Bioinformatics, 2016, 1–3
doi: 10.1093/bioinformatics/btw306
Advance Access Publication Date: 29 June 2016
Applications Note

OXFORD

Structural bioinformatics

Accounting for pairwise distance restraints in FFT-based protein–protein docking

Bing Xia[1], Sandor Vajda[1,2,*] and Dima Kozakov[3,4,*]

[1]Department of Biomedical Engineering, [2]Department of Chemistry, Boston University, Boston, MA 02215, USA, [3]Department of Applied Mathematics and Statistics and [4]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA

*To whom correspondence should be addressed.

Associate Editor: Burkhard Rost

Received on January 19, 2016; revised on April 15, 2016; accepted on May 9, 2016

Abstract

Summary: ClusPro is a heavily used protein–protein docking server based on the fast Fourier transform (FFT) correlation approach. While FFT enables global docking, accounting for pairwise distance restraints using penalty terms in the scoring function is computationally expensive. We use a different approach and directly select low energy solutions that also satisfy the given restraints. As expected, accounting for restraints generally improves the rank of near native predictions, while retaining or even improving the numerical efficiency of FFT based docking.

Availability and Implementation: The software is freely available as part of the ClusPro web-based server at http://cluspro.org/nousername.php

Contact: midas@laufercenter.org or vajda@bu.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The fast Fourier transform (FFT) correlation approach has proved to be very useful for macromolecular docking. The major advantage of the method is the extremely fast evaluation of scoring functions, enabling global systematic sampling of the entire conformational space defined by the relative orientations of two molecules. Although the scoring function must be expressed as a sum of correlation functions (Katchalski-Katzir *et al.*, 1992), it is possible to represent complex energy functions, based on molecular mechanics and including solvation terms, in a form suitable for FFT. FFT-based methods such as ZDOCK (Chen *et al.*, 2003), GRAMM (Tovchigrechko and Vakser, 2006), PIPER (Kozakov *et al.*, 2006), DOT (Mandell *et al.*, 2001) and DOCK/PIERR (Viswanath *et al.*, 2014) consistently perform well both as protocols of individual groups and as automatic servers. In particular, during the latest rounds of the blind protein docking experiment CAPRI (Lensink and Wodak, 2010, 2013), performance of our docking server ClusPro (Comeau *et al.*, 2004), based on the PIPER program, was comparable with top human predictor groups (Kozakov *et al.*, 2013). The server is heavily used: by December 2015 the Cluspro 2.0 website registered over 15 000 unique user IPs, and completed almost 150 000 docking calculations, currently adding about 5000 per month. Models built by the server have been reported in over 400 publications.

In spite of the significant progress, docking generally needs to be complemented by experimental validation. The main reason is that the current scoring functions are not accurate enough for finding the best models among the ones generated by the sampling. Thus, additional information can be very useful for improving the reliability of structure determination. Accordingly, ClusPro has the option to apply extra attraction terms to residues that are *a priori* known to be in the interface. Conversely, repulsion terms are applied to residues that are not expected to be in the interface. However, what ClusPro was lacking so far was the ability to define distance restraints between pairs of atoms or residues. Such restraints can be derived, e.g. from NMR Nuclear Overhauser Effect (NOE) experiments or by chemical crosslinking. The use of restraints is central to the popular High Ambiguity Driven biomolecular DOCKing (HADDOCK) server (Dominguez *et al.*, 2003), which incorporates the interaction restraints into the scoring function to guide the search toward regions of the conformational space in which the restraints are satisfied.

While the extra terms in the scoring function due to the restraints do not significantly increase the computational burden if the sampling is based on Monte Carlo or molecular dynamics algorithms, a similar approach is very costly when used with FFT based sampling. The problem is that each pairwise restraint in the scoring function requires a new correlation function term, and thus an additional Fourier transform, thereby reducing the numerical efficiency of the method. Thus, it is not surprising that none of the successful FFT-based docking programs has the option of accounting for pairwise restraints. However, since FFT performs global sampling, there is no need for guiding the search toward feasible regions. Based on this observation, we solve the problem by directly selecting low energy solutions that also satisfy the restraints. As will be shown, this implies that frequently only portions of conformational space needs to be examined, and hence the computational efforts can actually be reduced. A further advantage is that the scoring function is not affected, and thus we retain the favorable properties of the ClusPro server, validated in many rounds of the CAPRI docking experiment. We demonstrate the method using both simulated and experimentally determined distance restraints.

2 Methods

A pairwise distance restraint can be defined by two sets of atoms, S_1 and S_2 and a distance range, $d_{\min}$ to $d_{\max}$. The restraint is considered satisfied if there are at least one atom in S_1 and at least one atom in S_2 such that the distance between them falls in this range. While the implementation allows for arbitrary sets of atoms to be used to define a restraint, most frequently these involve a single atom or residue on each side of the interface. Given multiple restraints, users may want to require a certain number of restraints out of a group to be satisfied. In addition, restraints may be based on sources with varying reliability, requiring different cutoff values. Our implementation allows for grouping restraints into restraint groups, and restraint groups into restraint sets. Restraint groups are considered satisfied when more than a user specified number of restraints in the group are satisfied, and a restraint set is satisfied when more than a user specified number of its groups are satisfied. We have developed a Javascript Object Notation (JSON) based file format for specifying groups of restraints used by our restraint library, as well a script for converting data in the NOE format into our JSON format. A full description of the file format is provided in the Supplementary Data.

Docking is performed using PIPER, which samples all translations and rotations of a ligand protein with respect to a receptor protein. When a restraint set is provided, PIPER will only report solutions that satisfy the restraints. We note that an FFT based approach has been reported for visualizing such regions of the conformational space. (van Zundert and Bonvin, 2015). To do this efficiently, for each rotation we first generate the set of translations that satisfy each individual restraint, called the feasible translation set for the particular restraint. We then consider the intersection of feasible translation sets for the restraints in each restraint group, and select the translations that appear more often than the cutoff for the restraint group. The selected feasible translation sets for each restraint group are merged in a similar way to generate the feasible translation set for an entire restraint set.

We note that providing restraints can actually decrease the running times by using the restraint set to generate a feasible translation set for each rotation. For each feasible translation the van der Waals interaction energy is computed, and it is used to filter out translations that result in unacceptable clashes. If there are no feasible translations leading to an acceptable van der Waals energy, the rotation is skipped and no other energy terms are evaluated. In practice, this often results in skipping detailed energy calculation for the majority of the rotations. When the cost of generating the feasible translations is less than the cost of evaluating the additional energy terms, fast rotation skipping results in overall speedup. After selecting the solutions that satisfy the restraints, 1000 structures with the lowest PIPER energies are clustered and minimized to remove steric overlaps as customary in ClusPro (Kozakov *et al.*, 2013).

3 Examples of docking with restraints

We first demonstrated the method using simulated distance restraints for a large set of complexes and using different requirements for the number of restraints to be satisfied (see Supporting Material). Here, we present two examples of docking problems with restraints based on experimental data, with details also given as Supporting Material. The first example is constructing the phosphoryl transfer complex between the signal transducing proteins HPr and $IIA^{Glucose}$ (E2A) of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system, starting from the free form protein structures using 20 distance restraints based on intermolecular NOE measurements (Wang *et al.*, 2000). This problem was also studied using HADDOCK, defining a set of Ambiguous Interaction Restraints (AIRs) based on the NOE data (Dominguez *et al.*, 2003). We compared the results of docking without any restraint to the results of docking using the restraint set. ClusPro works very well for this problem even without any restraint, as the center of the second ranked cluster has interface RMSD (iRMSD) of 3.78 Å from the native state. Accounting for the restraints further improves the result, and the center of the largest cluster is shifted to 2.88 Å iRMSD (Supplementary Fig. S8).

The second example is based on the study of the recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex (Bentley *et al.*, 2011). Bmi1 and Ring1b are critical components of the polycomb repressive complex 1 (PRC1) that binds and ubiquitinates the nucleosome in histone H2A on Lys119. We note that the docking of the UbcH5c subunit of the PRC1 complex to histone H2A of the nucleosome was target 95 of the CAPRI docking experiment (Lensink and Wodak, 2013). Restraints were generated after examining the evidence available from the literature. The catalytically competent geometry requires that Cys85 of UbcH5c be located relatively close to the H2A acceptor lysine, Lys119, and hence we created one restraint that had to be satisfied between these two residues. The required range, 0–8 Å, was fairly large, because these residues were located in flexible tail regions of the proteins. Based on mutation data Lys97 and Arg98 of Ring1b interact with the histone in the nucleosome, and hence we created a second restraint group with multiple restraints, from Lys97 to the set of surface residues on the histone. In this second group we only require one of the restraints to be satisfied, since we do not know which of the residues on the surface of the histone interact with Lys97. In this case the best structure without restraints has the iRMSD of 38.68 Å whereas docking using this restraint set produced a near-native pose ranked 2 with the iRMSD of 4.53 Å (Supplementary Fig. S9).

4 Conclusions

We describe implementation of pairwise restraints in the FFT sampling approach. Unlike other approaches that bias the energy

function to steer the docking towards satisfying the restraints, we leave the energy function intact and restrain the search space. Thus, the restraints provide additional information but docking results can be obtained without them. Users can vary the confidence in the restraints by varying the number of restraints to be satisfied, and by specifying restraints in groups. The approach generally improves docking results even with relatively spurious restraints. The method is freely available as part of the ClusPro protein docking server.

Funding

This research was supported by NIH/NIGMS under grants GM093147, GM061867 R35 GM118078, and U01HL127522 and by NSF under grant DBI-1147082 and AF-1527292.

Conflict of Interest: none declared.

References

Bentley,M.L. *et al.* (2011) Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *EMBO J.*, **30**, 3285–3297.

Chen,R. *et al.* (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, **52**, 80–87.

Comeau,S.R. *et al.* (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, **20**, 45–50.

Dominguez,C. *et al.* (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.

Katchalski-Katzir,E. *et al.* (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. USA*, **89**, 2195–2199.

Kozakov,D. *et al.* (2013) How good is automated protein docking? *Proteins*, **81**, 2159–2166.

Kozakov,D. *et al.* (2006) PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins*, **65**, 392–406.

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

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

Mandell,J.G. *et al.* (2001) Protein docking using continuum electrostatics and geometric fit. *Protein Eng.*, **14**, 105–113.

Tovchigrechko,A. and Vakser,I.A. (2006) GRAMM-X public web server for protein–protein docking. *Nucleic Acids Res.*, **34**, W310–W314.

Viswanath,S. *et al.* (2014) DOCK/PIERR: web server for structure prediction of protein–protein complexes. *Methods Mol. Biol.*, **1137**, 199–207.

Wang,G.S. *et al.* (2000) Solution structure of the phosphoryl transfer complex between the signal transducing proteins HPr and IIA(Glucose) of the Escherichia coli phosphoenolpyruvate: sugar phosphotransferase system. *Embo. J.*, **19**, 5635–5649.

van Zundert,G.C. and Bonvin,A.M. (2015) DisVis: quantifying and visualizing accessible interaction space of distance-restrained biomolecular complexes. *Bioinformatics.*, **31**, 3222–3224.