated into the FFT-based docking process and post-docking stage in the HDOCK server (Step 7). During the docking process, to increase computational efficiency, the binding site restraints are mapped onto grids. Specifically, an identification (ID) number of 1 or n (>1) is assigned to those grid points occupied by the binding site residues or distance-restrained residues, respectively. Thus, the restraints can be included by retaining those binding modes that satisfy the following criteria: at least one of the grid points with an ID number of 1 for one molecule must be within a certain distance from the grid point of the other molecule, and the grid points with an ID of n (>1) for one molecule should be within the defined distance from the grid points with the same ID for the other molecule. During the post-docking stage, only a few thousand top binding modes are retained, and thus the computational cost is no longer an issue. As such, the residue restraints are directly incorporated by examining the distances between the residues from the receptor and ligand molecules in this stage.

Incorporating experimental SAXS data

The HDOCK server is equipped to incorporate the SAXS intensity profile into the docking process (Step 6). This is done through a weighted score during the post-docking stage, given the low-resolution nature of SAXS data. Specifically, the server first conducts a comprehensive quality check of user-input SAXS data and determines the useful SAXS data range using the ATSAS package[70] (See Box 2 and Fig. 3 for details). Next, with the SAXS data, the chi values of all the 4,392 binding models, which are predicted by FFT-based global docking, are calculated by using our in-house program *fox_hdock* according to the FoXS algorithm[71]. Then, the 4,392 binding modes are re-ranked according to their weighted scores of docking energy scores and SAXS chi values, $E_{\mathrm{ITScorePP}} + w \cdot E_{\mathrm{saxs}}$, where the weight w has been optimized on an independent training set with experimental SAXS data prepared by our group and validated on 11 cases with experimental SAXS data from the literature[72] (Table 3). The top-ranked models are then output as the final predictions. The quality report from the previous SAXS data check will also be shown on the result page. Users should be cautious in using the SAXS data if the report contains some WARNING messages (see Box 2 for details).

HDOCK for protein–RNA/DNA docking

In addition to protein–protein interactions, protein–RNA/DNA interactions are also crucial in many cellular processes. Therefore, protein–RNA/DNA docking is also pressingly needed. Traditional macromolecular docking approaches were originally designed for predicting protein–protein complex structures, although they were later adapted to accept RNA/DNA. As shown in our previous study, the scoring functions for protein–protein interactions and protein–RNA interactions are not transferable[37], although those for protein–RNA interactions and protein–DNA interactions are[27]. Thus, it is necessary to have a protein–nucleic acid docking algorithm with an intrinsic scoring function for protein–RNA/DNA interactions. Meeting the need, protein–RNA/DNA docking was also supported by the HDOCK server when the 3D structure of RNA/DNA is provided as input[27], though the server is being developed to support sequence input for RNA/DNA by integrating current state-of-the-art RNA/DNA structure prediction methods[73–80] (see Box 1 for details).

Quality check of docking structures

As the accuracy of predicted complex structures is strongly dependent on the quality of individual docking components, the HDOCK server will try to conduct a quality check for user-input structures or modeled structures before docking and provide users the corresponding message or quality report on the docking result page, so that users can evaluate the reliability of docking results. Specifically, for user-input structures, if the input is a protein, the ProQ program is used to evaluate the quality[81,82], where the LGscore and MaxSub are measured as the quality parameters. According to the LGscore or MaxSub, the accuracy of a protein structure is classified into three levels: 'correct', 'good' and 'very good'[81] (Table 4). If the protein structure is below the 'correct' quality according to both the LGscore and MaxSub parameters, a warning message of 'The quality of user-input protein structure is low' will be printed out. For the protein structures generated from sequences by homology modeling, their homology quality is checked by TMalign[83]. The homology quality of models can be classified into

Box 2 | Quality assessment of SAXS data

SAXS is an experimental technique for structural characterization of biological macromolecules in solution. Intensities from SAXS contain low-resolution information on the sample's size and shape. The SAXS data are obtained by subtracting the buffer scattering from the solution scattering. As the scattering intensities are collected over a period of exposure time, the SAXS profile is actually an averaged measurement of all the molecules in different orientations/conformations over time.

Pitfalls in SAXS data

As the SAXS profile is an averaged measurement, the monodispersity of the sample, which means that the SAXS data come from one type of molecule, is critical for correct data interpretation and SAXS-assisted modeling (for a detailed protocol to create monodisperse samples, see ref. [104]). In addition, structural disorder or flexibility of the sample will also cause difficulty for the SAXS data interpretation because it is challenging to model the large conformational changes of macromolecules in solution. Moreover, molecular weight (MW) and radius of gyration (RG) should also be checked from the SAXS data (for a detailed protocol, see refs. [105,106]) to ensure that the data are consistent with those of the studied complex.

Quality checks in HDOCK

Good quality of SAXS data is necessary for accurate SAXS-assisted modeling. Therefore, the HDOCK server conducts a comprehensive quality check of SAXS data based on the Kratky representation, the Guinier approximation and the distance distribution function, P(r), through the SAXS analysis package ATSAS[70].

Monodispersity check. First, the linearity of the Guinier plot is checked using the *autorg* program in the ATSAS package[70] as the linearity of the Guinier region is a vital prerequisite to ensure monodispersity[107]. Nonlinearity of the data suggests aggregation or interparticle interference, which will be returned as a warning to users. Second, the RG and intensity at zero scattering momentum, I(0), are calculated from both the Guinier approximation and the pair distance distribution function, P(r). Disagreements between two sets of RGs and I(0)s indicate small amounts of aggregation[108]. If there is a >20% disagreement between two sets of RGs or I(0)s, a warning will be generated. Third, the monodispersity will also be checked by comparing the MW from the SAXS data and the expected MW from the complex[108]. If there is a disagreement of >40% between the two MWs, a warning will be returned.

Disorder/flexibility check. The structural disorder/flexibility of the sample is checked according to the normalized Kratky plot of SAXS data. Globular proteins will show a bell-shaped peak with a height of 3/*e* (i.e., 1.104) at position $\sqrt{3}$ (i.e., 1.732), and disorder/flexible proteins will have a wider peak position and higher height. Then, according to the product (PK_{xy}) of peak position and height in the normalized Kratky plot, the sample can be considered as stable if $PK_{xy} \leq 1.2\times 3\sqrt{3}/e$, partially flexible if $1.2\times 3\sqrt{3}/e<PK_{xy} \leq 4\times 3\sqrt{3}/e$ or very flexible if $PK_{xy}>4\times 3\sqrt{3}/e$. In addition, the RGs for receptor (RG_{rec}) and ligand (RG_{lig}) are also calculated from the individual structures to be docked. If the RGs from the SAXS data are smaller than 80% of RG_{rec} or RG_{lig}, or >1.2 times of the sum of the RG_{rec} and RG_{lig}, it may indicate structural disorder/flexibility in the sample and thus generate a warning message.

Data filtering. Finally, the useful SAXS data range is determined using the *SHANUM* program in the ATSAS package through the Shannon sampling formalism. The filtered data are then used for calculating the SAXS chi scores (E_{saxs}) of generated complex models.

Quality report. The calculated parameters, assessment results and related plots from the quality checks are shown in the 'Quality of Docking Structures/Input Data' section on the bottom of the HDOCK result page.

Impact of data quality

Two examples of good- and poor-quality SAXS data and their impacts on model ranking are shown in Fig. 3. For the good-quality SAXS data, all the quality control conditions are completely satisfied (Fig. 3a), whereas the poor-quality data show a nonlinearity in the Guinier region, which means aggregation in the sample (Fig. 3b). With the good-quality SAXS data, we obtained two correct models (Models 4 and 5) within the top 10 predictions with a small interface RMSD (I_{rmsd}) of 2.67 Å for the best model (Model 5) when only using the SAXS chi score E_{saxs} (Fig. 3a). However, for the poor-quality SAXS data, no correct model was predicted within the top 10 predictions, and the first correct model is ranked beyond 2,000 (Fig. 3b).

! CAUTION If the input SAXS data pass all the quality control procedures, it is considered as trustworthy for SAXS-assisted modeling. Otherwise, users should take caution when using the SAXS data in the HDOCK server.

three categories: 'Low', 'Medium' and 'High' according to the criteria of sequence identity (Seq_ID)[61,84,85] or TMscore[86–88] (Table 5). If the homology quality is low for both the Seq_ID and TMscore criteria, a warning message of 'The quality of the homology model is low' will be printed out. With these quality check processes, users may get a basic idea about the quality of docking structures and then choose how to proceed with the predicted complex structures.

Job management

The hardware of the HDOCK server is a compute node of two Intel Xeon E5-2690 v4 2.60 GHz central processing units with 28 cores and 256 GB of memory. The operating system is Linux. We used the SLURM Workload Manager (https://slurm.schedmd.com/)[89] to automatically manage jobs on the HDOCK server. Once a docking job is submitted, the job will be put in queue on the server, waiting its turn to run. At the same time, users will be redirected to a status web page showing the job ID and running status of 'QUEUED', 'PENDING' or 'RUNNING'. Currently, the HDOCK server can run 20 jobs simultaneously and does not have a limit for the number of queued jobs in the background. The status page is refreshed every 10 s and will turn to the result page showing the docking results once the job is completed. Users may bookmark the job status page and retrieve the docking results through the bookmarked page at a later time. An email containing the result page link will also be sent to users when the job is completed if an email address is given at the time of job submission.

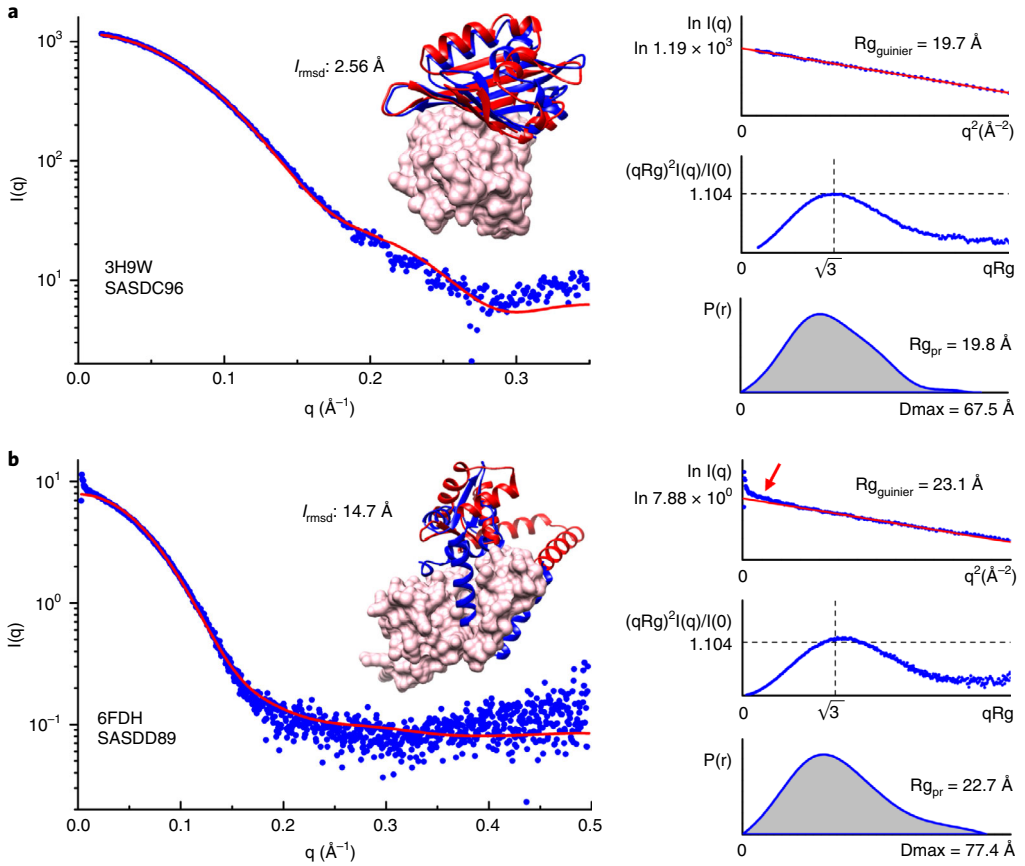

Fig. 3 | Examples of SAXS data with their best predicted models for the top 10 predictions. a and **b**, Examples of good-quality data (**a**) and poor-quality data (**b**). The poor quality in **b** is due to aggregation of the sample as shown by the nonlinearity of the Guinier plot near the red arrow. The receptor protein is represented by the molecular surface and colored in pink. The native and predicted ligand structures are colored in blue and red, respectively. In the left panels, the original SAXS data are represented by blue dots, and the computed profile from the predicted model by the FoXS program is shown by a red line. In the right panels, from top to bottom are the Guinier plot, the normalized Kratky plot and the pair distance distribution function from the experimental SAXS data. The experimental SAXS data are obtained from the SASBDB[99], and the corresponding crystal structures are obtained from the PDB.

Materials

Equipment

- Computer: A personal computer with an Internet connection and a web browser with JavaScript enabled. The HDOCK server is compatible with currently popular web browsers like Google Chrome, Firefox, Microsoft Internet Explorer and Microsoft Edge
- Data: The amino acid sequences or structures of receptor and ligand proteins to be docked
  ▲ **CRITICAL** The amino acid sequences should be in FASTA format and contain standard amino acids only. The structures should be in PDB format.
- (Optional) Experimental data: The residue restraints for protein–protein binding or SAXS data for the complex (see Box 2 for details)

Software

- A web browser like Google Chrome, Firefox, Microsoft Internet Explorer or Microsoft Edge
- An optional molecular visualizing software, such as UCSF Chimera[90] (https://www.cgl.ucsf.edu/chimera/) or PyMOL (https://pymol.org/), if users want to view the predicted complex structures locally

Table 3 | Performances of HDOCKlite for template-free unbound docking with experimental SAXS data

PDB code	Database ID	Type	E_{saxs}[a]		$E_{ITScorePP}$		$E_{ITScorePP} + w \cdot E_{saxs}$[b]	
			Rank[c]	L_{rmsd} (Å)	Rank[c]	L_{rmsd} (Å)	Rank[c]	L_{rmsd} (Å)
4V07	SASDA58	Dimer	131	4.01	11	4.62	11	4.63
1RYX	SASDAA2	Monomer	463	9.75	100	9.75	56	9.75
2R15	SASDAK5	Dimer	2	6.25	43	6.01	1	6.01
3F7L	APSODP	Dimer	101	4.52	87	4.48	82	4.48
3K3K	1PYR1P	Dimer	31	3.29	135	5.65	49	3.29
3V03	SASDA32	Monomer	80	8.03	11	5.98	9	8.24
1ZAH	SASDA68	Tetramer	86	4.86	198	7.00	98	7.00
4ZD3	SASDA28	Heterodimer	584	5.31	1	3.53	1	3.53
4W6Z	SASDA52	Tetramer	22	6.05	3	5.45	3	5.45
4BLC	SASDA92	Tetramer	64	9.31	126	7.92	80	8.54
1FA2	SASDA62	Tetramer	201	9.12	192	8.00	250	8.00
Average			160.45		82.45		58.18	

Shown are the results predicted by the SAXS score (E_{saxs}), ITScorePP ($E_{ITScorePP}$) and weighted score ($E_{ITScorePP} + w \cdot E_{saxs}$) on the 11 cases with experimental SAXS data from the literature, respectively, where the unbound structures were constructed through homology modeling with the templates of sequence identity <70%. [a]E_{saxs} is represented by the chi value. [b]The weight w was set to an optimal value of 15.0 based on an independent training set of cases with experimental SAXS data. [c]The rank of the first correct model with a ligand RMSD of <10 Å.

Table 4 | Quality criteria of input protein structures by ProQ

Correct	Good	Very good
1.5 < LGscore	3.0 ≤ LGscore < 5.0	5.0 ≤ LGscore
0.1 < MaxSub	0.5 ≤ MaxSub < 0.8	0.8 ≤ MaxSub

Table 5 | Quality criteria of homology-modeled protein structures by TMalign

Low	Medium	High
Seq ID < 30%	30% ≤ Seq ID < 50%	50% ≤ Seq ID
TMscore < 0.5	0.5 ≤ TMscore < 0.8	0.8 ≤ TMscore

Procedure

Visiting the HDOCK home page ● Timing <1 min

1 Go to the HDOCK server website at http://hdock.phys.hust.edu.cn/.
? TROUBLESHOOTING

Inputting receptor and ligand ● Timing 5 min

2 Provide the input (either an amino acid sequence in FASTA format or a PDB structure) for the receptor molecule in the 'Input Receptor Molecule' section (see label 1 in Fig. 4) using one of the four options (see Box 3 for details).
▲CRITICAL STEP For amino acid sequence input, only standard amino acids are supported, and the protein should contain only one chain. Users are recommended to input structures for multi-chain proteins.
▲CRITICAL STEP For docking efficiency, it is recommended that the larger one of two molecules is used as the receptor if one molecule is much larger than the other one.

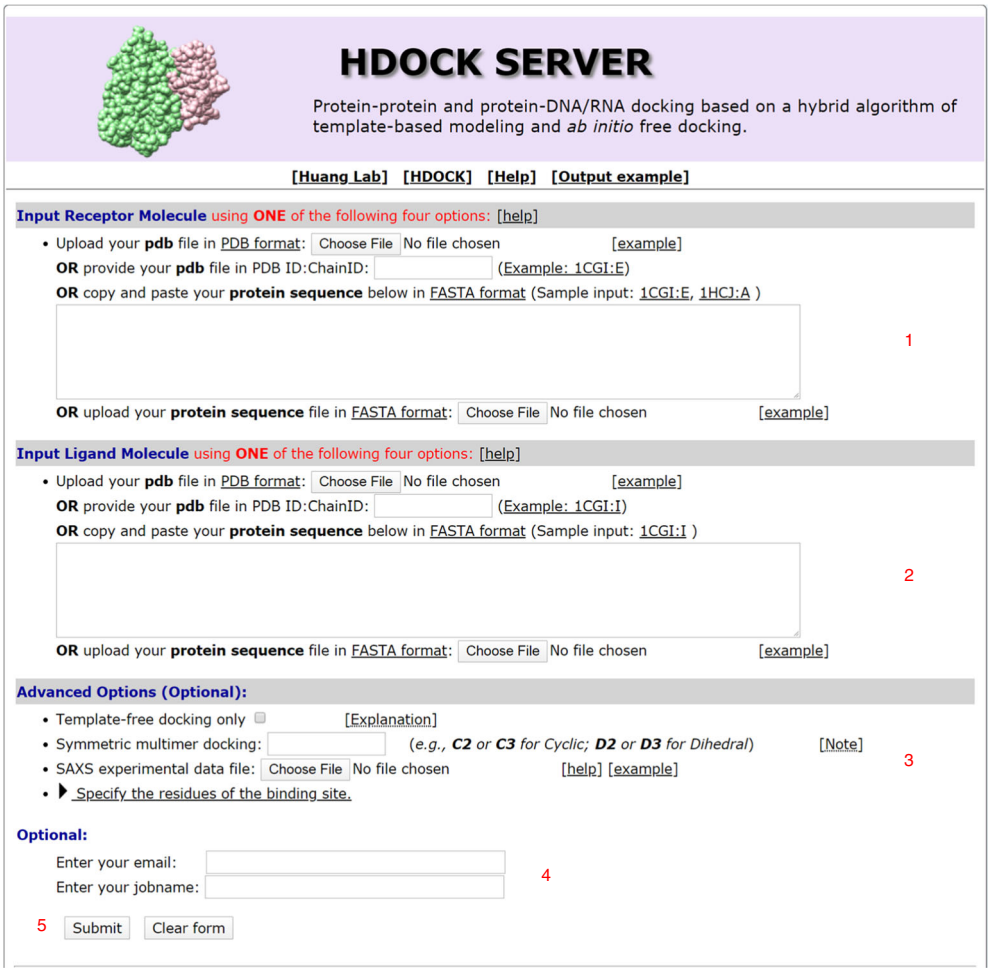

Fig. 4 | Home screen of the HDOCK server. Users can provide job input for receptor molecule (1), ligand molecule (2), advanced options (3) and optional email address or jobname (4), and then submit the job (5).

Box 3 | Four options to provide input for receptor or ligand

Users can provide the input for receptor and ligand in terms of structures or sequences by one of the following four options. If more than one type of input is provided, the first one will be used, and the rest are ignored.

1 Upload your structure in PDB format by clicking on the 'Browse' button to select a pdb file from your local computer. Users can download an example pdb file by clicking on the 'example' link at the end of the line.

2 Provide your structure by PDB_ID:ChainID. An example input is 1CGI:E, which represents the chain E of the PDB entry 1CGI. An example input can be shown through a click on the 'Example' link on the line.
! CAUTION No space should be entered between the PDB ID and the Chain ID.

3 Copy and paste your sequence in the box in FASTA format. An example input can be shown through a click on the 'Sample input' link on the line.
! CAUTION The maximum number of residues allowed for proteins is 5,000.

4 Upload your sequence file in FASTA format. Users can download an example sequence file by clicking on the 'example' link on the line.

3 Provide the input (either an amino acid sequence in FASTA format or a PDB structure) for the ligand molecule in the 'Input Ligand Molecule' section (see label 2 in Fig. 4) using one of the four options (see Box 3 for details).
▲ CRITICAL STEP For amino acid sequence input, only standard amino acids are supported, and the protein should contain only one chain. Users are recommended to input structures for multi-chain molecules.
▲ CRITICAL STEP The HDOCK server is primarily developed for protein–protein docking, but the procedure has also been adapted for protein–RNA/DNA docking by integrating an intrinsic scoring

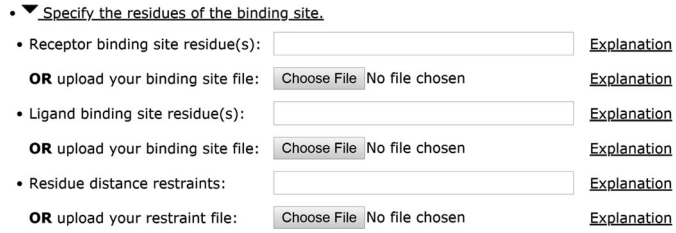

Fig. 5 | Expanded interface for providing the residues of the binding site. Direct pasting or file upload is supported for two types of restraints: the binding site residues on the receptor or ligand and the residue constraints between the receptor and ligand.

function for protein–RNA interactions (see Box 1 for details). For protein–RNA/DNA docking, the RNA/DNA structure can be input as either receptor or ligand.

(Optional) Providing advanced options ● Timing 5 min

4 (Optional) Choose template-free docking (see label 3 of Fig. 4). By default, HDOCK performs the hybrid protocol of template-based modeling and template-free docking. However, in some cases, users may want to perform template-free docking only without using the complex information available in the PDB. This can be done by checking this option.

5 (Optional) Perform symmetric multimer docking (see label 3 of Fig. 4). Users can provide the symmetry in the text box if they want to perform symmetric docking to obtain a symmetric homo-multimer complex structure. The server supports both Cn and Dn symmetric docking

! CAUTION In the symmetric multimer docking mode, only the receptor molecule is needed as the monomer input. The ligand molecule is ignored whether or not its information is provided. In addition, only template-free docking is conducted in this mode.

6 (Optional) Provide the experimental SAXS data (see label 3 of Fig. 4). Upload the SAXS data file by clicking on the 'Browse' button to select the file from your local computer. The SAXS data file should be in three-column format, including the angle (q) in units of 1/Å, the scattering intensity (I), and the experimental error. Users may download a SAXS sample file by clicking on the 'example' link.

7 (Optional) Specify the residues of the binding site as restraints for docking. Click on the 'Specify the residues of the binding site' to show the options (Fig. 5). Two types of restraints about the binding site are supported, which can be directly input in the text box or uploaded as a text file (see Box 4 for details).

▲ CRITICAL STEP Only a few key residue restraints should be provided, and the distance cutoff is recommended to be ≥8 Å, or the number of binding modes that meet the restraints may be very limited due to too many/strict constraints.

(Optional) Providing job information ● Timing 1 min

8 (Optional) Enter your email address in the text box (see label 4 in Fig. 4) to receive an email notification when the job is completed.

▲ CRITICAL STEP It is recommended for users to provide an email address for notification of job completion if they do not want to keep the job status page alive or bookmark the job status page in Step 11.

9 (Optional) Enter your jobname in the text box (see label 4 in Fig. 4). The jobname will be shown on the job status page if provided; otherwise, the jobname will be a jobid generated by the server system.

Submitting the job ● Timing <1 min

10 Submit the job by clicking on the 'Submit' button on the bottom of the input page (see label 5 in Fig. 4).

? TROUBLESHOOTING

Monitoring the job status ● Timing 10–60 min

11 Monitor the job status. Once the job is submitted, the web interface will be redirected to a page showing the jobname and running status (see Box 5 for an example). The job status is updated

Box 4 | Specifying the residues of the binding site

Two types of binding site information can be provided as restraints for docking as follows.

1 Binding site residues on the receptor or ligand.

(A) The binding site residues can be directly pasted in the text box like this:

```
195:A, 203-206:A, 108:B
```

which stand for the residues 195, 203–206 of chain A, and 108 of chain B. Note that the residues in a line must be separated by a comma.

(B) The binding site residues may also be uploaded as a text file that looks like this:

```
195:A
203-206:A
108:B
```

where the residues should be put on different lines in the file.

2 Distance restraints between the receptor and ligand

(A) The constraints can be separated by a comma for direct input on one line like this:

```
195:A 236:B 8, 215-218:A 306:B 6
```

which mean that the distance between residue 195 of chain A on the receptor and residue 236 of chain B on the ligand should be within 8 Å, and the distance between residues 215–218 of chain A on the receptor and residue 306 of chain B on the ligand must be within 6 Å.

(B) Likewise, the above distance restraints can also be uploaded as a text file in one-restraint-a-line format that looks like this:

```
195:A 236:B 8
215-218:A 306:B 6
```

In case of multiple chains and/or the same chain ID for receptor and ligand, the restraints will be defined like this:

```
126:A 196:A 8
78:A 108:B 8
58-60:B 198:A 10
```

▲ **CRITICAL STEP** For each restraint, the first field is for receptor, the second field is for ligand and the third field is for the constrained distance. The residue representation must be in num:chainID or num1-num2:chainID format, where the residue number and chain ID refer to the input structure if the input is a structure, or the modeled structure if the input is a sequence.

! **CAUTION** For the 3D structure modeled by the server, the chain ID is set to 'A' for a single-chain molecule. The numbering of residues is consistent with that in the input sequence.

Box 5 | Job status information for HDOCK runs

Once a job is submitted, the job will be put in queue and wait its turn to run. A running job status page will show information like the following:

Your HDOCK job *5dc66f93368e0* is RUNNING...

This page is automatically refreshed every 10 s. If the page hangs, you may press F5 to manually refresh this page anytime.
Your results will be shown on this page once your job is finished.
You can bookmark this page and access the results later.
If you have provided a valid email address, you will also be notified by email when your job is finished.
Your HDOCK results are stored for 2 weeks.

Your job will be finished in about 176.88 seconds.

every 10 s on the page. The docking results will be automatically shown on the page when the job is done. The docking process typically takes ~10 min.

? TROUBLESHOOTING

▲ **CRITICAL STEP** Users are recommended to bookmark the job status page for retrieving their docking results at a later time if they do not provide an email address in Step 8.

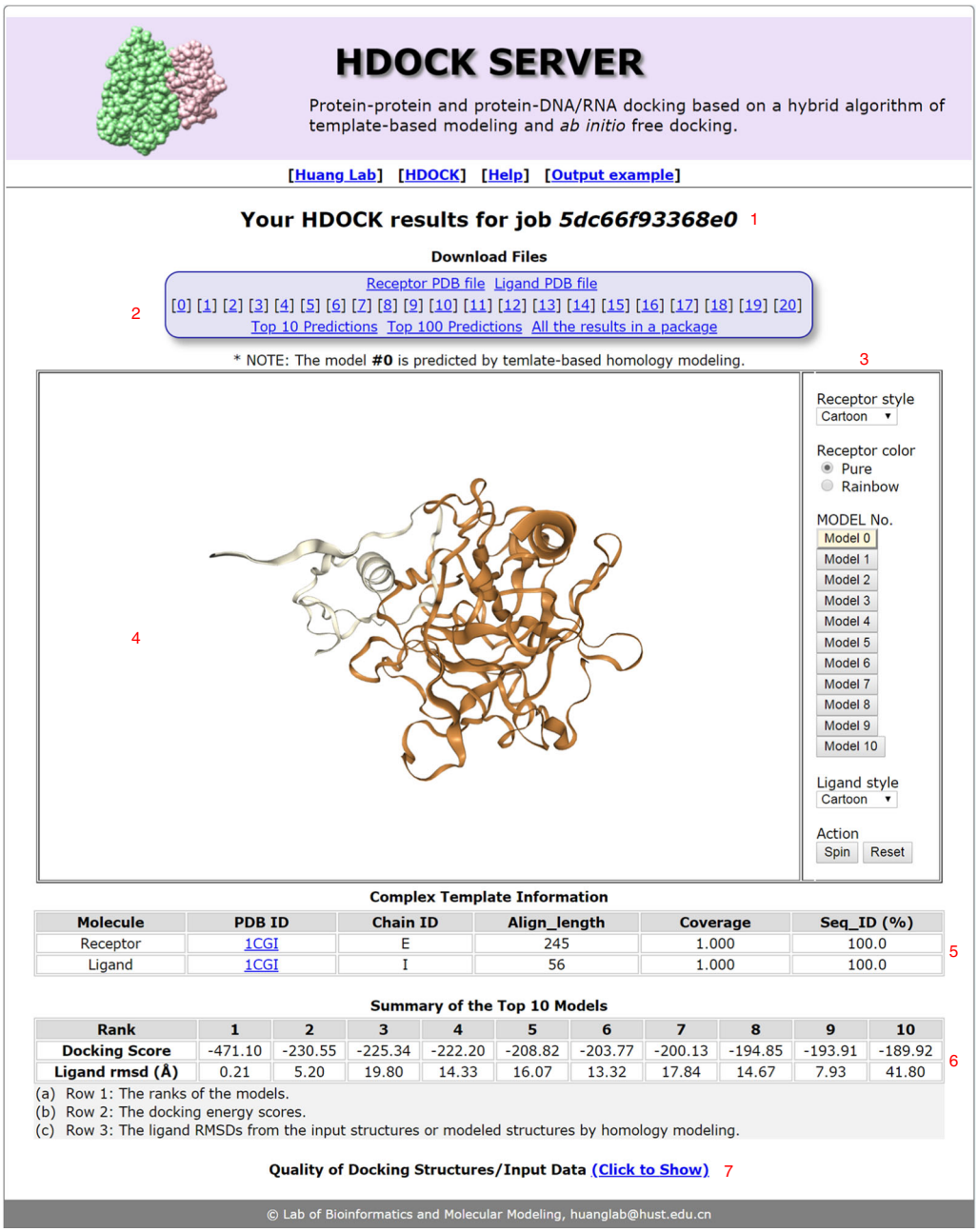

Fig. 6 | HDOCK result page. On the top of the page is a title with the jobname or a unique job ID (1) and the files for download (2). Optional buttons on the right (3) can control the NGL viewer to visualize the binding model on the left (4). The template information is listed below the NGL viewer (5) if available. The docking summary of the top 10 models is shown in a table (6). Users can click to show the quality report of docking structures/input data on the bottom (7).

PAUSE POINT Once the docking job is successfully submitted, the job will be put in queue, waiting for its turn to run. Users may bookmark the job status page and close the job status page. The docking results can be accessed through the bookmarked page or the link contained in the email notification after the job is done.

Analyzing the HDOCK results ● Timing 5–20 min

12 Retrieve the HDOCK results. If the job status page is kept alive, the HDOCK results will be automatically shown on the page when the job is completed; otherwise, visit the HDOCK result page by clicking on the link provided in the email notification or open the link bookmarked in Step 11. The docking results will be shown on the web page (Fig. 6).

? TROUBLESHOOTING

! CAUTION Users are recommended to download their docking results as soon as possible after their job is done, as the job results will be stored on the HDOCK server for only 2 weeks.

13 Visualize the top 10 binding models. Users can interactively view the top 10 binding modes (Models 1–10) in the NGL viewer by clicking on the corresponding buttons on the left of the page (see label 4 in Fig. 6). Users can use the mouse to rotate the displayed structure. The corresponding residue and atom names can be shown when the mouse cursor hovers over a place on the image.

? TROUBLESHOOTING

14 Visualize the model (Model 0) built through template-based modeling. The corresponding complex template information used to construct Model 0 is also shown below the NGL viewer on the page (see label 5 in Fig. 6), where a template with a sequence Coverage of >0.7 and a Seq_ID of >30% for both Receptor and Ligand is thought to be reliable.

! CAUTION This template-based model and corresponding template information will not be available on the result page if no homologous complex template is found in the PDB or in the template-free docking mode.

15 Examine the summary of the top 10 models (see label 6 in Fig. 6). The summary typically contains three rows: the ranks, docking energy scores, and ligand RMSDs. One additional row may also be shown as the SAXS chi-square values of the models if a SAXS data file is provided.

! CAUTION Our scoring function was not calibrated to experimental binding data. Therefore, the docking scores here do not reflect the true binding affinities, but a relative ranking of different binding models for the same receptor and ligand.

! CAUTION The ligand RMSD cannot be used as an indication of docking accuracy because it reflects only how far the predicted model is from the input structure or homology-modeled structure.

16 Download the HDOCK results by clicking on the corresponding links in the download box (see label 2 in Fig. 6). Users can download the Receptor PDB file, Ligand PDB file or 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.

▲ CRITICAL STEP Users should download their HDOCK results as soon as possible after their job is done, as the job results are stored on our server for only 2 weeks.

17 Check the quality of docking structures/input data (see label 7 in Fig. 6). HDOCK evaluates the quality of user-input or modeled receptor and ligand structures as well as the provided SAXS data and generates a report for the quality checks. Users can click to show the quality report on the bottom of the result page (see Supplementary Fig. 1 for an example).

▲ CRITICAL STEP As the quality of predicted complex structures is strongly dependent on the quality of docking components and input data, it is important for users to check the quality report so that they can get some idea about how accurate the structures/data are.

Analyzing the results locally (optional) ● Timing 10–30 min

18 Process the downloaded files. After using compression software to uncompress the downloaded gzip file (e.g., all_results.tar.gz), users should see a directory with the name of jobid, in which are a receptor pdb file, a ligand pdb file, a docking output file and the top 100 models with the file names ranging from model_1.pdb to model_100.pdb.

19 Use molecular visualization software, such as UCSF Chimera or PyMOL, to open and view the predicted complex models. Users may also use a text editor like 'vi' to view and edit the pdb file.

20 (Optional) Generate complex models locally. This step is for advanced users if they want to obtain >100 predicted complex models or filter the docked complex models with their own experimental information. Users can follow the instructions on the help page of the HDOCK server (http://hdock.phys.hust.edu.cn/help.php#post) to generate their own binding models.

▲ CRITICAL STEP Although the experimental information about the binding site and SAXS can be applied during this post-docking step, the binding site information would be better incorporated during the docking and post-docking processes in Step 7.

Troubleshooting

Troubleshooting advice can be found in Table 6.

Table 6 | Troubleshooting table

Step	Problem	Possible reason	Solution
1	The web page cannot be found	The HDOCK server is temporarily down for regular maintenance	Try to visit the HDOCK home page again at a later time
10	Error message: no atoms found in 'xxxx.pdb'	The uploaded pdb file does not contain valid atom records in PDB format. The file may be corrupted or is not in PDB format	Check the 'xxxx.pdb' to make sure that it is a correct pdb file containing valid atom records in PDB format
	Error message: invalid PDB ID/ Chain ID: XXXX:Y	The entry XXXX' is not found in the PDB, or the chain ID 'Y' is not found in the PDB entry of 'XXXX'	Check the PDB ID 'XXXX' and chain ID 'Y' to make sure that the corresponding chain XXXX:Y does exist in the PDB
	Error message: non-standard amino acids in the sequence	This error will occur for sequence input that contains non-standard amino acids	Check and edit the sequence to make sure that it contains only standard amino acids for proteins
	Error message: number of residues exceeds 5,000	The maximum number of residues is 5,000 for the receptor or ligand	Check the sequence to make sure that the receptor or ligand does not contain >5,000 residues
11	The job is in the 'PENDING' status for >24 h	The server is overloaded by other users, or the job hangs due to some internal issues	Users may send us an email with the jobid in it. We will normally check into the problem and get back to the user within 1 d
	The job status page hangs or is not responding	The network connection is slow or down	Do not worry. As long as the job status page is shown, it means that the job has been submitted successfully. The user can manually refresh the job status page or bookmark the page for access to the results later
12	The result page gives no models with a message: something is wrong with your job xxxx!!!	The server experiences some internal problems, or the server is not able to construct a 3D structure from the input sequence due to lack of appropriate templates in the PDB	Users may send us an email with the jobid in it for us to check. The user may try to use a specialized third-party structure prediction method to build the 3D structure and submit the structure for docking
13	The chain is broken in the predicted model	For sequence input, it could be due to the low-quality template. For structure input, the structure may contain long non-structured wrong insertions, which have been removed by the server	Users may send us an email with the jobid in it for us to check. Users may also check the structures to make sure that the predicted models are reasonable

Timing

Step 1, visiting the HDOCK homepage: <1 min
Steps 2 and 3, inputting receptor and ligand molecules: 5 min
Steps 4–7, providing advanced options (optional): 5 min
Steps 8 and 9, providing job information (optional): 1 min
Step 10, submitting the job: <1 min
Step 11, monitoring the job status: 10–60 min
Steps 12–17, analyzing the HDOCK results: 5–20 min
Steps 18–20, analyzing the results locally (optional): 10–30 min

Anticipated results

General docking results

As an example, we have docked the two proteins of target 1CGI from the protein–protein docking benchmark[91]. We provided the protein sequences as input by clicking on the '1CGI:E' and '1CGI:I' of 'Sample input' in the 'Input Receptor Molecule' and 'Input Ligand Molecule' sections, respectively, while keeping the other options unchanged. Once the job is completed, the result page contains:

- The title of the job: your HDOCK results for *jobid/jobname* (see label 1 in Fig. 6).
- The files for download: the Receptor and Ligand PDB files uploaded by users or constructed from the sequences by the server, the template-based complex model (Model 0) and the top 20 docked models

(Models 1–20), a compressed file of the top 10 predictions, a compressed file of the top 100 predictions and a compressed file for all the results (see label 2 in Fig. 6). Please note that Model 0 may not be available if no complex template is found.

- The NGL viewer-embedded display interface through which users can interactively view the top 10 predicted models on the left (see label 4 in Fig. 6) by using the optional buttons on the right (see label 3 in Fig. 6).
- The complex template information that contains the PDB ID, Chain ID, aligned sequence length, coverage of aligned sequence and sequence identity between target and template (see label 5 in Fig. 6). It is important to note that this template information may or may not show here depending on the availability of a complex template.
- A summary of the top 10 models that typically contains three rows: the ranks, docking scores and ligand RMSDs from the input structure or homology model (see label 6 in Fig. 6), as well as an additional row of SAXS chi-square values if a SAXS data file is provided (See Supplementary Fig. 3 for an example).
- Quality report of the docking structures/input data. Users can click to show the quality report of docking structures/input on the bottom of the result page (see label 7 in Fig. 6). The report may include the quality checking results of receptor, ligand and/or SAXS data, depending on the input information (see Supplementary Fig. 3 for an example).

The above anticipated results are generic to other docking categories/modes, except for some differences in specific structures, scores, quality report, etc. More docking examples can also be obtained from other public docking benchmarks[92]. It is important to note that all the anticipated results in the following sections are for docking demonstration and control only. The actual results may vary depending on the specific input and selected docking mode.

Template-free docking

As a template-free docking example, we also docked the two proteins of target 1CGI from the protein–protein docking benchmark[91]. We first provided the sequences of two proteins as the input by clicking on the '1CGI:E' and '1CGI:I' of 'Sample input' in the 'Input Receptor Molecule' and 'Input Ligand Molecule' sections, respectively. Then, we checked the 'Template-free docking only' option in the 'Advanced Options' section (see Step 4 of the Procedure). The docking results for this example are shown in Supplementary Fig. 1. It can be seen from the figure that the HDOCK server identified the proper complex templates 1ACB:E and 1CGI:I for the input sequences and constructed high-quality homology models for the receptor (Seq_ID: 99.6%; TMscore: 0.99614) and ligand (Seq_ID: 100%; TMscore: 0.97447). No template-based complex model (i.e., Model 0) is constructed because it is a template-free docking run. The template-free docking with the modeled structures resulted in three successful models (Models 2, 3 and 8) in the top 10 predictions, all of which are of acceptable accuracy according to the CAPRI criteria[93].

Symmetric multimer docking

As an example, we conducted a symmetric docking run to construct the D2 symmetric complex structure of target 1HCJ, a wild-type GFP, from our symmetric protein docking benchmark[94]. Users may also try other test cases in the benchmark (http://huanglab.phys.hust.edu.cn/SDBenchmark/). To start, we first provided the sequence of the monomer as input, which was done by clicking on the '1HCJ:A' of 'Sample input' in the 'Input Receptor Molecule'. Then, we provided the global symmetry 'D2' into the 'Symmetric multimer docking' textbox of the 'Advanced Options' section (see Step 5 of the Procedure). The docking results are shown in Supplementary Fig. 2. It can be seen from the figure that the HDOCK server successfully identified the proper complex template 1HCJ:A for the monomer sequence and constructed a high-quality homology model for the protein (Seq_ID: 100%; TMscore: 0.99566). The template-free docking resulted in two successful complex models (Models 1 and 7) in the top 10 predictions, where the Model 1 has a high accuracy with an interface RMSD of 0.374 Å.

Incorporating SAXS data

Based on the above template-free docking example of target 1CGI, we selected a SAXS data file from a local computer by clicking on the 'Browse' button in the 'SAXS experimental data file' line of the 'Advanced Options' section (see Step 6 of the Procedure). A SAXS data sample file can be downloaded by clicking on the 'example' link on the line. As shown in the docking results of

Supplementary Fig. 3, an additional row is added to show the 'SAXS chi-square' values in the 'Summary of the Top 10 Models' table. A block about 'SAXS data' is also added in the 'Quality of Docking Structures/Input Data' part, showing the quality report of input SAXS data as well as related plots. Incorporating the SAXS data improved the performance of template-free docking, resulting in more successful predictions and higher quality with five correct models (Models 3, 5, 6, 8 and 10) in the top 10 predictions and a medium-accuracy for Model 8, compared to pure template-free docking (Supplementary Fig. 1).

Applying residue constraints

Based on the template-free docking settings of target 1CGI, we incorporated a residue distance constraint between receptor and ligand by providing the restraint of '`192:A 18:A 8`' in the 'Residue distance restraints' textbox of 'Specify the residues of the binding site' in the 'Advanced Options' (see Step 7 of the Procedure). The docking results are shown in Supplementary Fig. 4. It can be seen from the figure that the models in the 'Summary of the Top 10 Models' table are very different from those in the corresponding table for template-free docking (Supplementary Fig. 1), due to the restriction of the distance restraints. Incorporating the distance restraint considerably improved the performance of template-free docking and resulted in a high-accuracy model for the top prediction (Model 1).

Protein–RNA/DNA docking

As an example for protein–RNA/DNA docking, we docked the protein and RNA molecules of target 1C0A, a complex between an aspartyl tRNA synthetase and an aspartyl tRNA, from the protein–RNA docking benchmark[95]. Users may also try other cases in other public benchmarks[96–98]. Here, we provided the 3D structures of protein and RNA as input by giving '1C0A:A' and '1C0A:B' in the receptor and ligand PDB_ID:ChainID textboxes. The docking results are shown in Supplementary Fig. 5. It can be seen from the figure that the HDOCK server did identify the accurate complex template 1C0A:A/B for the protein and RNA. Accordingly, the server generated a correct template-based complex 'Model 0'. In addition, the template-free docking also resulted in three successful complex models (Models 1, 2 and 10) within the top 10 predictions, where the Model 1 has a high accuracy with an interface RMSD of 0.496 Å from the crystal structure.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw data and example files can be downloaded on our HDOCK server (http://hdock.phys.hust.edu.cn/) or are available from the corresponding author upon request.

Code availability

The HDOCK service is freely available for academic use at http://hdock.phys.hust.edu.cn/.

References

1. Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
2. Vajda, S., Hall, D. R. & Kozakov, D. Sampling and scoring: a marriage made in heaven. *Proteins* **81**, 1874–1884 (2013).
3. Vakser, I. A. Protein-protein docking: from interaction to interactome. *Biophys. J.* **107**, 1785–1793 (2014).
4. Huang, S. Y. 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 (2015).
5. Huang, S. Y. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discov. Today* **19**, 1081–1096 (2014).
6. Wodak, S. J. & Janin, J. Computer analysis of protein-protein interaction. *J. Mol. Biol.* **124**, 323–342 (1978).
7. Katchalski-Katzir, E. et al. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl Acad. Sci. USA* **89**, 2195–2199 (1992).
8. Gabb, H. A., Jackson, R. M. & Sternberg, M. J. E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272**, 106–120 (1997).
9. Vakser, I. A. Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex. *Proteins* (Suppl 1), 226–230 (1997).
10. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737 (2003).

11. De Vries, S. J., van Dijk, M. & Bonvin, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897 (2010).
12. van Zundert, G. C. P. et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725 (2016).
13. Yu, J. C. et al. InterEvDock: a docking server to predict the structure of protein-protein interactions using evolutionary information. *Nucleic Acids Res.* **44**, W542–W549 (2016).
14. Lensink, M. F., Nadzirin, N., Velankar, S. & Wodak, S. J. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins* 1–23 (2020).
15. Janin, J. et al. CAPRI: a Critical Assessment of PRedicted Interactions. *Proteins* **52**, 2–9 (2003).
16. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528 (2019).
17. Hopf, T. A. et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **3**, e3430 (2014).
18. Zeng, H. et al. ComplexContact: a web server for inter-protein contact prediction using deep learning. *Nucleic Acids Res.* **46**, W432–W437 (2018).
19. Huang, S. Y. & Zou, X. Q. An iterative knowledge-based scoring function for protein-protein recognition. *Proteins* **72**, 557–579 (2008).
20. Lensink, M. F. et al. Blind prediction of homo- and hetero-protein complexes: the CASP13-CAPRI experiment. *Proteins* **87**, 1200–1221 (2019).
21. Yan, Y. M., Wen, Z. Y., Wang, X. X. & Huang, S. Y. Addressing recent docking challenges: a hybrid strategy to integrate template-based and free protein-protein docking. *Proteins* **85**, 497–512 (2017).
22. Yan, Y. & Huang, S.-Y. Protein-protein docking with improved shape complementarity. In Intelligent Computing Theories and Application (eds. Huang, D.-S., Bevilacqua, V.,Premaratne, P. & Gupta, P.) 600–605 (Springer International Publishing, Cham, Switzerland, 2018).
23. Huang, S. Y. & Zou, X. Q. MDockPP: a hierarchical approach for protein-protein docking and its application to CAPRI rounds 15-19. *Proteins* **78**, 3096–3103 (2010).
24. Huang, S. Y. et al. Inclusion of the orientational entropic effect and low-resolution experimental information for protein-protein docking in Critical Assessment of PRedicted Interactions (CAPRI). *Proteins* **81**, 2183–2191 (2013).
25. Lensink, M. F. et al. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins* **86**(Suppl 1), 257–273 (2018)..
26. Lensink, M. F. et al. Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment. *Proteins* **84**, 323–348 (2016).
27. Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S.-Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373 (2017).
28. Porter, K. A., Desta, I., Kozakov, D. & Vajda, S. What method to use for protein-protein docking? *Curr. Opin. Struct. Biol.* **55**, 1–7 (2019).
29. Nithin, C., Ghosh, P. & Bujnicki, J. M. Bioinformatics tools and benchmarks for computational docking and 3D structure prediction of RNA-protein complexes. *Genes (Basel)* **9**, E432 (2018).
30. Macalino, S. J. Y. et al. Evolution of in silico strategies for protein-protein interaction drug discovery. *Molecules* **23**, E1963 (2018).
31. Dudenhoeffer, B. R., Schneider, H., Schweimer, K. & Knauer, S. H. SuhB is an integral part of the ribosomal antitermination complex and interacts with NusA. *Nucleic Acids Res.* **47**, 6504–6518 (2019).
32. Fux, A., Korotkov, V. S., Schneider, M., Antes, I. & Sieber, S. A. Chemical cross-linking enables drafting ClpXP proximity maps and taking snapshots of in situ interaction networks. *Cell Chem. Biol.* **26**, 48–59.e7 (2019).
33. Deep, A. et al. Structural, functional and biological insights into the role of *Mycobacterium tuberculosis* VapBC11 toxin-antitoxin system: targeting a tRNase to tackle mycobacterial adaptation. *Nucleic Acids Res.* **46**, 11639–11655 (2018).
34. Kostareva, O. et al. Fab fragment of VHH-based antibody netakimab: crystal structure and modeling interaction with cytokine IL-17A. *Crystals* **9**, 177 (2019).
35. Sharma, S., Ahmed, M. & Akhter, Y. The molecular link between tyrosol binding to tri6 transcriptional regulator and downregulation of trichothecene biosynthesis. *Biochimie* **160**, 14–23 (2019).
36. Rose, A. S. et al. NGL viewer: web-based molecular graphics for large complexes. *Bioinformatics* **34**, 3755–3758 (2018).
37. Huang, S. Y. & Zou, X. Q. A knowledge-based scoring function for protein-RNA interactions derived from a statistical mechanics-based iterative method. *Nucleic Acids Res.* **42**, e55 (2014).
38. Kozakov, D. et al. The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–278 (2017).
39. Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **32**, W96–W99 (2004).
40. Tovchigrechko, A. & Vakser, I. A. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.* **34**, W310–W314 (2006).
41. Lesk, V. I. & Sternberg, M. J. 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 (2008).
42. Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D. & Ritchie, D. W. HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **38**, W445–W449 (2010).

43. Torchala, M., Moal, I. H., Chaleil, R. A. G., Fernandez-Recio, J. & Bates, P. A. SwarmDock: a server for flexible protein-protein docking. *Bioinformatics* **29**, 807–809 (2013).
44. Pierce, B. G. et al. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773 (2014).
45. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).
46. Lyskov, S. & Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36**, W233–W238 (2008).
47. de Vries, S. J., Schindler, C. E. M., de Beauchene, I. C. & Zacharias, M. A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys. J.* **108**, 462–465 (2015).
48. Cheng, T. M. K., Blundell, T. L. & Fernandez-Recio, J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins* **68**, 503–515 (2007).
49. Jimenez-Garcia, B., Pons, C. & Fernandez-Recio, J. pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics* **29**, 1698–1699 (2013).
50. Jimenez-Garcia, B., Pons, C., Svergun, D. I., Bernado, P. & Fernandez-Recio, J. pyDockSAXS: protein-protein complex structure by SAXS and computational docking. *Nucleic Acids Res.* **43**, W356–W361 (2015).
51. Tuszynska, I., Magnus, M., Jonak, K., Dawson, W. & Bujnicki, J. M. NPDock: a web server for protein-nucleic acid docking. *Nucleic Acids Res.* **43**, W425–W430 (2015).
52. Remmert, M., Biegert, A., Hauser, A. & Soding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
53. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
54. Marti-Renom, M. A. et al. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).
55. Yan, Y. & Huang, S.-Y. Pushing the accuracy limit of shape complementarity for protein-protein docking. *BMC Bioinformatics* **20**(Suppl 25), 696 (2019).
56. Venkatraman, V., Yang, Y. F. D., Sael, L. & Kihara, D. Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinformatics* **10,**, 47 (2009).
57. Xu, X. J. et al. Performance of MDockPP in CAPRI rounds 28-29 and 31-35 including the prediction of water-mediated interactions. *Proteins* **85**, 424–434 (2017).
58. Baek, M., Park, T., Heo, L., Park, C. & Seok, C. GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic Acids Res.* **45**, W320–W324 (2017).
59. Weng, G. et al. HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.* **47**, W322–W330 (2019).
60. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K. & Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XIII. *Proteins* **87**, 1011–1020 (2019).
61. Baker, D. & Sali, A. Protein structure prediction and structural genomics. *Science* **294**, 93–96 (2001).
62. Yan, Y. et al. Challenges and opportunities of automated protein-protein docking: HDOCK server versus human predictions in CAPRI Rounds 38-46. *Proteins* 1–15 (2020).
63. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010).
64. Yang, J. Y. et al. The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
65. Yang, J. Y. & Zhang, Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181 (2015).
66. Case, D. A. et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
67. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
68. Yan, Y. M., Tao, H. Y. & Huang, S. Y. HSYMDOCK: a docking web server for predicting the structure of protein homo-oligomers with Cn or Dn symmetry. *Nucleic Acids Res.* **46**, W423–W431 (2018).
69. Yan, Y. & Huang, S.-Y. CHDOCK: a hierarchical docking approach for modeling C_n symmetric homo-oligomeric complexes. *Biophys. Rep.* **5**, 65–72 (2019).
70. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr* **50**, 1212–1225 (2017).
71. Schneidman-Duhovny, D., Hammel, M. & Sali, A. Macromolecular docking restrained by a small angle X-ray scattering profile. *J. Struct. Biol.* **173**, 461–471 (2011).
72. Schindler, C. E. M., de Vries, S. J., Sasse, A. & Zacharias, M. SAXS data alone can generate high-quality models of protein-protein complexes. *Structure* **24**, 1387–1397 (2016).
73. Li, S. X., Olson, W. K. & Lu, X. J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **47**, W26–W34 (2019).
74. Wang, J. & Xiao, Y. Using 3dRNA for RNA 3-D structure prediction and evaluation. *Curr. Protoc. Bioinformatics* **57**, 5.9.1–5.9.12 (2017).
75. Zhao, Y. J. et al. Automated and fast building of three-dimensional RNA structures. *Sci. Rep.* **2**, 734 (2012).
76. Liu, J. H., Wang, J. T. L., Hu, J. & Tian, B. A method for aligning RNA secondary structures and its application to RNA motif detection. *BMC Bioinformatics* **6**, 89 (2005).
77. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
78. Rother, M., Rother, K., Puton, T. & Bujnicki, J. M. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* **39**, 4007–4022 (2011).

79. Wang, J., Zhao, Y. J., Zhu, C. Y. & Xiao, Y. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res.* **43**, e63 (2015).
80. Wang, J. et al. Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res.* **45**, 6299–6309 (2017).
81. Wallner, B. & Elofsson, A. Can correct protein models be identified? *Protein Sci.* **12**, 1073–1086 (2003).
82. Larsson, P., Skwark, M. J., Wallner, B. & Elofsson, A. Assessment of global and local model quality in CASP8 using Pcons and ProQ. *Proteins* **77**(Suppl 9), 167–172 (2009).
83. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
84. Capriotti, E. & Marti-Renom, M. A. Quantifying the relationship between sequence and three-dimensional structure conservation in RNA. *BMC Bioinformatics* **11**, 322 (2010).
85. Gardner, P. P., Wilm, A. & Washietl, S. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res.* **33**, 2433–2439 (2005).
86. Zhang, Y. & Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**, 702–710 (2004).
87. Xu, J. & Zhang, Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**, 889–895 (2010).
88. Gong, S., Zhang, C. & Zhang, Y. RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-scoreRNA. *Bioinformatics* **35**, 4459–4461 (2019).
89. Yoo, A. B., Jette, M. A. & Grondona, M. SLURM: Simple Linux Utility for Resource Management. In *Job Scheduling Strategies for Parallel Processing*. (eds. Feitelson, D., Rudolph, L. & Schwiegelshohn, U.) 44–60 (Springer, Berlin, Heidelberg, 2003).
90. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
91. Hwang, H., Vreven, T., Janin, J. & Weng, Z. P. Protein-protein docking benchmark version 4.0. *Proteins* **78**, 3111–3114 (2010).
92. Kundrotas, P. J. et al. Dockground: a comprehensive data resource for modeling of protein complexes. *Protein Sci.* **27**, 172–181 (2018).
93. Mendez, R., Leplae, R., De Maria, L. & Wodak, S. J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **52**, 51–67 (2003).
94. Yan, Y. & Huang, S. A non-redundant benchmark for symmetric protein docking. *Big Data Mining and Analytics* **2**, 92–99 (2019).
95. Nithin, C., Mukherjee, S. & Bahadur, R. P. A non-redundant protein-RNA docking benchmark version 2.0. *Proteins* **85**, 256–267 (2017).
96. Huang, S. Y. & Zou, X. A nonredundant structure dataset for benchmarking protein-RNA computational docking. *J. Comput. Chem.* **34**, 311–318 (2013).
97. Perez-Cano, L., Jimenez-Garcia, B. & Fernandez-Recio, J. A protein-RNA docking benchmark (II): extended set from experimental and homology modeling data. *Proteins* **80**, 1872–1882 (2012).
98. van Dijk, M. & Bonvin, A. M. J. J. A protein-DNA docking benchmark. *Nucleic Acids Res.* **36**, e88 (2008).
99. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357–D363 (2014).
100. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
101. Miao, Z. et al. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA* **21**, 1066–1084 (2015).
102. Miao, Z. et al. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23**, 655–672 (2017).
103. Miao, Z. C. & Westhof, E. RNA structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* **46**, 483–503 (2017).
104. Jeffries, C. M. et al. Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nat. Protoc.* **11**, 2122–2153 (2016).
105. Grant, T. D. et al. The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 45–56 (2015).
106. Bernado, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8**, 151–167 (2012).
107. Jacques, D. A. & Trewhella, J. Small-angle scattering for structural biology—expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19**, 642–657 (2010).
108. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40**, 191–285 (2007).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant no. 31670724), the National Key Research and Development Program of China (grant nos. 2016YFC1305800 and 2016YFC1305805) and the startup grant of Huazhong University of Science and Technology.

Author contributions

S.-Y.H. conceived and supervised the project. Y.Y., H.T., J.H. and S.-Y.H. designed and performed the experiments. Y.Y., H.T., J.H. and S.-Y.H. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41596-020-0312-x.

Correspondence and requests for materials should be addressed to S.-Y.H.

Reprints and permissions information is available at www.nature.com/reprints.

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

Received: 1 September 2019; Accepted: 3 February 2020;
Published online: 08 April 2020

Related links

Key references using this protocol

Yan, Y. et al. *Nucleic Acids Res.* **45**, W365–W373 (2017): https://doi.org/10.1093/nar/gkx407
Yan, Y. et al. *Proteins* **85**, 497–512 (2017): https://doi.org/10.1002/prot.25234
Yan, Y. et al. *Nucleic Acids Res.* **46**, W423–W431 (2018): https://doi.org/10.1093/nar/gky398
Yan, Y. and Huang, S.-Y. *BMC Bioinformatics* **20**(Suppl 25), 696 (2019): https://doi.org/10.1186/s12859-019-3270-y

nature research

Corresponding author(s): Sheng-You Huang

Last updated by author(s): Jan 29, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NGL Viewer (ver. 0.9.3) (http://nglviewer.org/) for model visualization.
Data analysis	Apache (ver.2.2.15) (https://www.apache.org/) and PHP (ver. 5.6.26) (https://www.php.net/) for web server service. SLURM (ver. 16.05.6) (https://slurm.schedmd.com/) for job scheduling. HHSuite (ver. 3.0.0)(https://github.com/soedinglab/hh-suite), BLAST (ver. 2.4.0+) (https://blast.ncbi.nlm.nih.gov/) and FASTA (ver. 36.3.8e) (https://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml) for sequence similarity search. ViennaRNA/RNAfold (ver. 2.4.10) (https://www.tbi.univie.ac.at/RNA/) for RNA secondary structure prediction. RADAR/RSmatch (ver. 2.0) (https://web.njit.edu/~wangj/RSmatch/) for RNA homology search. MODELLER (ver. 9.16)(https://salilab.org/modeller/) for homology modeling of proteins. ModeRNA (ver. 1.7.1) (http://genesilico.pl/moderna/) for RNA comparative modeling. 3DNA (version 1.6.9-2017aug09) (http://home.x3dna.org/) for RNA secondary structure analysis and building DNA/RNA duplex or strand. 3dRNA (version 1.7) (https://github.com/hust220/nsp/releases) for template-free RNA 3D structure prediction. AMBER (ver. 16) (http://ambermd.org/) for molecular dynamics simulations. ProQ (version 1.2) (https://proq.bioinfo.se/ProQ/ProQ.html) for protein model quality check. TMalign (https://zhanglab.ccmb.med.umich.edu/TM-align/) for protein homology quality check. RNAalign (https://zhanglab.ccmb.med.umich.edu/RNA-align/) for RNA homology quality check. ATSAS (version 3.0.0) (https://www.embl-hamburg.de/biosaxs/software.html) for small-angle scattering data analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sample data for docking tests can be obtained or downloaded from the HDOCK home page. The data that support the findings of this study are available from the corresponding author upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sequence and PDB structure sample inputs used for testing can be downloaded or shown through the links on the HDOCK home page. An example output can be shown through the "Output Example" link.
Data exclusions	No data were excluded from the analysis.
Replication	This work is for a molecular docking server and no biological experimental data involved. The server may not give the same results due to the updated PDB database. However, the results are reproducible on the same PDB database.
Randomization	This work is to give a protocol of our HDOCK server and no biological experimental data involved; thus the randomization of data is not applicable.
Blinding	This work is for a molecular docking server and no medical trials involved; thus blinding is not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging