PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Ranking multiple docking solutions based on the conservation of inter-residue contacts

Romina Oliva,[1]* Anna Vangone,[2] and Luigi Cavallo[2,3]

[1] Department of Applied Sciences, University "Parthenope" of Naples, Centro Direzionale Isola C4 80143, Naples, Italy
[2] Department of Chemistry and Biology, University of Salerno, Via ponte don Melillo 84084, Fisciano, (SA), Italy
[3] KAUST Catalysis Center (KCC), King Abdullah University of Science & Technology, 23955-6900 Thuwa, Saudi Arabia

ABSTRACT

Molecular docking is the method of choice for investigating the molecular basis of recognition in a large number of functional protein complexes. However, correctly scoring the obtained docking solutions (decoys) to rank native-like (NL) conformations in the top positions is still an open problem. Herein we present CONSRANK, a simple and effective tool to rank multiple docking solutions, which relies on the conservation of inter-residue contacts in the analyzed decoys ensemble. First it calculates a conservation rate for each inter-residue contact, then it ranks decoys according to their ability to match the more frequently observed contacts. We applied CONSRANK to 102 targets from three different benchmarks, RosettaDock, DOCKGROUND, and Critical Assessment of PRedicted Interactions (CAPRI). The method performs consistently well, both in terms of NL solutions ranked in the top positions and of values of the area under the receiver operating characteristic curve. Its ideal application is to solutions coming from different docking programs and procedures, as in the case of CAPRI targets. For all the analyzed CAPRI targets where a comparison is feasible, CONSRANK outperforms the CAPRI scorers. The fraction of NL solutions in the top ten positions in the RosettaDock, DOCKGROUND, and CAPRI benchmarks is enriched on average by a factor of 3.0, 1.9, and 9.9, respectively. Interestingly, CONSRANK is also able to specifically single out the high/medium quality (HMQ) solutions from the docking decoys ensemble: it ranks 46.2 and 70.8% of the total HMQ solutions available for the RosettaDock and CAPRI targets, respectively, within the top 20 positions.

Proteins 2013; 81:1571–1584

Key words: ranking; docking decoys; consensus; inter-molecular contacts; protein–protein interactions; structure prediction; CAPRI; RosettaDock; DOCKGROUND; CONS-COCOMAPS; COCOMAPS.

INTRODUCTION

Although most proteins fulfil their functions through interaction with other proteins, a dramatic disproportion still exists between the number of experimental structures solved for protein complexes and the number of structures available for single proteins.[1] In this scenario, molecular docking, that is, the prediction of a protein complex structure starting from the two separate components, is the method of choice for investigating the molecular basis of recognition in many functional biological systems. Molecular docking consists in the generation of a very large number of possible conformations (docking decoys), followed by the selection of possible native-like (NL) solutions, that is, solutions close to the native structure. Unfortunately, correctly scoring the generated decoys to rank NL solutions before the incorrect ones is still an open problem, also object of assessment in the Critical Assessment of PRedicted Interactions (CAPRI), a community-wide blind docking experiment.[2] Groups participating in CAPRI as scorers are asked to rank a series of models, containing at least one NL solution, and to re-submit the 10 top ranked models, with the obvious goal to have NL solutions among them. The result of the last CAPRI edition is that even the best performing scorers were able to include one or more NL solutions in the 10 top-ranked models only in half of the cases, thus outlining the complexity of the problem.[3]

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

Abbreviations: AUC, Area Under Curve; CAPRI, Critical Assessment of PRedicted Interactions; EF, Enrichment factor; FFP, fraction of false positives; FTP, fraction of true positives; HMQ, high/medium quality; NL, native-like; ROC, receiver operating characteristic

*Correspondence to: Romina Oliva; Department of Applied Sciences, University "Parthenope" of Naples, Centro Direzionale Isola C4 80143, Naples, Italy;
E-mail: romina.oliva@uniparthenope.it
Received 17 January 2013; Revised 16 March 2013; Accepted 8 April 2013
Published online 23 April 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24314

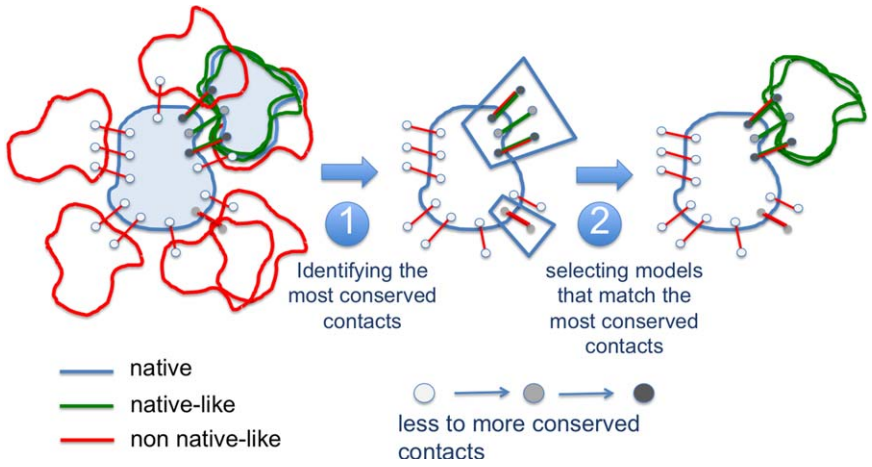

Figure 1

Scheme of the CONSRANK workflow. First, the most conserved residue-residue contacts are identified, then decoys are ranked according to their ability to match the most frequently observed contacts.

In this context, herein we present a novel approach to the problem of ranking multiple docking solutions, embodied in an algorithm we have named CONSRANK (CONSensus-RANKing). CONSRANK deeply differs from other valuable algorithms developed to the aim,[4–19] as it uses neither knowledge-based nor physics-based energy functions. Rather, it relies on the conservation (or frequency) of inter-residue contacts in the analyzed decoys ensemble.

The importance of inter-residue contacts when analyzing docking decoys is well established. In the CAPRI experiment, for instance, the correctness of a prediction, that is, its similarity to the native structure, is assessed based on a combination of root mean square deviation (RMSD) criteria and of conservation of inter-residue contacts, as compared to the native structure.[20] Interestingly, the fraction of common inter-residue contacts among a set of docking decoys has been recently shown by Bonvin and colleagues[21] to successfully apply to their clustering, and a similar concept, that is, the atom contact frequency in a set of predictions, has also been recently added to the ZRANK docking pipeline.[22]

Here we introduce for the first time the use of the conservation of inter-residue contacts to the different task of ranking multiple docking solutions. In particular, we first measure the conservation of each inter-residue contact in a given decoys ensemble, then we rank decoys according to their ability to match the more frequently observed contacts (Fig. 1). The idea is based on the results we obtained when analyzing several CAPRI targets to extract the most conserved inter-residue contacts for visualization in a "consensus map" (i.e. an intermolecular contact map where the conservation of inter-residue contacts is reported on a gray scale).[23] Quite strikingly, we indeed observed that a clear NL consensus in terms of inter-residue contacts emerged from the background noise even in the cases where only a small fraction of NL solutions was present in the decoys ensemble. More importantly, a significant fraction of native contacts was included within the 10 contacts with highest conservation rate. This finding clearly indicates that also incorrect solutions may contribute to the conservation of correct inter-residue contacts and is in line with results of the analysis that Lensink and Wodak performed on 20 CAPRI targets to evaluate the ability of docking protocols to predict the interface in protein–protein complexes.[24] They showed that a sizable fraction (24%) of the interfaces in models ranked as incorrect in the CAPRI assessment are actually correctly predicted and that these models contribute 70% of all the correct docking-based interface predictions. On the other hand, analogously to the scattering of X-ray beams by regular or irregular arrays of atoms, correct contacts, even from incorrect solutions, add constructively toward the native pattern, whereas incorrect contacts are expected to be wrong in a different way (unless the underlying docking algorithm is biased toward a specific wrong interface), thus giving destructive interference.

With these considerations on mind, we developed a simple and fast algorithm to rank docking decoys according to their ability to match the most conserved inter-residue contacts in the analyzed decoys ensemble. In the following, we illustrate the algorithm and demonstrate its very good performance on over 100 targets from three benchmarks: RosettaDock,[25] DOCKGROUND[26] and CAPRI.[27,28]

METHODS

The algorithm we have implemented is split into two sections. In the first one, we analyze the decoys ensemble

to find the most conserved inter-residue contacts. In the second one, we rank the decoys in the ensemble according to their ability to match the most conserved contacts (see Fig. 1).

Given an ensemble of N models of the same biomolecular complex, for each inter-residue contact we define the conservation rate, CR_{kl}, as in Eq. 1,

$$CR_{kl} = nc_{kl}/N \quad (1)$$

where nc_{kl} is the total number of models where residues k and l are in contact. The conservation rate thus ranges between $CR_{kl} = 0$, if the contact between residues k and l is never observed, to $CR_{kl} = 1$, if the contact is observed in all the models. Once the conservation rates have been calculated, the models in the ensemble are ranked according to their ability to match the most conserved inter-residue contacts. To this aim, for each model i we first calculate a score as in Eq. 2:

$$S_i = \sum_1^{Mi} CR_{kl} \quad (2)$$

where M_i is the total number of contacts in model i. Then, we calculate a normalized score, $\bar{S}_i$, as in Eq. (3):

$$\bar{S}_i = S_i/M_i \quad (3)$$

Note that the normalized score $\bar{S}_i$ of Eq. 3 coincides with the average conservation of the inter-residue contacts in each model. Models are ranked according to their $\bar{S}_i$ value. Within this work, two residues are defined in contact if any pair of atoms belonging to the two residues is closer than a cut-off distance of 5 Å. However, to test the stability of our approach to the cut-off distance, test calculations were performed on a selected number of cases by varying the cut-off distance in the range 4–10 Å. Conservation rates were additionally plotted in the form of consensus contact maps, as in the CONS-COCOMAPS program,[23] for visualization purposes. Contact maps for the corresponding native structures were obtained by COCOMAPS.[29]

All the programs under CONSRANK have been written in python, taking advantage of python libraries such as SciPy and Matplotlib. The program is freely available upon request from the authors.

RosettaDock benchmark

A total of 6270 decoys for the 35 targets of the Global-Unbound RosettaDock benchmark having at least one NL solution[25] (available at http://graylab.jhu.edu/docking/decoys/) have been downloaded and analyzed. Models having a ligand RMSD (Lrmsd) $\leq$5 Å were classified as high/medium quality (HMQ). All models having a ligand RMSD (Lrmsd) $\leq$10 Å, i.e. HMQ plus acceptable ones, were classified as NL. On average, each target presented 179.1 decoys, including 10.6 HMQ and 28.5 NL models.

DOCKGROUND benchmark

A total of 6605 decoys for the 61 targets of the DOCKGROUND benchmark[26] (available at http://dockground.bioinformatics.ku.edu/) have been downloaded and analyzed. For the decoys classification into HMQ and NL, see the above section. Each target presented on average 108.3 decoys, including 8.3 HMQ and 9.7 NL models.

CAPRI models

The docking models for recent CAPRI targets[27,28] were downloaded from the official web site (available at: ftp://ftp.ebi.ac.uk/pub/databases/msd/capri/). We analyzed all the 6 recent protein–protein targets having at least one acceptable quality prediction (T24, T25, T26, T29, T32, and T36), for which the docking models were made available to the public at July 31st 2012. A total of 1810 CAPRI models have been analyzed, 300 for target 24, round 9, 300 for target 25, round 9, 310 for target 26, round 10, 350 for target 29, round 13, 350 for target 32, round 15, and 200 for target 36, round 15. Models were classified as incorrect, acceptable, medium quality or high quality, according to the CAPRI assessment.[2,20] Each target presented on average 301.7 decoys, including 10.8 HMQ and 20.3 NL models.

For all the targets in the three examined benchmarks, the Receiver Operating Characteristic (ROC) curve was obtained by plotting the fraction of true positives (FTP) against the fraction of false positives (FFP).

Data analysis

Enrichment factor (EF) was calculated as the ratio between the fraction of HMQ or NL solutions ranked by CONSRANK in the top 5, 10 or 20 positions, and the fraction of HMQ or NL solutions in the examined ensemble.

Finally, as the RosettaDock and CAPRI decoys present polar hydrogens, whereas the DOCKGROUND decoys have no hydrogen atoms, the program was also run on a subset of 10 targets (the six from CAPRI and four from RosettaDock, 1A0O, 1ACB, 1BRS and 1CSE, chosen to present a range of different AUC values) upon excluding hydrogens. The result is that the decoys ranking is virtually unaffected by the presence of such polar hydrogens (AUC values change on average by 0.66 %), therefore they were kept in the analysis.

RESULTS

We tested CONSRANK on three different benchmarks: RosettaDock (global-unbound), DOCKGROUND and CAPRI. We remind the reader that decoys in the RosettaDock benchmark were obtained by Rosetta global docking searches,[25] those in DOCKGROUND were generated

Table I
Summary of the ranking of RosettaDock targets. N-decoys is the total number of decoys; N-HMQ is the number of high or medium quality models; N-NL is the number of native-like, that is, acceptable or better, models; R5-HMQ and R5-NL are the number of HMQ and NL models ranked in the top five positions, respectively; R10-HMQ and R10-NL are the number of HMQ and NL models ranked in the top 10 positions, respectively; R20-HMQ and R20-NL are the number of HMQ and NL models ranked in the top 20 positions, respectively. *EF* is the enrichment in NL or HMQ solutions in the corresponding top positions.

Target	N-decoys	N-HMQ	N-NL	R5-HMQ	R5-NL	R10-HMQ	R10-NL	R20-HMQ	R20-NL	AUC
1A0O	184	1	36	0	4	0	9	0	16	0.931
1ACB	181	1	5	0	0	0	0	0	0	0.186
1AHW	171	0	9	0	0	0	3	0	5	0.923
1ATN	185	9	13	5	5	9	10	9	13	1.000
1AVW	177	1	12	0	0	0	0	0	1	0.856
1AVZ	177	0	2	0	0	0	0	0	0	0.174
1BQL	178	18	31	4	5	9	10	17	20	0.991
1BRS	179	2	28	1	2	2	4	2	8	0.864
1BVK	99	0	76	0	5	0	10	0	20	0.930
1CGI	182	18	37	4	5	8	10	13	20	0.979
1CHO	175	5	34	0	0	0	0	0	0	0.644
1CSE	190	0	19	0	0	0	0	0	0	0.492
1DQJ	197	0	2	0	0	0	0	0	0	0.015
1FBI	190	3	3	0	0	0	0	0	0	0.114
1FSS	179	6	7	0	0	0	0	0	0	0.718
1JHL	186	0	12	0	0	0	0	0	0	0.609
1MAH	169	9	10	0	0	2	2	4	5	0.906
1MEL	181	9	36	2	5	5	10	8	18	0.984
1MLC	187	5	7	0	0	0	0	0	0	0.501
1PPE	179	43	150	1	5	2	10	8	20	0.945
1QFU	176	2	3	0	0	0	0	2	2	0.886
1SPB	174	8	14	0	0	0	0	0	0	0.686
1STF	184	16	18	5	5	10	10	14	15	0.938
1TAB	185	25	43	5	5	10	10	20	20	0.920
1TGS	185	12	46	2	5	4	10	7	20	0.895
1UDI	163	10	18	4	5	7	10	10	16	0.994
1UGH	181	33	65	3	5	6	10	11	20	1.000
1WQ1	186	0	8	0	0	0	0	0	0	0.766
2JEL	192	31	78	4	5	6	10	13	20	0.945
2KAI	180	56	77	5	5	10	10	19	20	0.996
2PTC	192	4	7	0	0	0	0	0	0	0.202
2SIC	179	14	17	5	5	7	7	10	11	0.965
2SNI	184	5	18	0	1	1	4	1	6	0.822
2TEC	183	15	21	0	0	0	0	0	0	0.776
4HTC	180	9	36	0	5	0	9	3	19	0.994
TOT %	6270	370 5.9[a]	998 15.9[a]	50 13.5[b]	82 8.2[c]	98 26.5[b]	168 16.8[c]	171 46.2[b]	315 31.6[c]	
average	179.1	10.6	28.5	1.4	2.3	2.8	4.8	4.9	9.0	0.758 0.799[d]
ST-DEV	15.5	13.1	30.4	2.0	2.4	3.7	4.7	6.3	8.8	0.292 0.247[d]
EF				4.8	2.9	4.7	3.0	4.1	2.8	

[a]Compared to N-decoys.
[b]Compared to N-HMQ.
[c]Compared to N-NL.
[d]Values obtained by excluding the 2 targets having only two native-like solutions.

by the GRAMM-X docking procedure,[26] whereas the CAPRI decoys were submitted by different predictors using different programs and procedures. A total of 14685 models, corresponding to 102 targets, were downloaded (6270 from RosettaDock, 6605 from DOCKGROUND and 1810 from CAPRI) and analyzed.

Given an ensemble of multiple docking solutions for a specific target, CONSRANK first calculates the conservation rate, CR_{kl}, of each observed inter-residue contact in the ensemble (see Methods). Then, for each model, it calculates the normalized score, $\bar{S}_i$ which also corresponds to its average inter-residue contact conservation rate. Models are ranked according to their $\bar{S}_i$ values: the higher the $\bar{S}_i$, the better the rank (Fig. 1). For visualization purposes, a consensus map[23] has also been calculated for each analyzed target.

After ranking the models, the number of HMQ and NL (acceptable or better, NL) solutions ranked within the top 5, 10, and 20 positions was counted. To further investigate the performance of the method, we also calculated the ROCs curve for each target by plotting the FTP versus the FFP. The Area Under the ROC Curve (AUC), compared to the 0.5 value for a random function, was used to assess the overall performance of the method.

Results of the ranking of the RosettaDock, DOCKGROUND and CAPRI targets are summarized in Tables

Table II

Summary of the ranking of the DOCKGROUND targets. N-decoys is the total number of decoys; N-HMQ is the number of high or medium quality models; N-NL is the number of native-like, that is, acceptable or better, models; R5-HMQ and R5-NL are the number of HMQ and NL models ranked in the top 5 positions, respectively; R10-HMQ and R10-NL are the number of HMQ and NL models ranked in the top 10 positions, respectively; R20-HMQ and R20-NL are the number of HMQ and NL models ranked in the top 20 positions, respectively. *EF* is the enrichment in NL or HMQ solutions in the corresponding top positions.

Target	N-decoys	N-HMQ	N-NL	R5-HMQ	R5-NL	R10-HMQ	R10-NL	R20-HMQ	R20-NL	AUC
1A2K	102	2	2	0	0	0	0	0	0	0.295
1A2Y	110	10	10	0	0	0	0	0	0	0.536
1AKJ	110	10	10	0	0	1	1	9	9	0.894
1AVW	110	10	10	0	0	0	0	0	0	0.656
1BTH	101	1	1	0	0	0	0	0	0	0.190
1BUI[a]	110	10	10	2	2	4	4	5	5	0.854
1BUI[b]	110	10	10	0	0	1	1	4	4	0.822
1BVN	110	10	12	0	0	0	0	0	0	0.770
1CHO	110	10	15	0	0	0	0	0	0	0.688
1DFJ	109	9	10	0	0	0	0	1	1	0.774
1E96	110	10	10	0	0	0	0	5	5	0.839
1EWY	110	10	10	0	0	0	0	0	0	0.624
1EZU	110	10	10	0	0	0	0	0	0	0.190
1F51	110	10	10	0	0	0	0	2	2	0.735
1F6M	110	10	10	0	0	0	0	0	0	0.360
1FM9	110	10	13	5	5	9	10	10	13	1.000
1G20	110	10	10	0	0	0	0	0	0	0.746
1G6V	108	8	8	0	0	0	0	0	0	0.286
1GPQ	110	10	10	2	2	4	4	7	7	0.911
1GPW	110	10	18	3	5	7	10	10	18	1.000
1HE1	110	10	13	5	5	9	9	10	12	0.979
1HE8	101	1	1	0	0	0	0	0	0	0.000
1HXY	102	2	2	0	0	0	0	0	0	0.145
1JPS	110	10	10	0	0	0	0	4	4	0.829
1KU6	110	10	10	0	0	0	0	0	0	0.706
1L9B	110	10	10	0	0	0	0	0	0	0.557
1MA9	110	10	10	0	0	0	0	6	6	0.849
1NBF	110	10	10	0	0	0	0	0	0	0.713
1OOK	104	4	4	0	0	0	0	0	0	0.552
1OPH	110	10	10	0	0	1	1	7	7	0.888
1P7Q	104	4	4	0	0	0	0	0	0	0.440
1PPF	110	10	10	5	5	9	9	10	10	0.992
1R0R	110	10	12	0	0	0	0	0	0	0.722
1R4M	101	1	1	0	0	0	0	0	0	0.070
1S6V	104	4	4	0	0	0	0	0	0	0.390
1T6G	110	10	57	1	5	1	10	1	18	0.951
1TMQ	110	10	10	0	0	0	0	0	0	0.586
1TX6	110	10	10	0	0	0	0	0	0	0.482
1U7F	110	10	10	0	0	2	2	8	8	0.847
1UEX	101	1	1	0	0	0	0	0	0	0.210
1UGH	110	10	12	5	5	9	10	10	12	1.000
1W1I	104	4	4	0	0	0	0	0	0	0.742
1WEJ	110	10	10	0	0	0	0	0	0	0.729
1WQ1	110	10	12	0	0	0	0	0	0	0.697
1XD3	110	10	10	0	0	0	0	0	0	0.722
1XX9	102	2	2	0	0	0	0	0	0	0.255
1YVB	110	10	10	0	0	0	0	0	0	0.650
1ZY8[c]	110	10	10	5	5	9	9	10	10	0.999
1ZY8[d]	110	10	10	5	5	5	5	8	8	0.955
2A5T	101	1	1	0	0	0	0	0	0	0.090
2BKR	110	10	11	0	0	0	0	2	2	0.788
2BNQ	101	1	1	0	0	0	0	0	0	0.000
2BTF	110	10	10	5	5	9	9	10	10	0.998
2CKH	110	10	10	0	0	0	0	0	0	0.722
2FI4	110	10	10	0	0	1	1	3	3	0.837
2GOO	110	10	10	0	0	0	0	1	1	0.734
2KAI	110	10	11	0	0	0	0	0	0	0.700
2SNI	110	10	10	0	0	0	0	0	0	0.692
3FAP	110	10	10	0	0	0	0	3	3	0.734
3PRO	110	10	17	4	5	6	10	8	14	0.967

TABLE II.
Continued

Target	N-decoys	N-HMQ	N-NL	R5-HMQ	R5-NL	R10-HMQ	R10-NL	R20-HMQ	R20-NL	AUC
3SIC	110	10	10	0	0	0	0	2	2	0.817
TOT %	6605	505 7.6[e]	589 8.9[e]	47 9.3[f]	54 9.2[g]	87 17.2[f]	105 17.8[g]	156 30.9[f]	194 32.9[g]	
Average	108.3	8.3	9.7	0.8	0.9	1.4	1.7	2.6	3.2	0.654 0.743[h]
ST-DEV	3.3	3.3	7.3	1.7	1.9	2.9	3.4	3.7	4.9	0.280 0.191[h]
EF				2.0	2.0	1.9	1.9	1.7	1.7	

[a]1BUI_A:C.
[b]1BUI_B:C.
[c]1ZY8_AB:K1.
[d]1ZY8_AB:K2.
[e]Compared to N-decoys.
[f]Compared to N-HMQ.
[g]Compared to N-NL.
[h]Values obtained by excluding the nine targets having only one or two native-like solutions.

I, II and III, respectively, and will be described below. We note here that we were consistently able to significantly enrich the fraction of NL and HMQ solutions in the 5, 10, and 20 top ranked positions, compared to the fraction of NL and HMQ solutions in the examined ensemble. In particular, the fraction of NL solutions in the top 10 positions in the RosettaDock, DOCKGROUND and CAPRI benchmarks is enriched on average by a factor of 3.0, 1.9, and 9.9, respectively (see *EF* in Tables I). Considering only the HMQ solutions, the enrichment raises to 4.7 for the RosettaDock and to 12.1 for the CAPRI targets.

Ranking of decoys in the Global-Unbound RosettaDock benchmark

Results of the ranking of the 6270 RosettaDock decoys are summarized in Table I. They correspond to 35 targets, including 20 enzyme-inhibitor, 10 antibody–antigen, and 5 other complexes.

CONSRANK proves to be effective in correctly ranking the docking solutions in the benchmark. It is indeed able to rank 168 (16.8%) of all the 998 NL solutions among the top 10 positions, and 315 (31.6%) of them among the top 20 positions (Table I). Furthermore, CONSRANK proves able to specifically single out the HMQ solutions. In fact, as it is apparent from Table I (columns 5–10), the percentage of HMQ solutions ranked within the top 5, 10, and 20 positions is consistently larger than the percentage of NL solutions in the same positions (13.5 vs. 8.2 %, 26.5 vs. 16.8 % and 46.2 vs. 31.6 %, in the top 5, 10, and 20 positions, respectively). Remarkably, almost half of the HMQ solutions are ranked within the top 20 positions.

For 17 out of the 35 analyzed targets (shadowed in the table), the performance of our ranking method is excellent, with AUC values above 0.9. Except for 1PPE, having almost all correct solutions (150 out of the total 179 ones), these targets, including examples of enzyme-inhibitor, antibody–antigen and other complexes, present a total of NL solutions ranging from 9 to 78, corresponding to 5.3% and 41% of the total solutions, respectively. In two cases, targets 1ATN and 1UGH, all the NL

Table III
Summary of the ranking of the CAPRI targets. N-decoys is the total number of decoys; N-HMQ is the number of high or medium quality models; N-NL is the number of native-like, that is, acceptable or better, models; R5-HMQ and R5-NL are the number of HMQ and NL models ranked in the top 5 positions, respectively; R10-HMQ and R10-NL are the number of HMQ and NL models ranked in the top 10 positions, respectively; R20-HMQ and R20-NL are the number of HMQ and NL models ranked in the top 20 positions, respectively. *EF* is the enrichment in NL or HMQ solutions in the corresponding top positions.

Target	N-decoys	N-HMQ	N-NL	R5-HMQ	R5-NL	R10-HMQ	R10-NL	R20-HMQ	R20-NL	AUC
T24	300	0	4	0	0	0	0	0	0	0.811
T25	300	13	32	2	5	5	10	9	19	0.990
T26	310	15	34	3	5	8	10	13	20	0.986
T29	350	9	17	4	5	5	10	9	16	0.997
T32	350	28	34	5	5	9	10	15	16	0.969
T36	200	0	1	0	0	0	0	0	0	0.467
TOT %	1810	65 3.9[a]	122 7.2[a]	14 21.5[b]	20 16.4[c]	27 41.5[b]	40 32.8[c]	46 70.8[b]	71 58.2[c]	
Average	301.7	10.8	20.3	2.3	3.3	4.5	6.7	7.7	11.8	0.870
ST-DEV	54.9	10.5	15.2	2.1	2.6	3.8	5.2	6.4	9.3	0.210
EF				12.8	9.8	12.5	9.9	10.7	8.7	

[a]Compared to N-decoys.
[b]Compared to N-HMQ.
[c]Compared to N-NL.

solutions are correctly ranked before any incorrect solution, leading to an AUC value of 1. In these two cases the correct solutions are 13 and 65, corresponding to 7.0 and 36 % of the total decoys, respectively.

The average AUC value over the 35 RosettaDock targets is 0.758, with 22 of them having an AUC value higher than 0.8. The average AUC value is improved to 0.799 if targets 1AVZ and 1DQJ, having only two NL solutions, are excluded from the analysis. For these targets, our method ranked the two correct solutions at positions 76th–107th and 88th–169th, respectively, resulting in the particularly low AUC values of 0.174 and 0.015. A bad AUC value (0.114) was also obtained for target 1FBI, having only 3 NL solutions.

The method performed badly in only four additional cases. In particular, AUC values worse that random (i.e. below 0.5) were obtained for the 1ACB and 2PTC targets, whereas AUC values around 0.5 were obtained for the 1CSE and 1MLC targets. 1CSE presents the disappointingly low AUC value of 0.492, although having 19 NL solutions (10% of the total). However, this is a special case because, beside the 19 NL models with a Lrmsd $<$10 Å, there are other 28 models with a Lrmsd $<$ 12 Å. As a matter of fact, many of the solutions classified as incorrect are in fact quite close to the native structure, although not close enough to be classified as NL. Therefore, it is not surprising that CONSRANK ranks several of these "almost NL" solutions in the top positions, thus decreasing the AUC value.

1ACB and 1MLC/2PTC present instead 5 and 7 NL solutions, corresponding to 2.8 and 3.7/3.6 % of the total solutions. In Figure S1, a comparison between the consensus map for the 2PTC target (seven NL solutions, AUC value of 0.202) and the native contact map is reported, from which it is apparent that the available solutions are biased because they point to a consensus that does not correspond to the native contacts.

However, we note that for other targets having a comparable or even lower fraction of NL solutions, for instance 1QFU, 1FSS or 1WQ1, significantly better results are obtained. As an example, the AUC value for the 1QFU target, having only three NL solutions, is as high as 0.886.

Ranking of decoys in the DOCKGROUND benchmark

Results of the ranking of the 6605 DOCKGROUND decoys are summarized in Table II. They correspond to a total of 61 targets.

The overall performance of the method is quite good, since it is able to rank 54 (9.2%) of all the 589 NL solutions within the top five positions, and 105 (17.8%) and 194 (32.9%) of them, respectively, within the top 10 and 20 positions (Table II). As in this benchmark the 505 HMQ solutions represent almost the totality of all the 589 NL ones, comparable results were obtained when considering only the HMQ solutions (Table II, columns 5–10)

The method performs excellently, with AUC values above 0.9, on 11 targets (shadowed in the table), and very well for additional 10 targets, with AUC values between 0.8 and 0.9. Except for 1T6G, having more than half NL solutions (57 out of the total 110 ones), these targets, present a number of NL solutions ranging from 10 to 18, corresponding to 9–16% of all the solutions.

Also for this benchmark, in three cases, targets 1FM9, 1GPW, and 1UGH, all the NL solutions are correctly ranked by CONSRANK before any incorrect solution, leading to an AUC value of 1. In these three cases the NL solutions are 13, 18 and 12, corresponding to 12, 16 and 11% of the total decoys, respectively.

The average AUC value over the 61 targets is 0.654 and rises to 0.743 when the nine targets having only one or two NL solutions are excluded from the analysis. Analogously to results on the RosettaDock benchmark, the method performs badly on targets having only 1–2 NL solutions (maximum AUC value 0.295). AUC values around 0.5 (ranging from 0.390 to 0.552) are also obtained for the 1OOK, 1P7Q and 1S6V targets, having only four NL solutions (3.8% of the total solutions). Low AUC values are obtained in four additional cases, in particular for the 1EZU, 1F6M, 1G6V and 1TX6 targets, having 10/10/8/10 NL solutions out of 110/110/108/110. In all these cases, the decoys in the benchmark are biased toward a wrong solution. This can easily be seen for the 1TX6 and 1F6M targets from Figure S1, where corresponding consensus and native contact maps are reported.

Ranking of CAPRI targets

Results of the analysis of the 1810 models for the 6 CAPRI targets are summarized in Table III.

Performance of CONSRANK on the CAPRI targets is strikingly good. It ranks 40 of the total 122 NL solutions (32.8 %) in the top 10 positions, and 71 of them (58.2 %) in the top 20 positions (Table III). Like in the ranking of the RosettaDock targets, the method specifically singles out the HMQ solutions. Analysis of the data in Table III (columns 5–10) indicates that the percentage of HMQ solutions ranked among the top 5, 10, and 20 positions is consistently larger than the percentage of NL ones (21.5 vs. 16.4 %, 41.5 vs. 32.8 % and 70.8 vs. 58.2 %, in the top 5, 10, and 20 positions, respectively). Therefore, about three quarters of all the HMQ solutions are ranked within the top 20 positions.

The average AUC value is 0.870, and only for target T36, having one NL solution out of 199 (0.5%), the performance of the method is not better than random (AUC value of 0.467). Instead, AUC values approximate to 1 for the T25, T26, T29, and T32 targets, having a number of NL solutions ranging from 17 to 34 (from 5

to 12% of the total solutions). For targets T25, T26, and T32, it is pretty clear that also incorrect solutions point to native contacts, as can be easily seen from the corresponding consensus maps (see Figs. 2, S2, and S3). In the case of target T29, the map is pretty spread and it is not easy to visually distinguish the native contacts from the background noise. However, we have previously shown that 5 out of the 10 best conserved inter-residue contacts are native, that is, correspond to distances within 5 Å in the native structure (while the remaining five are within a maximum distance of 7.7 Å).[23]

Target T24, having only four NL solutions (1.3 % of the total) also has an AUC value as high as 0.811 (the four NL solutions are ranked in positions 38, 59, 68, and 69 out of 296). Also in this case, we have previously shown that the ten best conserved contacts among the available models correspond to an average distance in the native structure below 7 Å.[23] Therefore, these and other contacts with high conservation rate can correctly drive the ranking of the docking decoys.

Dependence of the method performance on the percentage of native-like solutions

In Figure 3a, the obtained AUC value for each examined target is reported vs. the percentage of NL solutions available for it. As expected, AUC values are low for those targets having a very low percentage of NL solutions and significantly increase for targets having a higher percentage of correct solutions. As a general rule, a percentage of 10% or better is guarantee of a performance better than random, although AUC values approaching to 1 have also been found for many targets, especially in the RosettaDock and CAPRI benchmarks, having a percentage of NL solutions as low as 5% or below. Specifically, for the CAPRI targets 1.3 % or more native solutions lead to AUC values above 0.80, a value that increases to more than 0.96 when only targets with at least 5% native solutions are considered.

We also tried to correlate the maximum score obtained for each target, that is, the $\bar{S}_i$ score of the top ranked decoy (ranging from 0.04 to 0.66), with the percentage of NL solutions available. As it can be seen from Figure 3b, however, a linear correlation seems to emerge only for $\bar{S}_i$ values above 0.35 and percentages of NL solutions higher than 40%. At lower values, instead, no clear correlation emerges and $\bar{S}_i$ values of 0.2 or 0.3 may correspond to a range of NL percentages from 1 to 40%. Therefore, unless assuming very high values (above 0.35), the $\bar{S}_i$ absolute value alone is not sufficient to recognize decoy ensembles containing a significant fraction of correct solutions.

Analysis of merged decoys from RosettaDock and DOCKGROUND

The previous analysis clearly indicates that our method outperforms on the CAPRI targets as compared to the RosettaDock and DOCKGROUND ones. This can reasonably depend on the larger number of models available for the CAPRI targets and on the fact that the CAPRI decoys have been obtained by several docking algorithms, whereas decoys in each of the other two analyzed benchmarks are generated by a single docking program. In case these hypotheses are correct, merging decoys from different programs should improve the performance of the method. This could be tested as the DOCKGROUND and RosettaDock benchmarks have six common targets. In two cases, targets 1CHO and WQ1, AUC values below 0.8 were obtained when analysing decoys from the single benchmarks, and no NL solution was ranked within the top 20 positions. Therefore, we collected all the available decoys for these two targets and analyzed the augmented number of decoys together (285 for target 1CHO and 296 for the target 1WQ1). Results summarized in Table IV, clearly show a significant improvement over results obtained when the single benchmarks were analyzed (Tables I and II). In particular, the 20 available NL solutions for 1WQ1 are now ranked between positions 21 and 70, leading to an AUC value of 0.862 (it was 0.766 and 0.697 for RosettaDock and DOCKGROUND alone, respectively). For 1CHO the prediction power of the method improves even more, with 18 NL solutions, out of the total 49, ranked in the top 20 positions and an AUC value as high as 0.898 (it was 0.644 and 0.688 for RosettaDock and DOCKGROUND, respectively).

A comparison of the native 1CHO intermolecular map with consensus maps obtained from the single RosettaDock and DOCKGROUND benchmarks and from the merged decoys is reported in Figure 4. It is pretty clear that in the RosettaDock and DOCKGROUND maps false contacts emerge, whose conservation competes with that of NL ones. As hypothesized, such false contacts are different for the two benchmarks, and their conservation is consequently weakened when the decoys are analyzed all together, allowing the native consensus to more easily emerge.

Dependence of the method performance on the cut-off distance

To test the robustness of the method, we varied the value of the cut-off distance to define an intermolecular contact, from the CAPRI "gold standard" of 5 Å, to 4, 4.5, 5.5, 6, 8 and 10 Å, and applied it to a subset of targets. The chosen targets subset includes all the six analyzed CAPRI targets, plus six representative RosettaDock targets. In particular, the two RosettaDock targets with highest and lowest CONSRANK AUC values were selected for each biological class of antibody–antigen (1BQL and 1DQJ, respectively), enzyme-inhibitor (1UGH and 1ACB, respectively) and other (1ATN and 1AVZ, respectively) complexes.

At the CAPRI standard of 5 Å, the average AUC value of the six CAPRI target is 0.870 (see Table III), while the

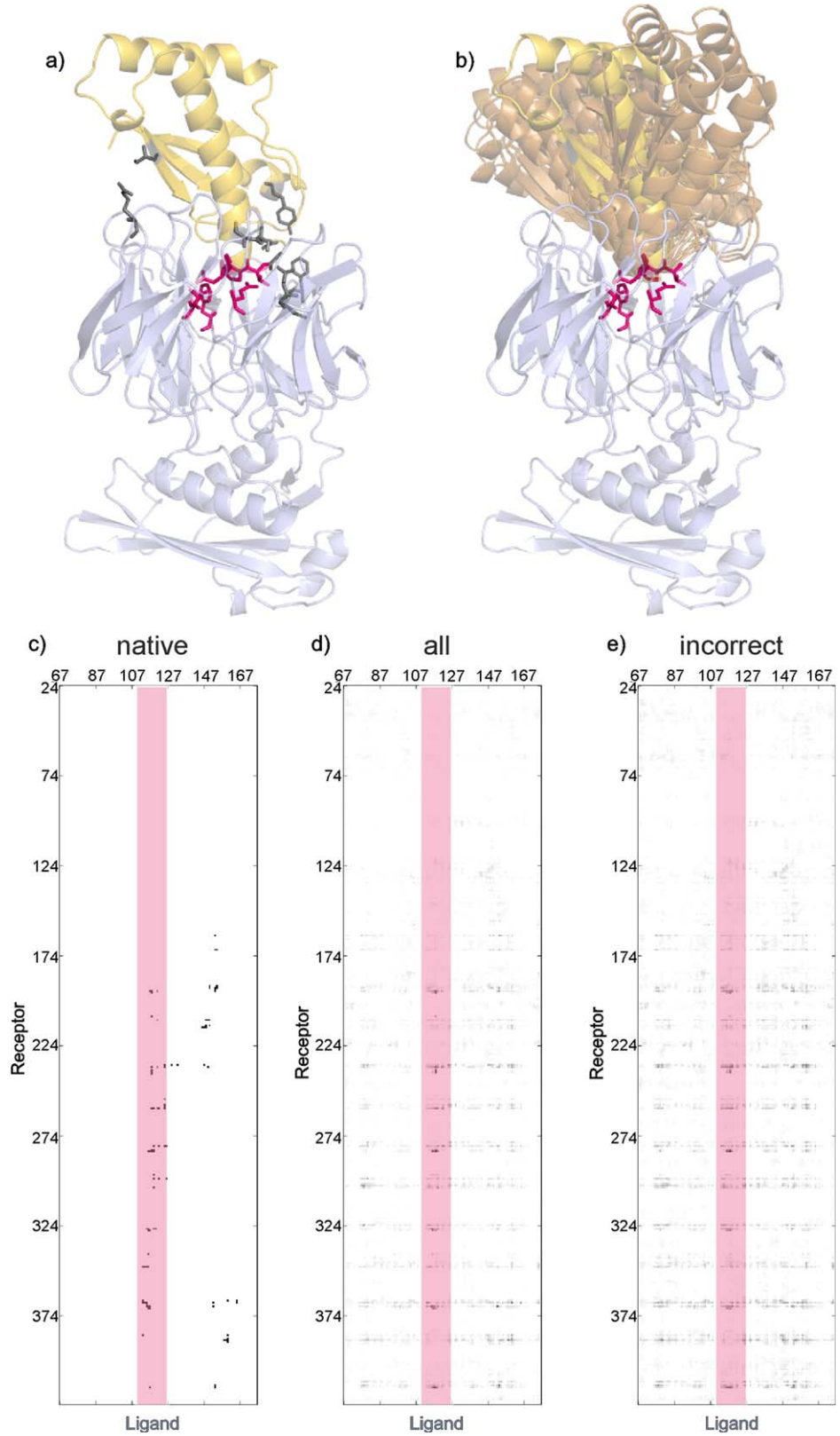

Figure 2

The CAPRI target T26. (**a**) Cartoon representation of the native structure (PDB code: 2HQS[39]), with stick representation of residues involved in the five most and least conserved intermolecular contacts, colored in hotpink and grey, respectively. (**b**) Cartoon representation of the native structure and of nine CAPRI incorrect models (in copper) matching the five most conserved residue–residue contacts (T26_P41.M02, T26_P81.M06, T26_P55.M08, T26_P40.M03, T26_P82.M05, T26_P82.M04, T26_P26.M07, T26_P40.M08, T26_P78.M02). (**c,d,e**) Comparison between the intermolecular contact map of the native structure (c) and the consensus maps obtained from all the 310 models (d) and from only the 276 incorrect models (e) submitted to CAPRI. The map region including the most conserved contacts is shadowed in pink to facilitate the comparison. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

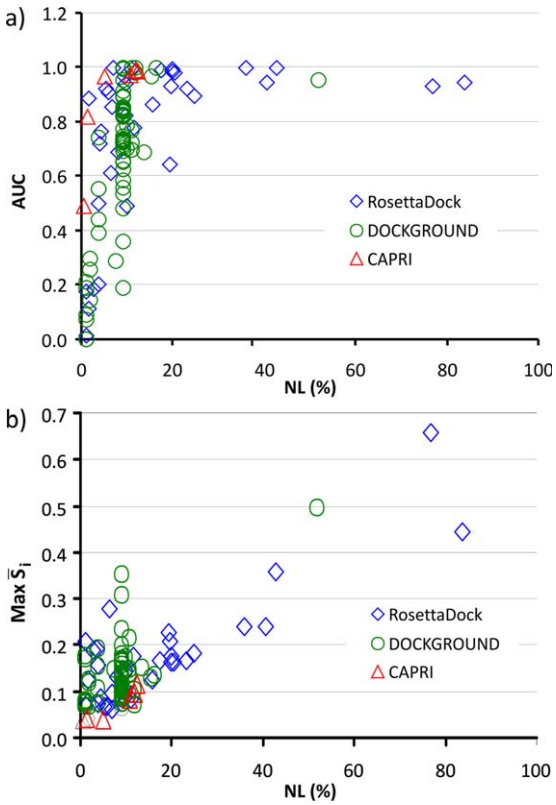

Figure 3

AUC value and Maximum score vs. the number of native-like (NL) solutions (%). Charts of: (**a**) the AUC-ROC value and (**b**) the calculated maximum score ($\bar{S}_i$) versus the percentage of NL solutions for the analyzed targets in the RosettaDock, DOCKGROUND and CAPRI benchmarks. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

average values of the selected RosettaDock targets with highest and lowest AUC are 0.997 and 0.125, respectively (Table I). As shown in Figure S4, the performance of CONSRANK, in terms of AUC values, is extremely stable to the variation of the cut-off distance, independently from the nature of the examined target. This can also be appreciated by considering that, in the examined range of distances, the average AUC value is 0.878 ± 0.007 for the six CAPRI targets, 0.945 ± 0.004 for the RosettaDock targets with highest AUC values and 0.176 ± 0.068 for the RosettaDock targets with lowest AUC values.

Comparative performance of the method

All the CAPRI targets we analyzed, but for T24, were also used in scoring experiments in recent CAPRI editions.[2,3] Groups participating in the experiments as scorers downloaded a number of models, containing at least one NL solution, and used their own scoring function to re-rank them and select the ten top-ranked ones.

Therefore, the performance of CONSRANK can be compared to that achieved by the scorer groups in CAPRI, at least in terms of number of NL solutions in the top ten positions. Note however that CAPRI scorers ranked models submitted by uploaders, whereas we worked on models submitted by predictors, which are the only ones publicly available for download. The former ensemble is typically larger and can also differ from the latter in the percentage of NL solutions it contains. With this *caveat* on mind, for each target the number of NL solutions ranked in the top ten positions (with specification of their possible medium/high quality) is reported in Table V , both for CONSRANK and for the best performing scorer group in CAPRI. Analysis of data in Table V clearly indicates that, with the only exception of T36, for which neither CONSRANK, nor CAPRI scorers could rank any NL solution within the top ten positions, CONSRANK performs significantly better than the best CAPRI scorer. Only in the case of T29, results of the best performing CAPRI scorer are comparable with those obtained by CONSRANK, that is, 9 versus 10 NL solutions in the top 10 positions, including five medium quality ones.

In some cases, the performance of CONSRANK can also be compared with that of other recently proposed methods. For instance, Fink *et al.*[8] applied their PROCOS program to a subset of 40 DOCKGROUND and 14 RosettaDock targets. They obtained an average AUC value of 0.711 for the DOCKGROUND targets, which compares with the average AUC value of 0.714 we obtained for the same targets. AUC values for the RosettaDock targets were not reported, but a comparison can be made by considering the number of correct solutions

Table IV

Summary of the ranking of merged decoys from RosettaDock and DOCKGROUND for the 1CHO and 1WQ1 targets. N-decoys is the total number of decoys; N-HMQ is the number of high or medium quality models; N-NL is the number of native-like, that is, acceptable or better, models; R5-HMQ and R5-NL are the number of HMQ and NL models ranked in the top 5 positions, respectively.; R10-HMQ and R10-NL are the number of HMQ and NL models ranked in the top 10 positions, respectively.; R20-HMQ and R20-NL are the number of HMQ and NL models ranked in the top 20 positions, respectively.

Target	N-decoys	N-HMQ	N-NL	R5-HMQ	R5-NL	R10-HMQ	R10-NL	R20-HMQ	R20-NL	AUC
1CHO	285	15	49	1	5	2	9	6	18	0.898
%		5.3[a]	17[a]							
1WQ1	296	10	20	0	0	0	0	0	0	0.862
%		3.4[a]	6.8[a]							

[a]Compared to N-decoys.

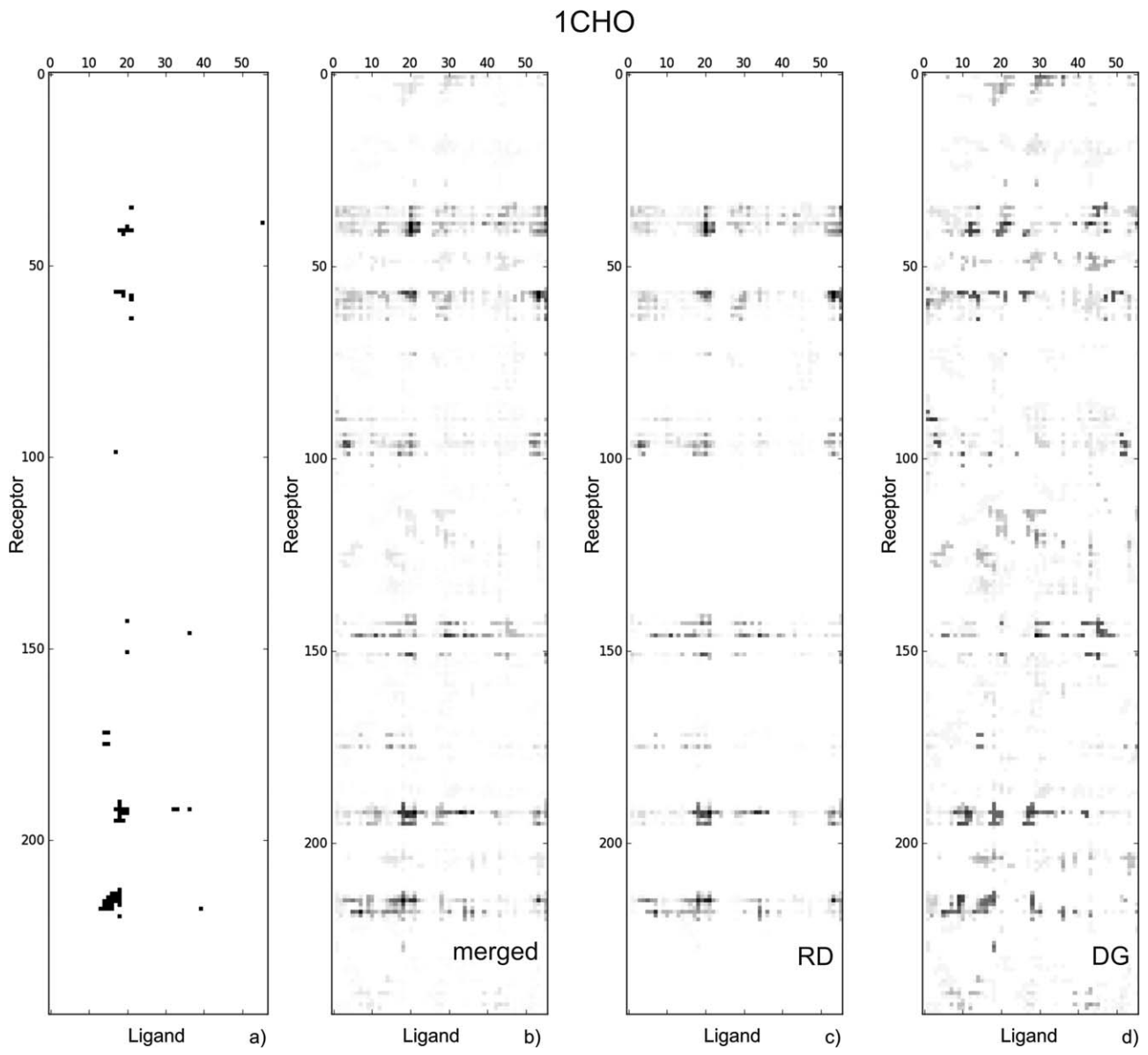

Figure 4

Comparison between the COCOMAPS[29] intermolecular contact map of the 1CHO native structure (**a**), the consensus map obtained from all the 175 models in the RosettaDock Global-Unbound benchmark (**c**), the consensus map obtained from the 110 models in the DOCKGROUND benchmark (**d**) and the consensus map obtained from the 285 merged decoys from the two benchmarks (**b**).

within the top 10 and 20 ranked positions. Our method performs significantly better than PROCOS for 7 out of the 14 targets, namely 1BRS, 1CGI, 1STF, 1TAB, 1TGS, 1UGH, 2SNI. For these targets we ranked on average 16 correct solutions within the first 20 positions, *versus* the average six correct solutions of PROCOS. In other four cases, 1ACB, 1CSE, 2PTC, and 2PTC, both methods failed to rank any correct solution within the first twenty solutions, whereas in the remaining three cases, 1FSS, 1MAH and 2SIC, our method performed slightly worse (average number of correct solutions within the first

Table V

Number of native-like solutions included in the top ten positions by the best performing CAPRI scorer group and by CONSRANK.

Target	Best CAPRI scorer[2,3]	CONSRANK
T25	6/2** (Fernandez-Recio)	10/5**
T26	4/3** (Weng)	10/8**
T29	9/5** (Bonvin)	10/5**
T32	2 (Vakser)	10/1**/9***
T36	0	0

**Medium quality solutions.
***High quality solutions.

twenty solutions 5 *vs.* the 8 of PROCOS). They also applied PROCOS to the CAPRI targets T29 and T32, ranking only 1 and 0 correct solutions within the first ten positions, versus 10 and 9 ranked in the top 10 positions by CONSRANK. The same authors also ran ZRANK[10] and DFIRE[30] on the same targets, finding a performance comparable to PROCOS for ZRANK and slightly worse for DFIRE.[8]

Recently, a novel ranking method, based on a probabilistic multi-classifiers adaptation and genetic algorithm, has also been proposed and tested on some CAPRI targets, including T25, T29, and T32.[14] The number of correct solutions (according to the CAPRI criteria) ranked in the top 10 positions was 2, 6, and 0 for the three targets, respectively, which compares with the 10 correct solutions ranked by CONSRANK for the same CAPRI targets (Tables III and IV).

DISCUSSION

It is commonly accepted that, on the way to the formation of a functional protein–protein complex, encounter (or transient) complexes form as a result of nonspecific collisions, mainly driven by electrostatic interactions.[31–34] In analogy with the funnel-shaped energy well, already described for the folding of single proteins,[35] some encounter complexes can be productive in the sense that they reorient to optimize van der Waals interactions, guided by desolvation, thus coming closer to the functional orientation. In this context, Li *et al.*,[36] evidenced a significantly enhanced encounter complex formation at or near the biological interface by rigid-body Langevin dynamics, while Brownian dynamics simulations of the barnase-barnstar association allowed Gabdoulline and Wade to conclude that the best way to define encounter complexes formation is the satisfaction of a small subset of intermolecular contacts, which are seen in the final structure.[37]

As already noticed, rigid-body docking calculations, based on optimization of the binding energy of interacting molecules, can in principle correctly describe the pool of encounter complexes, provided that the used energy description and sampling method are sufficiently accurate and efficient, respectively.[31] Recently, Valencia and co-workers[38] challenged the impossibility to use docking algorithms to predict protein interactors. They indeed showed that a standard docking program can distinguish true protein interactors from false ones, by possibly reflecting the favorable binding energy profile of a functional complex, due not only to the NL but also to the encounter complexes.

On these bases, we hypothesize that the good performance of our ranking method may be due to the success of current docking methods to capture in many cases favourable solutions, also in the form of encounter complexes. These solutions differ from each other and also from the functional structure, but present some native contacts. This seems to be particularly true when different docking approaches are applied to the prediction of a given target, as in CAPRI. While the presence of encounter complexes in the analyzed decoys ensemble can only be speculated, we could indeed verify that many incorrect solutions, for the analyzed targets where CONSRANK performs particularly well (with AUC values close to 1), were concentrated at or near the biological interface, thus presenting part of the native inter-residue contacts. In the following, we will discuss in detail the case of the *E. coli* TolB-Pal complex, that is, CAPRI target T26 (similar results were obtained for other targets, see for instance the CAPRI target T25 shown in Fig. S2). Inspection of the consensus contact maps for T26, reported in Figure 2 bottom panel, clearly indicates that the pattern of native residue–residue contacts also emerges when analysis is restricted only to the incorrect CAPRI models. The conservation of native inter-residue contacts in the pool of provided models is also visualized on the corresponding crystal structure, in the top panel of the Figure 2. In particular, residues involved in the five most and five least conserved native contacts are shown as pink and gray sticks, respectively. Interestingly, nine models classified as incorrect in the CAPRI assessment match all the five most conserved native residue-residue contacts. These incorrect models are shown, colored in copper, in the top right panel of the figure. Despite only 34 models of this target are classified as NL, the five most conserved native contacts are in fact found on average in 64 models (Table S1). Figure 2 clearly shows that, whereas the poorly conserved inter-residue contacts are peripheral, the most conserved inter-residue contacts are located at the center of the biological interface and can also be maintained by the incorrect models. It is also interesting that, among these five most conserved contacts, three out of the total ten inter-molecular H-bonds of the complex are found (Glu293-Thr114, Ala249-Glu116, and Ser205-Glu116). Furthermore, the three Pal residues involved in the most conserved contacts, at the N-terminus of helix III (Thr114, Pro115, and Glu116), coincide perfectly with the ligand region undergoing conformational change upon binding, with a helix turn unwinding and a dramatic reorientation of the Thr114 and Glu116 side chains to give inter-molecular H-bonds.[39] In the hypothesis that such incorrect models may represent in fact encounter complexes, it might be speculated that, upon establishing such key contacts, the system can easily fall in a narrow energy funnel towards the functional complex.

CONCLUSIONS

We presented CONSRANK, a simple and effective method to rank multiple docking solutions. The novelty

and strength of the method is that it is based on the conservation of contacts at the complex interface and decoys are ranked according to their ability to match the most conserved contacts. Although the fraction of common inter-residue contacts among a set of docking decoys and the atom contact frequency in a set of predictions have recently been applied in previous steps of the docking procedure (that is, clustering[21] and conformational sampling,[22] respectively), our method is the first one using the conservation of inter-residue contacts in the final step of the decoys ranking.

We tested CONSRANK in the analysis of 102 targets from three different benchmarks, finding it to perform consistently well. CONSRANK also proved able to specifically single out the HMQ solutions from the docking decoys ensemble. Remarkably, 46.2% and 70.8% of the total HMQ predictions available for the RosettaDock and CAPRI targets, respectively, were ranked within the top twenty positions. Interestingly, CONSRANK outperforms CAPRI scorers in all the cases where a comparison is feasible,[2,3] and for the remaining target T24, having only 1.3% of acceptable solutions, reaches an AUC value as high as 0.818. It is reasonable to assume that our method performs particularly well when applied to decoys coming from multiple docking programs and procedures, as in the case of CAPRI targets, because in such cases the noise of the incorrect solutions from the different docking procedures cancels, whereas the signal of the correct contacts gets stronger. We proved this concept on the 1WQ1 and 1CHO targets, which are common to the RosettaDock and DOCKGROUND benchmarks. Analysis of the merged decoys from the two benchmarks indeed offers a clear improvement over analysis of the single benchmarks (AUC from 0.766/0.697 to 0.862 for 1WQ1 and from 0.644/0.688 to 0.989 for 1CHO), due to an increased signal to noise ratio in the analysis of the conserved contacts. We have also shown that the CONSRANK performance is substantially invariant to changes in the cut-off distance used to define inter-residue contacts.

Therefore, the method we have developed to rank docking solutions is very well performing and robust, thus offering a valid alternative to the ranking methods already available. Our approach can be particularly useful to analyze docking solutions collected from different docking procedures. Analysis is extremely fast, and hundreds of docking decoys can be reliably ranked in minutes on a standard PC.

ACKNOWLEDGMENTS

Funding: RO has been supported by the Italian MIUR (Ministero dell'Istruzione, dell'Università e della Ricerca; Grant PRIN2008). RO thanks Dr. Angelo Ciaramella and Dr. Antonino Staiano for helpful discussions.

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucl Acid Res 2000;28:235–242.
2. Lensink MF, Mendez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd edition. Proteins 2007;69:704–718.
3. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins 2010;78:3073–3084.
4. Bernauer J, Aze J, Janin J, Poupon A. A new protein–protein docking scoring function based on interface residue properties. Bioinformatics 2007;23:555–562.
5. Camacho CJ, Gatchell DW, Kimura SR, Vajda S. Scoring docked conformations generated by rigid-body protein–protein docking. Proteins 2000;40:525–537.
6. Tress M, de Juan D, Grana O, Gomez MJ, Gomez-Puertas P, Gonzalez JM, Lopez G, Valencia A. Scoring docking models with evolutionary information. Proteins 2005;60:275–280.
7. Cheng TM, Blundell TL, Fernandez-Recio J. pyDock: electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. Proteins 2007;68:503–515.
8. Fink F, Hochrein J, Wolowski V, Merkl R, Gronwald W. PROCOS: computational analysis of protein–protein complexes. J Comput Chem 2011;32:2575–2586.
9. Andrusier N, Nussinov R, Wolfson HJ. FireDock: fast interaction refinement in molecular docking. Proteins 2007;69:139–159.
10. Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. Proteins 2007;67:1078–1086.
11. Huang SY, Zou X. An iterative knowledge-based scoring function for protein–protein recognition. Proteins 2008;72:557–579.
12. Martin O, Schomburg D. Efficient comprehensive scoring of docked protein complexes using probabilistic support vector machines. Proteins 2008;70:1367–1378.
13. Liang S, Meroueh SO, Wang G, Qiu C, Zhou Y. Consensus scoring for enriching near-native structures from protein–protein docking decoys. Proteins 2009;75:397–403.
14. Bourquard T, Bernauer J, Aze J, Poupon A. A collaborative filtering approach for protein–protein docking scoring functions. PLoS One 2011;6:e18541.
15. Liu S, Vakser IA. DECK: Distance and environment-dependent, coarse-grained, knowledge-based potentials for protein–protein docking. BMC Bioinformatics 2011;12:280.
16. Mitra P, Pal D. Using correlated parameters for improved ranking of protein–protein docking decoys. J Comput Chem 2011;32:787–796.
17. Khashan R, Zheng W, Tropsha A. Scoring protein interaction decoys using exposed residues (SPIDER): A novel multibody interaction scoring function based on frequent geometric patterns of interfacial residues. Proteins 2012;80:2207–2217.
18. Gong X, Wang P, Yang F, Chang S, Liu B, He H, Cao L, Xu X, Li C, Chen W, Wang C. Protein–protein docking with binding site patch prediction and network-based terms enhanced combinatorial scoring. Proteins 2010;78:3150–3155.
19. Zimmermann MT, Leelananda SP, Kloczkowski A, Jernigan RL. Combining statistical potentials with dynamics-based entropies improves selection from protein decoys and docking poses. J Phys chem B 2012;116:6725–6731.
20. Mendez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. Proteins 2003;52:51–67.
21. Rodrigues JP, Trellet M, Schmitz C, Kastritis P, Karaca E, Melquiond AS, Bonvin AM. Clustering biomolecular complexes by residue contacts similarity. Proteins 2012;80:1810–1817.
22. Hwang H, Vreven T, Pierce BG, Hung JH, Weng Z. Performance of ZDOCK and ZRANK in CAPRI rounds 13–19. Proteins 2010; 78:3104–3110.

23. Vangone A, Oliva R, Cavallo L. CONS-COCOMAPS: a novel tool to measure and visualize the conservation of inter-residue contacts in multiple docking solutions. BMC Bioinformatics 2012;13 Suppl 4:S19.
24. Lensink MF, Wodak SJ. Blind predictions of protein interfaces by docking calculations in CAPRI. Proteins 2010;78:3085–3095.
25. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J Mol Biol 2003;331:281–299.
26. Liu S, Gao Y, Vakser IA. DOCKGROUND protein–protein docking decoy set. Bioinformatics 2008;24:2634–2635.
27. Janin J, Wodak S. The third CAPRI assessment meeting Toronto, Canada, April 20–21, 2007. Structure 2007;15:755–759.
28. Janin J. The targets of CAPRI Rounds 13–19. Proteins 2010;78:3067–3072.
29. Vangone A, Spinelli R, Scarano V, Cavallo L, Oliva R. COCOMAPS: a web application to analyse and visualize contacts at the interface of biomolecular complexes. Bioinformatics 2011;27:2915–2916.
30. Yang Y, Zhou Y. Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 2008;72:793–803.
31. Blundell TL, Fernandez-Recio J. Cell biology: brief encounters bolster contacts. Nature 2006;444:279–280.
32. Tang C, Iwahara J, Clore GM. Visualization of transient encounter complexes in protein–protein association. Nature 2006;444:383–386.
33. Alsallaq R, Zhou HX. Electrostatic rate enhancement and transient complex of protein–protein association. Proteins 2008;71:320–335.
34. Camacho CJ, Weng Z, Vajda S, DeLisi C. Free energy landscapes of encounter complexes in protein–protein association. Biophys J 1999;76:1166–1178.
35. Shortle D, Simons KT, Baker D. Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci USA 1998;95:11158–11162.
36. Li X, Moal IH, Bates PA. Detection and refinement of encounter complexes for protein–protein docking: taking account of macromolecular crowding. Proteins 2010;78:3189–3196.
37. Gabdoulline RR, Wade RC. Simulation of the diffusional association of barnase and barstar. Biophys J 1997;72:1917–1929.
38. Wass MN, Fuentes G, Pons C, Pazos F, Valencia A. Towards the prediction of protein interaction partners using physical docking. Mol Syst Biol 2011;7:469.
39. Bonsor DA, Grishkovskaya I, Dodson EJ, Kleanthous C. Molecular mimicry enables competitive recruitment by a natively disordered protein. J Am Chem Soc 2007;129:4800–4807.