

Selecting molecular fragments for flexible ligand docking

Peter Kolb and Amedeo Caflisch

Department of Biochemistry, University of Zurich, Switzerland

[pkolb, caflisch]@bioc.unizh.ch



STRUCTURAL BIOLOGY
National Center of Competence in Research

Introduction

Our strategy to dock flexible ligands (SEED (Solvation Energy for Exhaustive Docking)-FFLD (Fragment-Based Flexible Ligand Docking) [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 to unambiguously place a ligand in the binding site. 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*. This is a weighted sum of several feature counts (e.g., atom number, heteroatom number, number of rings, ...)
- An initial *cutoff* is applied, which eliminates the smallest fragments.
- Core 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. 130 complexes in which the ligand had four or more fragments (according to DAIM) and 20 or fewer rotatable bonds were selected for the contact analysis. In the redocking analysis, 48 complexes with 10 or fewer rotatable bonds and a MW smaller than 550 g/mol were initially selected. In 26 of these test cases, appropriate SEED points were close to the fragments and at least one fragment combination yielded a pose with an RMSD of less than 2 Å. These cases were analysed in more detail.

DAIM-SEED-FFLD

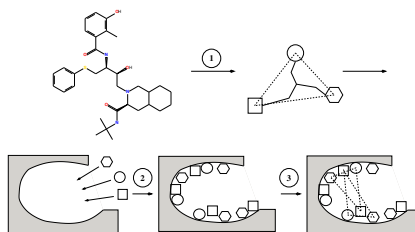


Fig. 1: Scheme of the docking process: after defining the fragments of a ligand ①, these are docked rigidly in the binding site of the receptor ②. The clustered fragment poses are then used as anchor points for the positioning of the entire ligand ③.

Results - Contact analysis

The triplet selected by DAIM was compared to the triplet of fragments with the highest numbers of contacts in the X-ray structure. A contact is defined as a protein heavy atom that lies within 4.5 Å of any ligand heavy atom. The DAIM selection, based only on the 2D-structure of the ligand, is significantly better than a random approach and correctly identifies 279 out of 390 (71.5%) fragments.

Results - Redocking

Redockings of the 26 ligands were performed, one for each possible triplet combination. This showed that in 17 cases, when using the DAIM selection as anchor fragments, an RMSD of better than 2 Å relative to the native structure was achieved [5]. The DAIM triplet also yielded a very low RMSD compared to the other triplets, regardless of the number of fragments (Fig. 2).

Triplet selection

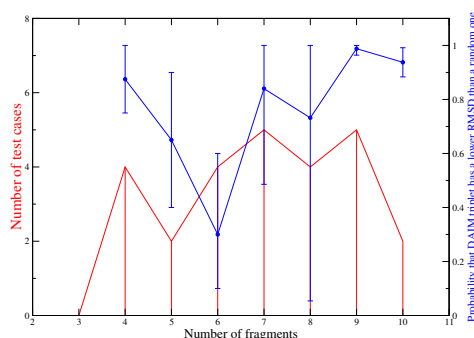


Fig. 2: DAIM selection compared to a random triplet.

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. Selecting appropriate anchors for docking based on 2D-information. Manuscript in preparation.