PrankWeb3 - binding site predictions for experimental and predicted protein structures

David Jakubec[1], Petr Škoda[1], Radoslav Krivak[1], Marian Novotný[2], David Hoksza[1]

[1] *Faculty of Mathematics and Physics, Charles University*
[2] *Faculty of Science, Charles University*

PrankWeb is a web-based state-of-the-art ligand-binding site prediction tool. We introduce a new version with two major and an array of minor improvements. The major improvements involve a new, faster and more accurate evolutionary conservation estimation pipeline and the ability to carry out LBS predictions in situations where no experimental structure is available.

The original version of PrankWeb utilized an evolutionary conservation calculation pipeline which attempted to reproduce the evolutionary history of the queried sequence and could take up to several hours to finish. In PrankWeb 3.0, we replaced the evolutionary rate-based conservation scores with an entropy-based metric. The new conservation calculation pipeline utilizes the HMMER3 package for fast and sensitive sequence similarity searches against the UniRef50 sequence database. This change led to more accurate predictions and enabled us to reduce the average time required for the conservation score calculations down to a few minutes.

To extend the functionality of PrankWeb to proteins with no experimental structures available, we have pre-computed the LBS predictions for the structural models from the newly developed AlphaFold database (ADB) and integrated the results into our web server. The user has now the option to enter a UniProt accession number; then, if available, the corresponding model is fetched from the ADB, LBSs are predicted and presented to the user. This required training a new machine learning model specialized for the AlphaFold structures and adapting the interface to visualize specific features such as residue-level confidence scores.

The minor improvements include the ability to deploy PrankWeb as a Docker container, support for the mmCIF file format, improved public REST API access, or the ability to batch download the LBS predictions for the whole PDB archive and parts of the ADB.