**Supplementary data for:**

**MetaScore: A novel machine-learning based approach to improve traditional scoring functions for scoring protein-protein docking conformations**

**Yong Jung1,2,5, Cunliang Geng8, Alexandre M. J. J. Bonvin8, Li C. Xue8,9\*, Vasant G. Honavar1,2,3,4,5,6,7\***

1Bioinformatics & Genomics Graduate Program, Pennsylvania State University, University Park, PA 16802, USA;

2Artificial Intelligence Research Laboratory, Pennsylvania State University, University Park, PA 16802, USA;

3Center for Big Data Analytics and Discovery Informatics, Pennsylvania State University, University Park, PA 16823, USA;

4Institute for Computational and Data Sciences, Pennsylvania State University, University Park, PA 16802, USA;

5Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA;

6Clinical and Translational Sciences Institute, Pennsylvania State University, University Park, PA 16802, USA;

7College of Information Sciences & Technology, Pennsylvania State University, University Park, PA 16802, USA;

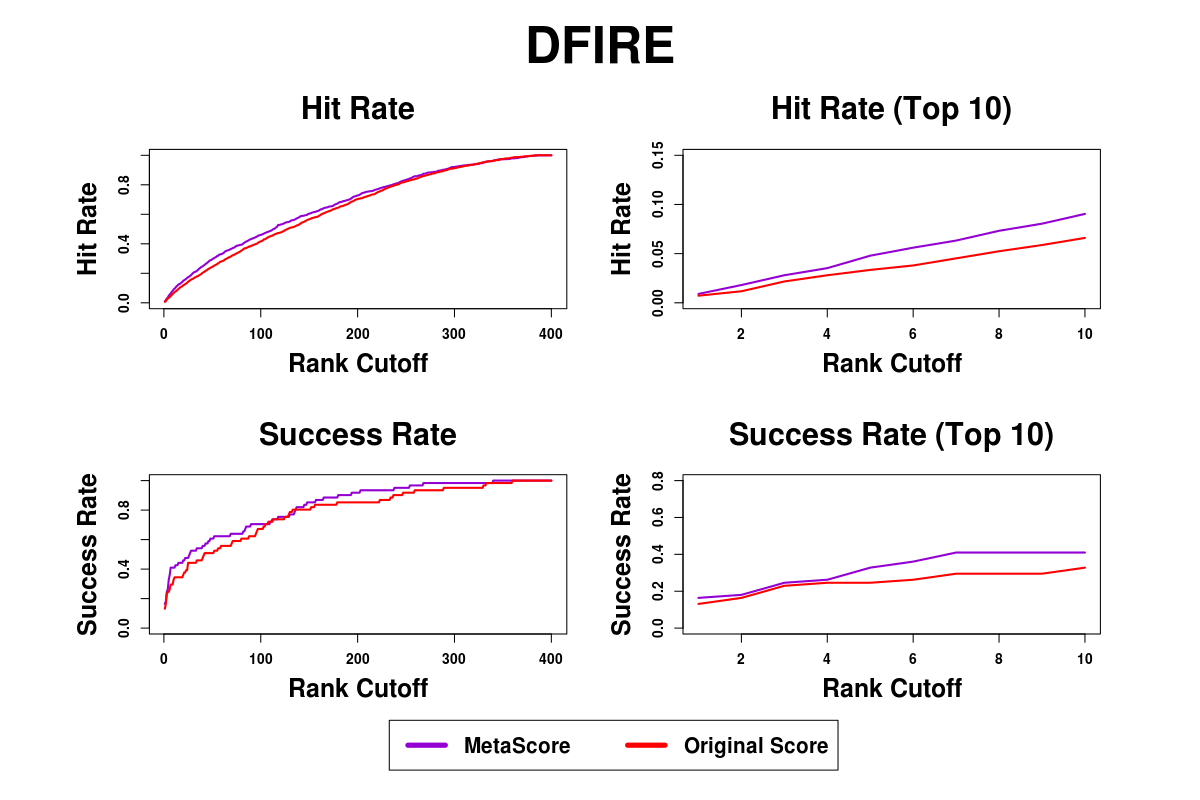
8Bijvoet Centre for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands;

9Center for Molecular and Biomolecular Informatics, Radboudumc, Greet Grooteplein 26-28, 6525 GA Nijmegen, the Netherlands.

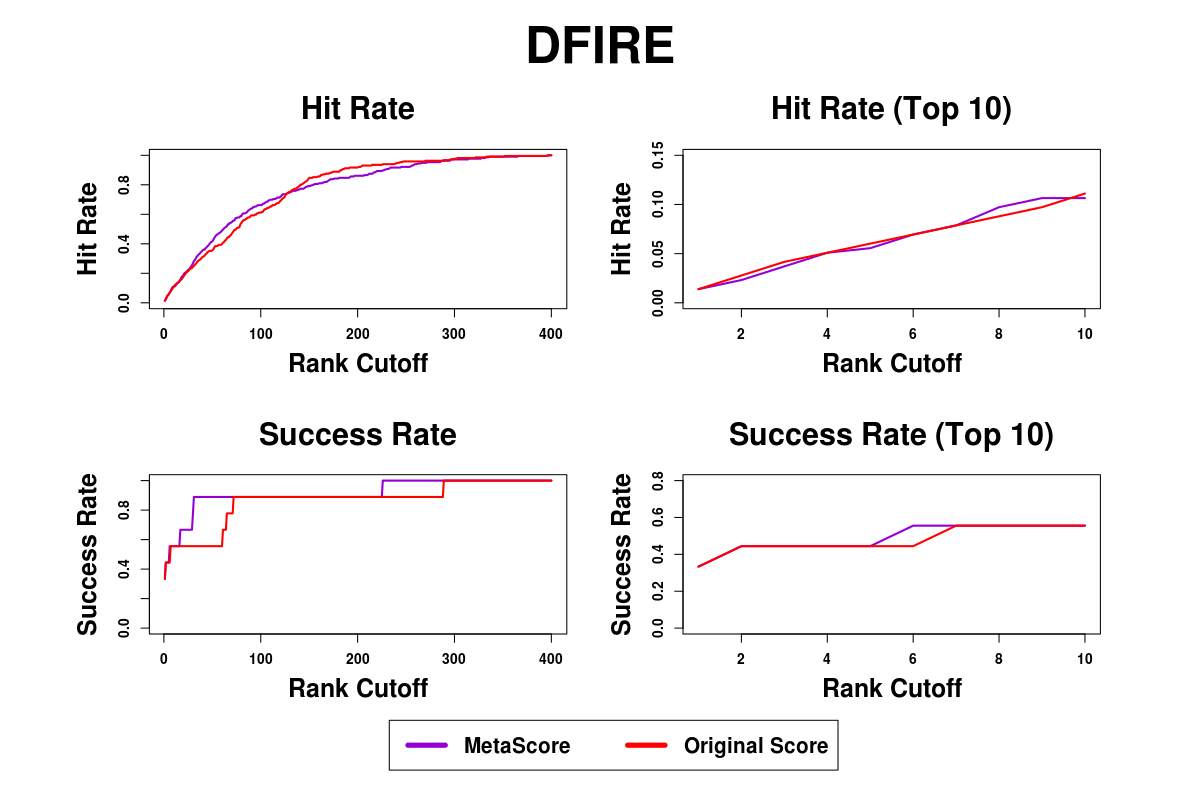
\*Correspondence: vhonavar@psu.edu. [Li.Xue@radboudumc.nl](mailto:Li.Xue@radboudumc.nl)

Supplementary Table S1. BLAST parameters.

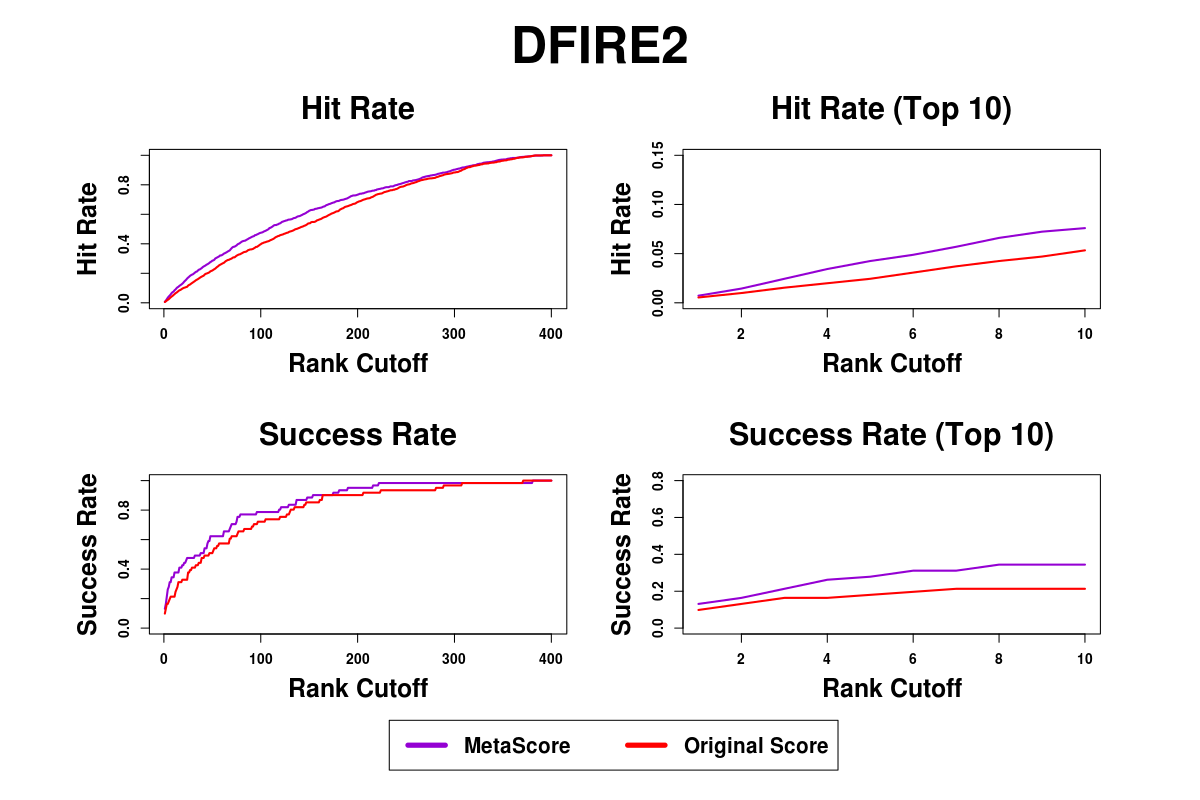
|  |  |  |  |
| --- | --- | --- | --- |
| Query Length | Substitution Matrix | Gap Open Cost | Gap Extend Cost |
| <35 | PAM-30 | 9 | 1 |
| 35-50 | PAM-70 | 10 | 1 |
| 50-85 | BLOSUM-80 | 10 | 1 |
| >85 | BLOSUM-62 | 11 | 1 |



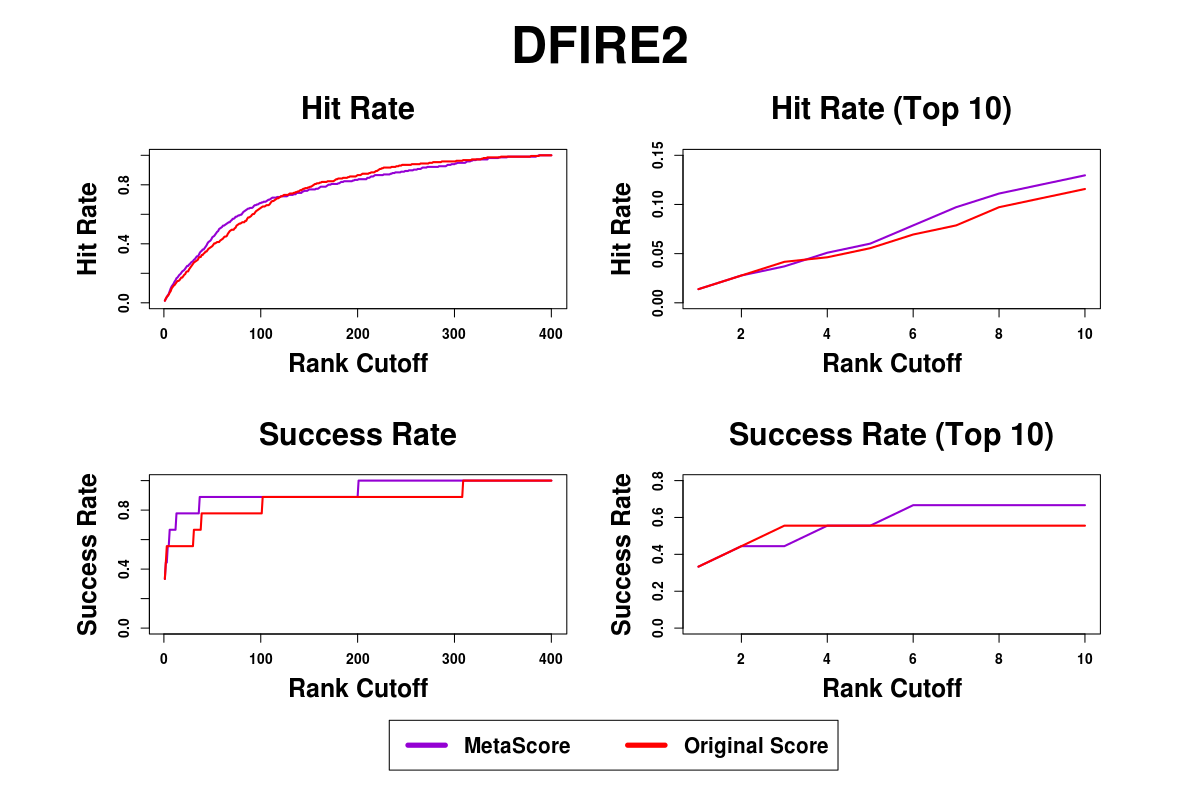
Supplementary Figure S1. Success rates and hit rates plotted against the top m conformations for a classical scoring method (DFIRE), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



Supplementary Figure S2. Success rates and hit rates plotted against the top m conformations for a classical scoring method (DFIRE), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



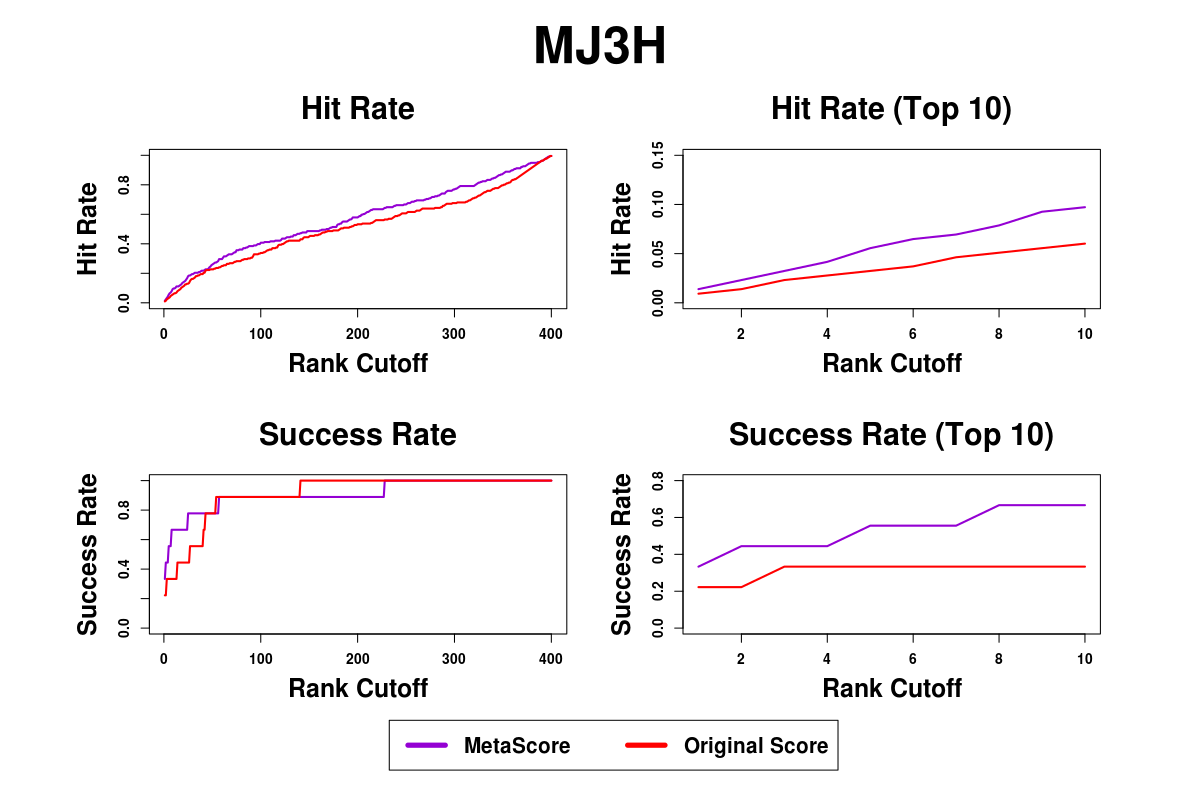
Supplementary Figure S3. Success rates and hit rates plotted against the top m conformations for a classical scoring method (DFIRE2), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



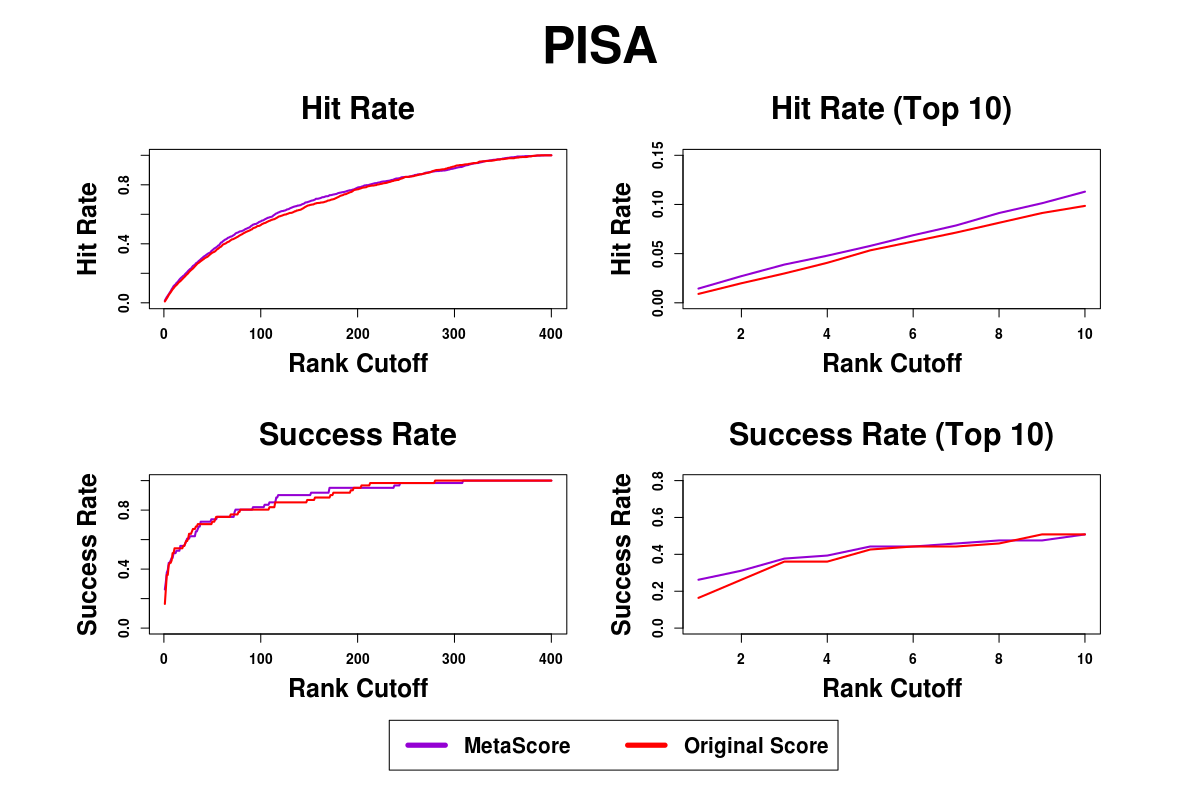
Supplementary Figure S4. Success rates and hit rates plotted against the top m conformations for a classical scoring method (DFIRE2), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



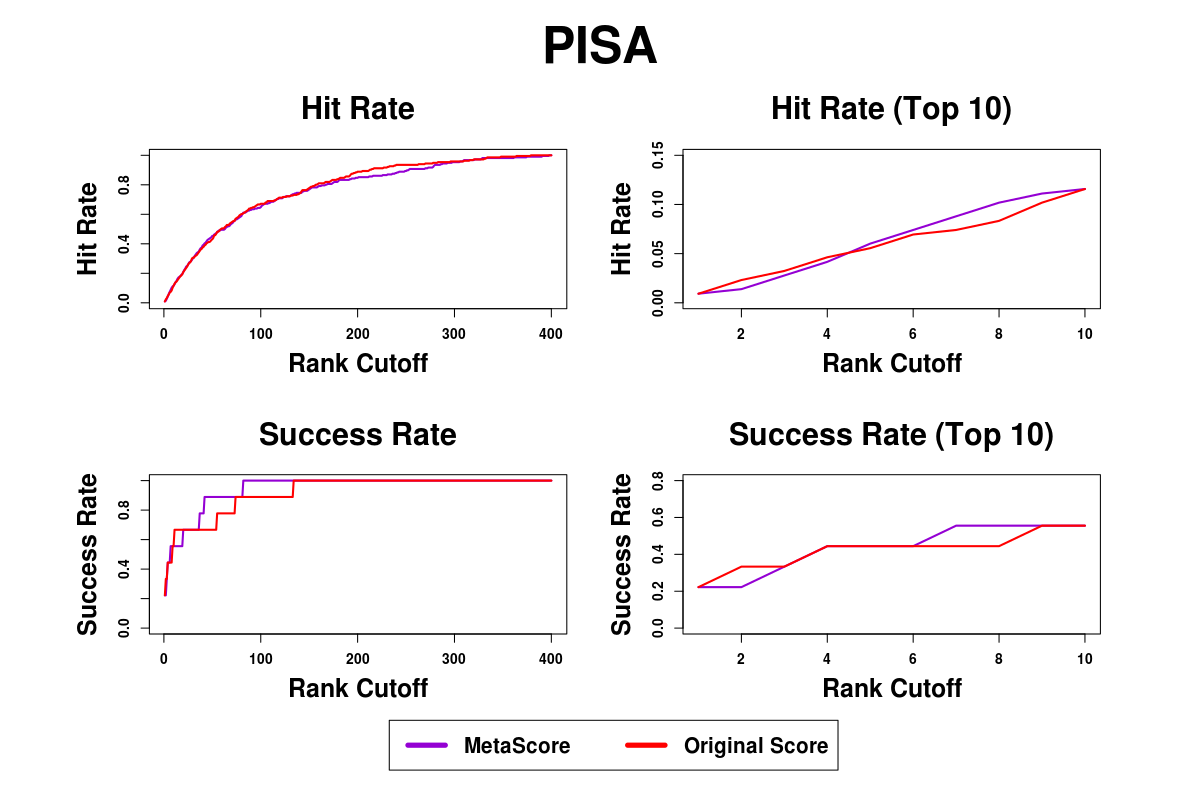
Supplementary Figure S5. Success rates and hit rates plotted against the top m conformations for a classical scoring method (MJ3H), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



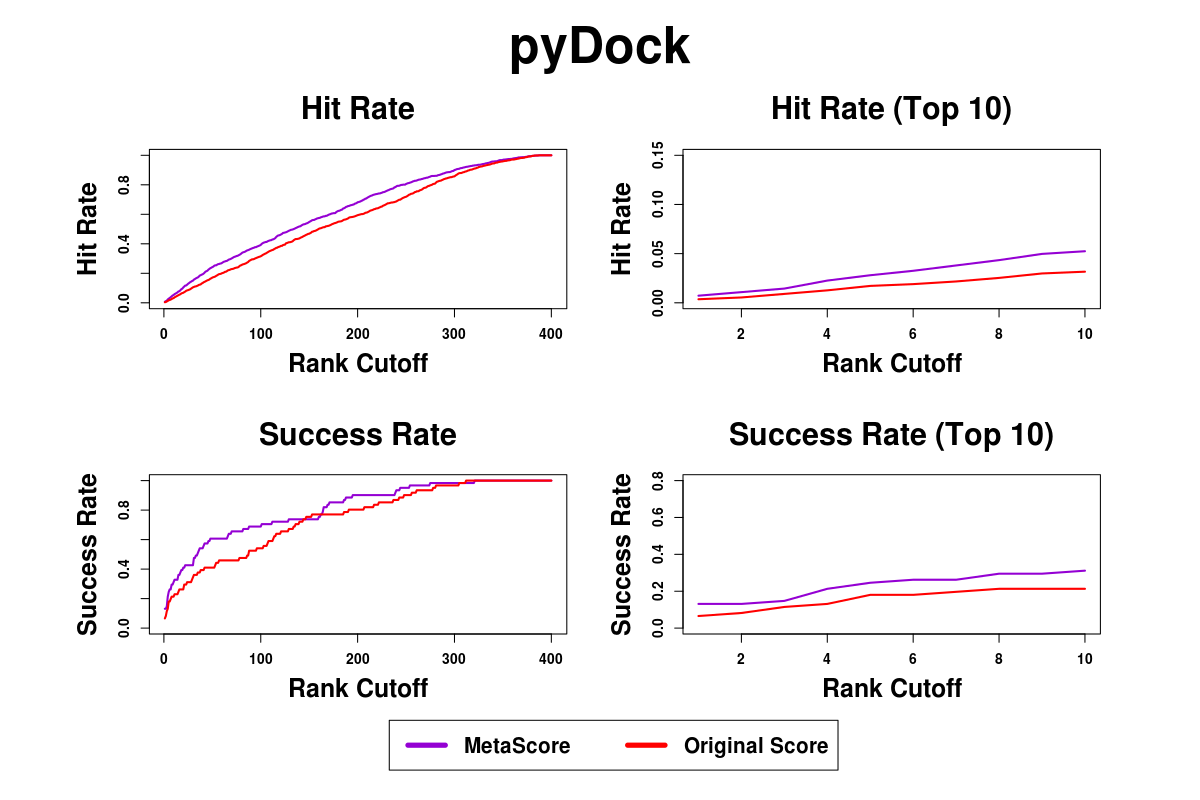
Supplementary Figure S6. Success rates and hit rates plotted against the top m conformations for a classical scoring method (MJ3H), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



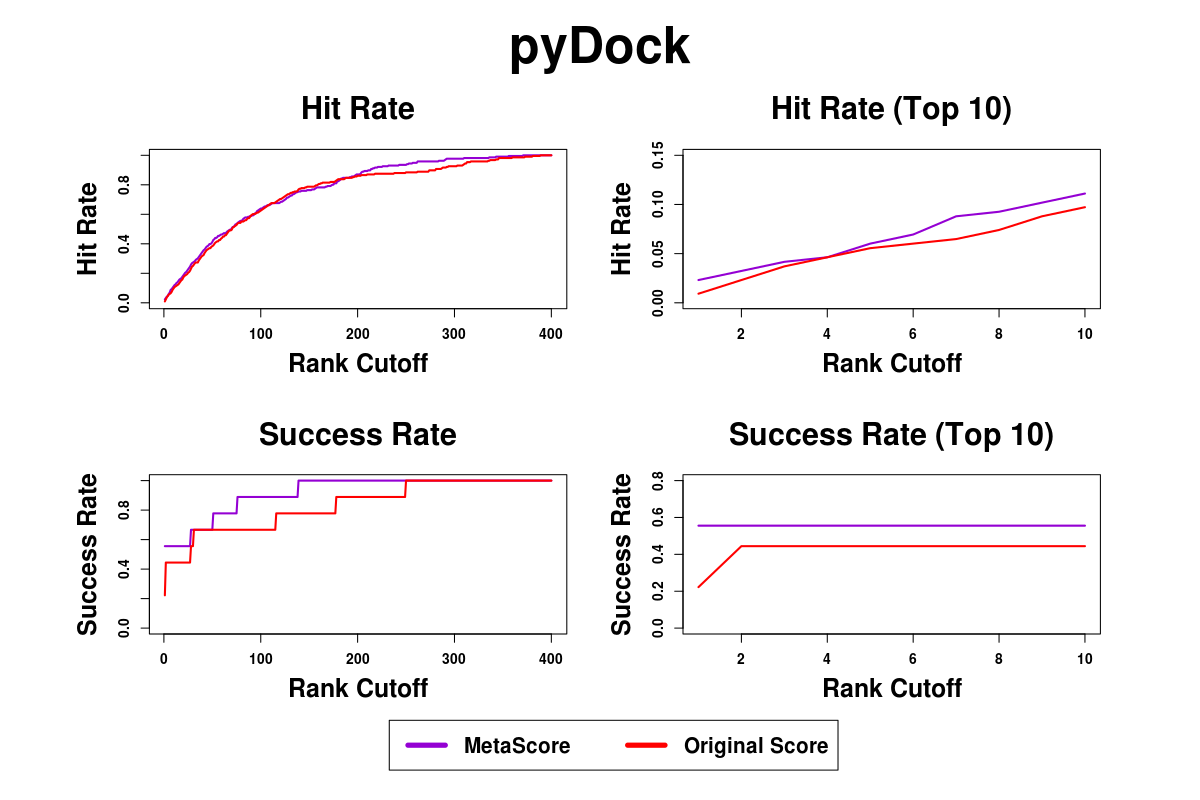
Supplementary Figure S7. Success rates and hit rates plotted against the top m conformations for a classical scoring method (PISA), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



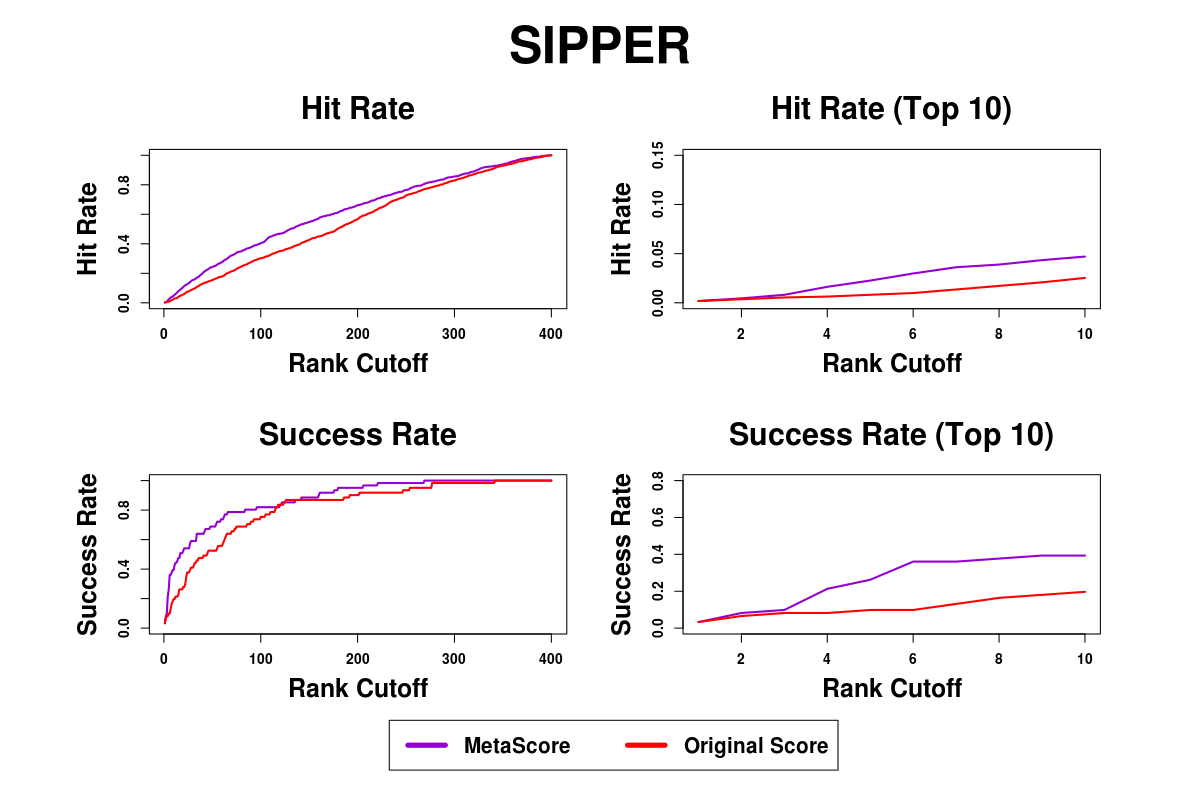
Supplementary Figure S8. Success rates and hit rates plotted against the top m conformations for a classical scoring method (PISA), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



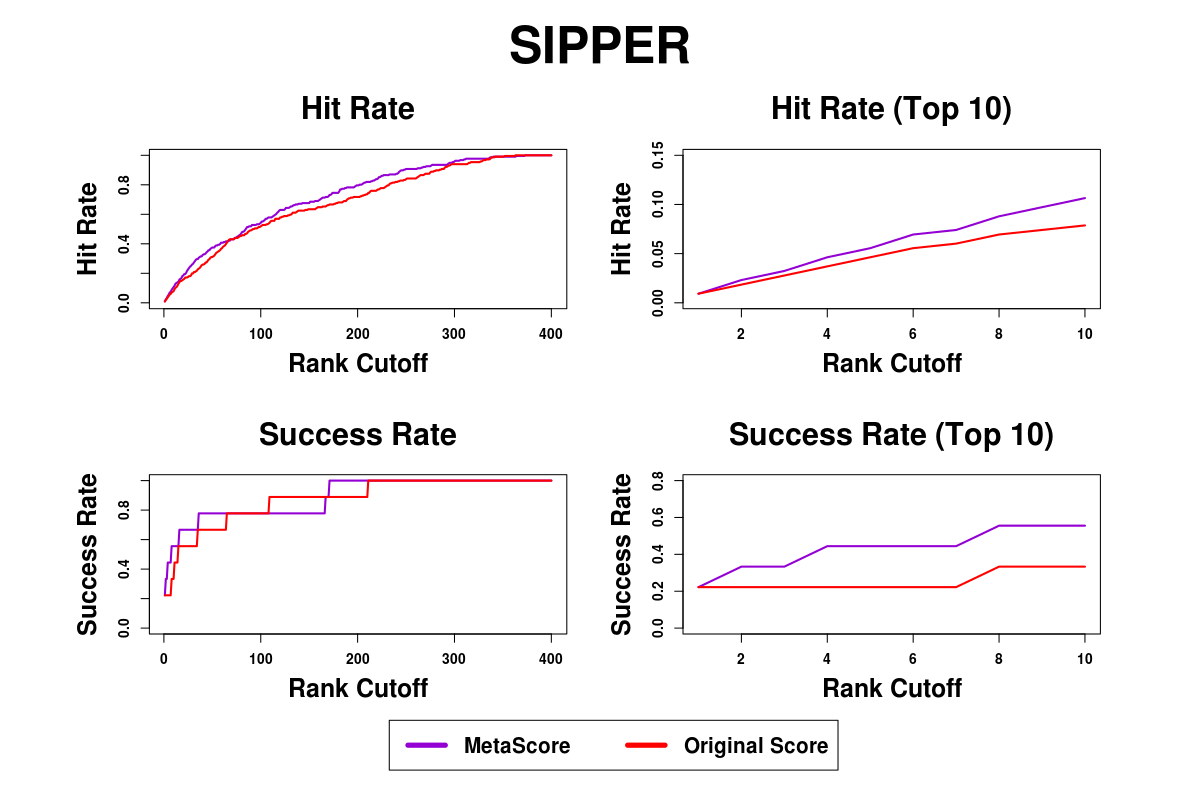
Supplementary Figure S9. Success rates and hit rates plotted against the top m conformations for a classical scoring method (pyDock), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



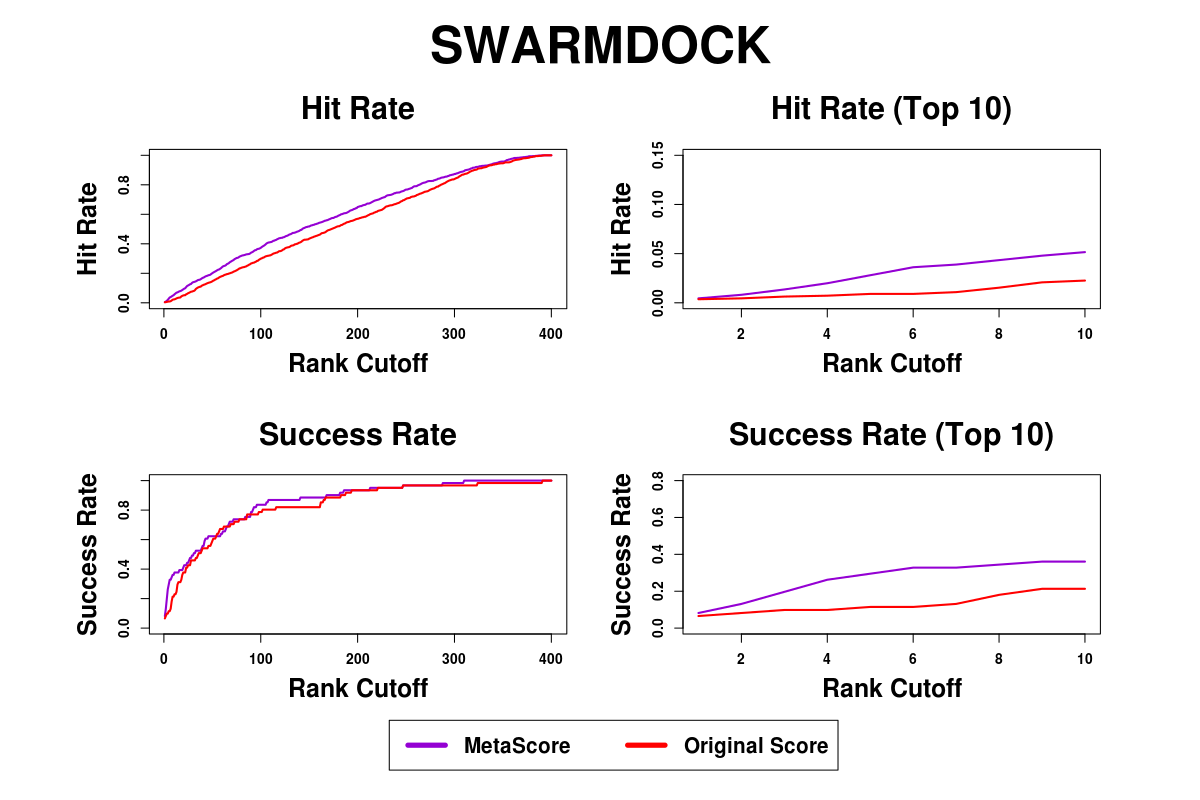
Supplementary Figure S10. Success rates and hit rates plotted against the top m conformations for a classical scoring method (pyDock), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



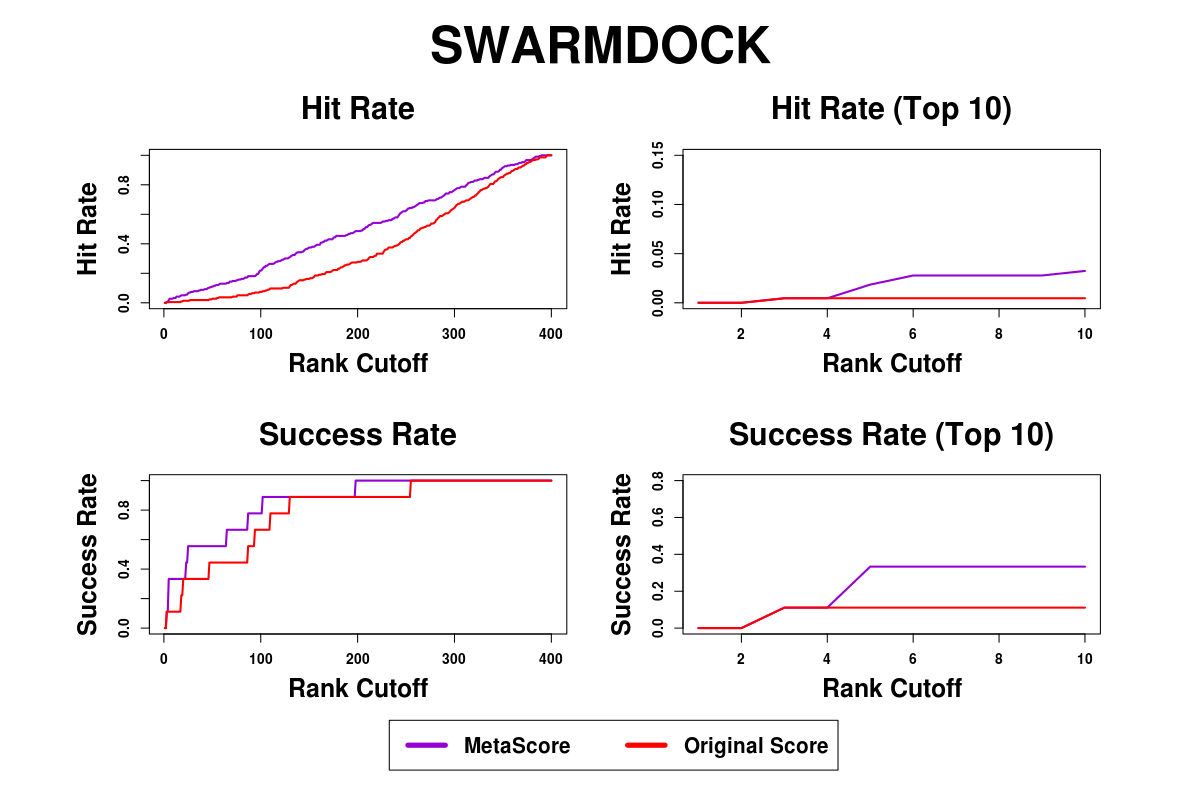
Supplementary Figure S11. Success rates and hit rates plotted against the top m conformations for a classical scoring method (SIPPER), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



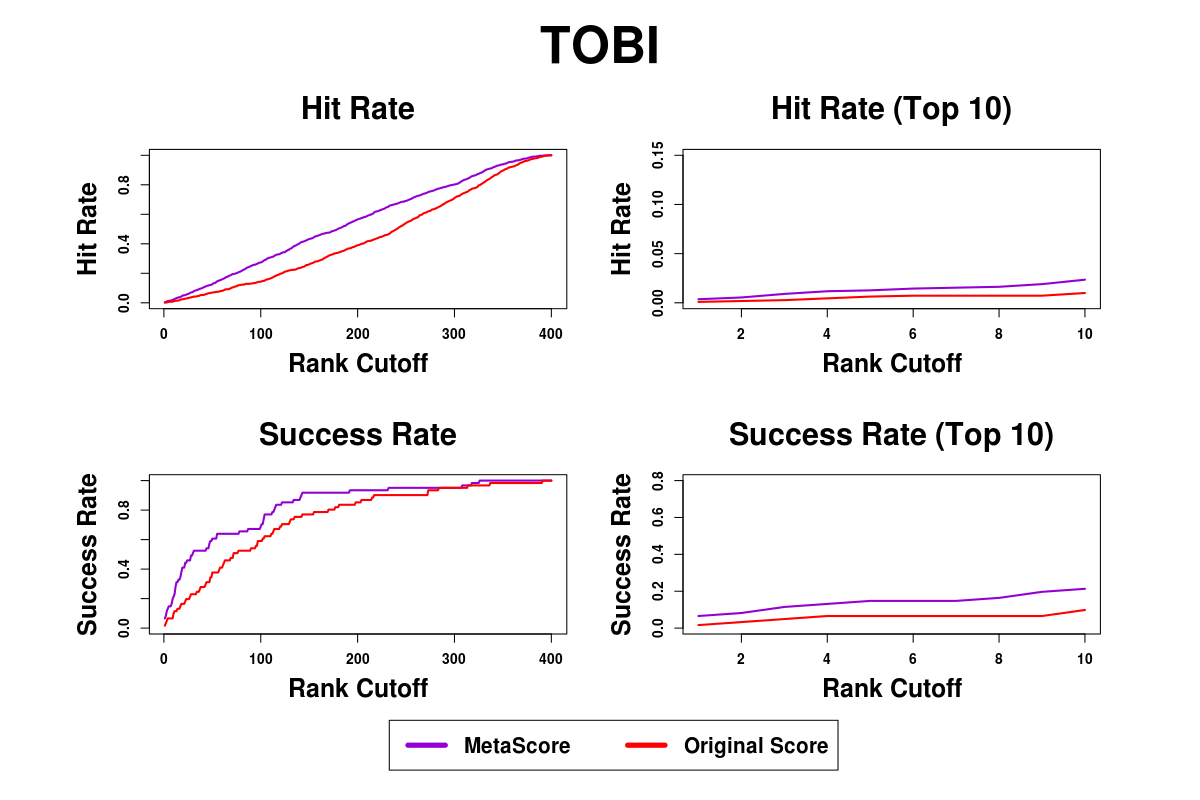
Supplementary Figure S12. Success rates and hit rates plotted against the top m conformations for a classical scoring method (SIPPER), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



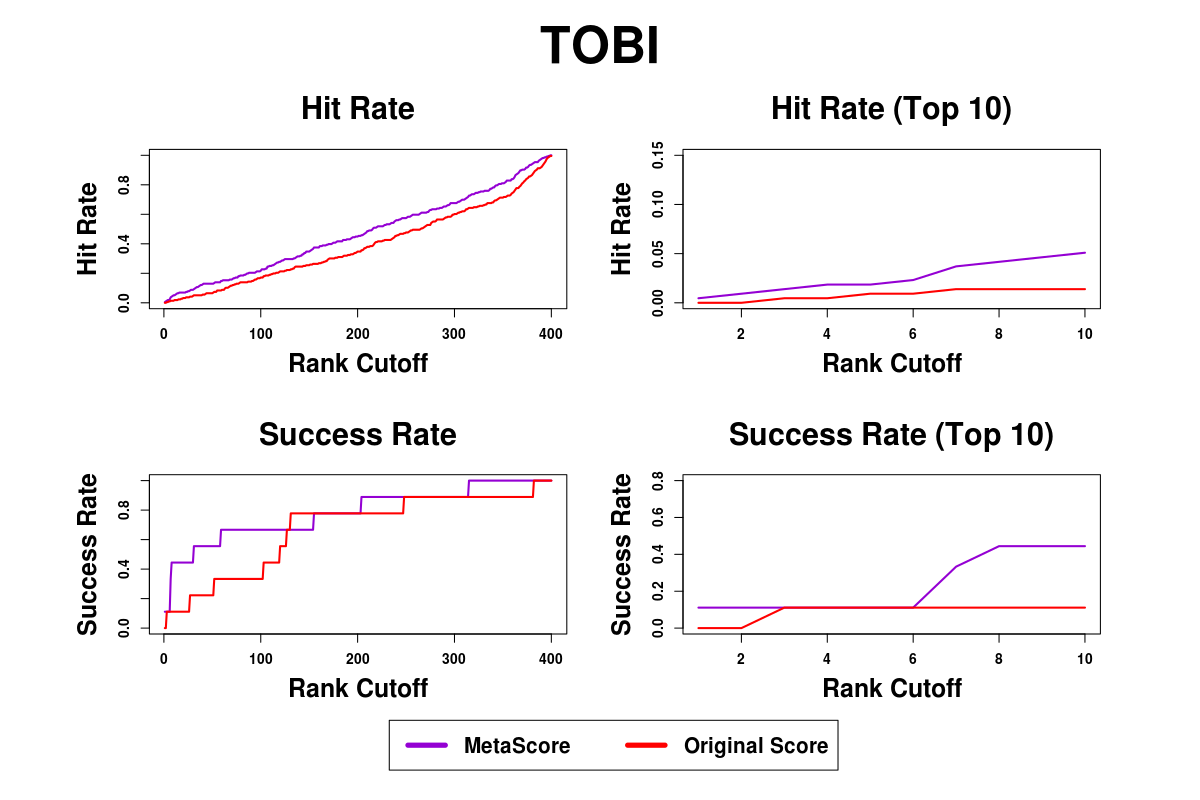
Supplementary Figure S13. Success rates and hit rates plotted against the top m conformations for a classical scoring method (SWARMDOCK), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



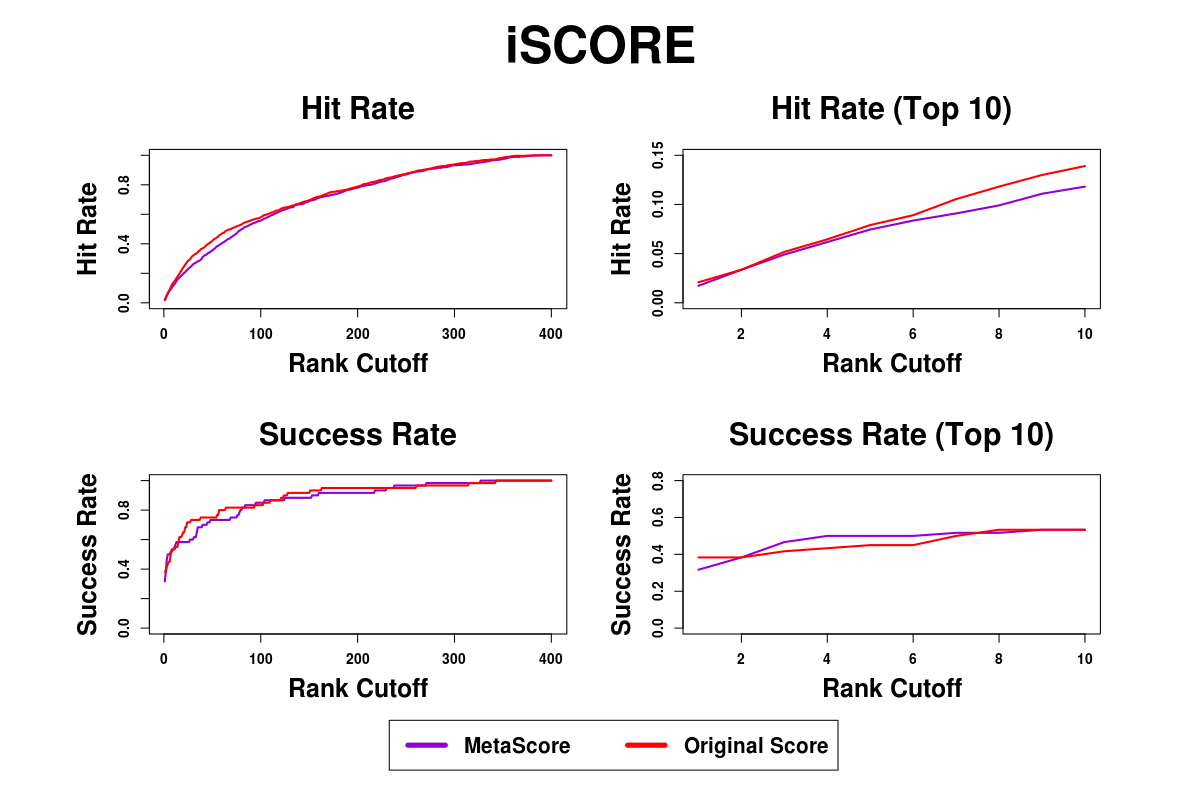
Supplementary Figure S14. Success rates and hit rates plotted against the top m conformations for a classical scoring method (SWARMDOCK), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



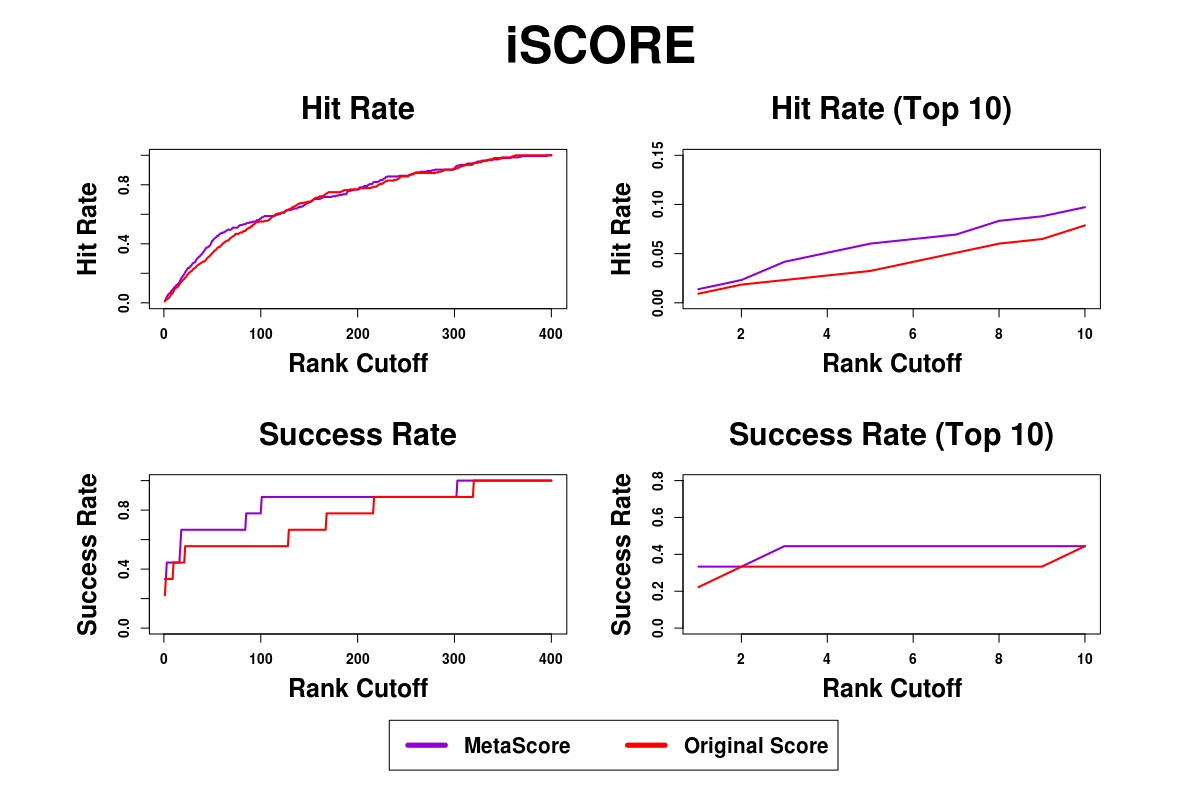
Supplementary Figure S15. Success rates and hit rates plotted against the top m conformations for a classical scoring method (TOBI), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



Supplementary Figure S16. Success rates and hit rates plotted against the top m conformations for a classical scoring method (TOBI), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



Supplementary Figure 17. Success rates and hit rates plotted against the top m conformations for a classical scoring method (iSCORE), machine learning-based method using RF, and the combined method of the two methods using BM4 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.



Supplementary Figure 18. Success rates and hit rates plotted against the top m conformations for a classical scoring method (iSCORE), machine learning-based method using RF, and the combined method of the two methods using BM5 decoy set. There are four panels. Top-left panel shows hit rates for conformations of top m ranging from 1 to 400; Top-right panel shows hit rates for conformations of top m ranging from 1 to 10; Bottom-left panel shows success rates for conformations of top m ranging from 1 to 400; Bottom-right panel shows success rates for conformations of top m ranging from 1 to 10.