

Egor Marin, ML Engineer

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
- specialized in building generative and predictive models (diffusion/PLMs with pytorch & jax) 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 ENPICOM B.V.	May 2024 — Present <i>Den Bosch, Netherlands</i>
<ul style="list-style-type: none">• 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	
Open-source software engineer MDAnalysis via Google Summer of Code	May 2023 — September 2023 <i>Remote</i>
<ul style="list-style-type: none">• designed and wrote backward-compatible parallelization with dask or multiprocessing for the core analysis library	
PhD Researcher University of Groningen	June 2021 — Dec 2023 <i>Groningen, Netherlands</i>
<ul style="list-style-type: none">• 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 Moscow Institute of Physics and Technology	March 2017 — September 2021 <i>Moscow, Russia</i>
<ul style="list-style-type: none">• 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 <i>CryoRhodopsins: a comprehensive characterization of a new clade of microbial rhodopsins from cold environments</i>	July 2025
Pre-print on biorxiv <i>Structural basis for no retinal binding in flotillin-associated rhodopsins</i>	29 April 2025

Nature Communications	4 October 2024
<i>A subgroup of light-driven sodium pumps with an additional Schiff base counterion</i>	
Journal of Chemical Information and Modeling	December 2023
<i>Regression-Based Active Learning for Accessible Acceleration of Ultra-Large Library Docking</i>	
Nature Communications Chemistry	2 May 2023
<i>Mirror proteorhodopsins</i>	
Nature Communications	August 2022
<i>Structural basis for receptor selectivity and inverse agonism in S1P5 receptors</i>	
Biomolecules	December 2020
<i>Structure-Based Virtual Screening of Ultra-Large Library Yields Potent Antagonists for a Lipid GPCR</i>	
Structural basis of ligand selectivity and disease mutations in cysteinyl leukotriene receptors	December 2019
<i>Nature Communications</i>	
Science Advances	October 2019
<i>Structure-based mechanism of cysteinyl leukotriene receptor inhibition by antiasthmatic drugs</i>	

EDUCATION

University of Groningen	2019 — 2023
<i>PhD</i>	<i>Groningen, Netherlands</i>
<ul style="list-style-type: none"> 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
<i>MSc in applied mathematics and physics</i>	<i>Moscow, Russia</i>
<ul style="list-style-type: none"> graduated <i>summa cum laude</i> 	
Moscow Institute of Physics and Technology	2013 — 2017
<i>BSc in applied mathematics and physics</i>	<i>Moscow, Russia</i>
<ul style="list-style-type: none"> graduated <i>magna cum laude</i> 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
<ul style="list-style-type: none"> polars: contributed to polars (issue #25383: extending <code>replace_many</code> with <code>leftmost</code> 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
<ul style="list-style-type: none"> MDAnalysis Core Developer <ul style="list-style-type: none"> 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 <code>gemmi</code> 	
Self-hosting	Jan 2021 — Present
<ul style="list-style-type: none"> self-hosting multiple applications (paperless, llama.cpp, openwebui) for family use under a Tailscale VPN 	
Data Science Competitions	2018 — Present
<ul style="list-style-type: none"> 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 	