CHARLES UNIVERSITY PRAGUE
faculty of mathematics and physics

Review of the doctoral thesis “Representation of chemical compounds and its utilization in similarity search” by Mgr. Petr Škoda

The presented thesis “Representation of chemical compounds and its utilization in similarity search” by Petr Škoda deals with a relatively broad range cheminformatics topics with drug discovery being the common denominator. This is illustrated in the first chapter of the thesis overviewing the drug discovery process and showing which aspects of the drug discovery pipeline are being addressed by the research topics covered in the thesis.

Although the range of the topics is relatively diverse, sometimes going beyond the cheminformatics domain into bioinformatics, the overview shows that the studied problems are actually relatively well connected addressing most the issues along the whole drug discovery pipeline. As seen from Figure 1, which visually summarizes the thesis contribution, although the author applied the expertise gained during his PhD into related projects of chemical space exploration and protein-ligand binding site detection, the core of his research lies in virtual screening, i.e. the development of molecular representations and methods for identification of small molecules with the desired activity. In this field, the candidate was studying the possibilities to include information about intended molecular target into the design of the molecular representations and screening methods. This research direction was followed by several methods which the candidate developed and which were tackling the problem from different angles. These ranged from the extension of existing methods to incorporate additional information, through the development of new representation to application of machine learning. Working on these topics revealed deficiencies in benchmarking existing approaches of ligand-based virtual screening which naturally resulted in the development of new benchmarking platform.

The author also proved that he is able and willing to invest time into seeing the computational approaches in the broader application context as proven in many

places of the thesis, but especially in the first chapter describing the full drug discovery process and showing how each piece of the carried out research fits into the bigger picture. This is not necessarily the case with a lot of application-oriented computer science research where often the domain is taken rather as a hostage of the computer science technique which needs to be proven worthy.

The quality of the presented research is backed up by several reviewed conference papers as well as several papers in journals with impact factor. The candidate has shown that he is able to carry out his own individual research and come up with and validate new ideas as well as work as part of a larger team.

I believe that the candidate proved to have an ability to perform research and to achieve scientific results. Therefore, I do recommend the thesis for defence.

RNDr. David Hoksza, Ph.D., thesis supervisor