New Algorithms for Chemical Space Search and Structure-Based Molecular Design

Matthias Rarey[1]

[1]*University of Hamburg, ZBH – Center for Bioinformatics, Albert-Einstein-Ring 8-10, 22761 Hamburg, Germany*

Due to the rise of machine learning technology over the last decade, Cheminformatics and Computational Molecular Design play an increasing role in modern, early-phase drug discovery. While many machine learning methods are applicable in many application fields including chemistry and structural biology, the pre-processing of data for training and validation as well as post-processing of results for connecting to real-world chemistry often require highly specialized algorithmic solutions. In this talk we will focus on two research directions, structure-based molecular design and chemical space search.

In structure-based design, the reliable prediction of binding affinities remains the key challenge despite several decades of research. A significant portion of the difficulties can be attributed to the data situation. For many machine learning methods, the amount, quantity and quality of protein-ligand complexes with associated bioactivity data is insufficient. We will present a pipeline which automatically crosslinks PDB and ChEMBL data based on their original data content, namely sequences and small molecules structures, and will showcase why atomistic detail is important in this process [1,2].

Chemical fragment spaces form the data infrastructure enabling large make-on-demand compound catalogs like Enamine REAL Space, SAVI Space as well as tailor-made combinatorial inhouse spaces like BICLAIM [3-5]. Today, most pharmaceutical research activities build on these resources, simply because they allow to decouple synthetic accessibility from computational molecular design. Over the past decade, we developed a complete infrastructure for managing and searching chemical fragment spaces covering 2D and 3D approaches. In the second half of the talk, more recent developments towards space analytics and integrated search protocols will be presented [6,7].

Bibliography :

[1] Ehmki, E.S.R. et al., J.Chem.Inf.Model. 66 (2026), 1013ff, https://doi.org/10.1021/acs.jcim.5c02505
[2] Gutermuth, T. et al., J.Chem.Inf.Model. (2026), accepted for publication
[3] Warr, W.A. et al., J.Chem.Inf.Model 62 (2022), 2021ff, https://doi.org/10.1021/acs.jcim.2c00224
[4] Korn, M. et al., Curr.Opin.Struct.Biol. 80 (2023), 102578ff, https://doi.org/10.1016/j.sbi.2023.102578
[5] Korn, M. et al., Nat. Sci. Data, 12 (2025), https://doi.org/10.1038/s41597-025-05384-z
[6] Lübbers, J. et al., J.Chem.Inf.Model. 64 (2024), 2008ff, https://doi.org/10.1021/acs.jcim.4c00147
[7] Lübbers, J. et al., J.Chem.Inf.Model. 66 (2026), 2220ff, https://doi.org/10.1021/acs.jcim.5c02952