

Egor Marin

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

COVER PARAGRAPH

Scientist with strong computational and biophysical background and tendency towards writing reusable code. I have eight years of experience as a structural biology/bioinformatics scientist, with h-index of 13 and experience ranging from data processing in serial crystallography and cryoEM to wet-lab sample preparation and virtual ligand screening.

As a structural biologist, have 400+ hours of crystallography beamtime, 20+ Tb of processed crystallography data and **80 processed projects in CryoSPARC** instance that I also set up and managed for 15 users on multiple workstations and a SLURM cluster. As of January 2026, have **38 deposited PDB structures** across 16 different publications. Notably, for faster processing of the projects wrote a [web-server](#) for personalised CryoSPARC notifications.

I also have good knowledge of Linux systems, supported by 4 years of management of various workstations and servers for more than 30 users. I regularly contribute to structural biology and data science open-source (projects such as polars, MDAnalysis, reciprocalspaceship).

Now employed as Machine Learning Scientist, working on full-cycle predictive and generative model development for an antibody drug-discovery platform startup.

SELECTED PROJECTS

Scientist, Structure-function relationship of cysteinyl leukotriene receptors Sep 2017 — Sep 2020

- did last-mile crystallography service from membrane protein crystals to refined PDB structures and wrote manuscripts collaborating with multiple research groups
- publications in Nature Communications, Science Advances and Biomolecules

Structural biologist, Structural characterisation of microbial rhodopsins Sep 2020 — Sep 2023

- reconstituted membrane proteins in nanodiscs and processed cryoEM data subsequently
- publications in Nature Communications and Science Advances

Software developer, Implementation of parallel analysis in MDAnalysis May 2023 — Sep 2023

- implemented backward-compatible parallelization for the most popular molecular dynamics analysis package; continue contributing as a core developer

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)

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](#).

- Science Advances** July 2025
CryoRhodopsins: a comprehensive characterization of a new clade of microbial rhodopsins from cold environments
- Pre-print on biorxiv** 29 April 2025
Structural basis for no retinal binding in flotillin-associated rhodopsins
- Nature Communications** 4 October 2024
A subgroup of light-driven sodium pumps with an additional Schiff base counterion
- Journal of Chemical Information and Modeling** December 2023
Regression-Based Active Learning for Accessible Acceleration of Ultra-Large Library Docking
- Nature Communications Chemistry** 2 May 2023
Mirror proteorhodopsins
- Nature Communications** August 2022
Structural basis for receptor selectivity and inverse agonism in S1P5 receptors
- Biomolecules** December 2020
Structure-Based Virtual Screening of Ultra-Large Library Yields Potent Antagonists for a Lipid GPCR
- Structural basis of ligand selectivity and disease mutations in cysteinyl leukotriene receptors** December 2019
Nature Communications
- Science Advances** October 2019
Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs

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
- 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`

Self-hosting

Jan 2021 — Present

- self-hosting multiple applications (vikunja, llama.cpp wrappers, openwebui) for family use under a 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