

Automatic and efficient decomposition of 2D-structures of small molecules for fragment-based high-throughput docking.



KTI/CTI

Peter Kolb and Amedeo Caflisch [pkolb, caflisch]@bioc.unizh.ch

Department of Biochemistry, University of Zurich, Switzerland

Introduction

Our strategy to dock flexible ligands (SEED-FFLD [1,2,3,4]) uses the binding modes of small rigid fragments to place the entire molecule in the binding site of a target receptor (Fig. 1). For geometrical reasons, at least three fragment positions are required for an unambiguous placement. The fragment identification and the selection of the three most suitable fragments for docking has been automatized and implemented in the program DAIM (Decomposition and Identification of Molecules).

Identification of fragments

Fragments are formed by atoms that are connected by non-rotatable bonds. These are double, triple, amidic and terminal bonds as well as bonds in rings. To obtain chemically more relevant fragments, small functional groups (such as -OH, -CH₃, -NH₂, ...) are reconnected. To reconstitute the appropriate valence for every atom, hydrogen atoms or methyl groups are added.

Selection of the three fragments

In order to find the three fragments which are most suitable for docking, DAIM employs the following selection scheme:

- Every fragment is assigned a score, which is the sum of several feature counts (e.g., number of atoms, heteroatoms, rings, H-bond donors and acceptors, ...)
- Highly substituted fragments are deselected and peripheral fragments are favoured.
- Finally, the three fragments with the highest scores are chosen.

The test set

The Ligand-Protein Database (LPDB, <http://lpdb.scripps.edu>) was used in the redocking study. 48 complexes in which the ligand had four or more fragments (according to DAIM), 10 or fewer rotatable bonds and a molecular weight of less than 550 g/mol were initially selected. In 36 of these test cases, a pose with an RMSD of less than 2 Å with respect to the X-ray structure was obtained with at least one of the fragment triplets used as anchors. For these cases, the calculations based on the triplet suggested by DAIM were compared to calculations based either on randomly chosen triplets or triplets consisting of the three largest fragments.

DAIM-SEED-FFLD

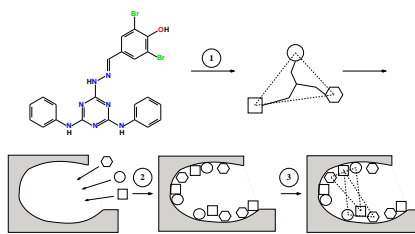


Fig. 1: Scheme of the docking process: after defining the fragments of a ligand ①, these are docked rigidly in the binding site of the receptor ②. The clustered fragment poses are then used as anchor points for the positioning of the entire ligand ③.

Results

Redockings of the 36 ligands were performed, one for each possible triplet combination. When using the three fragments selected by DAIM as anchor fragments, at least one pose with an RMSD of less than 2 Å relative to the X-ray structure was obtained in 20 cases. When using the size-based selection, there are only 14 cases that fulfil this criterion. The expectancy value for randomly selected triplets is six [5]. Hence, fragments selected by DAIM are more appropriate for docking.

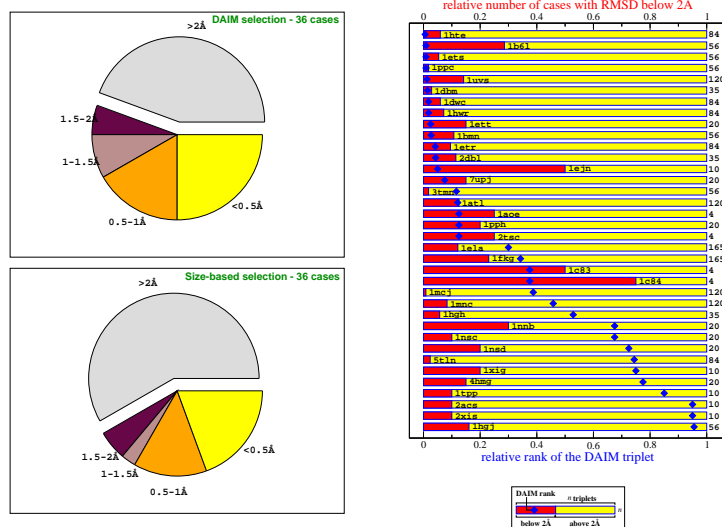


Fig. 2: DAIM triplet selection compared to size-based (left) and random (right) triplet selection.

References

- [1] MAJEUX, N., SCARSI, M., APOSTOLAKIS, J., EHRHARDT, C., AND CAFLISCH, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* 37 (1999), 88.
- [2] MAJEUX, N., SCARSI, M., AND CAFLISCH, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* 42 (2001), 256.
- [3] BUDIN, N., MAJEUX, N., AND CAFLISCH, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* 382 (2001), 1365.
- [4] CECCHINI, M., KOLB, P., MAJEUX, N., AND CAFLISCH, A. Automated docking of highly flexible ligands by genetic algorithms: A critical assessment. *J Comput Chem* 25 (2004), 412-422.
- [5] KOLB, P., AND CAFLISCH, A. Automatic and efficient decomposition of 2D-structures of small molecules for fragment-based high-throughput docking. Manuscript in preparation.