

**CONTACT  
INFORMATION**

Department of Software Engineering  
Faculty of Mathematics and Physics  
Charles University in Prague

Malostranské nám. 25  
118 00 Prague, Czech Republic

Phone: +420-951-554-406  
Fax: +420-221-914-323  
E-mail: david.hoksza@matfyz.cuni.cz  
WWW: <https://bioinformatics.cuni.cz/hoksza>

**EDUCATION**

**Charles University in Prague**, Czech Republic  
Faculty of Mathematics and Physics, School of Computer Science

Doc., Software Systems, July 2021

- Thesis: *Similarity-Based Approaches to Molecular Function Discovery*

Ph.D., Software Systems, September 2010

- Thesis: *Similarity Search in Protein Databases*
- Advisor: Tomáš Skopal, PhD.

RNDr., Software Systems, September 2008

M.Sc., Data Engineering, August 2006

- Thesis: *Multidimensional Indexing in Relational Databases*
- Advisor: Tomáš Skopal, PhD.

**RESEARCH  
INTERESTS**

**bioinformatics** : structural bioinformatics, protein sequence and structure databases, rna sequence and structure databases, visualization

**cheminformatics** : chemical space exploration and visualization, virtual screening

**databases** : indexing, metric indexing, non-metric indexing

**ACADEMIC  
APPOINTMENTS**

Associate Professor 2021 – present  
Department of Software Engineering, School of Computer Science, FMP,  
Charles University in Prague

Postdoc researcher May 2017 – April 2020  
Bioinformatics core, Luxembourg Centre from Systems Biomedicine ,  
University of Luxembourg

Assistant Professor June 2012 – 2021  
Department of Software Engineering, School of Computer Science, FMP,  
Charles University in Prague

Researcher September 2010 – June 2012  
Department of Software Engineering, School of Computer Science, FMP,  
Charles University in Prague

Lecturer September 2010 – February 2015  
Laboratory of Informatics and Chemistry, Faculty of Chemical Technology,  
Institute of Chemical Technology Prague

Lecturer September 2009 – September 2011  
Department of Software Engineering, Faculty of Information Technology,  
Czech Technical University in Prague

Lecturer September 2006 – August 2008

Department of Computer Science and Mathematics,  
The University of Finance and Administration

## RESEARCH STAYS

- Luxembourg Centre for Systems Biomedicine, University of Luxembourg 2017-2020  
Bioinformatics Core  
Development of methods and software tools for protein structure visualization and visualization and disease maps analysis.
- Georg-August-Universität Göttingen January 2015  
Institut für Informatik  
Knowledge-based modification of conditional random fields protein-protein identification method
- Helmholtz Zentrum München January 2012 – April 2012  
Institute of Bioinformatics and Systems Biology, Chemoinformatics & Chemical Biology group  
Application of similarity search in protein active sites databases to the drug side effects prediction

## SELECTED PUBLICATIONS

- Feidakis, C., Krivak, R., Hoksza, D., Novotny, M. (2022) AHOJ: rapid, tailored search and retrieval of apo and holo protein structures for user-defined ligands. *Bioinformatics*
- Iqbal, S., ..., Hoksza, D., ... (2022) Delineation of functionally essential protein regions for 242 neurodevelopmental disorders. *Brain*
- Jakubec, D., Skoda, P., Krivak, R., Novotny, M., Hoksza, D. (2022) PrankWeb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures. *Nucleic Acids Research* 50(W1): W593-W597
- PDBe-KB consortium, David Hoksza. PDBe-KB: collaboratively defining the biological context of structural data, *Nucleic acids research*, Oxford Journals, ISSN: 1362-4962, 2022
- Blake A Sweeney, David Hoksza, Eric P Nawrocki, Carlos Eduardo Ribas, Fábio Madeira, Jamie J Cannone, Robin Gutell, Aparna Maddala, Caeden D Meade, Loren Dean Williams, Anton S Petrov, Patricia P Chan, Todd M Lowe, Robert D Finn, Anton I Petrov R2DT is a framework for predicting and visualising RNA secondary structure using templates, *Nature communications*, Nature Publishing Group, ISSN: 2041-1723, 2021
- RNAcentral consortium, David Hoksza. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases, *Nucleic acids research*, Oxford Journals, ISSN: 1362-4962, 2021
- Sumaiya Iqbal, Eduardo Pérez-Palma, Jakob B Jespersen, Patrick May, David Hoksza, Henrike O Heyne, Shehab S Ahmed, Zaara T Rifat, M Sohel Rahman, Kasper Lage, Aarno Palotie, Jeffrey R Cottrell, Florence F Wagner, Mark J Daly, Arthur J Campbell, Dennis Lal. Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants, *Proceedings of the National Academy of Sciences*, National Academy of Science, ISSN: 1091-6490, 2020
- Sumaiya Iqbal, David Hoksza, Eduardo Perez-Palma, Patrick May, Jakob Berg Jespersen, Shehab Sarar Ahmed, Zaara Tasnim Rifat, Henrike O. Heyne, M. Sohel Rahman, Jerrey R. Cottrell, Florence F. Wagner, Mark J. Daly, Arthur J. Campbell and Dennis Lal. MISCAS: MIssense variant to protein StruCTure Analysis web SuiTe, *Nucleic Acids Research*, Oxford Journals, Oxford Journals, ISSN: 1362-4962, 2020

- David Hoksza, Piotr Gawron, Marek Ostaszewski, Jan Hausenauer, Reinhard Schneider. Closing the gap between formats for storing layout information in systems biology, *Briefings in Bioinformatics*, Oxford Journals, ISSN: 1477-4054, 2019
- PDBe-KB consortium, David Hoksza, Radoslav Krivák, Petr Škoda. PDBe-KB: a community-driven resource for structural and functional annotations, *Nucleic Acids Research*, Oxford Journals, ISSN: 1362-4962, 2019
- Lukáš Jendele, Radoslav Krivák, Petr Škoda, Marian Novotný, David Hoksza. PrankWeb: a web server for ligand binding site prediction and visualization, *Nucleic Acids Research*, Oxford Journals, ISSN: 1362-4962, 2019
- David Hoksza, Piotr Gawron, Marek Ostaszewski and Reinhard Schneider. MolArt: A molecular structure annotation and visualization tool, *Bioinformatics*, Oxford Journals, 2018
- Radoslav Krivák, David Hoksza. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure, *Journal of Cheminformatics*, 10(39), Springer, ISSN: 1758-2946, 2018
- Richard Eliáš, David Hoksza. TRAVeLer: a tool for Template-based RnA secondary structure visualization, *BMC Bioinformatics*, 18(847), Springer Nature, ISSN: 1471-2105, 2017
- J. Velkoborsky, D. Hoksza. Scaffold analysis of PubChem database as background for hierarchical scaffold-based visualization. *Journal of Cheminformatics*, 8(74), 2016
- R. Krivák, D. Hoksza. Improving protein-ligand binding site prediction accuracy by classification of inner pocket points using local features. *Journal of Cheminformatics*, 7(1), 2015
- P. Čech, D. Hoksza, D. Svozil. MultiSETTER: web server for multiple RNA structure comparison. *BMC Bioinformatics*, 16(253), 2015
- D. Hoksza, D. Svozil. Multiple 3D RNA Structure Superposition Using Neighbor Joining. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014, PP(99)
- D. Hoksza, P. Škoda, P. Voršilák, D. Svozil. Molpher: a software framework for systematic chemical space exploration. *Journal of Cheminformatics*, 6(7), 2014
- P. Čech, D. Svozil, D. Hoksza. SETTER: web server for RNA structure comparison. *Nucleic Acids Res.* 2012, 40(Web Server issue):W42-8
- D. Hoksza, D. Svozil. Efficient RNA pairwise structure comparison by SETTER method. *Bioinformatics* 2012, 28(14):1858-64
- T. Skopal, D. Hoksza, Improving the Performance of M-tree Family by Nearest-Neighbor Graphs In *ADBIS*, volume 4690 of *Lecture Notes in Computer Science*, pp. 172-188. Springer, 2007.
- T. Skopal, D. Hoksza, J. Pokorný, Construction of Tree-based Indexes for Level-Contiguous Buffering Support. In *Database System for Advanced Applications (DASFAA)*, volume 4443 of *Lecture Notes in Computer Science*, pp. 361–373, Springer, 2007.

REVIEWING AND  
ORGANIZATION  
SERVICE*Reviewing:*

- *Nature Machine Intelligence* , *Nucleic Acids Research* , *Bioinformatics* , *BMC Bioinformatics* , *BMC Genomics* , *BMC Structural Biology* , *Journal of Chemical Information and Modeling* , *Journal of Cheminformatics* , *DKE Journal* , *Journal of Computer Mathematics* , *Journal of Molecular Graphics and Modelling* , *Journal of Molecular Biology* , *Current Bioinformatics* , *SISAP (2008, 2009, 2010, 2012, 2013)* , *ADBIS (2011)* , *DATESO (2008, 2012)* , *PACBB (2014, 2015, 2016, 2017)*

*Organization:*

- Establishment of the bioinformatics study program at Charles University
- *ENBIK 2012, 2014* — National Bioinformatics Conference
- *CSPUG seminars 2011* — co-organizing seminars of the Czech and Slovak PostgreSQL Group
- *DATESO 2007*

## MEMBERSHIP

Grant Agency of Charles University - head of the computer science panel (2013 – present)

Founding member of the Charles University Structural Bioinformatics Group and SIRET Research Group

## GRANTS

*Czech Science Foundation* **2023–2026**

- Project ID: 23-07349S
- Topic: *Targeting protein cryptic binding sites with machine learning*
- Role: *Principal Investigator*

*Ministry of Education, Youth and Sports of the Czech Republic* **2022–2024**

- Project ID: *NPO\_UK\_MSMT-16602/2022*
- Topic: *Transformation for Universities at Charles University: Doctoral Programme in Bioinformatics and Computational Biology*
- Role: *Principal Investigator*

*Czech Science Foundation* **2022–2025**

- Project ID: 22-21696S
- Topic: *Deep Visual Representations of Unstructured Data*
- Role: *Team member*

*Fond National de la Recherche Luxembourg* **2020–2020**

- Project ID: *UCoVis*
- Topic: *A unified web-based platform for viral phylogeny, proteomics and genomics research in real-time*
- Role: *Team member*

*Czech Science Foundation* **2015–2018**

- Project ID: 15-00885S

- Topic: *Novel methods for computational prediction and visualization of secondary structures of ribosomal ribonucleic acids - an integrated solution*
- Role: *Co-Investigator*

*Czech Science Foundation* **2014–2016**

- Project ID: *14-29032P*
- Topic: *Efficient chemical space exploration using multi-objective optimization*
- Role: *Principal Investigator*

*Technology Agency of Czech Republic* **2012–2015**

- Project ID: *TA02010212*
- Topic: *ReceptorX: Integrated Platform for Identification and Development of new Drugs*
- Role: *Team Member*

*Czech Science Foundation* **2011–2014**

- Project ID: *P202/11/0968*
- Topic: *Large-scale Nonmetric Similarity Search in Complex Domains*
- Role: *Team Member*

*Czech Science Foundation* **2009–2011**

- Project ID: *201/09/0683*
- Topic: *Similarity Searching in Very Large Multimedia Databases*
- Role: *Team Member*

*Grant Agency of Charles University* **2007–2008**

- Project ID: *57907*
- Topic: *Similarity Search in Biological Databases*
- Role: *Principal Investigator*

PROFESSIONAL  
EXPERIENCE

**ASP a.s.**, Prague, Czech Republic

*Analyst of Information Systems, Database Expert* **2004–2010**

- Design of information system (conceptual modeling)
- SQL programming (Microsoft SQL Server — TSQL, Oracle — PL/SQL)
- SQL optimization (Microsoft SQL Server, Oracle)
- Database administration (MSSQL, Oracle)
- Web programming (ASP.NET, C#)

TEACHING  
EXPERIENCE

**Charles University in Prague**, Prague, Czech Republic

*Teaching Assistant* **September 2006 – present**

- Data visualization
- Bioinformatics algorithms
- Data organization and processing
- Administration of MS SQL Server
- Database Systems
- Database Applications

**Institute of Chemical Technology**, Prague, Czech Republic

*Teaching Assistant*

**September 2011 – February 2015**

- Bioinformatics Algorithms
- Database Systems

**Czech Technical University in Prague**, Prague, Czech Republic

*Teaching Assistant*

**September 2009 – September 2011**

- Database Systems
- Retrieval of Multimedia Content on the Web
- Searching Web and Multimedia Databases

**The University of Finance and Administration**, Prague, Czech Republic

*Lecturer*

**September 2006 – August 2008**

- Introduction to Unix

AWARDS

**Bernard Bolzano Foundation in Computer science Award**

- Awarded for the paper *Density-Based Classification of Protein Structures Using Iterative TM-score*, 2010

**Werner Von Siemens Excellence Award**

- Honorable mention for the PhD thesis *Similarity Search in Protein Databases*, 2010