Abstract

Ligand binding site prediction from protein structure is a fundamental problem in the field of structural bioinformatics that has many applications related to the elucidation of protein function and structure-based drug discovery. The first focus of this thesis was the application of machine learning to this and related problems. The second focus was the development of practically usable tools based on our research. The machine learning based tools produced as a result of the work on this thesis include the pocket re-scoring method PRANK, a stand-alone ligand binding site prediction method P2Rank (together with its extended web interface PrankWeb) and the peptide binding prediction method P2Rank-Pept. We have shown that our methods outperformed available state-of-the-art tools while providing other benefits like prediction speed and stability. Furthermore, we have developed AHoJ, a flexible tool for the search and alignment of Apo-Holo protein pairs in the PDB. AHoJ that is ideal for creating Apo-Holo datasets which can in turn help to better evaluate binding site prediction methods in the future.