Evaluation of CONSRANK-Like Scoring Functions for Rescoring Ensembles of Protein–Protein Docking Poses

Guillaume Launay[1†], Masahito Ohue[2*†], Julia Prieto Santero[1], Yuri Matsuzaki[3], Cécile Hilpert[1], Nobuyuki Uchikoga[4], Takanori Hayashi[2] and Juliette Martin[1*]

[1] CNRS, UMR 5086 Molecular Microbiology and Structural Biochemistry, University of Lyon, Lyon, France, [2] Department of Computer Science, School of Computing, Tokyo Institute of Technology, Tokyo, Japan, [3] Tokyo Tech Academy for Leadership, Tokyo Institute of Technology, Tokyo, Japan, [4] Department of Network Design, School of Interdisciplinary Mathematical Sciences, Meiji University, Tokyo, Japan

Scoring is a challenging step in protein–protein docking, where typically thousands of solutions are generated. In this study, we ought to investigate the contribution of consensus-rescoring, as introduced by Oliva et al. (2013) with the CONSRANK method, where the set of solutions is used to build statistics in order to identify recurrent solutions. We explore several ways to perform consensus-based rescoring on the ZDOCK decoy set for Benchmark 4. We show that the information of the interface size is critical for successful rescoring in this context, but that consensus rescoring in itself performs less well than traditional physics-based evaluation. The results of physics-based and consensus-based rescoring are partially overlapping, supporting the use of a combination of these approaches.

Keywords: protein–protein interaction, docking, scoring, prediction, interface

INTRODUCTION

Protein–protein docking aims at predicting the structure of a complex starting from the structures of isolated components (Melquiond et al., 2012; Vakser, 2014). The CAPRI community-wide initiative allows a blind assessment of the participant methods on common data sets and evaluation criteria, offering an updated view of progress in the field since 2001 (Lensink et al., 2007, 2017; Lensink and Wodak, 2010). Protein–protein docking methods typically generate thousands of potential solutions for a particular complex. Scoring the models to discriminate near-native solutions is a known bottleneck of docking methods (Moal et al., 2013a,b; Malhotra et al., 2015). Most scoring functions are physics-based, attempting to capture the determinants underlying the stability of protein–protein complexes, e.g., shape complementary, electrostatics and desolvation potential (Dominguez et al., 2003; Cheng et al., 2007; Pierce and Weng, 2007, 2008; Moal and Bates, 2010; Ritchie and Venkatraman, 2010; Ohue et al., 2014). Knowledge-based functions, on the other hand, aim at taking advantage of the information from available structures, *via* pair potentials (Lu et al., 2003; Huang and Zou, 2008; Mezei, 2017), or multibody potentials (Khashan et al., 2012). Docking methods often use scoring functions that combine physical terms with knowledge-based

terms (Kozakov et al., 2006; Liang et al., 2009; Feliu et al., 2011; Vreven et al., 2011). More recently, evolutionary information has been successfully used for scoring (Andreani et al., 2013; Yu et al., 2017).

Another approach consists in relying on the recurrences observed in the set of solutions, i.e., consensus-based scoring. Consensus-based scoring functions seek to identify solutions with features that are the most frequent in the solution set, independently of any physics-based or evolutionary evaluation. The CONSRANK scoring function, proposed by Oliva et al. (2013, 2015), Vangone et al. (2013), and Chermak et al. (2015, 2016) has shown very good results, based on the conservation of interface contacts.

In this study, we compare several CONSRANK-like scoring functions on large sets of docking poses generated by ZDOCK, including CONSRANK. We then explore how to combine consensus-based rescoring with the native scoring function of ZDOCK.

METHODS

Docking Decoy Set

The ZDOCK3.0.2 decoy set (Pierce et al., 2011) (6 degree sampling, fixed receptor format) for Benchmark4 (Hwang et al., 2010a) was retrieved from https://zlab.umassmed.edu/zdock/decoys.shtml. This data set encompasses 176 protein–protein complexes, with 54,000 docking poses for each complex. For each pose, the interface Cα RMSD, with respect to the bound structure, is given. A near-native docking hit is defined as a prediction with interface Cα RMSD < 2.5 Å.

Consensus-Based Rescoring Schemes

Following the CONSRANK method (Oliva et al., 2013; Chermak et al., 2015), docking poses are rescored using the frequencies of interface contacts in the set of docking poses. Interface contacts are defined using a distance cut-off of 5 Å between the heavy atoms of receptor and ligand proteins.

For each contact C_{ij} between residue i from receptor and residue j from ligand the relative frequency in the decoy set is defined by:

$$S\left(C_{ij}\right) = \frac{F\left(C_{ij}\right)}{N} \in [0, 1] \qquad (1)$$

where $F(C_{ij})$ denotes the frequency of C_{ij} at the protein–protein interface in the set of N decoys. These relative frequencies are then averaged, i.e., normalized by the interface size, to compute the CONSRANK score of each pose P:

$$CONSRANK\_score\,(P) = \frac{\sum_{C_{ij} \in P} S(C_{ij})}{N_{cont}(P)}, \qquad (2)$$

where $N_{cont}(P)$ denotes the number of interface contacts in docking pose P.

Variations of CONSRANK Scores

First, we considered the un-normalized version of CONSRANK scores (Oliva et al., 2013), denoted as CONSRANK_U, where the relative frequencies of interface contacts are only summed, and not averaged:

$$CONSRANK\_U(P) = \sum_{C_{ij} \in P} S(C_{ij}). \qquad (3)$$

Then, we implemented two other variations, by replacing relative frequencies of contacts $S(C_{ij})$ by relative frequencies of residues:

$$S(R_i) = \frac{F(R_i)}{N} \in [0, 1] \qquad (4)$$

where $F(R_i)$ denotes the frequency of residue i at the protein–protein interface (distance between heavy atoms lower than 5 Å) in the set of N decoys. The two related scores are respectively defined by:

$$Residue\_Average\,(P) = \frac{\sum_{R_i \in P} S(R_i)}{N_{res}(P)}, \qquad (5)$$

$$Residue\_Sum\,(P) = \sum_{R_i \in P} S(R_i). \qquad (6)$$

where $N_{res}(P)$ denotes the number of interface residues in pose P. Here, interface residues are simply those involved in contacts at the interface.

Note that is it possible to compute the contact and residue frequencies (Eqs 1 and 4) on a given set of docking poses and then to evaluate another set of docking poses (with Eqs 2, 3, 5, and 6).

Clustering

We implemented the BSAS clustering procedure (Basic Sequential Algorithmic Scheme) (Koutroumbas and Theodoridis, 2008; Jiménez-García et al., 2018) to reduce the structural redundancy of docking poses. The principle of BSAS is the following. Docking poses are ranked according to a score in decreasing order. The pose with the highest score initiates the first clusters. The other poses are sequentially compared to already clustered poses: they are included in a cluster if they are within a given cut-off of cluster members, otherwise they initiate a new cluster. At the end of the process, the pose with the highest score in each cluster is the representative of each cluster. In order to allow a fast clustering process, we do not compute the RMSD between ligand atoms. Instead we use a distance cut-off between the centers of mass of the ligands, here set to 8 Å.

Evaluation

The top 2,000 solutions according to the ZDOCK native scoring function were rescored using the rescoring schemes detailed below. We monitored the presence of near-native docking hits (interface Cα RMSD < 2.5 Å) in the top 10 solutions after re-ranking. Each protein–protein complex with a near-native docking hit in the top 10 solutions is counted as a success.

Implementation

The consensus-based rescoring functions are implemented in python code accessible on GitHub, which operates directly on

ZDOCK output files, and allows to treat rapidly thousands of docking poses (typically a few seconds for 2,000 poses, up to 1 min for 54,000 poses). In addition, the code allows to compute statistics on a given set of poses and re-score another of structures (see "Results" section). All the scripts necessary to reproduce the results shown in this article are available at: https://github.com/MMSB-MOBI/CHOKO.

RESULTS

In this study, we compare four consensus-based scores to identify the near-native solutions among the ensembles generated by ZDOCK. The traditional CONSRANK score (Oliva et al., 2013) is considered, as well as its un-normalized version, and two variations that consider residue statistics instead of contact statistics. We first evaluate each consensus score separately. Then, we combine these results with the native ZDOCK physics-based scoring function. Finally, we add a clustering step, to reduce structural redundancy and further improve the results.
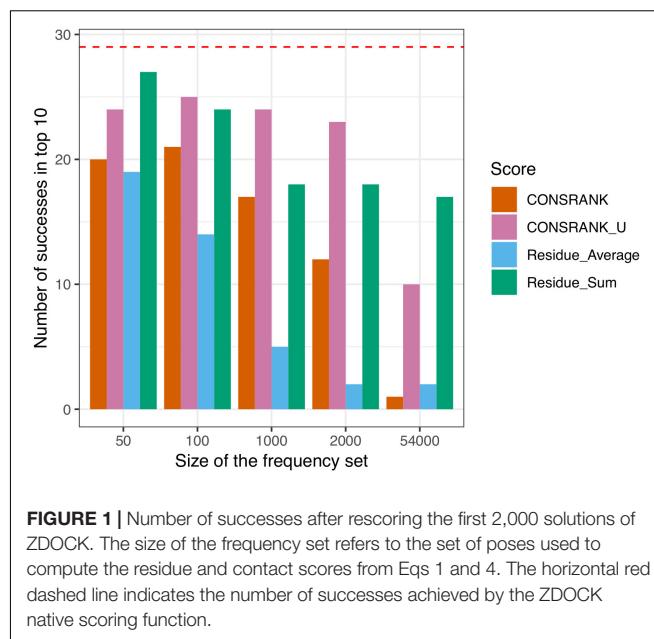
Quality of Decoys

In the initial data set of 176 protein–protein complexes, ZDOCK was able to generate at least one near-native docking hit (interface Cα RMSD < 2.5 Å) in the first top 2,000 solutions for 90 protein–protein complexes. These 90 protein–protein complexes thus constitute our reference data set for the rest of the study. We explore if and how consensus-based rescoring is efficient at scoring the decoys of these 90 protein–protein complexes.

Evaluation of Different Consensus-Based Rescoring Functions

First, we compare the four versions of consensus-based rescoring functions: either contact-based [following the CONSRANK (Oliva et al., 2013) scheme] or residue-based, with or without interface size normalization. We estimate the performance by counting the number of successes, i.e., number of complexes with at least one near-native hit (interface Cα RMSD < 2.5 Å) in the first 10 solutions after rescoring. We also tested the effect of varying the subset of docking poses used to compute the contact and residue frequencies (Eqs 1 and 4): we used either the first 50, 100, 1,000 or 2,000 first poses provided by ZDOCK, or the full set of 54,000 poses, referred as the frequency set. In any case, we rescored the first 2,000 poses provided by ZDOCK.

The results of this evaluation are shown in **Figure 1**. We can see that un-normalized rescoring functions (CONSRANK_U and Residue_Sum) constantly outperform the normalized rescoring functions (CONSRANK, Residue_Average). The size of the subset used to compute contact and residue frequencies (Eqs 1 and 4) has a major influence on the number of successes. Indeed, ZDOCK solutions are ranked by the ZDOCK native scoring function; hence the top of the list is, in many cases, enriched in near-native docking hits. Estimating contact and residue scores on a reduced subset of poses at the top of the list is logically more efficient. On the contrary, estimating contact and residue scores from the full list leads to a loss of information, and worsens the prediction. When frequencies were

FIGURE 1 | Number of successes after rescoring the first 2,000 solutions of ZDOCK. The size of the frequency set refers to the set of poses used to compute the residue and contact scores from Eqs 1 and 4. The horizontal red dashed line indicates the number of successes achieved by the ZDOCK native scoring function.

estimated on the 54,000 poses, the number of successes was 1 for CONSRANK, 10 for CONSRANK_U, 2 for Residue_Average, and 17 for Residue_Sum. In the best settings tested here, estimating the scores on the first 50 solutions to rescore the first 2,000 solutions allows to reach a number of successes equal to 27 with the Residue_Sum scoring function, *versus* 20 for the CONSRANK scheme. It is thus possible to rescore large sets of docking poses using consensus-based scoring functions, with better performance than the commonly used CONSRANK scheme.

Combination With ZDOCK Native Scoring Function

In this section, we explore how to combine rescoring functions with the native scoring function of ZDOCK. Out of the 90 protein–protein complexes with at least one near-native docking hit in the top 2,000 solutions, the ZDOCK native scoring function identifies 29 successes, i.e., 29 complexes with at least one near-native docking hit in the top 10, see **Figure 1**. This is indeed better than the four consensus-based rescoring functions tested here. One could wonder if it is then possible to improve the initial prediction of ZDOCK using rescoring.

We first analyzed the overlap of the successful cases by ZDOCK and each rescoring function, and found that many successful cases achieved by rescoring are well predicted by the ZDOCK scoring function, see **Supplementary Figure S1**. For example, when estimating scores on the first 50 solutions for the rescoring (**Supplementary Figure S1A**), all the successes identified by the normalized rescoring functions CONSRANK and Residue_Average are included in the successes identified by ZDOCK. Un-normalized scoring functions are able to identify 1 case not included in the ZDOCK successes for CONSRANK_U, and 4 for Residue_Sum.

We then tested a combination of ZDOCK poses and rescored poses by combining the first N_1 poses of ZDOCK with the first N_2 poses after rescoring, with $N_1 + N_2 = 10$, and no redundancy. Again, we vary the subset of docking poses used to compute the contact and residue frequencies with Eqs 1 and 4 (frequency set = top 50, 100, 1,000 or 2,000 poses) and in any case, we rescore the first 2,000 solutions provided by ZDOCK using Eqs 2, 3, 5, and 6. We estimate the performance by counting the number of successes, i.e., number of complexes with at least one near-native docking hit (interface Cα RMSD < 2.5 Å) in the first 10 solutions.

The results of this evaluation are shown in **Supplementary Figure S2**. Regardless of the size of the frequency set, the best combination is always obtained with the Residue_Sum scoring function. Combining the first six ZDOCK poses with the first four Residue_Sum rescored poses, and estimating the frequencies on the full set of 2,000 poses (bottom right panel in **Supplementary Figure S2**) allows to reach a number of successes equal to 32, compared to 18 with Residue_Sum alone and 29 with ZDOCK alone. This suggests the possibility to marginally improve the native results of ZDOCK by a simple combination of poses. It is interesting to note that, in this situation, the information about residues is more efficient in rescoring than the information about pairwise contacts.
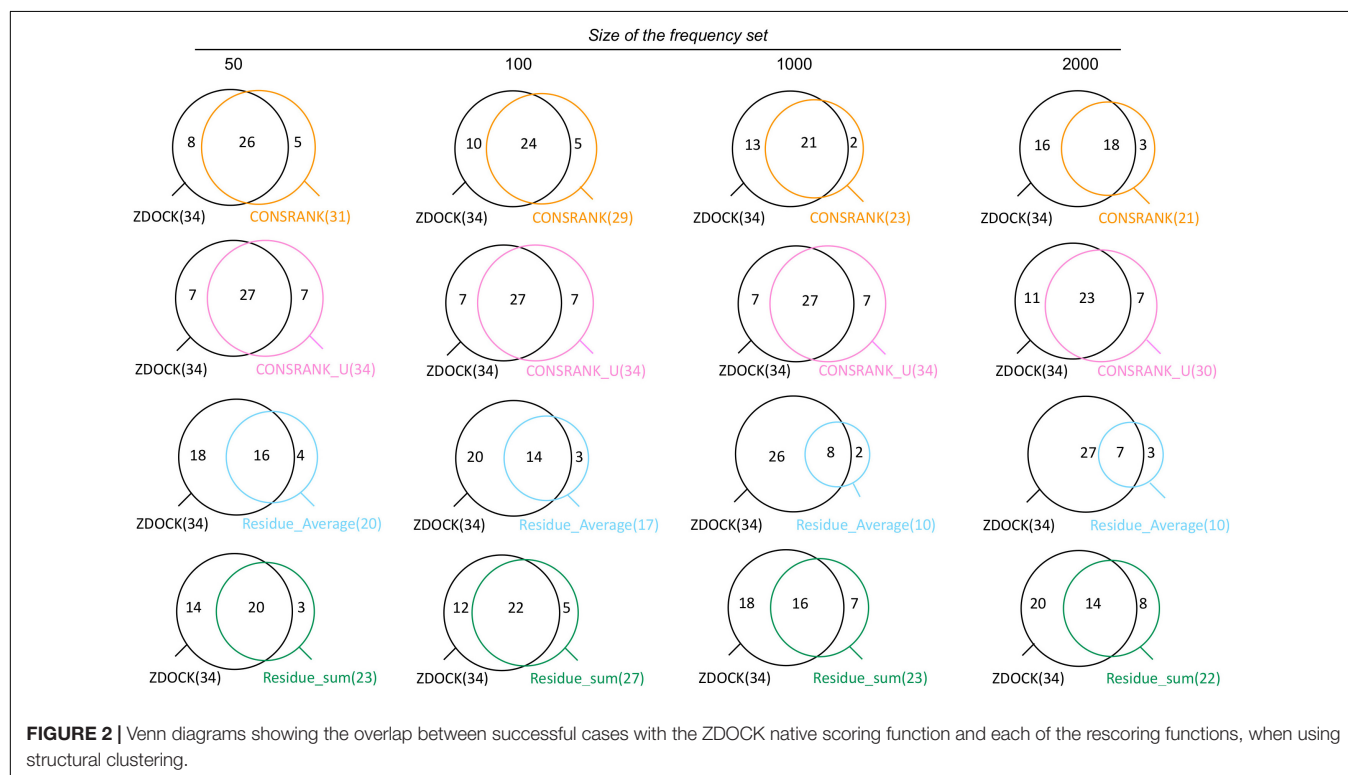
Combining Clusters

Clustering is classically used to improve the performance, by reducing the structural redundancy of docking solutions (Kozakov et al., 2005; Hwang et al., 2010b; Koukos et al., 2020). Here, we used the BSAS clustering algorithm, which takes into account the scores, to cluster poses by their ligand center of mass.

When applied to ZDOCK results, independently of rescoring, we obtained an improvement in terms of number of successes: 34 successes instead of 29, reflecting a structural redundancy of the ZDOCK set. We first analyzed the overlap between ZDOCK results and the results of each rescoring function when using structural clustering, see **Figure 2**.

As shown in **Figure 2**, the results of ZDOCK and rescoring functions are partially overlapping also after structural clustering. The rescoring functions are able to identify between 2 and 8 additional successful cases, with more additional cases brought by un-normalized scoring functions CONSRANK_U and Residue_Sum. This means that a perfect combination of ZDOCK and rescoring with no loss would reach a number of successes equal to 42.

We then explore how to combine ZDOCK results and rescoring results. We have tested a combination of clusters. On the one hand, we computed clusters from the poses ranked by their initial ZDOCK scores. On the other hand, we computed clusters from poses reordered after consensus rescoring. We then combine the representative poses of the first N_1 ZDOCK clusters, with the representative poses of the first N_2 poses after rescoring, with $N_1 + N_2 = 10$. We estimate the performance by counting the number of successes, i.e., number of complexes with at least one near-native hit (interface Cα RMSD < 2.5 Å) in the first 10 solutions.

The results of this evaluation are shown in **Figure 3**. In agreement with the Venn diagram analysis, the best combination is obtained using an un-normalized rescoring function, CONSRANK_U. It constantly outperforms CONSRANK, regardless of the frequency set. When using the first 1,000
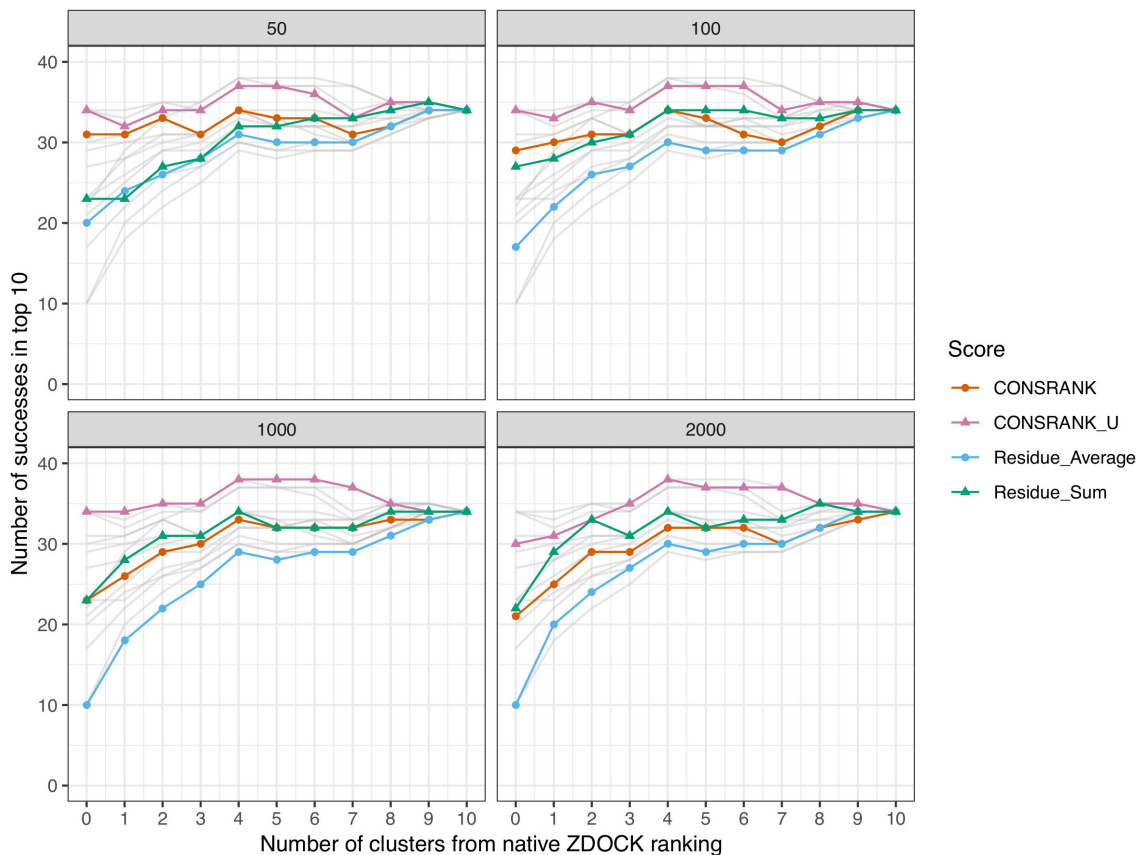

FIGURE 2 | Venn diagrams showing the overlap between successful cases with the ZDOCK native scoring function and each of the rescoring functions, when using structural clustering.

FIGURE 3 | Number of successes after combination of clusters with the ZDOCK native scoring function. Each panel corresponds to different frequency sets, i.e., sets of poses used to compute the residue and contact scores from Eqs 1 and 4. In any cases, the first 2,000 solutions of ZDOCK are rescored. Gray lines represent data from other panels for comparison.
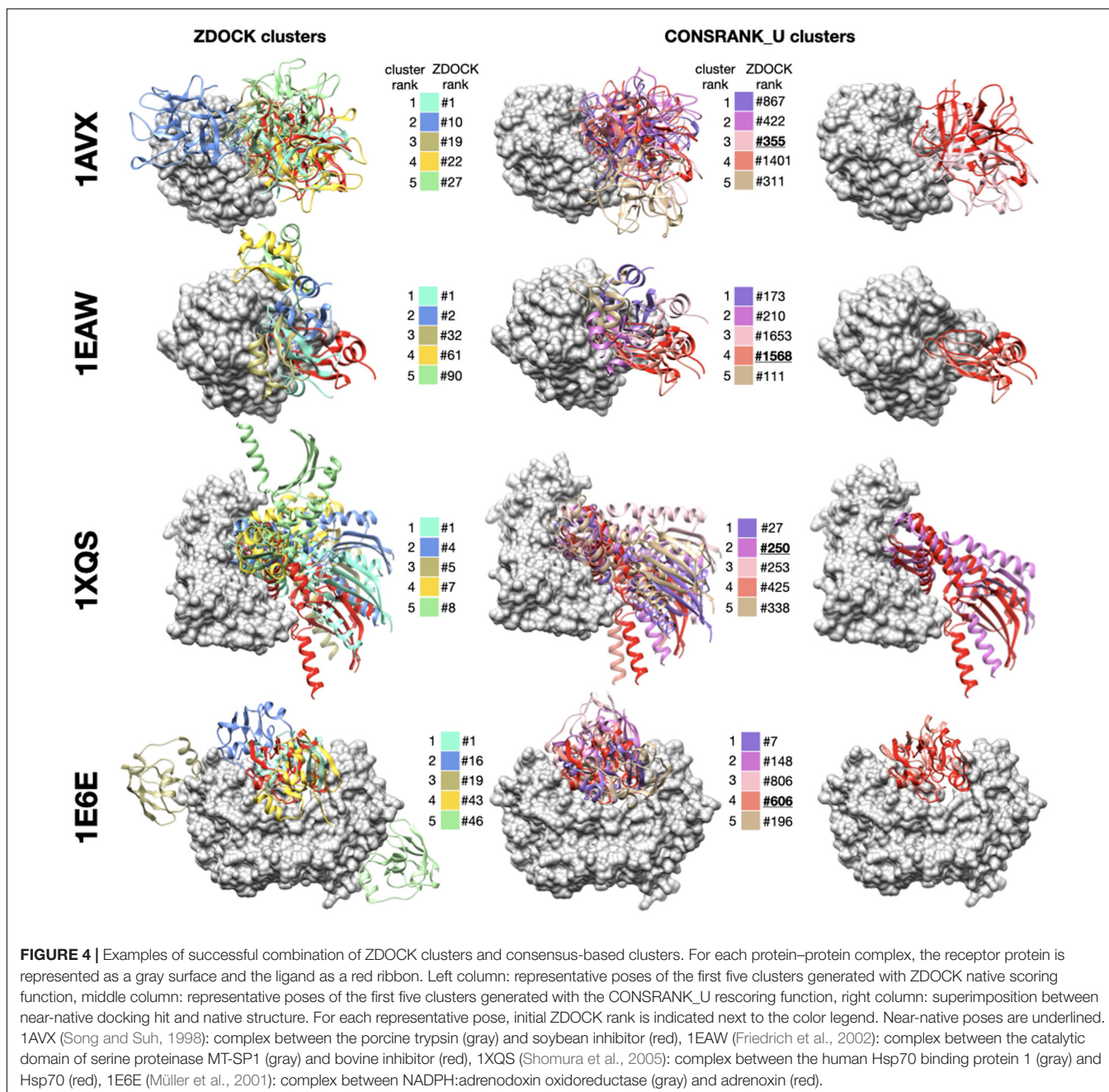
poses to estimate the frequencies (bottom left panel in **Figure 3**), the combination of five ZDOCK clusters and five CONSRANK_U clusters achieves a number of successes equal to 38, compared to 34 with ZDOCK alone and 34 with CONSRANK_U alone. This result suggests that CONSRANK-like rescoring could be used together with physics-based evaluation. Contrary to what was observed in simple pose combination (**Figure 2**), the most efficient rescoring scheme when dealing with clusters is based on contact frequencies, not residue frequencies. It seems that, after structural clustering, the information of pairwise contacts, which is more precise than residues, becomes more useful in discrimination.

Illustrative Examples

To complete this study, in this section, we present examples to illustrate the asset of consensus rescoring when used in combination with the native ZDOCK scoring function. We used the results generated using the CONSRANK_U function, with frequencies estimated on the first 1,000 poses, and combined five ZDOCK clusters and five consensus clusters. As explained in the previous section, this setting allows to reach 38 successes. We present four examples from the ZDOCK decoy set where

the use of consensus rescoring is critical in **Figure 4**. For all these protein–protein complexes, no near-native docking hit is observed in the first 10 clusters of ZDOCK (or in the first 10 poses). The use of CONSRANK_U rescoring in conjunction with clustering allows the identification of near-native docking poses in the top 10. In every case, these near-native poses do not belong to the top of the ZDOCK initial list: they are ranked 355 for 1AVX, 1568 for 1EAW, 250 for 1XQS, and 606 for 1E6E. These examples highlight the usefulness of consensus-based rescoring to rescue poses with poor initial ranks.

For the 38 successful complexes in this experiment, we systematically computed the number of near-native poses coming from ZDOCK clusters and the number of near-native poses coming from CONSRANK_U clusters. Detailed results are provided in **Supplementary Table S1** for the 38 complexes with at least one near-native pose in the top ten. In eight cases, the near-native poses were present only in ZDOCK clusters, in 10 cases, the near-native poses were present only in CONSRANK_U clusters and in the 20 remaining cases, near-native poses were present in both ZDOCK and CONSRANK_U clusters. We observed no significant bias in terms of functional category or interface size between

FIGURE 4 | Examples of successful combination of ZDOCK clusters and consensus-based clusters. For each protein–protein complex, the receptor protein is represented as a gray surface and the ligand as a red ribbon. Left column: representative poses of the first five clusters generated with ZDOCK native scoring function, middle column: representative poses of the first five clusters generated with the CONSRANK_U rescoring function, right column: superimposition between near-native docking hit and native structure. For each representative pose, initial ZDOCK rank is indicated next to the color legend. Near-native poses are underlined. 1AVX (Song and Suh, 1998): complex between the porcine trypsin (gray) and soybean inhibitor (red), 1EAW (Friedrich et al., 2002): complex between the catalytic domain of serine proteinase MT-SP1 (gray) and bovine inhibitor (red), 1XQS (Shomura et al., 2005): complex between the human Hsp70 binding protein 1 (gray) and Hsp70 (red), 1E6E (Müller et al., 2001): complex between NADPH:adrenodoxin oxidoreductase (gray) and adrenoxin (red).

protein complexes that were successful only with ZDOCK or only with CONSRANK_U. This indicates that ZDOCK and CONSRANK_U results are only partially overlapping, justifying the need to combine them.

CONCLUSION

We have implemented four variants of consensus-based rescoring functions: the CONSRANK score, the CONSRANK un-normalized score, and their equivalents based on residue frequencies and tested them on the rescoring of large sets of

docking poses of the ZDOCK benchmark. In this context, un-normalized scores that do take into account the size of the interfaces are in general more efficient than normalized scores. When used alone, consensus-based scoring functions degraded the initial performance of the physics-based ZDOCK scoring function. However, when both physics-based and consensus-based scoring functions were used in combination, we observed a marginal improvement. This calls for calibration when using consensus-based scoring functions to re-rank large sets of docking decoys, since they are, by definition, highly dependent on the docking decoy population.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/
Supplementary Material.

AUTHOR CONTRIBUTIONS

GL, JS, CH, and JM contributed to software. GL, YM, and
NU contributed to investigation. MO and JM contributed
to the methodology. TH contributed to the data curation.
JM conceptualized and wrote the manuscript. All authors
contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found
online at: https://www.frontiersin.org/articles/10.3389/fmolb.
2020.559005/full#supplementary-material

REFERENCES

Andreani, J., Faure, G., and Guerois, R. (2013). InterEvScore: a novel
coarse-grained interface scoring function using a multi-body statistical
potential coupled to evolution. *Bioinformatics* 29, 1742-1749. doi: 10.1093/
bioinformatics/btt260

Cheng, T. M.-K., Blundell, T. L., and Fernandez-Recio, J. (2007). pyDock:
electrostatics and desolvation for effective scoring of rigid-body
protein-protein docking. *Proteins* 68, 503–515. doi: 10.1002/prot.
21419

Chermak, E., Donato, R. D., Lensink, M. F., Petta, A., Serra, L., Scarano, V., et al.
(2016). Introducing a clustering step in a consensus approach for the scoring of
protein-protein docking models. *PLoS One* 11:e0166460. doi: 10.1371/journal.
pone.0166460

Chermak, E., Petta, A., Serra, L., Vangone, A., Scarano, V., Cavallo, L., et al.
(2015). CONSRANK: a server for the analysis, comparison and ranking of
docking models based on inter-residue contacts. *Bioinformatics* 31, 1481–1483.
doi: 10.1093/bioinformatics/btu837

Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: a protein-
protein docking approach based on biochemical or biophysical information.
J. Am. Chem. Soc. 125, 1731–1737. doi: 10.1021/ja026939x

Feliu, E., Aloy, P., and Oliva, B. (2011). On the analysis of protein-protein
interactions via knowledge-based potentials for the prediction of protein-
protein docking. *Protein Sci.* 20, 529–541. doi: 10.1002/pro.585

Friedrich, R., Fuentes-Prior, P., Ong, E., Coombs, G., Hunter, M., Oehler, R.,
et al. (2002). Catalytic domain structures of MT-SP1/Matriptase, a matrix-
degrading transmembrane serine proteinase. *J. Biol. Chem.* 277, 2160–2168.
doi: 10.1074/jbc.m109830200

Huang, S.-Y., and Zou, X. (2008). An iterative knowledge-based scoring function
for protein-protein recognition. *Proteins* 72, 557–579. doi: 10.1002/prot.
21949

Hwang, H., Vreven, T., Janin, J., and Weng, Z. (2010a). Protein-protein docking
benchmark version 4.0. *Proteins* 78, 3111–3114. doi: 10.1002/prot.22830

Hwang, H., Vreven, T., Pierce, B. G., Hung, J.-H., and Weng, Z. (2010b).
Performance of ZDOCK and ZRANK in CAPRI rounds 13-19. *Proteins* 78,
3104-3110. doi: 10.1002/prot.22764

Jiménez-García, B., Roel-Touris, J., Romero-Durana, M., Vidal, M., Jiménez-
González, D., and Fernández-Recio, J. (2018). LightDock: a new multi-scale
approach to protein–protein docking. *Bioinformatics* 34, 49–55. doi: 10.1093/
bioinformatics/btx555

Khashan, R., Zheng, W., and Tropsha, A. (2012). Scoring protein interaction
decoys using exposed residues (SPIDER): a novel multibody interaction scoring
function based on frequent geometric patterns of interfacial residues. *Proteins*
80, 2207–2217. doi: 10.1002/prot.24110

Koukos, P. I., Roel-Touris, J., Ambrosetti, F., Geng, C., Schaarschmidt, J., Trellet,
M. E., et al. (2020). An overview of data-driven HADDOCK strategies in CAPRI
rounds 38-45. *Proteins* 88, 1029–1036. doi: 10.1002/prot.25869

Koutroumbas, K., and Theodoridis, S. (2008). *Pattern Recognition*. Amsterdam:
Academic Press.

Kozakov, D., Brenke, R., Comeau, S. R., and Vajda, S. (2006). PIPER: an FFT-
based protein docking program with pairwise potentials. *Proteins* 65, 392–406.
doi: 10.1002/prot.21117

Kozakov, D., Clodfelter, K. H., Vajda, S., and Camacho, C. J. (2005). Optimal
clustering for detecting near-native conformations in protein docking. *Biophys.
J.* 89, 867–875. doi: 10.1529/biophysj.104.058768

Launay, G., Ohue, M., Santero, J. P., Matsuzaki, Y., Hilpert, C., Uchikoga, N.,
et al. (2020). Rescoring ensembles of protein-protein docking poses using
consensus approaches. *bioRxiv[Preprint]*, 2020.04.24.059469. doi: 10.1101/
2020.04.24.059469

Lensink, M. F., Méndez, R., and Wodak, S. J. (2007). Docking and scoring protein
complexes: CAPRI 3rd Edition. *Proteins* 69, 704–718. doi: 10.1002/prot.21804

Lensink, M. F., Velankar, S., and Wodak, S. J. (2017). Modeling protein–protein
and protein–peptide complexes: CAPRI 6th edition. *Proteins* 85, 359–377. doi:
10.1002/prot.25215

Lensink, M. F., and Wodak, S. J. (2010). Docking and scoring protein
interactions: CAPRI 2009. *Proteins* 78, 3073–3084. doi: 10.1002/prot.
22818

Liang, S., Meroueh, S. O., Wang, G., Qiu, C., and Zhou, Y. (2009). Consensus
scoring for enriching near-native structures from protein-protein docking
decoys. *Proteins* 75, 397–403. doi: 10.1002/prot.22252

Lu, H., Lu, L., and Skolnick, J. (2003). Development of unified statistical potentials
describing protein-protein interactions. *Biophys. J.* 84, 1895–1901. doi: 10.
1016/s0006-3495(03)74997-2

Malhotra, S., Mathew, O. K., and Sowdhamini, R. (2015). DOCKSCORE: a
webserver for ranking protein-protein docked poses. *BMC Bioinform.* 16:127.
doi: 10.1186/s12859-015-0572-6

Melquiond, A. S. J., Karaca, E., Kastritis, P. L., and Bonvin, A. M. J. J. (2012).
Next challenges in protein–protein docking: from proteome to interactome and
beyond. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2, 642–651. doi: 10.1002/
wcms.91

Mezei, M. (2017). Rescore protein-protein docked ensembles with an interface contact statistics. *Proteins* 85, 235–241. doi: 10.1002/prot.25209

Moal, I. H., and Bates, P. A. (2010). SwarmDock and the use of normal modes in protein-protein docking. *Int. J. Mol. Sci.* 11, 3623–3648. doi: 10.3390/ijms11103623

Moal, I. H., Moretti, R., Baker, D., and Fernández-Recio, J. (2013a). Scoring functions for protein–protein interactions. *Curr. Opin. Struct. Biol.* 23, 862–867. doi: 10.1016/j.sbi.2013.06.017

Moal, I. H., Torchala, M., Bates, P. A., and Fernández-Recio, J. (2013b). The scoring of poses in protein-protein docking: current capabilities and future directions. *BMC Bioinform.* 14:286. doi: 10.1186/1471-2105-14-286

Müller, J. J., Lapko, A., Bourenkov, G., Ruckpaul, K., and Heinemann, U. (2001). Adrenodoxin reductase-adrenodoxin complex structure suggests electron transfer path in steroid biosynthesis. *J. Biol. Chem.* 276, 2786–2789. doi: 10.1074/jbc.m008501200

Ohue, M., Shimoda, T., Suzuki, S., Matsuzaki, Y., Ishida, T., and Akiyama, Y. (2014). MEGADOCK 4.0: an ultra–high-performance protein–protein docking software for heterogeneous supercomputers. *Bioinformatics* 30, 3281–3283. doi: 10.1093/bioinformatics/btu532

Oliva, R., Chermak, E., and Cavallo, L. (2015). Analysis and ranking of protein-protein docking models using inter-residue contacts and inter-molecular contact maps. *Molecules* 20, 12045–12060. doi: 10.3390/molecules200712045

Oliva, R., Vangone, A., and Cavallo, L. (2013). Ranking multiple docking solutions based on the conservation of inter-residue contacts. *Proteins* 81, 1571–1584. doi: 10.1002/prot.24314

Pierce, B., and Weng, Z. (2007). ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* 67, 1078–1086. doi: 10.1002/prot.21373

Pierce, B., and Weng, Z. (2008). A combination of rescoring and refinement significantly improves protein docking performance. *Proteins* 72, 270–279. doi: 10.1002/prot.21920

Pierce, B. G., Hourai, Y., and Weng, Z. (2011). Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* 6:e24657. doi: 10.1371/journal.pone.0024657

Ritchie, D. W., and Venkatraman, V. (2010). Ultra-fast FFT protein docking on graphics processors. *Bioinformatics* 26, 2398–2405. doi: 10.1093/bioinformatics/btq444

Shomura, Y., Dragovic, Z., Chang, H.-C., Tzvetkov, N., Young, J. C., Brodsky, J. L., et al. (2005). Regulation of Hsp70 function by HspBP1: structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. *Mol. Cell* 17, 367–379. doi: 10.1016/s1097-2765(05)01010-5

Song, H. K., and Suh, S. W. (1998). Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator11Edited by R. *Huber. J. Mol. Biol.* 275, 347–363. doi: 10.1006/jmbi.1997.1469

Vakser, I. A. (2014). Protein-protein docking: from interaction to interactome. *Biophys. J.* 107, 1785–1793. doi: 10.1016/j.bpj.2014.08.033

Vangone, A., Cavallo, L., and Oliva, R. (2013). Using a consensus approach based on the conservation of inter-residue contacts to rank CAPRI models. *Proteins* 81, 2210–2220. doi: 10.1002/prot.24423

Vreven, T., Hwang, H., and Weng, Z. (2011). Integrating atom-based and residue-based scoring functions for protein–protein docking. *Protein Sci.* 20, 1576–1586. doi: 10.1002/pro.687

Yu, J., Andreani, J., Ochsenbein, F., and Guerois, R. (2017). Lessons from (co-)evolution in the docking of proteins and peptides for CAPRI Rounds 28-35. *Proteins* 85, 378–390. doi: 10.1002/prot.25180