Extension of web-based interface for protein binding sites prediction

Lukáš Polák

Introduction

Detection of protein ligand-binding sites is a vital aspect of nowadays drug discovery and development. Identification of the potential binding sites (also called pockets) allows an understanding of various molecule interactions, which is the first step of rational drug discovery pipelines.

PrankWeb is a web-based tool developed at MFF UK for visualizing predictions of protein binding sites. It is based on the P2Rank tool that predicts binding sites quickly and efficiently. The website is used by thousands of users every month.

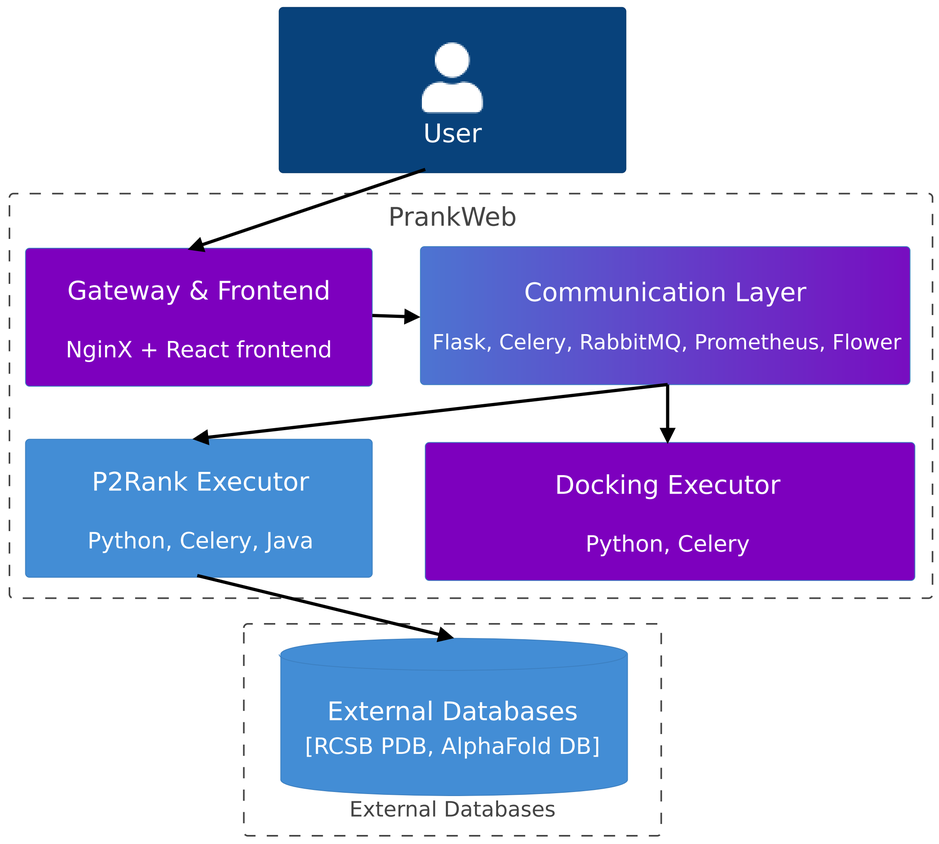
Thesis goals

There were two main goals of the thesis:

- Improve the web frontend and replace old and unsupported components with new ones.
- Extend the server architecture to enable the simple addition of modules (plugins) for post-processing of the predicted binding sites.

Server architecture

The architecture of PrankWeb is rather complex and consists of several components. For simplicity, the architecture may be divided as follows:



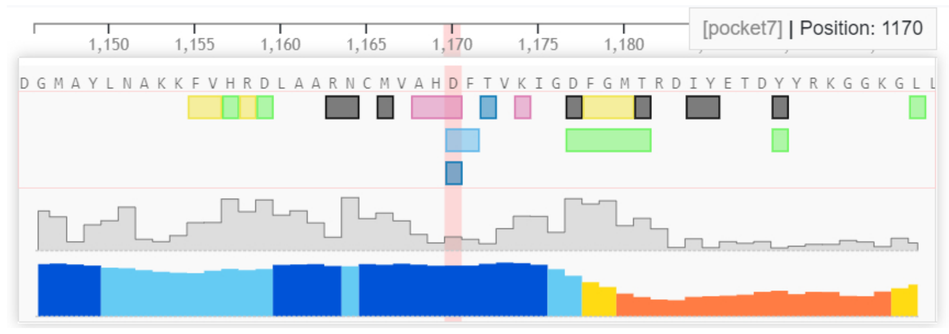
The nodes of the diagram do not represent actual containers, but logical units of the system. Blue nodes represent units that were not modified during the thesis, purple nodes represent units that were substantially modified or added during the implementation.

Thesis-specific technologies

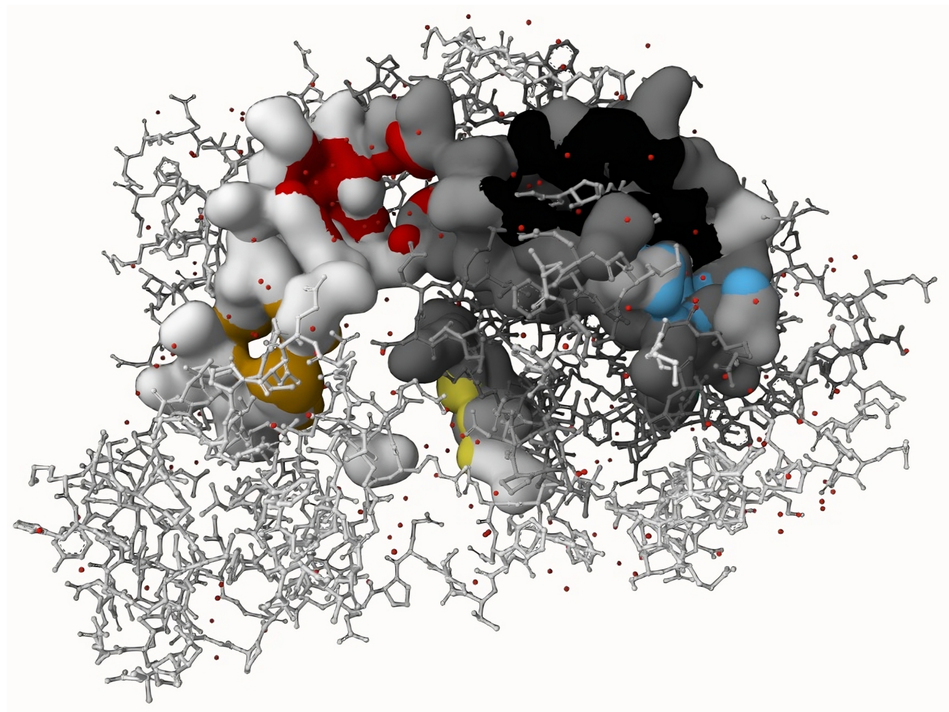
- **Frontend** - React (TypeScript), specialized open-source libraries Mol* and RCSB Saguaro 1D Viewer
- **Backend** - Python, Flask, Celery

The entire project is dockerized and deployed via Docker-compose. Some of the containers use different technologies as well (e.g. Java), but those were not affected by the thesis.

Frontend



One of the two replaced components is the 1D viewer. We employed the RCSB Saguaro 1D Viewer. In the picture, the viewer shows the human insulin receptor protein sequence (UniProt ID P06213) with the predicted binding sites colored in the same theme as in Mol*. The conservation and AlphaFold confidence scores are shown in the form of a histogram.



In the second picture, the 3D representation of the protein 2SRC is shown. As the 3D viewer, we used the Mol* library, which is actively developed. The viewer is interactive and there are plenty of options to customize the visualization. In this picture, the ball-and-stick representation is used for the protein, the pockets are shown in the surface representation and each of the pockets is colored differently. Moreover, the atoms are colored by their conservation score in gray.

Backend plugins

We introduced a plugin system to PrankWeb. Newly created components allow the developers to create new plugins that can be used to post-process the predicted binding sites. As an example of the plugin system, we implemented a plugin that allows molecular docking into the predicted binding sites. The implementation may be integrated into the existing server architecture, so there are minimal changes to the existing codebase and deployment.

Source codes & Contact

Thesis supervisor: **doc. RNDr. David Hoksza, Ph.D.**, Department of Software Engineering

GitHub repository (see the branches): **github.com/luk27official/prankweb**
Running instance: **prankweb.cz**

Try it out yourself!

I'd also like to thank the CUSBG members.