Closing the loop: combining active learning with sequence or structure-based method for peptide optimisation

Jérôme Eberhardt, Markus Lill, Torsten Schwede

Biozentrum, SIB Swiss Institute of Bioinformatics, University of Basel, Basel
jerome.eberhardt@unibas.ch

The ability to probe and alter interactions between macromolecules and small molecule entities with high specificity forms one of the foundations of modern life science. The design of small molecules binding to specific targets proceeds through iterative cycles of hypothesis formulation, molecule generation, biological testing, data analysis, leading to an updated structure-activity relationship hypothesis to initiate a new cycle. Traditionally, this process involves scientists and facilities from several disciplines with manual “hand-over” of the results from each step.

Recent developments in chemistry, biology, AI and robotics allow for the first time to build such an automated closed-loop design-make-test (DMT) environment to design bioactive molecules for life-science and clinical applications. Guided by modern AI algorithms, the platform will combine fully automated modules for molecular synthesis, purification, and activity testing, to enable the iterative variation and optimization of a bioactive molecule.

As a first step toward fully automated small molecules design, we will start with established automated solid phase chemistry for peptidic compounds, which provides a molecule class of broad biomedical relevance and structural variability. This involves, at the beginning, benchmarking existing methods and checking their applicability, in terms of computational time and prediction accuracy, in the context of a closed-loop environment with no human intervention. For the design and optimisation of peptidic sequences a combination of an active learning method, Bayesian optimisation, with structured-based methods implemented in the Rosetta software was explored. Preliminary results of the Major HistoCompatibility (MHC) complex class I show encouraging results. Further investigations will be pushed in that direction, including benchmarks and development for de novo peptide prediction, cyclisation and integration of nonstandard amino acid residues.