



Protein-Ligand Docking Including Protein Flexibility: A Hierarchical Approach

I. Antes, Ch. Hartmann, Th. Lengauer

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 65-68, 2008.

© 2008 by John von Neumann Institute for Computing
Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

Protein-Ligand Docking Including Protein Flexibility: A Hierarchical Approach

Iris Antes, Christoph Hartmann, and Thomas Lengauer

Max-Planck-Institut für Informatik,
Campus E1 4, 66123 Saarbrücken, Germany
E-mail: antes@mpi-inf.mpg.de

To describe protein-ligand interactions realistically, it is necessary to account for the structural changes in both the receptor and the ligand during complex formation. However, in classical docking approaches the receptor is treated either rigid or semi-rigid. We developed new program, DynaDock⁶, which allows full flexibility for both, the ligand and the receptor during docking. For this purpose we combined two new methods: IRECS (Iterative REduction of Conformational Space)⁴ and OPMD (Optimized Potential Molecular Dynamics), which together allow an efficient sampling of ligand and receptor conformations.

1 Introduction

The major application for protein-ligand docking is the development of new drugs for efficient and safe treatment of diseases. Molecular docking methods developed for this purpose are very efficient approaches, which are specifically designed for rapid identification and characterization of protein-ligand interactions. The main application area of these methods is virtual screening. Virtual screening places tight CPU time constraints on the applied methodology due to the enormous number of potential drug molecules to be tested (around 100.000 to 1.000.000 molecules). Due to these time constraints one very rigorous approximation is commonly used: the structural changes in the receptor upon ligand binding are neglected. However, for a growing number of target structures receptor flexibility is crucial. Thus the efficient treatment of protein flexibility during docking is one of the major challenges within the field.

2 Methods

2.1 IRECS

The most commonly used methods for an efficient treatment of side chain flexibility during docking are discrete optimization approaches based on predefined side chain rotamers, like e.g. ensemble methods (e.g. FlexE¹). Due to the large number of possible side chain conformations at the receptors binding interface, these docking approaches often face the problem of a combinatorial explosion if all possible side chain conformations are considered. Thus it is crucial to preselect a small number of flexible side chains in the binding site for which alternative conformations are used. IRECS was developed for this purpose. It is a new tool for side chain placement, which is especially tailored to the needs of molecular docking. In contrast to other side chain placement tools, which predict the same number of conformations (mostly one) to all side chains of the protein, our tool is able to predict

an ensemble of the most probable conformations for each side chain of a protein. The relative numbers of rotamers that are assigned to each side chain correspond to the side chains flexibility. The absolute level of flexibility can be defined by the user as the final rotamer density (average number of rotamers per residue in the protein) of the output structures. Thus IRECS leads to a minimal, flexibility optimized set of binding site side chain conformers. The predicted side chain ensembles can be used for ensemble based docking (FlexE). The typical application areas of IRECS/FlexE are drug design projects for which a relatively fast algorithm is required and for which a discrete approximation of side chain flexibility is sufficient.

2.2 DynaDock

Although the number of target structures for which molecular docking can be performed increases considerably by the use of discrete, rotamer based flexible docking approaches, there are several limitations of these methods for which no satisfying solution has been yet found. The main restriction is that backbone movement can be considered only to a very limited extent. Another limitation is that, to be efficient enough, a rather coarse grained definition of the conformations must be used. Thus induced fit effects upon ligand binding, which are too small to lead to new side chain conformations, but are necessary for successful binding, are not considered. Last but not least, all docking methods still have a sampling problem if the number of the rotational degrees of freedom to be considered is above a certain limit. Thus efficiently docking a very large, flexible ligand into a flexible binding site is still a challenge for the field.

We have developed a new docking algorithm, DynaDock, which is especially suited for such cases. DynaDock combines discrete search algorithms from bio- and cheminformatics with continuous biophysical simulation methods. The underlying idea behind the algorithm is to first perform a flexible ligand-rigid protein docking on the basis of a modified interaction energy function, which allows for overlapping conformations (can be performed with IRECS). These conformations are then refined by regaining the correct physical interaction energy landscape during a fully flexible simulation of the whole system. For this purpose a new simulation method was developed, Optimized Potential Molecular Dynamics (OPMD), which allows for an efficient molecular dynamics based sampling of the conformational space of both the protein and the ligand. The combined DynaDock-OPMD algorithm leads to a mutual adaptation of the 3D shapes of the ligand and the protein's binding pocket.

3 Results

The performance of IRECS was evaluated on a set of 160 crystal structures of proteins. First, final structures were predicted with a rotamer density equal one, i.e. with a single conformation per side chain. This corresponds the output of other side chain prediction programs. The results were compared to two other side-chain prediction tools, SCWRL² and SCAP³. IRECS achieved a χ_1 accuracy of 84.7% and a χ_{1+2} accuracy of 74.3%, using a 40° cutoff. The average side chain RMSD from the crystal structures was 0.78 Å. This is comparable to the performance of SCWRL² ($\chi_1 = 82.3\%$, $\chi_{1+2} = 68.0\%$, av. RMSD = 0.85 Å) and SCAP³ ($\chi_1 = 84.0\%$, $\chi_{1+2} = 80.6\%$, av. RMSD = 0.82 Å).

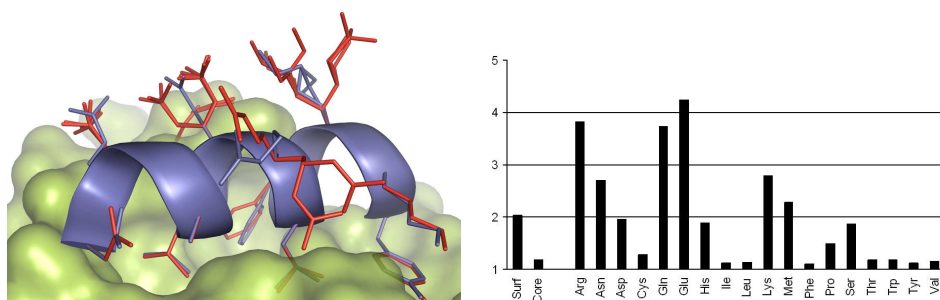


Figure 1. (a) A helix (chain B, position 69-80) of human UDP-galactose 4-epimerase (PDB: 1EK6). The side chains predicted with IRECS are red, the side chains of the crystal structure are blue. (b) Average number of rotamers assigned to the different amino acids in the final structures predicted by IRECS (rotamer density of 2).

In addition, final structures with a rotamer density of two were predicted and the average number of rotamers per amino acid type was analysed. These numbers (Figure 1b) correspond very well to the internal flexibility of the individual amino acids. Thus the ensembles of side chain rotamers assigned by IRECS are representative for the internal flexibility of individual residues.

The DynaDock method was tested on a set of 20 X-ray structures of protein-peptide complexes obtained from the PDB database. The structures were chosen to cover a wide range with respect to the peptide's length (3 to 11 residues) and its surface exposure (20 to 60 %). Starting with randomly disturbed peptide conformations around the binding site with a peptide RMSD larger than 3.5 Å, we were able to obtain refined structures with RMSD values smaller than 2.0 Å for all complexes in our test set using the OPMD approach. Analysing the RMSD of the lowest energy structures, it was observed that also these values were smaller than 2.1 Å for 11 cases and only one case was above 3.0 Å, namely 3.5 Å. These are very promising results, especially considering that also the longest peptide (11 residues) could be refined to 1.08 Å.

References

1. H. Claussen, C. Buning, M. Rarey, and T. Lengauer, *FlexE: efficient molecular docking considering protein structure variations*, J. Mol. Biol. **308(2)**, 377–395, 2001.
2. A. Adrian, A. A. Canutescu, A. A. Shelenkov, and R. L. Dunbrack, *A graph-theory algorithm for rapid protein side-chain prediction*, Protein Sci **12(9)**, 2001–2014, 2003.
3. Z. Xiang and B. Honig, *Extending the accuracy limits of prediction for side-chain conformations*, J. Mol. Biol. **311(2)**, 421–430, 2001.
4. C. Hartmann, I. Antes, and T. Lengauer, *IRECS: A new algorithm for the selection of most probable ensembles of side-chain conformations in protein models*, Protein Science **16(7)**, 1294–1307, 2007.
5. C. Hartmann, I. Antes, and T. Lengauer, *Virtual screening with modeled or flexible side chain conformations*, Proteins, in press.
6. I. Antes, *DynaDock: A new docking approach for protein-peptide docking including receptor flexibility*, in revision.

