[L7] Computational Methods for Cheminformatics and Structure-Based Design

Matthias Rarey

Universität Hamburg, ZBH – Center for Bioinformatics, Bundesstraße 43, 20146 Hamburg, Germany, *http://www.zbh.uni-hamburg.de*

Structure-based drug design always starts with the selection and preprocessing of protein structures. The Protein Databank (PDB) is a highly valuable resource for structures. Besides the structures themselves it makes access to corresponding associated data from experiment to literature easy. Nevertheless, several important preprocessing steps are still required usually performed within a modelling tool suite. With more and more structures available, the integration of structure access and preprocessing gains importance. Overall, the task can be substantially simplified and performed on always up-to-date databases allowing better educated decisions at the end. ProteinsPlus is a web service unifying structure access, assessment, and preprocessing [1,2]. Important processes like exploring ligands, determining their tautomeric forms and protonation states, predicting hydrogen positions and optimizing the hydrogen bond network, estimating electron density support for substructures and molecules, detecting binding pockets and estimating their druggability are fully integrated. Highly efficient database searches for active sites allow to quickly create structure ensembles, getting an overview of known bound ligands, active site mutations and conformational flexibility. Combined with an easy-to-use interface, ProteinsPlus is an ideal starting point for structure-based design projects relying on public crystallographic structures.

The second part of the talk is dedicated to a so-far unsolved problem in cheminformatics. SMARTS expressions are widely used and the quasi-standard for describing molecular patterns in pharmaceutical research. Many tools exist able to match SMARTS expressions against molecules. In some cases, however, it make sense to compare SMARTS expressions with each other. For example, if two sets of filter rules for unwanted structures exist, a comparison gives insights related to the filtered chemical space. A new algorithm able to estimate the similarity between two patterns or the degree of overlap between the resulting chemical spaces after matching two patterns will be presented. Furthermore, the algorithm calculates a mapping between the atom nodes of the patterns substantially simplifying the comparison by eye. First results indicate several applications like rule set comparison, pattern error detection, construction of pattern hierarchies, or pattern database search.

Bibliography:

[1] Fährrolfes, R.; Bietz, S.; Flachsenberg, F.; Meyder, A.; Nittinger, E.; Otto, T.; Volkamer, A.; Rarey, M. (2017). ProteinsPlus: a web portal for structure analysis of macromolecules. Nucleic Acids Research, 45:W337-W343.

[2] Bietz, S.; Inhester, T.; Lauck, F.; Sommer, K.; von Behren, M.; Fährrolfes, R.; Flachsenberg, F.; Meyder, A.; Nittinger, E.; Otto, T.; Hilbig, M.; Schomburg, K.; Volkamer, A.; Rarey, M. (2017) From cheminformatics to structure-based design: Web services and desktop applications based on the NAOMI library . Journal of Biotechnology, 261:207-214.