BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 03/05 pp92–99
Volume 2, Number 2, June 2019
DOI: 10.26599/BDMA.2018.9020035

A Non-Redundant Benchmark for Symmetric Protein Docking

Yumeng Yan and Sheng-You Huang*

Abstract: Symmetric proteins play important roles in many biological processes, such as signal transduction and molecular transportation. Therefore, determining the symmetric oligomeric structure of subunits is crucial to investigate the molecular mechanism of the related processes. Due to the high cost and technical difficulties associated with many experimental methods, computational approaches, such as molecular docking, have played an important complementary role in the determination of symmetric complex structures, in which a benchmark data set is pressingly needed. In the present work, we develop a comprehensive and non-redundant benchmark for symmetric protein docking based on the structures in the Protein Data Bank (PDB). The diverse dataset consists of 251 targets, including 212 cases with cyclic groups symmetry, 35 cases with dihedral groups symmetry, 3 cases with cubic groups symmetry, and 1 case with helical symmetry. According to the conformational changes in the interface between bound and unbound structures, the 251 targets were classified into three groups: 176 “easy”, 37 “medium”, and 38 “difficult” cases. A preliminary docking test on the targets of cyclic groups symmetry with M-ZDOCK indicated that symmetric multimer docking remains challenging. The benchmark will be beneficial for the development of symmetric protein docking algorithms. The proposed benchmark data set is available for download at http://huanglab.phys.hust.edu.cn/SDBenchmark/.

Key words: benchmark; symmetric protein; molecular docking; scoring functions; protein binding

1 Introduction

Symmetry is an important and profound concept and has played an important role in science from its very origin. For example, physicists are constantly on the lookout for the symmetric properties of a physical system to determine the related conservation laws characterizing this system[1]. In chemistry, molecular symmetry is also a fundamental concept. Symmetry can help predict or explain the chemical properties of a molecule, such as its dipole moment and spectroscopic transitions. In addition, since the double-helical symmetric structure of DNA was reported by Watson and Crick in 1953, symmetry has played an extremely important role in biological science[2].

Proteins are among the most important biological macromolecules in cells and have evolved to conduct a variety of cellular functions, from reaction catalysts to signal transduction and cell regulation[3,4]. As part of the biomolecular science, proteins also possess the property of symmetry. Most soluble and membrane-bound proteins *in vivo* conduct their functions by forming symmetric oligomer complexes[2]. Therefore, determining the 3D structures of symmetric oligomeric complexes is crucial to investigate the mechanisms of self interactions and assembly, understand the related biological processes, and ultimately develop therapeutic drugs[5,6]. However, due to the high cost and technical difficulties associated with many experimental methods, the number of experimentally

• Yumeng Yan and Sheng-You Huang are with Institute of Biophysics, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: yanyumeng110@hust.edu.cn; huangsy@hust.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2018-07-14; accepted: 2018-10-10

determined symmetric complex structures is limited. As such, molecular docking[7–21], which computationally samples and ranks putative binding modes through their binding scores, has played an important role in predicting the structures of symmetric oligomeric complexes[22].

An important aspect in docking is the construction of a good benchmark consisting of appropriately selected structures, which is critical not only for the development of docking algorithms and scoring functions[23,24], but also for the comparative assessment of current algorithms to improve the existing programs and develop new methods[24,25]. Despite the significant progresses achieved in hetero-protein docking benchmarks, such as the protein-protein docking benchmarks developed by the Weng group[26,27] and Dockground developed by the Vakser group[28,29], however, little effort has been made toward benchmarking symmetric oligomeric protein docking. As symmetric docking algorithms are often developed independently from standard protein-protein docking algorithms, development of a benchmark dataset for symmetric homomultimer docking is also necessary. Therefore, in the present work, we have developed a non-redundant benchmark of 251 diverse targets for symmetric oligomeric proteins. Each target in the benchmark includes a symmetric bound complex and the unbound structure of its identical subunits. The unbound structure was obtained from asymmetric complexes or monomers. This benchmark will be beneficial for the development and improvement of homomultimeric protein docking algorithms.

2 Materials and Method

2.1 Symmetry in protein structures

In general, protein symmetry refers to the point group or helical symmetry of identical subunits and can be classified into four types: cyclic groups (C_n), dihedral groups (D_n), cubic groups, and helical symmetry (H). Cyclic groups have one axis of rotational symmetry and are formed via a ring of symmetrically arranged subunits. There are twenty subtypes of symmetry in cyclic groups: C_2–C_{15}, C_{17}, C_{22}, C_{24}, C_{31}, C_{38}, and C_{39}. Dihedral groups contain an additional perpendicular axis for forming a two-fold symmetry and have twelve subtypes of symmetry: D_2–D_9, D_{11}, D_{12}, D_{17}, and D_{48}. Cubic groups are formed with a three-fold symmetry combined with another non-perpendicular rotational axis and include three possibilities: tetrahedral (T), octahedral (O), and icosahedral (I). Helical symmetry (H) combines translation with rotation around the direction of translation to create extended filaments[2]. In this work, all of the symmetry information of the included proteins is collected from the Protein Data Bank (PDB)[30].

2.2 Data selection and curation

A good benchmark for molecular docking should possess the following features. First, the targets of the benchmark should be adequately diverse to enable testing of the robustness of docking programs. Second, the structures in the benchmark must preferably be experimentally determined structures rather than computational models to exclude computational errors. Third, each target in the benchmark should include both bound and unbound structures to reflect realistic conformational changes upon binding[25].

To meet these requirements, we adopted the following criteria to construct an appropriate benchmark for symmetric proteins. First, we queried all X-ray structures with a resolution cutoff of 2.5 nm but without any RNA/DNA chains for each protein symmetry type in the PDB[30]. Protein chains with less than 20 residues were excluded. As of May 2, 2017, the search of the PDB yielded a total of 48 168 entries for all types of symmetry. Proteins with C_n symmetry yielded the most entries (i.e., 37 351 entries) while those with helical symmetry yielded the fewest entries (i.e., 255 entries). Since each subtype of C_n/D_n symmetry features different rotational angles and numbers of identical subunits, different constraints and strategies should be applied during docking. Therefore, we used the protein symmetry subtype to classify our targets.

We used biological unit information from the PDB to distinguish crystal contacts from biological complexes. For consistency, biological units with a specific number of chains were retained for a certain type of symmetry. For example, biological units with two protein chains were selected for C_2 symmetry, and the corresponding number of protein chains for D_2 symmetry was four. The number of chains for each symmetry type was reasonably set to retain most of the biological units for each type. Then, the Structural Classification Of Proteins (SCOP) database (version 1.75)[31] was applied to remove redundancies at the family level for each symmetry type. If more than one target was found

in the same SCOP family, the target with the best resolution and the longest chain length was selected. To remove possible redundancies, targets without SCOP unique identifers-sunid[32] or with more than one sunid were excluded from the bound candidate list. We also checked the interface of bound structures to ensure the presence of interacting residues within 5 nm between adjacent chains. All structures were manually checked to ensure that the interfaces were reasonable and those ligand-mediated or covalently linked structures were excluded as our work focuses on protein-protein interactions. Finally, we obtained a candidate list of bound structures for each symmetry subtype, including a total of 1400 bound targets.

We then attempted to identify the corresponding unbound structures for the 1400 bound targets. We searched sequences of the bound chain of each target against all of the protein chains in the PDB using the Protein Basic Local Alignment Search Tool (BLASTP) algorithm of the Basic Local Alignment Search Tool (BLAST) package[33]. A protein chain was defined as a candidate for the unbound structure if it met the following criteria: it shared over 95% sequence identity with the bound structure, alignment covered 95% of the sequence of the bound subunit, the difference in length was no more than 5% of the length of the bound subunit, and the protein chain was from an asymmetric complex or monomer. If multiple unbound structure candidates were found for a certain bound structure, the one with the highest sequence similarity and highest structure resolution was chosen. If the highest resolution for X-ray structures was poorer than 0.4 nm, the Nuclear Magnetic Resonance (NMR) structure would be selected instead. All unbound structures were subjected to manual inspection based on bound structures. Finally, 251 bound targets with unbound structures available for identical subunits were obtained, and these targets formed our non-redundant benchmark for symmetric protein docking.

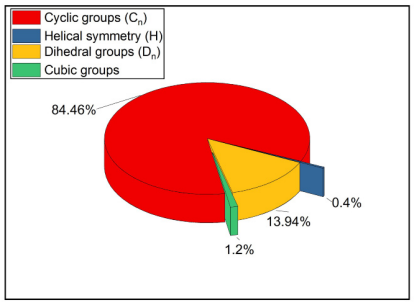
3 Results and Discussion

3.1 Benchmark dataset

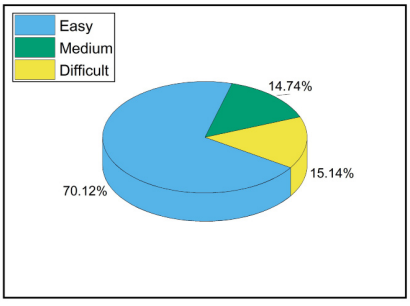
The 251 targets of the benchmark for symmetric proteins are listed on our website at http://huanglab.phys.hust.edu.cn/SDBenchmark/. For convenience, the unbound and bound structures of each target in the benchmark are named by their PDB code and chain ID(s). Each target is represented by the PDB code of its bound structure. The table reveals that the benchmark covers a wide range of proteins in terms of symmetric types, chain length, conformational changes, and complex types. For example, the benchmark contains all four types of symmetry, as well as most of their subtypes (Table 1, Fig. 1a). From Fig. 1a, it can be seen that the cases with the C_n symmetry are the most with a percentage of 84.46%. Figure 2 shows four representative examples of four types of symmetry. The chain length of the unbound structures ranges from 25 residues for the matrix protein M2 of target 3BKD to 843 residues for the glycogen

Table 1 Numbers of targets with different symmetries.

Symmetry type (Total number)	Symmetry subtype	Number of targets
Cyclic groups (212)	C_2	178
	C_3	20
	C_4	5
	C_5	4
	C_6	4
	C_7	1
Dihedral groups (35)	D_2	24
	D_3	6
	D_4	2
	D_5	1
	D_6	2
Cubic groups (3)	T	3
Helical symmetry (1)	H	1



(a) Different symmetry types



(b) Different flexility degrees

Fig. 1 Statistics of the symmetric targets in the benchmark.

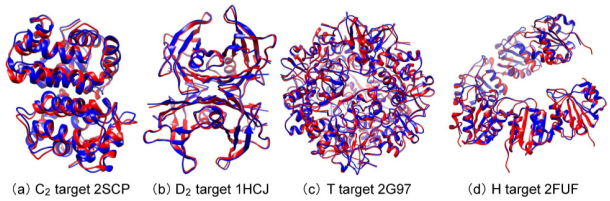


Fig. 2 Four representative targets for four symmetry types. Unbound structures (blue) are superimposed onto the subunits of the bound structures (red). The figure was prepared using UCSF Chimera[34].

phosphorylase of target 2GJ4. Root-Mean-Square Deviation (RMSD) and Template Modeling (TM)-score calculated by the TM-score program[35] were used to characterize conformational changes between bound and unbound structures. First, sequence alignment was performed using BLASTP[33]. Then according to the alignment result, the residue number in the bound and unbound structures was updated to ensure that the corresponding residues had the same number. Second, the TM-score program was used to calculate the RMSD and TM-score. To improve the usability of the benchmark, the "unbound complex structure" constructed by superimposing unbound protein chains onto bound chains was also provided in the benchmark. The unbound structures showed a wide range of conformational changes with a maximum RMSD of 2.279 nm for target 2CN4 and minimum TM-score of 0.30 for the target 1ET1. The structures in the benchmark only retained heavy atoms. Other atoms, such as water, hydrogen, and alternative atoms, were removed from the structures. Information on important HETATM atom, such as ligands and nonstandard amino acids, for the unbound structures are also listed in the benchmark.

3.2 Difficulty classification

To assign the difficulty levels of the targets, we classified the 251 targets into three categories of "easy", "medium", and "difficult" cases according to the interface RMSD (I_{rmsd}) between bound and unbound structures after optimal superimposition of their interfaces[36]. Here, I_{rmsd} was defined as the RMSD of the C_α atoms at the interface after optimal superimposition between bound and unbound structures; the interface refer to those residues within 1 nm of each other for the two partners of the bound complex. If multiple interfaces are available for a target, the interface with the largest conformational change was used to represent the conformational change of the target. The criteria used to categorize the targets and the statistics of targets with different properties are shown in Table 2 and Fig. 1b, respectively. Three representative examples corresponding to "easy", "medium", and "difficult" cases are shown in Fig. 3.

Figure 3 reveals that the conformational change is very small for the "easy" target 5P21 and that the backbones between the bound and unbound structures nearly overlap. For the "medium" target 1MKK, significant conformational changes are observed at the interface. A large conformational change can be seen in the backbones of the "difficult" target 1HRU, especially at the interface between the bound and unbound structures. For different levels of conformational changes, the degree of docking difficulty differs, and the appropriate docking strategies should be used. For "easy" targets, rigid-body docking may achieve good results; however flexibility must be considered when docking "difficult" targets.

A docking benchmark helps objectively evaluate the performance of docking algorithms. Through successful and, in particular, failed predictions and the different performances of existing docking programs, users may gain useful insights into the problems, find the advantages and disadvantages of different algorithms, and furthermore improve or develop novel docking programs. Therefore, the benchmark should be diverse, reflect realistic applications, and thus be challenging enough for the existing docking algorithms. Hence, we conducted a preliminary docking test

Table 2 Criteria for classifying targets by interface RMSD.

Category	Criterion	Number of targets
Easy	$I_{\mathrm{rmsd}} \leqslant 0.15$ nm	176
Medium	0.15 nm$\leqslant I_{\mathrm{rmsd}} \leqslant 0.3$ nm	37
Difficult	$I_{\mathrm{rmsd}} \geqslant 0.3$ nm	39

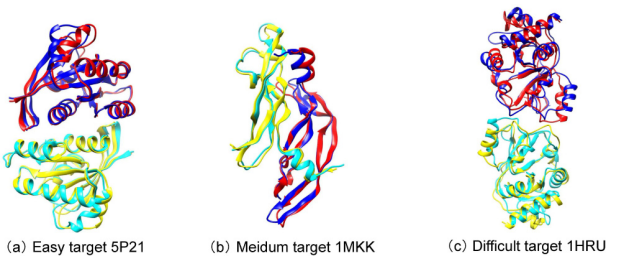


Fig. 3 Three representative targets for three difficulty levels. Bound structures are colored red/cyan and the corresponding unbound structures are colored in blue/yellow. (a) "Easy" target 5P21 ($I_{\mathrm{rmsd}} = 0.067$ nm); (b) "Medium" target 1MKK ($I_{\mathrm{rmsd}} = 0.218$ nm); and (c) "Difficult" target 1HRU ($I_{\mathrm{rmsd}} = 0.303$ nm).

for the targets of C_n symmetry in both bound and unbound docking by using M-ZDOCK, a Fast Fourier Transformation (FFT)-based approach for C_n symmetric multimer docking. The scoring function of M-ZDOCK consists of three energy terms: surface complementarity, electrostatics, and desolvation[8].

Given the predicted structures, we calculated the success rates and average number of hits (i.e., successful predictions) of M-ZDOCK in C_n symmetric binding mode predictions for both bound and unbound docking. Here, we used the criteria in the Critical Assessment of Prediction of Interactions (CAPRI) experiments[23,37,38] to evaluate the predicted binding modes. According to the criteria, the accuracies of the binding modes can be grouped into four categories: high, medium, acceptable, and incorrect. In this work, a prediction with an at least acceptable accuracy was considered as a "hit". The success rate was defined as the number of targets with at least one successful prediction divided by the total number of targets in the benchmark when a certain number of top predictions is considered.

Figure 4 shows the results of M-ZDOCK in both bound and unbound docking. Figure 5 shows the docking results of the three different types.

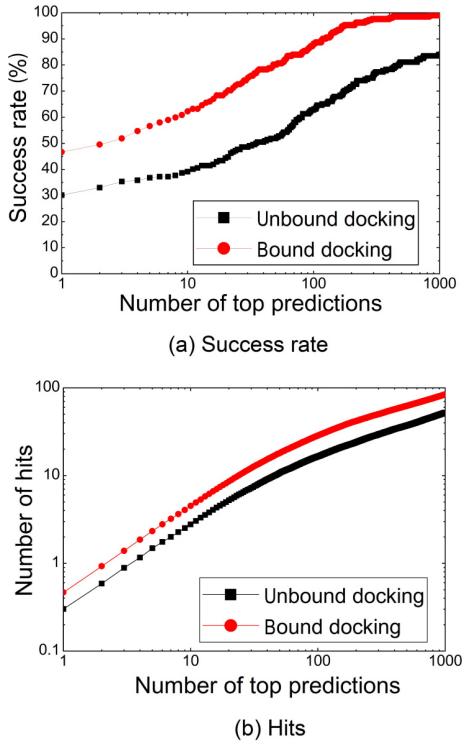


Fig. 4 Success rates (a) and average number of hits per complex (b) obtained by M-ZDOCK as a function of the number of top predictions for unbound and bound docking.

The bound docking results can serve as a primary test of the performance of scoring functions because no conformational change occurs in bound structures during docking. It can be observed from Fig. 4a that M-ZDOCK achieves a satisfactory performance in bound docking with a success rate of 46.7% for top 1 prediction and 62.3% for top 10 predictions. Figure 4a also shows that nearly all of the test cases are successfully predicted when the top 1000 predictions are considered. This result indicates that the sampling process was able to search near-native conformations for all targets but that the scoring function was not good enough to rank the correct binding modes within top predictions. Such a finding suggests that a more accurate scoring function for symmetric protein interactions is needed. Binding mode prediction for unbound docking was much more challenging with a success rate of only 30.2% for top 1 prediction and 39.2% for top 10 predictions. The poorer performance of unbound docking compared with that of bound docking indicates the significant impact of conformational changes and the necessity to consider protein flexibility during realistic docking. Figure 4b shows the average number of hits per complex for bound and unbound docking. The results of bound docking are clearly better than those of unbound docking, as expected. When the top 100 predictions were considered, M-ZDOCK obtained an average of 28.6 hits per complex for bound docking compared with 16.6 hits for unbound docking. The smaller number of hits of unbound docking compared with that of bound docking again indicates that the C_n symmetric multimer docking problem is challenging and requires the development of advanced docking programs with more accurate scoring functions.

Similar trends can also be seen in Fig. 5. For the three types of cases in the benchmark, the results of bound docking are better than those of unbound docking in terms of both success rate and average number of hits. In addition, the difference in performance between bound and unbound docking becomes more significant as the degree of docking difficulty increases in our benchmark. The difference in success rates for top 1 prediction between bound and unbound docking for "easy", "medium", and "difficult" cases is 4.76%, 21.88%, and 63.64%, respectively. This trend in performance is similar for average number of hits. The difference in number of hits for top 1 prediction for "easy", "medium", and "difficult" cases is 0.047, 0.219,

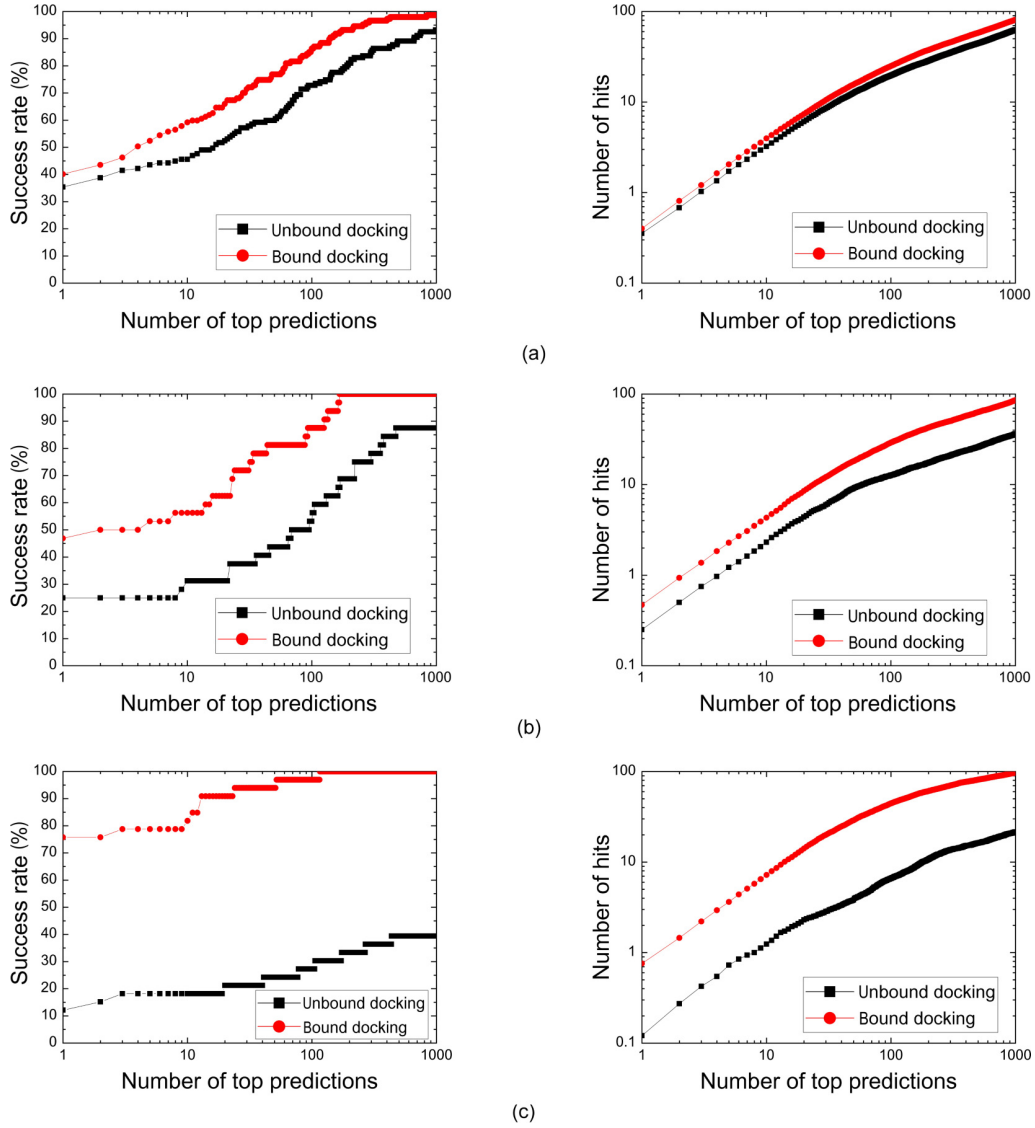


Fig. 5 Docking results of easy (a), medium (b), and difficult (c) cases. The left column shows the results of success rate for these three types, and the right column shows the corresponding results of average number of hits.

and 0.637, respectively. Such a trend indicates that larger conformational changes cause greater difficulties in docking and poorer docking results, which is consistent with the difficulty classification of our benchmark.

4 Conclusion

We have constructed a comprehensive and non-redundant benchmark of 251 diverse targets for symmetric protein docking. The benchmark consists of 212 test cases of C_n symmetry, 35 test cases of D_n symmetry, 3 test cases of cubic groups symmetry, and 1 test case of H symmetry. According to the conformational changes observed at their interfaces, the 251 targets of the benchmark were grouped into 176 "easy", 37 "medium", and 38 "difficult" cases. A preliminary docking test on the targets of C_n symmetry showed that the symmetric multimer docking problem remains a challenge and requires the development of more accurate docking algorithms and scoring functions. The benchmark also includes the targets of other types of symmetry, such as D_n symmetry and H symmetry. The present benchmark is of value for symmetric multimeric protein docking and scoring. The benchmark will be updated annually and is available at http://huanglab.phys.hust.edu.cn/SDBenchmark/.

Acknowledgment

This work was supported by the National Key R&D Program of China (Nos. 2016YFC1305800 and

2016YFC1305805), the National Natural Science Foundation of China (No. 31670724), and the startup grant of Huazhong University of Science and Technology.

References

[1] Y. Kosmann-Schwarzbach, *The Noether Theorems*. New York, NY, USA: Springer, 2011, pp. 55–64.

[2] D. S. Goodsell and A. J. Olson, Structural symmetry and protein function, *Ann. Rev. Biophys. Biomol. Struct.*, vol. 29, no. 1, pp. 105–153, 2000.

[3] I. A. Vakser, Protein-protein docking: From interaction to interactome, *Biophys. J.*, vol. 107, no. 8, pp. 1785–1793, 2014.

[4] G. Waksman and C. Sansom, Introduction: Proteomics and protein-protein interactions: Biology, chemistry, bioinformatics, and drug design, in *Proteomics and Protein-Protein Interactions*, G. Waksman, ed. Boston, MA, USA: Springer, 2005, pp. 1–18.

[5] S. Y. Huang, Search strategies and evaluation in protein-protein docking: Principles, advances and challenges, *Drug Discov. Today*, vol. 19, no. 8, pp. 1081–1096, 2014.

[6] R. J. Bienstock, Computational drug design targeting protein-protein interactions, *Curr. Pharm. Des.*, vol. 18, no. 9, pp. 1240–1254, 2012.

[7] Y. M. Yan, Z. Y. Wen, X. X. Wang, and S. Y. Huang, Addressing recent docking challenges: A hybrid strategy to integrate template-based and free protein-protein docking, *Proteins*, vol. 85, no. 3, pp. 497–512, 2017.

[8] B. Pierce, W. Tong, and Z. Weng, M-ZDOCK: A grid-based approach for C_n symmetric multimer docking, *Bioinformatics*, vol. 21, no. 8, pp. 1472–1478, 2004.

[9] B. G. Pierce, K. Wiehe, H. Hwang, B. H. Kim, T. Vreven, and Z. P. Weng, ZDOCK server: Interactive docking prediction of protein-protein complexes and symmetric multimers, *Bioinformatics*, vol. 30, no. 12, pp. 1771–1773, 2014.

[10] I. André, P. Bradley, C. Wang, and D. Baker, Prediction of the structure of symmetrical protein assemblies, *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 45, pp. 17656–17661, 2007.

[11] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson, PatchDock and SymmDock: Servers for rigid and symmetric docking, *Nucleic Acids Res.*, vol. 33, pp. 363–367, 2005.

[12] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson, Geometry-based flexible and symmetric protein docking, *Proteins*, vol. 60, no. 2, pp. 224–231, 2005.

[13] M. Baek, T. Park, L. Heo, C. Park, and C. Seok, GalaxyHomomer: A web server for protein homo-oligomer structure prediction from a monomer sequence or structure, *Nucleic Acids Res.*, vol. 45, no. 1, pp. 320–324, 2017.

[14] I. A. Vakser, Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex, *Proteins*, vol. 29, no. 1, pp. 226–230, 1997.

[15] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin, HADDOCK: A protein-protein docking approach based on biochemical or biophysical information, *J. Am. Chem. Soc.*, vol. 125, no. 7, pp. 1731–1737, 2003.

[16] D. W. Ritchie and G. J. L. Kemp, Protein docking using spherical polar Fourier correlations, *Proteins*, vol. 39, no. 2, pp. 178–194, 2000.

[17] M. Zacharias, Protein-protein docking with a reduced protein model accounting for side-chain flexibility, *Protein Sci.*, vol. 12, no. 6, pp. 1271–1282, 2003.

[18] M. Torchala, I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, and P. A. Bates, SwarmDock: A server for flexible proteinprotein docking, *Bioinformatics*, vol. 29, no. 6, pp. 807–809, 2013.

[19] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, and S. Vajda, The ClusPro web server for protein-protein docking, *Nat. Protoc.*, vol. 12, no. 2, pp. 255–278, 2017.

[20] T. M. Cheng, T. L. Blundell, and J. Fernandez-Recio, pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking, *Proteins*, vol. 68, no. 2, pp. 503–515, 2007.

[21] Y. M. Yan, D. Zhang, P. Zhou, B. T. Li, and S. Y. Huang, HDOCK: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy, *Nucleic Acids Res.*, vol. 45, no. 1, pp. 365–373, 2017.

[22] Y. Yan, H. Tao, and S. Y. Huang, HSYMDOCK: A docking web server for predicting the structure of protein homo-oligomers with C_n or D_n symmetry, *Nucleic Acids Res.*, vol. 46, no. 1, pp. 423–431, 2018.

[23] J. Janin, K. Henrick, J. Moult, L. T. Eyck, M. J. Sternberg, S. Vajda, I. Vakser, and S. J. Wodak, Critical assessment of predicted interactions, *Proteins*, vol. 52, no. 1, pp. 2–9, 2003.

[24] J. B. Jr. Dunbar, R. D. Smith, C. Y. Yang, P. M. Ung, K. W. Lexa, N. A. Khazanov, J. A. Stuckey, S. M. Wang, and H. A. Carlson, CSAR benchmark exercise of 2010: Selection of the protein-ligand complexes, *J. Chem. Inf. Model.*, vol. 51, no. 9, pp. 2036–2046, 2011.

[25] S. Y. Huang and X. Q. Zou, A non-redundant structure dataset for benchmarking protein-RNA computational docking, *J. Comput. Chem.*, vol. 34, no. 4, pp. 311–318, 2013.

[26] H. Hwang, T. Vreven, J. Janin, and Z. P. Weng, Protein-protein docking benchmark version 4.0, *Proteins*, vol. 78, no. 15, pp. 3111–3114, 2010.

[27] T. Vreven, I. H. Moal, A. Vangone, B. G. Pierce, P. L. Kastritis, M. Torchala, R. Chaleil, B. Jiménez-García, P. A. Bates, J. Fernandez-Recio, et al., Updates to the integrated protein-protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2, *J. Mol. Biol.*, vol. 427, no. 19, pp. 3031–3041, 2015.

[28] S. Y. Liu, Y. Gao, and I. A. Vakser, DOCKGROUND protein-protein docking decoy set, *Bioinformatics*, vol. 24, no. 22, pp. 2634–2635, 2008.

[29] I. Anishchenko, P. J. Kundrotas, A. V. Tuzikov, and I. A. Vakser, Protein models docking benchmark 2, *Proteins*, vol. 83, no. 5, pp. 891–897, 2015.

[30] H. M. Berman, J. Westbrook, Z. K. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The protein data bank, *Nucleic Acids Res.*, vol. 28, no. 1, pp. 235–242, 2000.

[31] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, SCOP: A structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.*, vol. 247, no. 4, pp. 536–540, 1995.

[32] L. Lo Conte, S. E. Brenner, T. J. P. Hubbard, C. Chothia, and A. G. Murzin, SCOP database in 2002: Refinements accommodate structural genomics, *Nucleic Acids Res.*, vol. 30, no. 1, pp. 264–267, 2002.

[33] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden, BLAST+: Architecture and applications, *BMC Bioinformatics*, vol. 10, p. 421, 2009.

[34] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis, *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.

[35] Y. Zhang and J. Skolnick, Scoring function for automated assessment of protein structure template quality, *Proteins*, vol. 57, no. 4, pp. 702–710, 2004.

[36] Y. M. Yan and S. Y. Huang, RRDB: A comprehensive and non-redundant benchmark for RNA-RNA docking and scoring, *Bioinformatics*, vol. 34, no. 3, pp. 453–458, 2018.

[37] R. Méndez, R. Leplae, M. F. Lensink, and S. J. Wodak, Assessment of CAPRI predictions in rounds 3–5 shows progress in docking procedures, *Proteins*, vol. 60, no. 2, pp. 150–169, 2005.

[38] M. F. Lensink, R. Méndez, and S. J. Wodak, Docking and scoring protein complexes: CAPRI 3rd edition, *Proteins*, vol. 69, no. 4, pp. 704–718, 2007.

Sheng-You Huang is a professor in the School of Physics at Huazhong University of Science and Technology. He received the PhD degree from Wuhan University in 2003. His research interests include computational biology, molecular modeling, and bioinformatics for complex interactions in biological systems with applications to drug design and discovery.

Yumeng Yan is currently a PhD student in the School of Physics at Huazhong University of Science and Technology. He got the bachelor degree from Huazhong University of Science and Technology in 2016. His research interests include molecular docking and scoring functions.