Egor Marin, ML Engineer/Scientist

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

Summary

- ML Scientist with a PhD in Structural Biology and 8+ years of research experience
- built and managed ML/DL codebase with data and model registry in a team of scientists and engineers
- specialized in building generative and predictive models for antibody drug discovery and structural biology
- managed complete research data lifecycles, from cryo-EM/crystallography data processing (20TB+) to deploying ML-driven analysis tools
- Core Developer for MDAnalysis, contributing major features like a parallel processing backend (dask/multiprocessing)

Experience

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

- develop and deploy generative and predictive models for antibody drug discovery using pytorch/lightning and custom discrete diffusion pipelines
- own the end-to-end ML lifecycle, from sourcing and cleaning biochemical data to building reproducible training pipelines with MLflow for experiment tracking
- manage codebase with data and training pipelines, ensuring regular code reviews and code reusability while maintaining modularity and compatibility of the models

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

- designed and wrote backward-compatible parallelization with dask or multiprocessing for the core analysis library

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

- led a research project as a corresponding author, applying regression-based active learning to accelerate ultra-large library docking, resulting in a publication in J. Chem. Inf. Model
- designed, deployed, and managed a high-throughput data processing infrastructure (CryoSPARC) for 15+ users, processing over 80 cryo-EM projects and 20TB of data
- developed a custom web server (ntfy-cryosparc) to provide personalized notifications, improving workflow efficiency for the research group
- solved 38+ protein structures using cryo-EM and X-ray crystallography, leading to publications in journals including Nature Communications and Science Advances

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

- managed the end-to-end protein crystallography pipeline for multiple research groups, from sample preparation to synchrotron and XFEL data collection
- processed, refined, and analyzed crystallographic data, successfully determining and depositing 38 protein structures to the PDB
- contributed structural data and analysis that formed the basis for 16 peer-reviewed publications

Skills

Technology: python | rust | pytorch | jax | scipy | numpy | sklearn | polars | pydantic | mlflow | docker | AWS | k8s | SLURM

Domain: cryoEM | X-ray crystallography | biophysics | drug discovery | cheminformatics | antibodies

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 2013 — 2017
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.

Open Source Contributions & Projects

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 (paperless, llama.cpp, 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-10% in Takeda competition at Signate
- 5th place in Tochka Bank graph ML competition