Ranking Protein–Protein Docking Results Using Steered Molecular Dynamics and Potential of Mean Force Calculations

Laura J. Kingsley,[a] Juan Esquivel-Rodríguez,[b] Ying Yang,[a] Daisuke Kihara,[b,c] and Markus A. Lill*[a]

Crystallization of protein–protein complexes can often be problematic and therefore computational structural models are often relied on. Such models are often generated using protein–protein docking algorithms, where one of the main challenges is selecting which of several thousand potential predictions represents the most near-native complex. We have developed a novel technique that involves the use of steered molecular dynamics (sMD) and umbrella sampling to identify near-native complexes among protein–protein docking predictions. Using this technique, we have found a strong correlation between our predictions and the interface RMSD (iRMSD) in ten diverse test systems. On two of the systems, we investigated if the prediction results could be further improved using potential of mean force calculations. We demonstrated that a near-native (<2.0 Å iRMSD) structure could be identified in the top-1 ranked position for both systems. © 2016 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24412

Introduction

Despite many advances in modeling, docking, and scoring, predicting protein–protein interactions is still riddled with challenges.[1] Selecting the final model(s) is typically considered one of the most difficult steps and is often the most critical. Here, we describe a novel, physics-based, multistep approach to identify near-native protein–protein complex structures from a set of top-ranked poses.

In our method, summarized in Figure 1, steered molecular dynamics (MD) simulations are used to estimate the force required to separate the partners of docked protein–protein complexes by pulling one partner away from the other. The top-10 complexes (those with the highest force required for separation) are selected for more detailed investigation using umbrella sampling. The umbrella sampling simulations combined with the weighted histogram analysis method (WHAM) provide an estimate of the potential of mean force (PMF) of protein dissociation. The difference in the PMF between the bound (starting configuration) and unbound (ending configuration) state is the calculated free energy of complex dissociation.

Results and Discussion

A set of 10 diverse protein–protein complexes was used to evaluate our method (Table 1). From ~54,000 poses produced using ZDOCK,[12] a set of ~100 representative poses were selected. The selected poses were then evaluated using steered MD and five standard scoring functions, zrank1,[13] zrank2,[13] zdock,[12] irad,[14] and a custom potential based on van der Waals, electrostatics and knowledge-based terms,[15] herein referred to as "stats." The scoring functions were independently evaluated using the interface RMSD (iRMSD), a commonly used metric to evaluate protein–protein docking poses.[16]

A prediction was considered "good" if the iRMSD $\leq$ 2.0 Å, a pose was considered "acceptable" if the iRMSD $\leq$ 4.0 Å, and a prediction was considered "poor" if the iRMSD $>$ 4.0 Å.

Plots displaying the number of actives recovered versus the percentage of complexes screened is shown in Figure 2A (tabulated values are shown in Supporting Information Table 1). Steered MD produced the best results of any scoring scheme tested producing at least one good pose for 7/10 systems tested and an acceptable pose for 10/10 systems within the top-10 predictions. The irad and stats scoring functions performed similarly to sMD, both produced good predictions in 6/10 cases and acceptable poses were predicted for 10/10 and 8/10 systems, respectively. In terms of

[a] L. J. Kingsley, Y. Yang, M. A. Lill
Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, 575 Stadium Mall Drivem, West Lafayette, Indiana 47907
E-mail: mlill@purdue.edu
[b] J. Esquivel-Rodríguez, D. Kihara
Department of Computer Science, Purdue University, 305 North University Street, West Lafayette, Indiana 47907
[c] D. Kihara
Department of Biological Sciences, Purdue University, 249 South Martin Jischke Drive, West Lafayette, Indiana 47907

Contract grant sponsor: National Institute of General Medical Sciences of the National Institutes of Health; Contract grant number: R01GM092855 (to M.A.L. and L.J.K.); Contract grant sponsor: Graduate Assistantship in Areas of National Need (to L.J.K.); Contract grant sponsor: National Institute of General Medical Sciences of the National Institutes of Health; Contract grant number: R01GM097528 (D.K. and J.E.R.); Contract grant sponsor: National Science Foundation; Contract grant numbers: IIS1319551; DBI1262189; IOS1127027

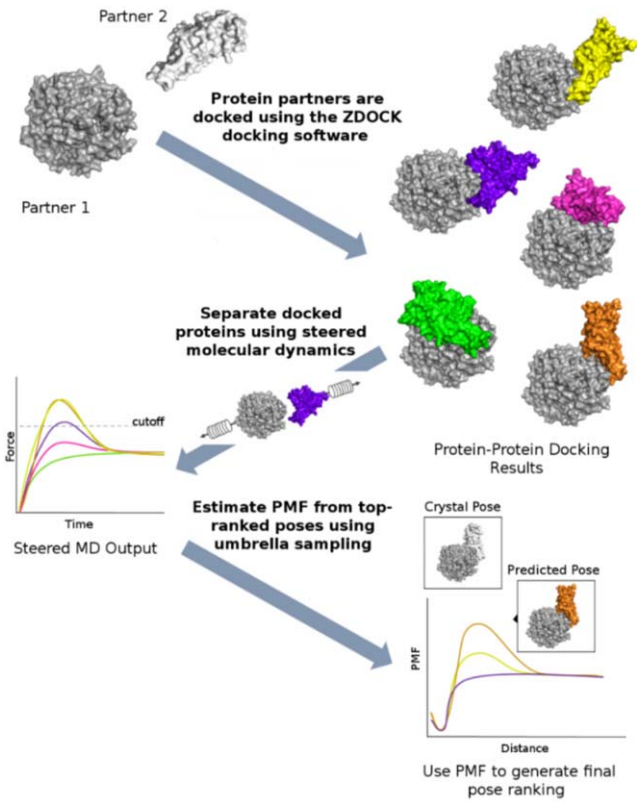

Figure 1. First, the target proteins (Partner 1 and Partner 2) are docked using ZDOCK, resulting in several thousand poses. The top-100 poses are selected based on the internal ZDOCK scoring function. These poses are then separated using sMD, and the force required for separation is computed. The top-10 poses from the force calculations are then reranked by PMF. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

enrichment, steered MD and stats performed the best. This is especially apparent in 1DJF, 1EZU, and 1UDI, where steered MD and stats significantly outperform the other scoring functions. Furthermore, both perform perfectly or nearly perfectly in 4/10 systems (1PPE, 1DJF, 1EAW, and 1UDI) as shown in Figure 2A (dotted line in inset indicates perfect prediction).

In general, most scoring functions that were tested produced a good or acceptable pose within the top-10 predictions, but oftentimes the top-10 predictions also included several poor poses. The inclusion of poor poses is less detrimental if they are ranked below good or acceptable poses, but this was not always the case. For instance, in 2HRK, only a single good pose (iRMSD: 1.98 Å) was identified in the top-10 predictions by steered MD and this pose was ranked tenth overall. Furthermore the three acceptable poses (iRMSDs: 2.49 Å, 2.13 Å, and 2.86 Å) were also ranked poorly (seventh, eighth, and ninth, respectively). In a blind prediction scenario, this type of result could easily lead to an unproductive final model. Thus, we attempted to further refine the top-10 predictions using umbrella sampling.

Ideally, in cases such as 2HRK, rescoring using PMF will result in the low iRMSD structures being reranked closer to the top. Alternatively, the 1VFB dataset contains several successful poses that are ranked near the top and 7 out of the top 10 poses are acceptable (iRMSD $\leq$ 4.0 Å). To ensure that reranking by PMF does not alter a successful screen, the top-10 predicted complexes from the 1VFB systems were also rescored using umbrella sampling.

Umbrella sampling is a technique where overlapping MD trajectories are utilized to produce an estimate of the potential of mean force (PMF) along a predefined reaction coordinate, in this case the distance describing the dissociation of the two protein units along the vector created by the centers of mass of each unit. These calculations, although computationally expensive, may provide a more accurate quantification of protein–protein interactions compared to steered MD alone. As a proof of concept, we selected the top-10 structures from 2HRK and 1VFB and used umbrella sampling to rerank these structures.

In both cases, reranking the top-10 poses using the PMF calculated by umbrella sampling improved the results. In 2HRK, the lowest iRMSD complex (1.98 Å) rose from a tenth place when ranked by steered MD alone to first when using PMF (Fig. 2B-left panel). Likewise in the 1VFB dataset, the 9.84Å structure fell from first ranked in the steered MD ranking down to one of the lowest ranked structures when ranked by PMF (Fig. 2B-right panel). In addition, in the 1VFB dataset all good poses (iRMSD $\leq$ 2.0Å) were ranked in the top-4 highest positions using PMF (Fig. 2B-right panel).

As a comparison, we also calculated the PMF of the crystal structures (shown in bolded black lines in Fig. 2B). In the case of 1VFB, the calculated PMF of crystal structure was in agreement with the low iRMSD ($\leq$2.0Å) structures. This finding suggests that the crystal structure and accurately predicted poses demonstrate similar behavior in the calculations. However, in the case of 2HRK, the PMF of the crystal structure was ~30 kJ/mol larger than the best ranked structure (1.98 Å). One possible explanation for this finding is that not all crystal contacts are adequately reproduced in the docked results; specifically the presence or absence of interfacial waters.

The hydration site analysis program WATsite,[17] was used to compare the number of hydration sites in or immediately adjacent to the interface of the protein–protein complex based on the x-ray conformation and pose with lowest iRMSD for 2HRK and 1VFB. A comparison between the 2HRK crystal structure and the equilibrated lowest iRMSD pose (1.98Å) revealed that not all contact-mediating hydration sites in the x-ray structure of the protein–protein interface were conserved in the low iRMSD pose (Supporting Information Fig. 1). Whereas 16 contact-mediating hydration sites were identified in the protein interface of the x-ray structure only 12 were found in the low iRMSD pose (Supporting Information Table 2). Repeating this analysis for 1VFB revealed that the same number of contact-mediating waters were identified in low iRMSD and x-ray structure supporting the observation that the PMF of the crystal structure and good poses were approximately equal. Thus, important water-mediated interactions are lost for the low iRMSD pose resulting in reduced complex stability compared to the x-ray structure of the complex.

To the best of our knowledge, this is the first time that steered MD and PMF calculations have been used to evaluate

Table 1. Summary of systems used to evaluate scoring functions.

System (PDB ID)	Residues and chains[a]	IRMSD range of docking results[b]	Number of poses tested	Reference
Ubiquitin ligase and ubiquitin (2OOB)	Total: 113 Pull: 42 Stationary: 71	1.62Å–9.97Å	96	2
Trypsin and CMTI-I peptide inhbitor (1PPE)	Total: 274 Pull: 29 Stationary: 245	0.65Å–9.67Å	100	3
Antibody and antigen (1VFB)	**Total: 352 Pull: 129 Stationary: 107 & 116**	**1.32Å–9.98Å**	**100**	4
aminoacyl-tRNA synthetase and tRNA aminoacylation cofactor Arc1p (2HRK)	**Total: 282 Pull: 102 Stationary: 180**	**1.96Å–9.94Å**	**98**	5
Ribonuclease A and a peptide inhibitor (1DFJ)	Total: 579 Pull: 124 Stationary: 455	1.19Å–9.81Å	99	6
Ferredoxin-NADP Reductase and ferredoxin (1EWY)	Total: 395 Pull: 98 Stationary:297	1.43 Å–9.98 Å	100	7
Matriptase and aprotinin (1EAW)	Total: 299 Pull:58 Stationary: 241	0.74 Å–9.89 Å	100	8
SARS- receptor binding domain and receptor (2AJF)	Total: 777 Pull:180 Stationary:597	1.63Å–10.0 Å	98	9
Ecotin and trypsin (1EZU)	Total:365 Pull:142 Stationary:223	1.37 Å–9.99 Å	97	10
Uracil-DNA Glycosylase and its protein inhibitor (1UDI)	Total:207 Pull:83 Stationary: 124	1.15 Å–9.89 Å	100	11

Those shown in bold were used in the PMF calculations.
[a] "Pull" refers to the length of the chain that was pulled during the steered MD simulation, "Stationary" refers to the chain that was restrained during the steered MD simulation.
[b] Results were prefiltered to remove any poses above 10 Å.

protein–protein docking poses. Furthermore the use of explicit solvent MD simulations allows for the incorporation of waters into the interface which are accounted for in our procedure, a feature that is very rarely included in traditional docking and scoring methods.

Despite the limited number of test cases, we believe the proposed stepwise method to be a promising approach, although there are some important considerations about the limitations of this method. Importantly, the time required for calculation of PMF profiles could present a significant limitation. In practice, we suggest that a more rapid scoring function might be used as a prefilter, prior to implementing the more computationally demanding umbrella sampling (both the stats and the irad scoring functions performed exceptionally well in our hands). In addition to the time required to calculate the PMF profiles, the calculations are sensitive to the reproducibility of the interactions in the interface. In an ideal case, protein partners would not change conformation on binding and interface interactions would be strictly between partners (i.e., not mediated by water or other cofactors). Caution should be exercised in cases where drastic conformational changes are thought to occur or in cases where protein interactions are extensively mediated by other molecules. Methods such as principal component analysis (PCA), may be employed to determine the best vector for sMD simulations in cases of intricate interfaces or where significant conformational change is anticipated.

In summary, the use of steered MD and umbrella sampling in ranking protein–protein docking conformations represents a novel approach in this field and has been found to be successful in the test cases presented here and elsewhere.[18] While there are some limitations to this approach, notably the computational cost, we believe that this approach may prove useful in a range of systems and be a complimentary approach to the currently used scoring functions for protein–protein docking.

Methods

Only a general outline of the procedure and tools used has been included here, a detailed methods section has been included in the Supporting Information.

All protein systems were docked using the ZDOCK algorithm producing ~54,000 conformations. From these 100 representative conformations were selected for steered MD. Gromacs 4.6.1 was used to prepare and equilibrate each system prior to sMD. From the sMD simulations, the total force was computed as the difference between the lowest and highest recorded force for each simulation. Umbrella sampling was performed on the top-

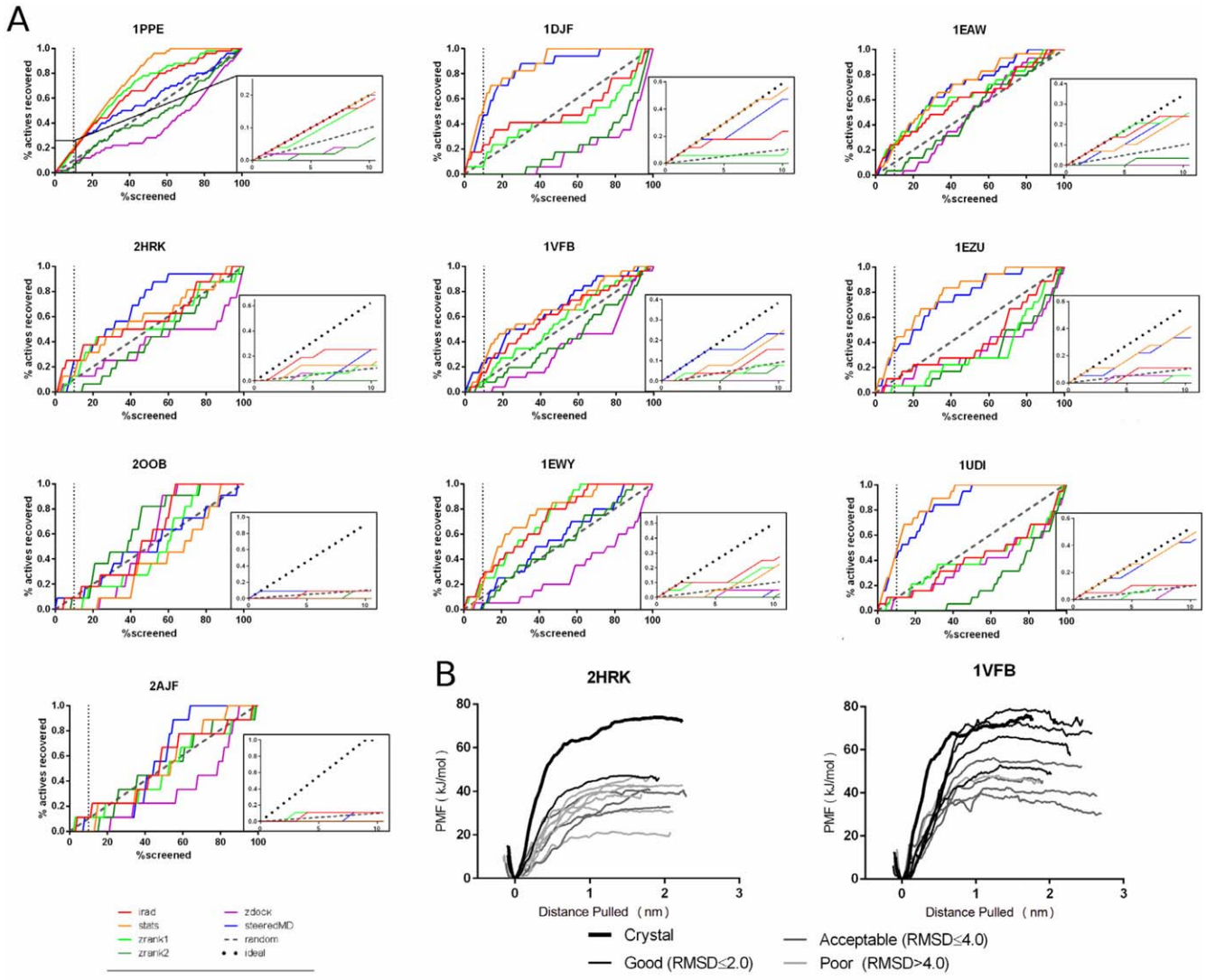

Figure 2. A) Plot describing the recovery of acceptable poses (iRMSD $\leq$ 4.0) versus number of poses screened for each scoring function. Insets show an expanded plot of the initial 10% of complexes screened. **B)** PMF curves from umbrella sampling in 2HRK (left) and 1VFB (right). The crystal structures are shown in the bolded black line. The poses have been color coded from dark to light, where the darkest lines represent low iRMSD structures ($\leq$2.0 Å), the slightly lighter lines represent acceptable structures (IRMSD $\leq$4.0Å), and the lightest lines represent unacceptable structures (iRMSD $>$ 4.0Å). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

10 structures from 1VFB and 2HRK and the g_wham program from Gromacs was used to estimate the PMF using the sampled windows. WATsite was used in the interfacial water analysis.

Keywords: protein–protein interaction · ZDOCK · steered molecular dynamics · potential of mean force · umbrella sampling

How to cite this article: L. J. Kingsley, J. Esquivel-Rodríguez, Y. Yang, D. Kihara, M. A. Lill. *J. Comput. Chem.* **2016**, *37*, 1861–1865. DOI: 10.1002/jcc.24412

Additional Supporting Information may be found in the online version of this article.

[1] I. Moal, M. Torchala, P. Bates, J. Fernandez-Recio, *BMC Bioinformatics* **2013**, *14*, 286.
[2] P. Peschard, G. Kozlov, T. Lin, I. A. Mirza, A. M. Berghuis, S. Lipkowitz, M. Park, K. Gehring, *Mol. Cell.* **2007**, *27*, 474.
[3] W. Bode, H. J. Greyling, R. Huber, J. Otlewski, T. Wilusz, *FEBS Lett.* **1989**, *242*, 285.
[4] T. N. Bhat, G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, W. Dall'Acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, R. J. Poljak, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1089.
[5] H. Simader, M. Hothorn, C. Köhler, J. Basquin, G. Simos, D. Suck, *Nucleic Acids Res.* **2006**, *34*, 3968.
[6] B. Kobe, J. Deisenhofer, *Nature* **1995**, *374*, 183.
[7] R. Morales, G. Kachalova, F. Vellieux, M. H. Charon, M. Frey, *Acta Crystallogr. Sect. D* **2000**, *56*, 1408.
[8] R. Friedrich, P. Fuentes-Prior, E. Ong, G. Coombs, M. Hunter, R. Oehler, D. Pierson, R. Gonzalez, R. Huber, W. Bode, E. L. Madison, *J. Biol. Chem.* **2002**, *277*, 2160.
[9] F. Li, W. Li, M. Farzan, S. C. Harrison, *Science* **2005**, *309*, 1864.
[10] S. A. Gillmor, T. Takeuchi, S. Q. Yang, C. S. Craik, R. J. Fletterick, *J. Mol. Biol.* **2000**, *299*, 993.
[11] R. Savva, L. H. Pearl, *Nat. Struct. Biol.* **1994**, *2*, 752.
[12] R. Chen, L. Li, Z. Weng, *Proteins* **2003**, *52*, 80.
[13] B. Pierce, Z. Weng, *Proteins* **2007**, *67*, 1078.

[14] T. Vreven, H. Hwang, Z. Weng, *Protein Sci.* **2011**, *20*, 1576.
[15] J. Esquivel-Rodríguez, Y. D. Yang, D. Kihara, *Proteins* **2012**, *80*, 1818.
[16] (a) M. F. Lensink, R. Méndez, S. J. Wodak, *Proteins* **2007**, *69*, 704; (b) Pons, C. Grosdidier, S. Solernou, A. Pérez-Cano, L. Fernández-Recio, J. **2010**, *78*, 95.
[17] B. Hu, M. A. Lill, *J. Comput. Chem.* **2014**, *35*, 1255.
[18] (a) A. G. Meyer, S. L. Sawyer, A. D. Ellington, C. O. Wilke, *PeerJ PrePrints* **2014**, *2*, e138v3; (b) L. S. L. Cheung, D. J. Shea, N. Nicholes, A. Date, M. Ostermeier, K. Konstantopoulos, *K. Scie. Rep.* **2015**, *5*, 8247; (c) W. Allen, M. Wiley, K. Myles, Z. Adelman, D. Bevan, *J. Mol. Model.* **2014**, *20*, 1; (d) R. A. Rodriguez, L. Yu, L. Y. Chen, *J. Chem. Theory Comput.* **2015**, *11*, 4427.

Received: 2 April 2016
Revised: 2 May 2016
Accepted: 6 May 2016
Published online on 27 May 2016