TECHNISCHE UNIVERSITÄT DRESDEN

Biotechnology Center of the TU Dresden (BIOTEC)

BIOTEC, Tatzberg 47/49, 01307 Dresden

BIOTEC
Biotechnology Center
TU Dresden

Prof. Dr. Michael Schroeder
Bioinformatics

Contact:	Sabine Zeissig
Telephone:	+49 351 463-40077
Fax:	+49 351 463-40061
Email:	sabine.zeissig@tu-dresden.de

Dresden, 24 June 2020

XXX

Dear Prof. Tvrdik,

it is a great pleasure to comment on the habilitation thesis of Dr. David Hoksza entitled „Similarity-based Approaches in molecular function discovery". I am a professor in Bioinformatics at the Biotechnology Center of TU Dresden. A computer scientist by training, I develop novel algorithms for protein structure, sequence, and text with applications in drug discovery.

The life sciences are rich in data. Recent advances across all areas of the field led to high throughput technologies, which generate data at an unprecedented rate. Vast amounts of data, require software for their analysis. Dr. David Hoksza has made significant contributions by developing novel algorithms and accessible software across a wide variety of data in the life sciences. He has covered many areas ranging from RNA sequence alignment to binding site prediction in protein structures.

He has published over 50 papers over the last 10 years and has shown a very steady performances. 9 of his papers have more than 9 citations and the well cited papers, which document the impact of his work are assembled in the thesis. Broadly, the thesis covers two different areas and data: RNA and protein structure. DNA, which carries genetic information, is organised in a double helix, which is very stable and long-lasting. In contrast, RNA comes single stranded and because of the ability to form base pairs, it is capable of building base-paired stems and single-stranded loop regions. Understanding the folding of RNA is vital to understand e.g. how it may bind to proteins and how it can carry out some functions. If RNA carries out function then it may be conserved both in sequence, but more important in structure. To address conservation in structure, structural alignments are important. Dr. Hoksza addresses this problem by the development of two algorithms with their implementation, SETTER and MULTISETTER. The work has made some impact and is taken up by other tool developers in the

area. IT would be interesting for future work to bring these tools closer to biological applications. The second part of the thesis concerns Dr Hoksza's work on binding site predictions.

The protein databank PDB is growing and computational structure prediction has seen big breakthroughs in the past decade by absorbing improved statistics of co-evolving residues and by employing machine learning techniques. With these increasing amounts of structural data available their downstream analysis becomes more and more important. One important problem is the identification of ligand binding sites. A number of approaches have been developed over the years, which mainly build on geometry and conservation. The former encompasses improved representations of the protein surface and refined algorithms to define candidate pockets. The latter proves useful as binding sites carry out a function and are therefore more likely to be conserved. Over the years many tools and approaches have emerged to tackle this difficult problem. Dr. Hoksza contributes by developing P2rank, which is innovative as it employs machine learning, and by comparing it to existing approaches. On two benchmark datasets he can show the superior performance of his tool.

Overall, it is interesting that Dr. Hoskza publishes in two closely connected, but separate, communities: bioinformatics and chemoinformatics. An achievement. Dr. Hoksza's work combines software engineering and development with state-of-the-art algorithms in two important areas of bio- and chemoinformatics. He has contributed steadily and with some impact in his area. In summary, I accept his habilitation thesis. He has very clearly shown independent and creative scientific thinking meriting a habilitation.

Sincerely,

Michael Schroeder