AUTOMATIC AND EFFICIENT DECOMPOSITION OF 2D-STRUCTURES OF SMALL MOLECULES FOR FRAGMENT-BASED HIGH-THROUGHPUT DOCKING

Peter Kolb and Amedeo Caflisch

Department of Biochemistry, University of Zurich
Winterthurerstr. 190, CH-8057 Zurich

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. For geometrical reasons, at least three fragment positions are required to unambiguously place a ligand in the binding site.

The identification of the fragments and the selection of the three most suitable ones for the docking procedure is performed by the program DAIM (Decomposition And Identification of Molecules). This approach is based entirely on the 2D structure of the ligand. After the definition of the fragments, the three most suitable fragments are chosen based on the following cutoffs: minimum size, maximum number of substituents and the "chemical richness", which is based on a fingerprint developed in-house.

The three fragments selected by DAIM should form the most significant interactions with the receptor. Consequently, usage of these three fragments should lead to a correctly docked structure more often than usage of other combinations of three fragments. To validate the DAIM approach, 130 complexes from the LPDB (Ligand-Protein Data Bank [5]) were analysed. The decomposition of all ligands resulted in an average number of 11.1 $\pm$ 4.6 fragments per ligand. For each fragment, the number of intermolecular heavy-atom contacts was determined. The set of three fragments with the highest number of contacts was compared to the DAIM selection. In 32 cases DAIM correctly predicted three out of three, in 72 cases two out of three, in 23 cases one out of three and in only 3 cases none was correctly predicted. The significance of this result was estimated by comparing with a selection based on fragment size alone and 1000 random selections. Furthermore, for a subset of 36 ligand-protein complexes with ligands of less than 11 rotatable bonds, the ligands were redocked using every possible combination of three fragments. Using the three fragments selected by DAIM, the redocked poses were within 2 Å root mean square deviation (RMSD) of the X-ray pose in 20 out of 36 cases. In contrast, the selection based on fragment-size alone yielded poses below 2 Å RMSD from the X-ray pose only in 14 cases. Moreover, the probability to predict 20 out of 36 cases correctly when choosing a random triplet for the calculation is below 0.28%. This comparison indicates that it is possible to predict the interaction pattern of small molecules from their 2D-structure, and that this information can be successfully used to dock ligands.

[1] Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive Docking of Molecular Fragments With Electrostatic Solvation. Proteins 1999; 3 7:88-105.

[2] Majeux N, Scarsi M, Caflisch A. Efficient Electrostatic Solvation Model for Protein-Docking. Proteins 2001; 42:256-268.

[3] Budin N, Majeux N, Caflisch A. Fragment-based flexible ligand docking by evolutionary optimization. Biol Chem 2001; 382:1365-1372.

[4] M. Cecchini, P. Kolb, N. Majeux, and A. Caflisch, Automated docking of highly flexible ligands by genetic algorithms: A critical assessment. J Comput Chem 25, 412-422, 2004.

[5] http://lpdb.scripps.edu/