Batuhan Kav

batuhankav [at] gmail [dot] com
GitHub LinkedIn

EXPERIENCE

Machine Learning Engineer, ariadne.ai GmbH (Heidelberg, Germany) **09.2022 – 06.2025**

- Research and development lead for biomedical imaging and spatial data analysis, managing the full lifecycle of machine learning projects with a focus on CNN-based models, and their integration into the company's software-as-a-service product.
- Created robust generalized object segmentation models using deep neural networks, which reduced the customer project delivery times from weeks to days. Automatizing the entire deep learning lifecycle, from data collection to curation, to model training and selection.
- Led and delivered multiple internal and customer projects involving statistical modeling of high dimensional image data, contributing to a scientific publication and overseeing front-end integration. Hands-on experience with dimension reduction and unsupervised clustering techniques, including graph neural networks.
- Led a small team in designing and implementing robust and scalable computer vision algorithms with supervised/self-supervised deep learning models, achieving a 50% reduction in computational costs and improved model robustness
- Project lead for integrating proteomics and spatial transcriptomics data into the existing internal workflows.

Postdoctoral Researcher, Forschungszentrum Jülich, Dept. of Structural Biology with Prof. Birgit Strodel (Jülich, Germany) **06.2019 - 09.2022**

- Implemented a virtual drug discovery data pipeline for compounds preventing disordered protein aggregation. Collected, parsed, and stored >1 billion chemical structures, calculated their physicochemical properties using classical and deep cheminformatics models, performed molecular docking and applied molecular dynamics simulations for selecting lead compounds.
- Developed non-linear optimization methods based on genetic algorithms to optimize molecular dynamics simulation force fields for amyloid-beta aggregation. Used enhanced sampling methods for binding free energy calculations.
- In a multi-institution project, investigated the effect of mutations in protein sequences and their relevance for renal diseases. Performed sequence-to-structure predictions, molecular dynamics force field optimization, and protein-protein docking.
- Supervised master's and PhD students.

PhD Student, Max Planck Institute of Colloids and Interfaces with Dr. Thomas Weikl and Prof. Emanuel Schneck (Potsdam, Germany) **08.2015 - 03.2019**

- Investigated the role of membrane-embedded small sugar molecules in facilitating the membrane adhesion using molecular dynamics simulations and statistical physics models. Computed binding constants and developed models to estimate physical forces involved during the adhesion process.
- Developed a full-atomistic molecular dynamics force field to accommodate different water types.

OPEN-SOURCE CONTRIBUTIONS AND PERSONAL PROJECTS

Contributor, scikit-mol **05.2025 - Present**

- https://github.com/EBjerrum/scikit-mol
  Project aims at bridging the gap between common Python tools for calculating molecular properties (RDKit) and building predictive/generative models (scikit-learn).
- Implemented new dataset splitting algorithm: StratifiedGroupShuffleSplit that creates stratified random splits while respecting the group ids.

Kaggle Contributor to Predict New Medicines with BELKA Challenge **05.2024 – 09.2024**

- https://www.kaggle.com/competitions/leash-BELKA
  Participated in the competition to develop new computational models for drug-discovery. Used deep learning models based on graph neural networks to classify the binding of small molecules to a set of proteins.

Kaggle contributor to predict polymer structure

- Competition aimed to create models to predict the physical properties of polymers using their sequence information. Worked with sequence, embedding, and molecular descriptor models and combined them with classical ML models like random forests and XGBoost.

Scientific Software Developer & Project Lead, NMRLipids (FAIRMD) Project Databank **02.2020 - Present**

- https://github.com/NMRLipids/Databank
  Core design and implementation of the open-source NMRLipids databank that stores molecular dynamics simulation data and automates its analysis and evaluation for its users. It is the world largest open and curated data repository for lipid simulations. Results published in Nature Communications.
- Conceptualized, coordinated, performed, and concluded an international research project with 10+ collaborators. Oversaw the project's entire lifecycle and its publication in a high-impact journal as a corresponding author

Scientific Software Developer, IDP Databank Project **02.2020 - Present**

- Core design and implementation of the open-source IDP databank that extends the scope of the NMRLipids databank to the intrinsically disordered proteins. Contributing to the development of evaluation metrics to measure the simulation-experiment comparisions.

COMPUTER SKILLS

Languages: Python
ML-frameworks: PyTorch, PyTorch-geometric, scikit-learn,
Python libraries: numpy, scikit-image, OpenCV, Pandas, RDKit
Cloud: AWS, Sagemaker StudioLab
Version Control: git
Databases: SQL
ML/Dev-Ops: MLFlow, Docker
Molecular Dynamics and Docking: AMBER, Gromacs, OpenMM, Autodock Vina

LANGUAGES

Turkish: Mother tongue
English: Full professional working proficiency
German: Advanced language skills (B2-level ongoing)

EDUCATION

PhD, Max Planck Institute of Colloids and Interfaces (Potsdam, Germany) **08.2015 - 03.2019**
MSc Physics, Koç University with (Istanbul, Türkiye) **09.2013 - 08.2015**
BSc Chemistry with minor in Physics, Bilkent University (Ankara, Türkiye) **09.2009 - 08.2013**
Summer Visiting Student, University of Cambridge (Cambridge, United Kingdom) 06.2011 - 09.2011

PEER REVIEWED SCIENTIFIC PUBLICATIONS

1) Guo, Z., Nguyen, K, Claudio, NM., Ziglari, T., **Kav, B.,** Pucci, F. et al. (2024) Sentinel node B cells drive epitope spreading and anti-tumor T cell immunity by recognition and presentation of extracellular vesicle-linked antigens. Submitted to Cell
2) Kiirikki, A., Antila, H., Bort, L., Buslaev, P., Fernando, F., Ferreira, T. M., Fuchs, P., Garcia-Fandino, R., Gushchin, I., **Kav, B.**, et al. (2024). Overlay databank unlocks data-driven analyses of biomolecules for all. Nature Communications, 15, 1136
3) Antila, HS., Dixit, S., **Kav, B.**, Madsen, JJ., Miettinen, MS., Ollila, OHS. (2024). Evaluating polarizable biomembrane simulations against experiments. Journal of Chemical Theory and Computation, 20, 10, 4325–4337
4) Fatafta, H., Khaled M., **Kav B.**, Olubiyi, O.O, Strodel, B. (2024) A brief history of amyloid aggregation simulations. WIREs Computional Molecular Science 2024;14:e1703.
5) **Kav, B.**, Weikl, T. R., & Schneck, E. (2023). Measuring pico-newton forces with lipid anchors as force sensors in molecular dynamics simulations. The Journal of Physical Chemistry B, 127(18), 4081–4089.
6) Smorodina, E., **Kav, B.**, Fatafta, H., & Strodel, B. (2023). Effects of ion type and concentration on the structure and aggregation of the amyloid peptide Aβ 16- 22. Proteins: Structure, Function, and Bioinformatics.
7) Antila, H. S., **Kav, B.**, Miettinen, M. S., Martinez-Seara, H., Jungwirth, P., & Ollila, O. S. (2022). Emerging era of biomolecular membrane simulations: Automated physically-justified force field development and quality-evaluated databanks. The Journal of Physical Chemistry B, 126(23), 4169–4183.
8) Castillo, S. R., Rickeard, B. W., DiPasquale, M., Nguyen, M. H., Lewis-Laurent, A., Doktorova, M., **Kav, B.**, Miettinen, M. S., Nagao, M., Kelley, E. G., et al. (2022). Probing the link between pancratistatin and mitochondrial apoptosis through changes in the membrane dynamics on the nanoscale. Molecular Pharmaceutics, 19(6), 1839–1852.
9) Fatafta, H., **Kav, B.**, Bundschuh, B. F., Loschwitz, J., & Strodel, B. (2022). Disorder-to-order transition of the amyloid-β peptide upon lipid binding. Biophysical Chemistry, 280, 106700.
10) **Kav, B.**, & Strodel, B. (2022). Does the inclusion of electronic polarisability lead to a better modelling of peptide aggregation? RSC advances, 12(32), 20829–20837.
11) **Kav, B.**, Dem´e, B., Gege, C., Tanaka, M., Schneck, E., & Weikl, T. R. (2021). Interplay of trans-and cis-interactions of glycolipids in membrane adhesion. Frontiers in molecular biosciences, 8, 754654.
12) **Kav, B.**, Grafmüller, A., Schneck, E., & Weikl, T. R. (2020). Weak carbohydrate–carbohydrate interactions in membrane adhesion are fuzzy and generic. Nanoscale, 12(33), 17342–17353.
13) Samantray, S., Yin, F., **Kav, B.**, & Strodel, B. (2020). Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. Journal of chemical information and modeling, 60(12), 6462–6475.
14) Antila, H., Buslaev, P., Favela-Rosales, F., Ferreira, T. M., Gushchin, I., Javanainen, M., **Kav, B.**, Madsen, J. J., Melcr, J., Miettinen, M. S., et al. (2019). Headgroup structure and cation binding in phosphatidylserine lipid bilayers. The Journal of Physical Chemistry B, 123(43), 9066–9079.
15) **Kav, B.**, Öztürk, M., & Kabakçıoğlu A. (2016). Function changing mutations in glucocorticoid receptor evolution correlate with their relevance to mode coupling. Proteins: Structure, Function, and Bioinformatics, 84(5), 655–665.
16) Herling, T. W., Garcia, G. A., Michaels, T. C., Grentz, W., Dean, J., Shimanovich, U., Gang, H., Müller, T., **Kav, B.**, Terentjev, E. M., et al. (2015). Force generation by the growth of amyloid aggregates. Proceedings of the National Academy of Sciences, 112(31), 9524–9529.