Egor Marin

Groningen, Netherlands | me@marinegor.dev | github.com/marinegor | linkedin.com/in/marinegor | marinegor.dev

Cover paragraph

Machine learning scientist coming from biophysics / computational biology background. Have lots experience with cheminformatics and small-molecule property prediction models, a paper in active learning for ultra-large virtual ligand screening, and a position as a core developer in MDAnalysis (most popular open-source package for molecular dynamics analysis).

Technology-wise, love (typed) python and recently Rust, jax data model, and toml over yaml for configs.

Work Experience

Machine Learning Scientist May 2024 — Present
ENPICOM B.V. *Den Bosch, Netherlands*

- full-cycle ML model development: from literature survey and data collection and cleaning to reproducible training and deployment
- working with both generative and predictive models for various tasks in the antibody development field

Open-source software engineer May 2023 — September 2023
MDAnalysis via Google Summer of Code *Remote*

- designed and wrote backward-compatible parallelization (dask / multiprocessing) for molecular dynamics trajectory analysis

PhD Researcher June 2021 — Dec 2023
University of Groningen *Groningen, Netherlands*

- full-cycle membrane protein biochemistry: protein expression, purification, nanodisc reconstitution or crystallization
- processed cryoEM data (80 CryoSPARC projects), set up cryoEM data processing & management infrastructure
- supervised a project as a corresponding author: "Regression-Based Active Learning for Accessible Acceleration of Ultra-Large Library Docking"

Junior Research Associate March 2017 — September 2021
Moscow Institute of Physics and Technology *Moscow, Russia*

- last-mile protein crystallography service: crystal harvesting, data management, synchrotron and XFEL data collection
- refined, analysed and deposited protein structures to PDB (38 structures as of September 2025)
- coordinated data analysis and manuscript preparation

Scientific Journalist Jun 2016 — Aug 2017
Moscow Institute of Physics and Technology *Moscow, Russia*

- wrote press-releases on published papers
- communicated with scientists & media.

Skills

- **Programming Languages**: Python, bash, Rust, C++, Typescript
- **Python**: uv/ruff/ty❤️‍🔥, pytest, hypothesis, pydantic, dask
- **Data Science**: polars🫶, huggingface🤗, pandas/numpy/sklearn/skrub
- **Visualization**: altair, marimo, matplotlib/seaborn/jupyter
- **Deep Learning**: pytorch, lightning, jax, mlflow
- **ML in biology**: protein language models, diffusion/discrete diffusion/flow matching, continuous diffusion models (Alphafold/Openfold/Boltz)
- **Structural biology**: cryoEM data processing and structure refinement, X-ray crystallography data collection, processing and refinement
- **Cheminformatics**: RDKit, polaris, molecular docking, structural bioinformatics (MDAnalysis/mdtraj/biotite)
- **Technologies**: AWS, Docker, k8s/SLURM, Modal, Airflow, Github actions

Selected publications

For full list, see google scholar.

Regression-Based Active Learning for Accessible Acceleration of Ultra-Large Library Docking
- *Journal of Chemical Information and Modeling*, December 2023
- proposed the project, supervised the study and wrote manuscript, performed ML benchmarks

Structural basis for receptor selectivity and inverse agonism in S1P5 receptors,
- *Nature Communications*, August 2022
- collected data at PAL XFEL, processed SFX data, refined model, performed Alphafold simulations, molecular docking & VLS benchmarks of available S1P models

Structural Aspects of Photopharmacology: Insight into the Binding of Photoswitchable and Photocaged Inhibitors to the Glutamate Transporter Homologue,
- *Journal of American Chemical Society*, January 2021
- performed molecular docking of photo-switchable compounds in crystallographic structure and compared docking results with functional data

Structure-Based Virtual Screening of Ultra-Large Library Yields Potent Antagonists for a Lipid GPCR,
- *Biomolecules*, December 2020
- prepared small-molecule libraries for docking, did large-scale docking and analyzed the results, wrote manuscript

Education

University of Groningen 2019 — 2023
PhD *Groningen, Netherlands*
- thesis 'On the methods of studying protein-ligand interaction dynamics'"
- methods: cryoEM, X-ray crystallography, molecular dynamics, protein biochemistry, protein biophysics
- publications in JACS, Crystal Growth & Design, Journal of Cheminformatics & Modelling

Computer Science Center 2020 — 2022
Full-time extracurricular educational program in computer science *St. Petersburg, Russia*
- relevant coursework: Python, C++, Algorithms and data structures, Data science, Intro to Linux systems, Rust

Moscow Institute of Physics and Technology 2017 — 2019
MSc in applied mathematics and physics *Moscow, Russia*
- graduated *summa cum laude*

Moscow Institute of Physics and Technology 2017 — 2019
BSc in applied mathematics and physics *Moscow, Russia*
- graduated *magna cum laude*
- related coursework: Calculus I-IV, Linear Algebra I-II, Complex Analysis, Differential Equations I-II, Analytical Mechanics I-II, Thermodynamics.

Extracurricular Activities

Open-source contributions Jan 2021 — Present
- polars: contributed to polars (issue #25383: extending `replace_many` with `leftmost` option.
- polars-distance: minor contribution to polars plugin for distance calculation
- reciprocalspaceship: wrote parser for serial crystallography data into binary dataframe-like class

MDAnalysis Core Developer February 2025 — Present
- MDAnalysis Core Developer
  - wrote a parallel backend for all analysis classes (dask/multiprocessing)
  - added a DSSP module for native secondary structure assignment
  - currently working on fast unified MMCIF parser based on `gemmi`

Data Science Competitions 2018 — Present
- top-10% in Kaggle "Predict Molecular Properties" (public notebooks + gradient boosting on self-written rotationally invariant features)
- top-1 in first round of "Learning How To Smell" at AIcrowd
- top-10% in Takeda competition at Signate
- 5th place in Tochka Bank graph ML competition