Bioinformatics, 31(9), 2015, 1481–1483
doi: 10.1093/bioinformatics/btu837
Advance Access Publication Date: 21 December 2014
Applications Note

OXFORD

Structural bioinformatics

CONSRANK: a server for the analysis, comparison and ranking of docking models based on inter-residue contacts

Edrisse Chermak[1,†], Andrea Petta[2,†], Luigi Serra[2,†], Anna Vangone[3,‡], Vittorio Scarano[2], Luigi Cavallo[1] and Romina Oliva[4,*]

[1]Kaust Catalysis Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, [2]Dipartimento di Informatica ed Applicazioni, [3]Department of Chemistry and Biology, University of Salerno, Fisciano, SA, Italy and [4]Department of Sciences and Technologies, University "Parthenope" of Naples, Naples, Italy

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.
‡Present address: Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands
Associate Editor: Anna Tramontano

Received on September 23, 2014; revised on December 5, 2014; accepted on December 14, 2014

Abstract

Summary: Herein, we present CONSRANK, a web tool for analyzing, comparing and ranking protein–protein and protein–nucleic acid docking models, based on the conservation of inter-residue contacts and its visualization in 2D and 3D interactive contact maps.
Availability and implementation: CONSRANK is accessible as a public web tool at https://www.molnac.unisa.it/BioTools/consrank/.
Contact: romina.oliva@uniparthenope.it

Algorithms for macromolecular docking and relative scoring functions, as monitored in the Critical Assessment of Predicted Interactions (CAPRI) experiment, are constantly progressing (Lensink and Wodak, 2013). However, to date no program can provide a single docking solution with a high-enough confidence to be correct. Docking programs instead generally provide the user with an ensemble of models, corresponding to a subset (usually refined) of the solutions they generated in the conformational sampling step. The user is thus left with the issue of analyzing the ensemble of obtained models, maybe from different docking programs, and scoring them to hopefully single out the correct ones.

In this context, we have previously proposed to derive a consensus from an ensemble of protein–protein docking models, based on the conservation within them of the inter-residue contacts (Vangone *et al.*, 2012). We have also proposed the visualization of such consensus in a 'consensus contact map', i.e. an intermolecular contact map where the contacts' conservation is reported on a gray scale. Interestingly, when analyzing several CAPRI targets, we observed that a significant fraction of native contacts was included within the contacts with highest conservation rate, even in the cases where only a small percentage of solutions were correct.

Based on this observation, we then developed a consensus approach to the scoring and ranking of docking models, CONSRANK (CONSensus-RANKing), which ranks models based on their ability to match the most conserved contacts in the ensemble they belong to (Oliva *et al.*, 2013). Application of CONSRANK to the ranking of over 110 targets from different sources showed a very good performance, as it was able to consistently enrich the top ranked positions in correct solutions, provided that they represented an appreciable fraction of the total models (3% was enough for the CAPRI prediction sets, coming from different programs and predictors). Importantly, CONSRANK was shown to perform significantly better than a simple RMSD-based consensus method (Vangone *et al.*, 2013). This means that although the consensus approach clearly has a potential *per se*, the frequency of inter-residue contacts used by CONSRANK is a particularly effective measure to highlight the consensus itself.

Herein, we present the CONSRANK server, which integrates the above analyses, extends them to protein–DNA and protein–RNA complexes, and makes them easily available to the scientific community through an advanced interactive web interface (at the URL: https://www.molnac.unisa.it/BioTools/consrank/). The CONSRANK input consists of an ensemble of docking models in the PDB format (Berman *et al.*, 2000). A user-friendly interface allows the user to upload the input PDB files separately or in an archive (.tar, .tar.gz or .zip) file. The user is also requested to specify the chain identifiers for the molecules involved in the interaction to be analyzed. Interactors made of two chains (e.g. antibodies) can also be dealt with. As we have shown that the CONSRANK approach is particularly effective in ranking models obtained by different docking algorithms, we also provide a tool to consistently renumber a set of diverse PDB files.

CONSRANK outputs are displayed on the results HTML page for one week and archived as downloadable compressed files. A link to the online resource is also emailed to the user, if requested. One thousand models with two molecular chains about 150-residues long are processed in 2 min. Performance scales linearly to 20 min for 10 000 models. CONSRANK output includes in the main page user-sortable and searchable tables reporting: (i) a list of the inter-residue contacts observed in at least 1% of the models, with relative conservation rate (Vangone *et al.*, 2012); (ii) the ranking of the submitted models based on the CONSRANK normalized score (Oliva *et al.*, 2013); (iii) parameters (C50, C70, C90) reflecting the overall conservation of inter-residue contacts in the models ensemble (Vangone *et al.*, 2012).

Further, CONSRANK shows in the output main page a static consensus map. Clicking on it, an interactive map is provided that can be zoomed and navigated to visualize the identity of the residue pairs corresponding to a given contact (dot) and its conservation rate. Finally, an interactive 3D representation of the consensus map, a '3D consensus map', where the third dimension is given by the conservation rate of each inter-residue contact, is also provided. Contacts with a conservation rate below 0.01 (i.e. present in <1% of the models) are not shown in the 3D map.

Once the general output has been generated, the user can choose to perform further analyses on single models, by clicking on the action boxes next to each model name, in the ranking table. In particular, by clicking on the relative box, CONSRANK will calculate the inter-molecular contact map for a given model and will color consequently the corresponding contacts in the interactive 2D and 3D consensus maps. This analysis takes few seconds and can be applied to as many models as the user likes. Results relative to the last three analyzed models are contemporarily reported, in different colors, in the maps. This helps the user to visualize at a glance how much the models resemble each other and how well each of them reflects the overall consensus. Furthermore, by clicking on the 'Cocomaps' box, the COCOMAPS server will be launched on the selected model, to provide detailed information on its interface, including the interface area, a list of the inter-molecular H-bonds and an online 3D visualization of the complex in JMol (http://www.jmol.org/) (Vangone *et al.*, 2011).

An example of CONSRANK 3D maps is shown in Figure 1 for the CAPRI target T46. T46 was a difficult target, as both components were to be homology built. Predictors were however able to include 13 correct solutions among the total 387 submitted models. CONSRANK can efficiently rank them, so that all its top 10 positions are occupied by correct models. In Figure 1, the CONSRANK 3D map is reported for the 'consensus', calculated on the total 387 models, with contacts matched by the first and 100th ranked models

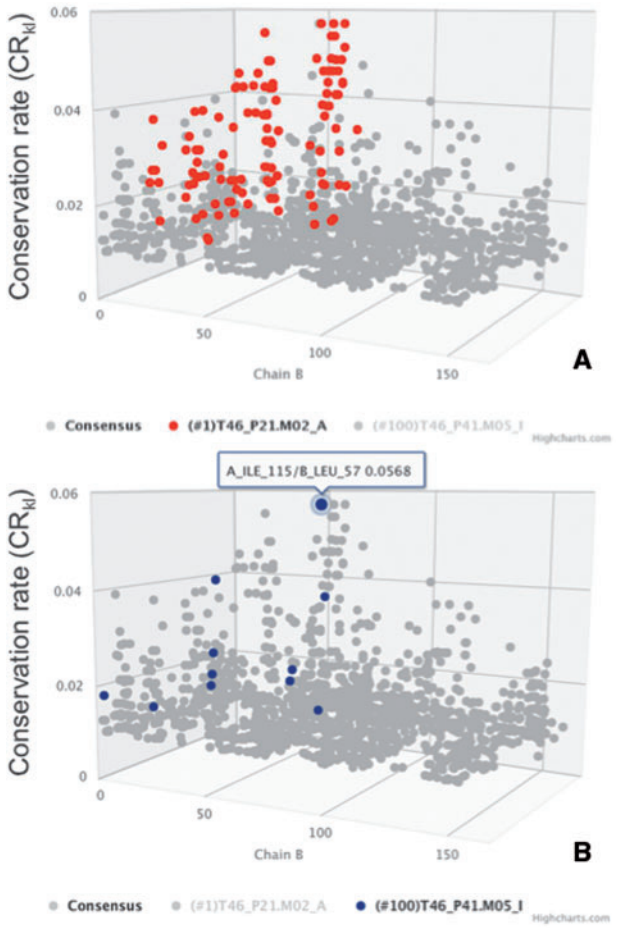

Fig. 1. CONSRANK 3D maps for the CAPRI target T46, showing the consensus contacts for the ensemble of 387 models (gray). (**A**) Contacts present in the model ranked first, and (**B**) contacts present in the model ranked 100th are highlighted

differently colored. As only about 3% of the models are correct, the contacts conservation rate is overall quite low, but still the native contacts clearly emerge from the background, correctly driving the ranking. Model ranked first, which is indeed a correct solution, obviously matches the most conserved contacts (Fig. 1A). As for the clearly incorrect model ranked 100th (Fig. 1B), as expected most of its contacts have a low conservation rate (note that more than half of them are not visualized in the 3D map because their conservation is too low). However, it also presents one of the best conserved contacts, between receptor Leu57 and ligand Ile115, that is indeed a native one, being present in the corresponding X-ray structure [PDB ID: 3Q87 (Liger *et al.*, 2011)]. This confirms that the conservation of native inter-residue contacts may be contributed by incorrect solutions (Oliva *et al.*, 2013; Vangone *et al.*, 2012), also in line with previous results by Lensink and Wodak (2010). By analyzing 20 CAPRI targets, they indeed showed that about one-quarter of the interfaces in models ranked as incorrect in the CAPRI assessment are actually correctly predicted. These findings highlight the importance of a thorough analysis of a set of docking models in terms of inter-residue contacts.

CONSRANK, the interactive and user-friendly web tool we propose here, has in fact most of the desirable features that can make this analysis easy and exhaustive. Starting from an ensemble of the

docking models, it ranks them according to their ability to match the most conserved contacts and offers a user-customizable graphical representation, including a straightforward and effective comparison of different docking solutions.

Funding

R.O. was supported by Regione Campania [LR5-2002-AF2008]. L.C. thanks the KAUST for support.

Conflict of Interest: none declared.

References

Berman,H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.

Lensink,M.F. and Wodak,S.J. (2010) Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins*, **78**, 3085–3095.

Lensink,M.F. and Wodak,S.J. (2013) Docking, scoring, and affinity prediction in CAPRI. *Proteins*, **81**, 2082–2095.

Liger,D. *et al.* (2011) Mechanism of activation of methyltransferases involved in translation by the Trm112 'hub' protein. *Nucleic Acids Res.*, **39**, 6249–6259.

Oliva,R. *et al.* (2013) Ranking multiple docking solutions based on the conservation of inter-residue contacts. *Proteins*, **81**, 1571–1584.

Vangone,A. *et al.* (2011) COCOMAPS: a web application to analyse and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, **27**, 2915–2916.

Vangone,A. *et al.* (2012) CONS-COCOMAPS: a novel tool to measure and visualize the conservation of inter-residue contacts in multiple docking solutions. *BMC Bioinformatics*, **13** (Suppl. 4), S19.

Vangone,A. *et al.* (2013) Using a consensus approach based on the conservation of inter-residue contacts to rank CAPRI models. *Proteins*, **81**, 2210–2220.