Egor Marin

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

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 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*

- wrote backward-compatible parallelization 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 2022 — Aug 2022
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
- **Deep Learning**: pytorch, lightning, jax, mlflow
- **ML in biology**: protein language models, diffusion/discrete diffusion/flow matching, Alphafold/Openfold/Boltz
- **Technologies**: AWS, Docker, SLURM, Modal, Airflow, Github actions
- **Cheminformatics**: RDKit, polaris, molecular docking, structural bioinformatics (MDAnalysis/mdtraj/biotite)

Publications

CryoRhodopsins: a comprehensive characterization of a new clade of microbial rhodopsins from cold environments, Science Advances 7 July 2025

- Prepared samples for cryoEM (nanodisc reconstitution), processed cryoEM data and organized data collection.

Regression-Based Active Learning for Accessible Acceleration of Ultra-Large Library Docking, Journal of Chemical Information and Modeling 29 December 2023

- Proposed the project, supervised the study and wrote manuscript, performed ML benchmarks.

Custom Design of a Humidifier Chamber for In Meso Crystallization, Crystal Growth & Design12 December 2023

- Purified and crystallized protein, performed in meso crystallization, processed data, refined the structure, analyzed cryo-EM data, wrote the manuscript.

Structural insights into thrombolytic activity of destabilase from medicinal leech, Scientific Reports24 April 2023

- Crystallized the protein, collected data, solved the structures, performed molecular dynamics simulations and analyzed the results.

Structural basis for receptor selectivity and inverse agonism in S1P5 receptors, Nature Communications 12 August 2022

- Collected data at PAL XFEL, processed SFX data, refined model, performed Alphafold simulations, molecular docking & VLS benchmarks of available S1P models.

Metabolic fate of human immunoactive sterols in Mycobacterium tuberculosis, Journal of Molecular Biology 19 February 2021

- Collected crystallography data, supervised model refinement, wrote manuscript

Structural Aspects of Photopharmacology: Insight into the Binding of Photoswitchable and Photocaged Inhibitors to the Glutamate Transporter Homologue, Journal of American Chemical Society 15 January 2021

- Did 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 3 December 2020

- Prepared small-molecule libraries for docking, did large-scale docking and analyzed the results, wrote manuscript.

Small-wedge synchrotron and serial XFEL datasets for cysteinyl leukotriene GPCRs, Scientific Data12 November 2020

- Organized, annotated and deposited raw data, developed robust re-processing algorithms, wrote manuscript.

Molecular mechanism of light-driven sodium pumping, Nature Communications 1 May 2020

- Processed serial synchrotron crystallography data using CrystFEL, deposited raw data.

Structural basis of ligand selectivity and disease mutations in cysteinyl leukotriene receptors, Nature Communications 6 December 2019

- Collected small-wedge serial synchrotron crystallography data, processed them, refined, deposited and analyzed structures.

Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs, Science Advances 1 October 2019

- Collected synchrotron crystallography data, processed XFEL and synchrotron data, refined, deposited and analyzed structures.

An outlook on using serial femtosecond crystallography in drug discovery, Expert Opinion on Drug Discovery11 June 2019

- Wrote sections about SFX data processing & phasing.

Structural insights into ion conduction by channelrhodopsin 2, Science 24 November 2017

- Data collection and processing for both WT and mutant proteins.

Education

University of Groningen 2019 — 2023

PhD *Groningen, Netherlands*

- thesis 'On the methods of studying protein-ligand interaction dynamics'"
- methods: cryoEM, X-ray crystallography, 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*

- managed bachelor and master students, created a course on modern protein crystallography
- publications in Science, Nature Communications, Science Advances, Scientific Data
- 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

- reciprocalspaceship: wrote parser for serial crystallography data into binary dataframe-like class
- ntfy-cryosparc: wrote web-server to parse CryoSPARC (tm) notifications and notify appropriate users
- polars-distance: minor contribution to polars plugin for distance calculation

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 an MMCIF parser based on `gemmi`

Self-hosting Jan 2021 — Present

- self-hosting bunch of open-source docker containers under Tailscale VPN

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