The long and rocky road from a PDB file to a protein-ligand docking score

Matthias Rarey, Stefan Bietz, Nadine Schneider, Sascha Urbaczek,

Center for Bioinformatics (ZBH), University of Hamburg, Bundesstr, 43, 20146 Hamburg, Germany

Estimating the ligand's binding affinity for a given protein-ligand complex remains a key challenge in structure-based molecular design. Scoring functions addressing this task are under development for more than 20 years, and although progress is made, there is still much room for further improvement. Looking at individual test cases, it becomes clear that it is crucial to look at the tiny details. A single unsatisfied hydrogen bond donor is usually enough to substantially reduce binding affinity. From the modeling perspective, this also means that a single hydrogen bond modeled incorrectly substantially reduces the prediction quality of a scoring function.

Structures are made available in PDB files reflecting the information which can be gained from experiment. Since in x-ray crystallography, hydrogen atoms cannot be resolved, a difficult and critical preprocessing phase is needed to complete the structure. Especially for ligand molecules, which can be arbitrary organic compounds, the correct chemical structure including hybridization states and bond types has to be perceived [1]. The problem of attaching hydrogens is connected to finding the correct tautomer and protonation state. These in turn cannot be predicted without taking the hydrogen bonding network including bridging water molecules into account. For this task, we developed ProToss, a tool for predicting the tautomeric and protonation state together with hydrogen positions in a single optimization run [2][3].

Eventually, the HYDE scoring function enables the estimation of the free energy of binding. Based on a statistical hydrogen bond network model, HYDE consistently describes hydrogen bonds, dehydration and the hydrophobic effect [4,5]. In order to avoid uncertainties from experimental data, octanol-water partition data of small molecules was exclusively used to derive dehydration energies. Several validation studies and a variety of applications demonstrate the performance of the ProToss-HYDE pipeline.

[1] Urbaczek, S., Kolodzik, A., Heuser, S., Groth, I., Rarey, M., J. Chem.Inf.Model. 53(1) (2013): 76-87
[2] Lippert, S.; Rarey, M., J. Cheminf. 1(13), (2009)
[3] Bietz, S.; Urbaczek, S., Schulz, B., Rarey, M., J. Cheminf. 6(12), (2014)
[4] Schneider, N., Lange, G., Hindle, S., Klein, R., Rarey, M., J.Comput.-Aided Mol.Des. 27(1) (2013): 15-29
[5] Lange, G., Klein, R., Albrecht, J. Rarey, M., Reulecke, I., Patent EP2084520, US8,688,387, (2010)