

# HawkDock: a web server to predict and analyze the protein–protein complex based on computational docking and MM/GBSA

Gaoqi Weng<sup>1</sup>, Ercheng Wang<sup>1</sup>, Zhe Wang<sup>1</sup>, Hui Liu<sup>1</sup>, Feng Zhu<sup>1</sup>, Dan Li<sup>1</sup> and Tingjun Hou<sup>1,2,\*</sup>

<sup>1</sup>Hangzhou Institute of Innovative Medicine, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China and <sup>2</sup>State Key Lab of CAD&CG, Zhejiang University, Hangzhou, Zhejiang 310058, China

Received February 22, 2019; Revised April 23, 2019; Editorial Decision May 01, 2019; Accepted May 01, 2019

## ABSTRACT

Protein–protein interactions (PPIs) play an important role in the different functions of cells, but accurate prediction of the three-dimensional structures for PPIs is still a notoriously difficult task. In this study, HawkDock, a free and open accessed web server, was developed to predict and analyze the structures of PPIs. In the HawkDock server, the ATTRACT docking algorithm, the HawkRank scoring function developed in our group and the MM/GBSA free energy decomposition analysis were seamlessly integrated into a multi-functional platform. The structures of PPIs were predicted by combining the ATTRACT docking and the HawkRank re-scoring, and the key residues for PPIs were highlighted by the MM/GBSA free energy decomposition. The molecular visualization was supported by 3Dmol.js. For the structural modeling of PPIs, HawkDock could achieve a better performance than ZDOCK 3.0.2 in the benchmark testing. For the prediction of key residues, the important residues that play an essential role in PPIs could be identified in the top 10 residues for ~81.4% predicted models and ~95.4% crystal structures in the benchmark dataset. To sum up, the HawkDock server is a powerful tool to predict the binding structures and identify the key residues of PPIs. The HawkDock server is accessible free of charge at <http://cadd.zju.edu.cn/hawkdock/>.

## INTRODUCTION

Protein–protein interactions (PPIs) are involved virtually in all cellular processes, such as signal transduction, protein expression regulation and DNA replication. Therefore, determination of their complex structures is critical to understand the underlying molecular mechanisms of crucial bio-

logical processes (1) and even design compounds that interfere with PPIs with pharmaceutical significance (2). However, only a tiny number of the 3D structures of protein–protein complexes have been determined experimentally and deposited into the released databases, such as Protein Data Bank (PDB). In this context, computational methods, especially the protein–protein docking, have been increasingly applied to the structural prediction of macromolecular assemblies, which is expected to be a valuable complement to the experimental methods.

Protein–protein docking algorithms could be roughly classified into two categories: template-based modeling and template-free docking (3). Template-based modeling is based on the observation that protein–protein complexes usually interact in the same way if their interacting pairs share >30% sequence identity (4). Hence, the near-native structure of a protein–protein complex can be predicted by this method if an appropriate template is available. However, the number of the available templates is still very limited at present, and thus the template-free docking is generally more popular. Most template-free docking algorithms consist of two stages: sampling stage and scoring stage. In the sampling stage, a large number of decoys are generated, and the decoys sampled from this stage are re-scored and ranked by various scoring functions in the subsequent scoring stage.

A number of template-free docking servers have been developed and released to the public, such as ATTRACT (5), ClusPro (6), HADDOCK (7), ZDOCK server (8), SwarmDock (9), pyDockSAXS (10), pyDockWeb (11), InterEvDock2 (12), GRAMM-X (13), RosettaDock server (14), PatchDock (15), Hex server (16), 3D-Garden (17), FRODOCK 2.0 (18), etc. Moreover, a hybrid strategy by combining template-based modeling and template-free docking was proposed in HDOCK (19). The scoring strategies used in most servers rely on the scoring functions with a few energy terms, such as van der Waals potential, electrostatic potential, hydrogen-binding potential, desolvation,

\*To whom correspondence should be addressed. Tel: +86 517 8820 8412; Email: tingjunhou@zju.edu.cn

etc. However, the desolvation energy, which plays a leading role in identifying correct binding poses, cannot be calculated accurately till now. In addition, most docking servers do not provide functions to identify and visualize the key residues for PPIs. It is quite challenging for non-expert users to identify the key residues in a protein–protein binding interface directly using some molecular visualization programs, as the binding interface usually involves a large number of residues (20).

In the past decade, two end-point free energy calculation methodologies, Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) (21), which are more theoretically rigorous than scoring functions, have been widely used to predict binding free energies and identify correct binding conformations for protein–protein systems (22–30). Furthermore, MM/PBSA and MM/GBSA with per-residue energy decomposition of the binding free energy of a protein–protein complex have been successfully employed to highlight the key residues in the binding interface (31–33). According to the assessment results reported by our previous study (25), MM/GBSA achieved better results not only in predicting the binding affinities but also in recognizing the near-native binding modes for PPIs than the tested scoring functions. However, MM/GBSA is much more computationally expensive than those commonly applied scoring functions due to the high computational cost of the polar desolvation term based on a Generalized Born (GB) model. In order to achieve a balance between computational efficiency and accuracy, we developed HawkRank, a physical scoring function with similar energy terms in MM/GBSA (34) by introducing a fast and effective method to calculate the desolvation potentials based on solvent accessible surface areas (35).

In this study, the HawkDock server was developed to predict and analyze the structures of PPIs by integrating the ATTRACT docking algorithm, the HawkRank scoring function and the MM/GBSA free energy decomposition analysis. In HawkDock, to predict the structure of a protein–protein complex, a large number of binding poses are first generated using rigid-body docking protocol of ATTRACT. Then, the near-native binding poses are recognized by the HawkRank scoring function combined with ATTRACT score. In addition, MM/GBSA was integrated into the HawkDock server to help users to analyze the key residues in a protein–protein binding interface and re-rank the top 10 models. All prediction and analysis functions in the HawkDock server are automated and the results are presented interactively through a user-friendly interface.

## MATERIALS AND METHODS

### Workflow of the HawkDock server

The HawkDock server is an integrated web server that combines the HawkRank program developed in our group (34) for re-ranking docking poses and several third-party programs, including ATTRACT (36) for protein–protein docking, MM/GBSA (21) for the identification of key residues and 3Dmol.js (37) for molecular visualization. The HawkDock pipeline is shown in Figure 1.

For the prediction of a protein–protein complex based on two unbound structures, the server first performs the grid-accelerated rigid-body docking based on a randomized global search algorithm implemented in ATTRACT (5). The maximum steps of minimization and the distance squared cutoff are set to 1000 and 50.0 Å<sup>2</sup>, respectively (5). In addition, the docking process can also incorporate the spatial constraints provided by the users. Then, the best 10,000 decoys generated by ATTRACT are re-scored by HawkRank. Subsequently, the best 1000 decoys given by ATTRACT and the best 1000 decoys given by HawkRank are clustered by the Fraction of Common Contacts (FCC) clustering method, respectively, whose fraction of common contacts threshold is set to 0.5 (38). The best-scored model in each cluster is extracted and all the extracted models are re-ranked by the HawkRank scoring. Besides, the representative models identified by both ATTRACT and HawkRank were labeled as the top-ranked structures. Moreover, considering the prediction accuracy and computational efficiency of MM/GBSA, it can be used to re-rank the top 10 models predicted by ATTRACT and HawkRank. In addition, the 10 most frequently occurring residues found in the top 10 models for the receptor and ligand can also be provided. Finally, the top 10 and 100 docking models are available for users to view interactively and download through a web page, respectively. Furthermore, the docking models can be analyzed by MM/GBSA to identify the key residues in the protein–protein binding interface.

For the docking models predicted by HawkDock or other docking programs and the complex structures determined by experimental techniques, MM/GBSA can be directly implemented for the analysis of key residues. The per-residue free energy contributions are summarized in the result page.

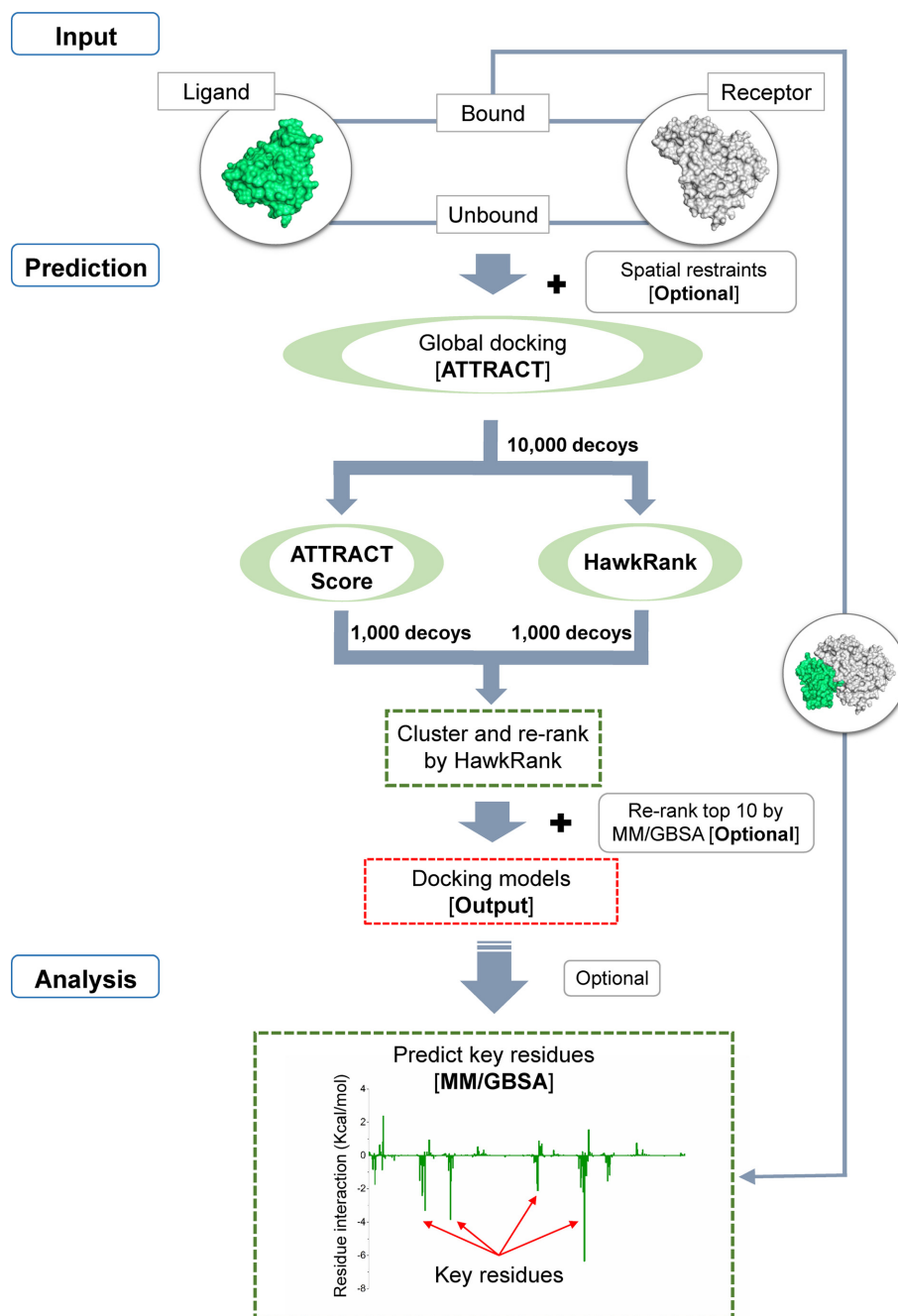
### The MM/GBSA free energy decomposition

In our previous study, the MM/GBSA calculation based on the ff02 force field (39) and the GB<sup>OBC1</sup> model (40) yielded the best predictions for protein–protein binding free energies. Therefore, these parameters were used in the analysis of the key residues in protein–protein binding interfaces. First, all missing hydrogens and heavy atoms of the protein–protein complex were added by the *tleap* module in Amber16 (41), and then the force field parameters of the ff02 force field were assigned to the proteins. Subsequently, the complex was optimized *in vacuo* by 2000 cycles of steepest descent and 3000 cycles of conjugate gradient minimizations. At last, the polar desolvation energy was calculated by the modified GB (GB<sup>OBC1</sup>) models developed by Onufriev et al. The exterior and interior/solute dielectric constants were set to 80 and 1, respectively.

### Input

The functions of the HawkDock server include the structural modeling of a protein–protein complex by using HawkDock and the prediction of the key residues in protein–protein binding interfaces using MM/GBSA.

Figure 2A shows the usage interfaces of HawkDock. The users are requested to upload PDB files or provide PDB



**Figure 1.** Workflow of the HawkDock server that is divided into three major steps: (i) input of unbound or bound protein structures; (ii) structural prediction of protein–protein complex by using the global docking algorithm implemented in ATTRACT and the HawkRank scoring function; (iii) identification of the key residues by MM/GBSA.

ID:chain ID (e.g. 1C3D:A). The non-standard amino acid residues, heteroatoms (HETATM records), original hydrogen atoms and residues with incomplete backbone atoms are removed and the missing side chains are added by PDB2PQR (42). In addition, the options of distance restraints and ‘re-rank top 10 models by MM/GBSA’ are also available for the users. A number of binding poses will be generated in the local domain if distance restraints are provided. When the MM/GBSA re-scoring function is chosen, our server will not only re-rank the top 10 models predicted

by ATTRACT and HawkRank, but also provide the top 10 most frequently occurring residues for the receptor and ligand found in the top 10 models. As the server runs, the job status information is displayed on the log page.

Figure 2B shows the usage interfaces of MM/GBSA. PDB files or PDB ID (e.g. 1SYX) should be input by the users. Similar to HawkDock, the non-standard amino acid residues, heteroatoms (HETATM records) and original hydrogen atoms are removed. The uploaded protein–protein complex structures need to be docked in advance or be de-

**A**

Job Name (optional)

Email (optional)

Input Receptor

Input Ligand

Advanced options (optional)

Restraints

Re-rank top10 models by MM/GBSA

**6** Results for job: test

Status of your job: **Waiting....**

Estimated time: **Estimate the run time when the job starts running.**

Submitted time: 2018-09-08 14:09 (UTC + 8 hours).

Start time: Unknown.

**Step 0:** Queue in cluster system.  
**Step 1:** Prepare.  
**Step 2:** Calculate receptor's grid.  
**Step 3:** Docking.  
**Step 4:** Preprocess the files of PDB.  
**Step 5:** Rescoring.  
**Step 6:** Finished.

**B**

Job Name (optional)

Email (optional)

Input Complex

Chain ID of Receptor

Chain ID of Ligand

**5** Results for job: test

Status of your job: **Waiting....**

Estimated time: **Estimate the run time when the job starts running.**

Submitted time: 2018-09-08 14:30 (UTC + 8 hours).

Start time: Unknown.

**Step 0:** Queue in cluster system.  
**Step 1:** Tleap.  
**Step 2:** Minimization.  
**Step 3:** MM/GBSA.  
**Step 4:** Finished.

**Figure 2.** Data input in the HawkDock server. (A) A HawkDock job needs: (1) optional job name and email address, (2) PDB files or PDB IDs, (3) optional distance restraints, (4) optional MM/GBSA re-scoring, (5) the 'submit' button and (6) the log of the HawkDock job. (B) A MM/GBSA job needs: (1) optional job name and email address, (2) PDB file or PDB ID, (3) the chain IDs of receptor and ligand, (4) the 'submit' button and (5) the log of the MM/GBSA job.

terminated by experimental techniques. Moreover, the users are supposed to specify the chain IDs of receptor (e.g. A or A;B) and ligand (e.g. C or C;D). And the job status information of MM/GBSA is also displayed on the log page.

## Benchmarks

The benchmark dataset I (Dataset I) to assess the structural modeling capability of HawkDock was obtained from ZDOCK benchmark 4.0 (43). ZDOCK benchmark 4.0 consists of 176 nonredundant protein-protein complexes, for which the NMR or X-ray unbound structures of the constituent proteins are also available, including the 124 cases in Benchmark 3.0 and 52 newly-added complexes. Since the 124 complexes in Benchmark 3.0 were used as the training set to develop HawkRank, the other 52 complexes were used as the benchmark dataset to evaluate the docking performance of HawkDock (Supplementary Table S1).

After carefully retrieving literatures, the benchmark dataset II (Dataset II) with 116 key residues in the complex interfaces identified by experiments for 43 proteins (receptors or ligands) in 32 complexes (a complex is separated into a receptor and a ligand) in Dataset I (Supplementary Table S2) was constructed to validate the prediction capability of MM/GBSA to determine the key residues for PPIs. The crystal structures were directly submitted to the MM/GBSA calculations and the key residues were then de-

terminated. As for a docking model, the binding pose with the smallest interface RMSD (I.RMSD) in the 10 000 decoys generated by HawkDock was analyzed by MM/GBSA.

## Evaluation criteria

Two major evaluation parameters in the CAPRI campaign, ligand RMSD (L.RMSD) and I.RMSD, were used as the criteria to evaluate the docking performance for Dataset I. When I.RMSD is <4 Å or L.RMSD is <10 Å, the prediction is considered to be successful (44). The success rate was defined as the percentage of the total cases with at least one correct model in the top *N* predictions.

For Dataset II, the success rate, defined as the percentage of the cases with at least one key residue within certain predictions, was employed to evaluate the key residue prediction capability of MM/GBSA.

## RESULTS

### HawkDock server

The HawkDock server based on the Python web framework of Tornado (an asynchronous networking library) is deployed on a Linux server of an Intel(R) Xeon(R) E5-2696 v3 2.30GHz CPUs with 36 cores and 64 GB of memory. After submitting jobs, the HawkDock server will create a HawkDock or MM/GBSA task and then put it into



the queue immediately. A HawkDock job usually needs 3 min (for proteins around 150 residues) to 25 min (for proteins ~1000 residues). When the option of 're-rank top 10 by MM/GBSA' is chosen, the running time of HawkDock will be extended to 25 min (for proteins ~200 residues) to 175 min (for proteins ~1000 residues). For MM/GBSA, the running time for the prediction of key residues is about 2 min (for proteins ~200 residues) to 15 min (for proteins ~1000 residues). Afterwards, the web interface will be redirected to the job status and result page. The URL of the page is unique for each job, and hence bookmarking the URL is highly recommended to the users. Moreover, an email will be sent to the user once the job is finished, if a valid email address has been provided by the user.

### Output

Once the calculation is finished, the job status page will be auto-refreshed to the result page. As shown in Figure 3A, the result of HawkDock contains seven components. The downloadable files include three types of files: (i) the receptor and ligand PDB files used as the docking input; (ii) a text file with the docking scores for the top 100 models; (iii) the compressed tar files for the top 10 and top 100 models, respectively. If necessary, the users can also contact us for downloading more models.

In addition, the top 10 models can be interactively displayed with 3Dmol.js, a WebGL-based molecular viewer. The brief instructions of 3Dmol.js are shown on the bottom of the result page. It should be noted that all residues will be renumbered from 1 and their numbers might be different from those in the original file. Furthermore, the option is offered in the HawkDock result page to facilitate the users to directly perform the MM/GBSA analysis for one or several docking models predicted by ATTRACT and HawkRank. In the sixth part, the models can be selected and submitted directly to perform the MM/GBSA calculations. Moreover, the result page also shows the job name and the docking scores of the top 10 models.

If the users choose the option of 'Re-rank top10 models by MM/GBSA', as shown in Supplementary Figure S1, the top 10 most frequently occurring residues identified in the top 10 models for the receptor and ligand are displayed in the third and fourth parts, respectively. In addition, the analysis results for each model are listed in the eighth part. The top 10 models are re-ranked based on the binding free energies predicted by MM/GBSA.

The result page of MM/GBSA is presented in Figure 3B, including the job name, the value of the binding free energy of each complex, the downloadable file, the table sorted by the per-residue free energy contributions from the largest to the smallest, molecular viewer and its brief instruction. The upper and lower tables show the results of the receptor and ligand, respectively. The detailed values of all energy terms are summarized into a csv file. In addition, the users can highlight important residues in 3Dmol.js by selecting these residues in the table.

### Docking performance on CAPRI targets

We participated in recent CASP13-CAPRI as a server group. According to the rules of CAPRI, the server groups

must submit their results within 2 days, but the human predictor groups have ~20 days and can use any additional information (6). Furthermore, only the sequences were provided for the docked systems so that the wrong monomer structure of proteins may be predicted from the beginning. Since it is the first time for us to participate in this challenge, we met some problems in the registration as a server group so that we missed some targets. In addition, HawkDock was designed for dimer docking and thus we ended up with participating in seven targets (T148, T152, T153, T154, T155, T156 and T157). As for the seven proteins in CASP13, taking the top 5 hits as the criterion, HawkDock could provide acceptable models for two targets (T152 and T153), compared with acceptable models for one target (T152) provided by LZERD (45) and acceptable models for two targets (T152 and T153) provided by other servers, such as ClusPro (6), SwarmDock (9), HADDOCK (7), HDOCK (19) and GalaxyPPDock (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=PPDOCK>). Even for the human predictor groups, the best-performed groups could only give successful predictions for three targets. When the top 10 models were considered in the server groups, HDOCK and GalaxyPPDock performed better and achieved success for three targets. These results suggest the robustness of HawkDock in protein-protein docking.

### Docking performance on benchmark

The results of HawkDock were also compared with those of ATTRACT and ZDOCK 3.0.2, a widely used rigid-body docking program (46), based on the analysis of Dataset I. 10 000 and 54 000 decoys were generated and ranked by ATTRACT and ZDOCK 3.0.2, respectively. The detailed ranking results are summarized in Supplementary Table S1 and the success rates from top 10 to top 1000 and top 1 to top 10 are shown in the main figure and the inserted zoomed view of Figure 4A, respectively. In terms of the top 10 level, HawkDock achieves a better performance with a success rate of 25.00%, while the success rates for ATTRACT and ZDOCK 3.0.2 are 13.46% and 21.15%, respectively. With more predictions considered, HawkDock performs significantly better than ATTRACT and ZDOCK 3.0.2 and yields the success rates of 42.31%, 50.00%, 69.23%, 80.77% and 88.46% at the top 50, 100, 200, 400 and 1000 levels, compared with 34.62%, 44.23%, 55.77%, 65.39% and 78.85% for ATTRACT and 30.77%, 42.31%, 50.00%, 65.39% and 71.15% for ZDOCK 3.0.2. At the top 1 level, HawkDock performs worse than ZDOCK 3.0.2, but its success rate can be improved from 5.76% to 11.54% by the MM/GBSA rescoring, which is similar to that of ZDOCK 3.0.2 and much higher than those of the other tested methods. Furthermore, the frequently occurring residues in the top 10 models can also be analyzed by this MM-GBSA refinement protocol. In summary, these results highlight the robustness and reliability of HawkDock in the structural prediction of PPIs.

### Key residue prediction

The prediction capability of MM/GBSA in identifying the key residues in complex interfaces was then assessed based

**A**

**Results**

Results for job: example **1**

Downloadable Files:

Receptor PDB File

Ligand PDB File **2**

HawkDock Output

Top 10 Predictions

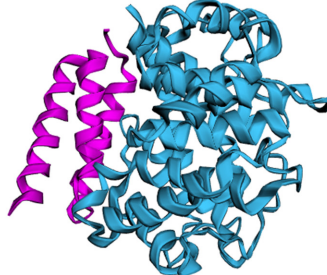
Top 100 Predictions

Summary of the top 10 models: **3**

Rank	Score
1	-5565.28
2	-5000.85
3	-4966.04
4	-4262.42
5	-4261.76
6	-4247.58
7	-4196.96
8	-3741.39
9	-3677.32
10	-3174.67

**Molecule visualization**

**4**



**5**

Model 1 Model 2 Model 3 Model 4 Model 5  
Model 6 Model 7 Model 8 Model 9 Model 10

Perform MM/GBSA analysis on the top 10 models (optional)

Model 1 Model 2 Model 3 Model 4 Model 5 **6**  
Model 6 Model 7 Model 8 Model 9 Model 10

**Submit**

model1 [http://cadd.zju.edu.cn/hawkdock/result/example\\_model1-1554908418547](http://cadd.zju.edu.cn/hawkdock/result/example_model1-1554908418547)

**Action Options: 7**

- Rotate: Press and hold the left mouse button.
- Zoom: Use mouse's scroll button.
- Translate: Press and hold mouse's scroll button.

**Browsers Requirements**

Please make sure that **WebGL** is enabled in your browser. To test for WebGL support, see <http://webglreport.com/?v=1>.

- Support: Chrome, Firefox (version 4 or later), Safari, Microsoft Edge, Internet Explorer (version 11).

**B**

**Results**

Results for job: example\_gbsa **1**

Binding free energy of complex: **-63.79** (kcal/mol) **2**

Downloadable Files:

example\_gbsa.csv

Receptor PDB File **3**

Ligand PDB File

**Rank 4** Residue (Rec) Binding Free Ener

1	<input checked="" type="checkbox"/>	A-ARG-127	-8.47
2	<input type="checkbox"/>	A-LYS-125	-5
3	<input type="checkbox"/>	A-SER-132	-4.71
4	<input type="checkbox"/>	A-PRO-133	-2.95
5	<input type="checkbox"/>	A-ARG-121	-2.82

Showing 1 to 5 of 135 rows 5 rows per page

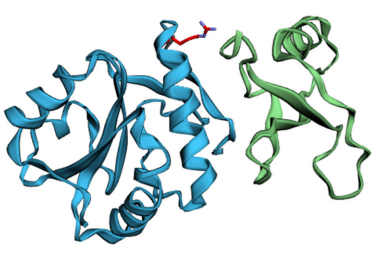
**Rank 5** Residue (Lig) Binding Free Ener

1	<input type="checkbox"/>	B-ARG-79	-7.94
2	<input type="checkbox"/>	B-TYR-75	-7.4
3	<input type="checkbox"/>	B-LEU-84	-5.65
4	<input type="checkbox"/>	B-TYR-85	-3.57
5	<input type="checkbox"/>	B-GLN-73	-3.04

Showing 1 to 5 of 62 rows 5 rows per page

**Molecule visualization**

**6**



**Action Options: 7**

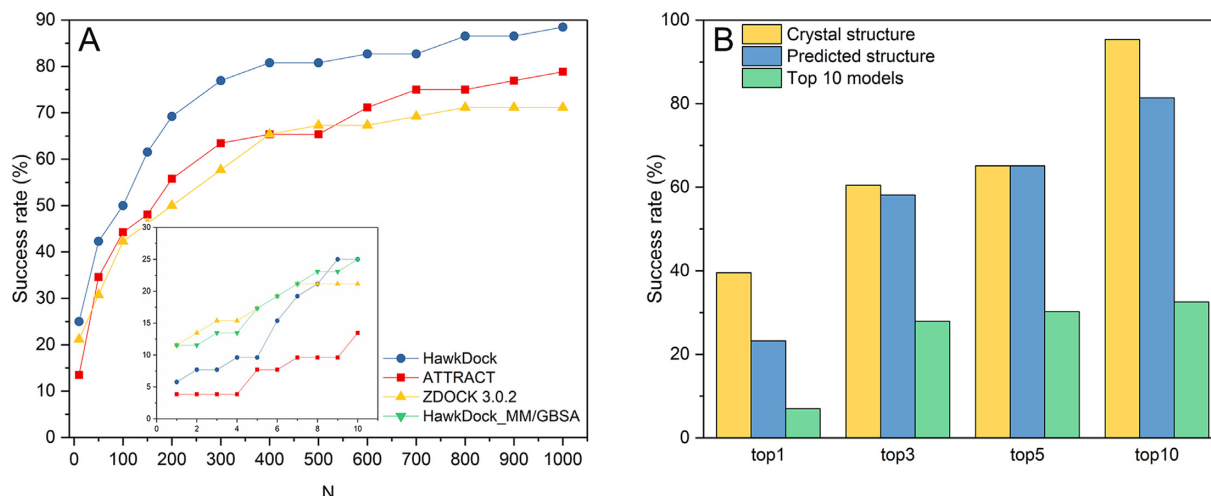
- Rotate: Press and hold the left mouse button.
- Zoom: Use mouse's scroll button.
- Translate: Press and hold mouse's scroll button.

**Browsers Requirements**

Please make sure that **WebGL** is enabled in your browser. To test for WebGL support, see <http://webglreport.com/?v=1>.

- Support: Chrome, Firefox (version 4 or later), Safari, Microsoft Edge, Internet Explorer (version 11).

**Figure 3.** The result pages of HawkDock and MM/GBSA. **(A)** At the top left of the page is (1) the job name, and under it are (2) the files for downloading. (3) A summary of the docking scores for the top 10 models is presented on the left bottom. (4) The top 10 models can be viewed in 3Dmol.js with (5) the optional buttons to control which model to display. (6) The table summarizes the submission and results of MM/GBSA and (7) the brief instructions are shown on the bottom. **(B)** At the top of page is (1) the job name, and under it is (2) the value of the binding free energy and (3) the files for downloading. The per-residue free energy contributions ordered from largest to smallest for (4) receptor and (5) ligand are displayed in the table, (6) through which the users can choose which residue to display in 3Dmol.js. (7) The brief instructions are also displayed on the bottom.

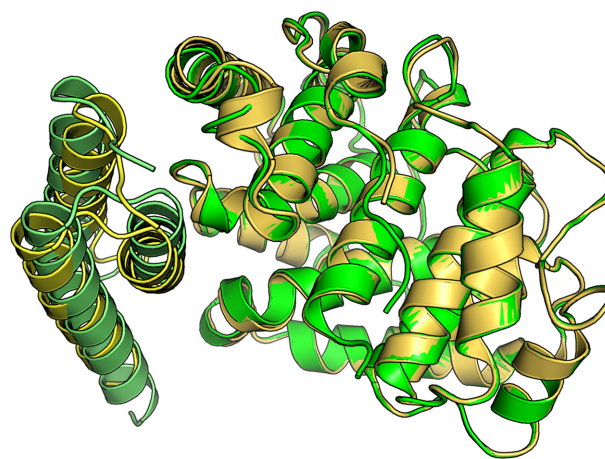


**Figure 4.** (A) The success rates from top 10 to top 1000 and top 1 to top 10 of HawkDock, ATTRACT, and ZDOCK 3.0.2, and (B) the success rates of the key residues predicted by MM/GBSA.

on the analysis of Dataset II. As expected, MM/GBSA showed an excellent performance in the key residue prediction (Supplementary Table S2 and Figure 4B). For the crystal structures in Dataset II, 39.53% (17) were successfully predicted at the top 1 level. Even for the predicted structures, MM/GBSA also reached a success rate of 23.26% at the top 1 level. As more predictions were considered in the evaluation, the success rates for both structures increased rapidly and were promoted to 95.35% and 81.40% in the top 10 level, respectively. Therefore, if a reliable model can be provided, an excellent performance can be achieved by MM/GBSA. In addition, analysis of the frequently occurring residues in the top 10 models were also executed by us. In dataset II, in terms of the top 10 level, the correct conformations can only be found in five proteins, four of which were predicted successfully at the top 5 level. For the proteins without an acceptable model at the top 10 level, 9 and 10 of 38 is also predicted successfully at the top 3 and 10 level, respectively, which provides an available solution when no reliable model is submitted. To sum up, MM/GBSA achieve a good performance in identifying the key residues in complex interfaces.

#### Examples of web server output

To illustrate the practicability of the HawkDock server, the complex 3D5S (47) from ZDOCK benchmark 4.0 predicted by the unbound structures of complement C3d fragment (C3d, PDB ID: 1C3D) (48) and fibrinogen-binding protein C-ter domain (Efb-C, PDB ID: 2GOM) (49) was used as an example. As shown in Figure 5, HawkDock generated the correct binding conformation with an l.RMSD of 1.337 Å between the crystal structure and the best hit. Then, this model was analyzed by MM/GBSA. The key residues of R131 and N138 from Efb-C (47) were successfully predicted in the top 10 predictions, ranking 4 and 10, respectively. Interestingly, the statistics of the frequently occurring residues in the top 10 models performed better than that based on the reliable model. R131 and N138 were successfully predicted at the one and eight rankings, respec-



**Figure 5.** The structural alignment of the crystal structure (PDB ID: 3D5S) colored yellow and the theoretical model predicted by HawkDock (colored green).

tively. For the crystal structure, it is clear to see that the MM/GBSA achieved an excellent performance and the key residue of N138 was successfully predicted as the best prediction (R131 was mutated to alanine in 3D5S).

#### CONCLUSION

Here, we present a user-friendly HawkDock server for the structural prediction and key residue analysis of PPIs. The combination of ATTRACT, HawkRank and MM/GBSA makes it very efficient for the identification of near-native docking models and key residues. The HawkDock server is an on-going project and further developments will be focus on the incorporation of additional information (e.g. co-evolutionary information or SAXS) and the integration with automatic modeling of protein structures.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.



## ACKNOWLEDGEMENTS

We are very grateful to Dr Martin Zacharias (ATTRACT) for making their methods available for implementation in the HawkDock server.

## FUNDING

National Key R&D Program of China [2016YFA0501701, 2016YFB0201700]; National Science Foundation of China [21575128, 81773632]. Funding for open access charge: National Natural Science Foundation of China.  
*Conflict of interest statement.* None declared.

## REFERENCES

- Gress, A., Ramensky, V. and Kalinina, O.V. (2017) Spatial distribution of disease-associated variants in three-dimensional structures of protein complexes. *Oncogenesis*, **6**, e380.
- Vakser, I.A. (2014) Protein-protein docking: from interaction to interactome. *Biophys. J.*, **107**, 1785–1793.
- Vreven, T., Hwang, H., Pierce, B.G. and Weng, Z. (2014) Evaluating template-based and template-free protein-protein complex structure prediction. *Brief. Bioinform.*, **15**, 169–176.
- Aloy, P., Ceulemans, H., Stark, A. and Russell, R.B. (2003) The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, **332**, 989–998.
- de Vries, S.J., Schindler, C.E., Chauvot de Beauchene, I. and Zacharias, M. (2015) A web interface for easy flexible protein-protein docking with ATTRACT. *Biophys. J.*, **108**, 462–465.
- Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhony, D., Yueh, C., Beglov, D. and Vajda, S. (2017) The ClusPro web server for protein-protein docking. *Nat. Protoc.*, **12**, 255–278.
- de Vries, S.J., van Dijk, M. and Bonvin, A.M.J.J. (2010) The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.*, **5**, 883–897.
- Pierce, B.G., Wiehe, K., Hwang, H., Kim, B.H., Vreven, T. and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, **30**, 1771–1773.
- Torchala, M., Moal, I.H., Chaleil, R.A., Fernandez-Recio, J. and Bates, P.A. (2013) SwarmDock: a server for flexible protein-protein docking. *Bioinformatics*, **29**, 807–809.
- Jimenez-Garcia, B., Pons, C., Svergun, D.I., Bernado, P. and Fernandez-Recio, J. (2015) pyDockSAXS: protein-protein complex structure by SAXS and computational docking. *Nucleic Acids Res.*, **43**, W356–W361.
- Jimenez-Garcia, B., Pons, C. and Fernandez-Recio, J. (2013) pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics*, **29**, 1698–1699.
- Quignot, C., Rey, J., Yu, J., Tuffery, P., Guerois, R. and Andreani, J. (2018) InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. *Nucleic Acids Res.*, **46**, W408–W416.
- Tovchigrechko, A. and Vakser, I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.*, **34**, W310–W314.
- Lyskov, S. and Gray, J.J. (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.*, **36**, W233–W238.
- Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. and Wolfson, H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.
- Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M.D. and Ritchie, D.W. (2010) HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.*, **38**, W445–W449.
- Lesk, V.I. and Sternberg, M.J. (2008) 3D-Garden: a system for modelling protein-protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. *Bioinformatics*, **24**, 1137–1144.
- Ramirez-Aportela, E., Lopez-Blanco, J.R. and Chacon, P. (2016) FRODOCK 2.0: fast protein-protein docking server. *Bioinformatics*, **32**, 2386–2388.
- Yan, Y., Zhang, D., Zhou, P., Li, B. and Huang, S.Y. (2017) HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.*, **45**, W365–W373.
- Clackson, T. and Wells, J.A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383–386.
- Gohlke, H. and Case, D.A. (2004) Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.*, **25**, 238–250.
- Ylilauri, M. and Pentikainen, O.T. (2013) MMGBSA as a tool to understand the binding affinities of filamin-peptide interactions. *J. Chem. Inf. Model.*, **53**, 2626–2633.
- Ayoub, A.T., Klobukowski, M. and Tuszynski, J.A. (2015) Detailed per-residue energetic analysis explains the driving force for microtubule disassembly. *PLoS Comput. Biol.*, **11**, e1004313.
- Xu, D., Si, Y.B. and Meroueh, S.O. (2017) A computational investigation of small-molecule engagement of hot spots at protein-protein interaction interfaces. *J. Chem. Inf. Model.*, **57**, 2250–2272.
- Chen, F., Liu, H., Sun, H., Pan, P., Li, Y., Li, D. and Hou, T. (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. *Phys. Chem. Chem. Phys.*, **18**, 22129–22139.
- Hou, T., Chen, K., McLaughlin, W.A., Lu, B. and Wang, W. (2006) Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain. *PLoS Comput. Biol.*, **2**, e1.
- Hou, T., Li, Y. and Wang, W. (2011) Prediction of peptides binding to the PKA RII $\alpha$  subunit using a hierarchical strategy. *Bioinformatics*, **27**, 1814–1821.
- Hou, T.J., Li, N., Li, Y.Y. and Wang, W. (2012) Characterization of domain-peptide interaction interface: Prediction of SH3 domain-mediated protein-protein interaction network in yeast by generic structure-based models. *J. Proteome Res.*, **11**, 2982–2995.
- Huang, D., Qi, Y., Song, J. and Zhang, J.Z. (2019) Calculation of hot spots for protein-protein interaction in p53/PMI-MDM2/MDMX complexes. *J. Comput. Chem.*, **40**, 1045–1056.
- Vreven, T., Hwang, H., Pierce, B.G. and Weng, Z. (2012) Prediction of protein-protein binding free energies. *Protein Sci.*, **21**, 396–404.
- Corrada, D., Soshilov, A.A., Denison, M.S. and Bonati, L. (2016) Deciphering dimerization modes of PAS Domains: Computational and experimental analyses of the AhR:ARNT complex reveal new insights into the mechanisms of AhR transformation. *PLoS Comput. Biol.*, **12**, e1004981.
- Pouplana, R. and Campanera, J.M. (2015) Energetic contributions of residues to the formation of early amyloid-beta oligomers. *Phys. Chem. Chem. Phys.*, **17**, 2823–2837.
- Xu, L., Li, Y.Y., Li, D., Xu, P., Tian, S., Sun, H.Y., Liu, H. and Hou, T.J. (2015) Exploring the binding mechanisms of MIF to CXCR2 using theoretical approaches. *Phys. Chem. Chem. Phys.*, **17**, 3370–3382.
- Feng, T., Chen, F., Kang, Y., Sun, H., Liu, H., Li, D., Zhu, F. and Hou, T. (2017) HawkRank: a new scoring function for protein-protein docking based on weighted energy terms. *J. Cheminform.*, **9**, 66.
- Hou, T.J., Qiao, X.B., Zhang, W. and Xu, X.J. (2002) Empirical aqueous solvation models based on accessible surface areas with implicit electrostatics. *J. Phys. Chem. B*, **106**, 11295–11304.
- Zacharias, M. (2003) Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.*, **12**, 1271–1282.
- Rego, N. and Koes, D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.
- Rodrigues, J.P., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A.S. and Bonvin, A.M. (2012) Clustering biomolecular complexes by residue contacts similarity. *Proteins*, **80**, 1810–1817.
- Cieplak, P., Caldwell, J. and Kollman, P. (2001) Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: Aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.*, **22**, 1048–1057.



40. Onufriev,A., Bashford,D. and Case,D.A. (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*, **55**, 383–394.
41. Case,D.A., Cheatham,T.E. 3rd, Darden,T., Gohlke,H., Luo,R., Merz,K.M. Jr, Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
42. Dolinsky,T.J., Nielsen,J.E., McCammon,J.A. and Baker,N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, **32**, W665–W667.
43. Hwang,H., Vreven,T., Janin,J. and Weng,Z. (2010) Protein-protein docking benchmark version 4.0. *Proteins*, **78**, 3111–3114.
44. Janin,J., Henrick,K., Moult,J., Eyck,L.T., Sternberg,M.J., Vajda,S., Vakser,I., Wodak,S.J. and Critical Assessment of, P.I. (2003) CAPRI: A critical assessment of predicted interactions. *Proteins*, **52**, 2–9.
45. Venkatraman,V., Yang,Y.F.D., Sael,L. and Kihara,D. (2009) Protein-protein docking using region-based 3D Zernike descriptors. *BMC Bioinform*, **10**, 407.
46. Pierce,B.G., Hourai,Y. and Weng,Z.P. (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One*, **6**, e24657.
47. Haspel,N., Ricklin,D., Geisbrecht,B.V., Kavraki,L.E. and Lambris,J.D. (2008) Electrostatic contributions drive the interaction between *Staphylococcus aureus* protein Efb-C and its complement target C3d. *Protein Sci.*, **17**, 1894–1906.
48. Nagar,B., Jones,R.G., Diefenbach,R.J., Isenman,D.E. and Rini,J.M. (1998) X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science*, **280**, 1277–1281.
49. Hammel,M., Sfyroera,G., Ricklin,D., Magotti,P., Lambris,J.D. and Geisbrecht,B.V. (2007) A structural basis for complement inhibition by *Staphylococcus aureus*. *Nat. Immunol.*, **8**, 430–437.