[marinegor.dev](#)
[All posts](#) [TIL](#) [Publications](#) [CV](#) [Tags](#) [All pages](#)

Curriculum Vitae aka CV aka Resume

Posted on Sep 20, 2025

Egor Marin

Machine Learning Scientist @ [ENPICOM B.V.](#), working with protein language models, bioinformatics, diffusion – you name it; `@marinegor` at most of the platforms.

PDF version of this CV is [here](#)

- [Profile](#)
- [Socials](#)
- [Career](#)
    - [2024-current: Machine Learning Scientist](#)
    - [2017-2023: Scientist](#)
    - [2016-2017: Scientific Journalist](#)
- [Software skills & activities](#)
- [Science skills & acitivities](#)
- [Education](#)

Profile

I have formal education in applied mathematics and physics (BSc + MSc), one year of full-time extracurricular education in computer science, PhD in biophysics and structural biology, and 8 years of computational lab experience.

I enjoy writing code and building complex systems, and want to do that for the things that get to run many times, and hence should be designed and written wisely. I know a lot about (computational) biology, mostly on molecular level (structural biology, protein biochemistry). And I can communicate with people, including mentoring or leading small teams, at least that's what I have had experience with so far.

Socials

- email: [me@marinegor.dev](#)
- [github](#)
- [twitter](#)
- [linkedin](#)
- [bluesky](#)
- [google scholar](#)
- [ORCID](#)

Career

I have been roughly 8 years in science, working with membrane proteins and their structure-function relationships: GPCRs, (microbal) rhodopsins, membrane transporters, P450 enzymes, antibodies and nanobodies.

I am currently employed at ENPICOM B.V., and before that have worked at the University of Groningen and Moscow Institute of Physics and Technology. Also, I have worked at many synchrotrons and XFELs, and also was a visiting research assistant at the University of Southern California.

2024-current: Machine Learning Scientist

Doing machine learning in biotech-oriented SaaS company.

- 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

2017-2023: Scientist

- conducted research, managed data, wrote publications, participated in conferences
- managed students (BSs & MSc diploma), created a course on modern protein crystallography

2016-2017: Scientific Journalist

- analyzed publicational activity of MIPT
- wrote press-releases on published papers
- communicated with scientists & media.

Software skills & activities

Bag of words: python, numpy/sklearn/pytorch/lightning, polars🐻‍❄️, huggingface🤗datasets and tokenizers, uv/ruff❤️‍🔥, pytest, docker/docker-compose, bash, mlflow, Ubuntu/nixos, HPC/SLURM/dask, prefect/modal/airflow.

🤓MDAnalysis Core Developer since February 2025. For MDAnalysis, 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.

🧑‍💻contributed to opensource: 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.

😎participated in Google Summer of Code contributing to MDAnalysis: introduced process-based parallelization to the library using `dask` or `multiprocessing` (see main PR).

🏆participated in data science competitions (top-10% in Kaggle "Predict Molecular Properties", top-1 in first round of "Learning How To Smell", top-10% in Takeda Signate competition, 5th place in Tochka Bank graph

ML competition).

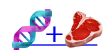💾administrated ~15 Linux workstations and servers with around 40 users, managing around 200 Tb of research data.

🍲performed large-scale calculations on SLURM and PBS, wrote bash scripts and pipelines or reliable and reproducible data processing of serial crystallography data.

🤷‍♂️self-hosted bunch of things: *arr, telegram bots, WebDAV, proxy & VPN servers, paperless, openwebui, you name it

🦀Wrote a python(pyo3)+Rust(pest) parser for crystallographic data, contributed to [polars-distance](#)

Science skills & acitivities

Bag of words: structural biology, crystallography, cryoEM, cheminformatics, computer vision, data science, molecular docking, drug discovery, protein structure, GPCRs, membrane proteins, structure-based drug discovery, antibodies, protein language models, discrete diffusion, flow matching, ~~Alpha~~Openfold

🧬➕🥩**structural biology**: co-published papers in Science, Nature Communications, JACS, Science Advances, Journal of Chemical Information and Modelling, Scientific Data. Performed data analysis, wrote texts, created figures, managed writing process – the normal stuff.

💊**structure-based drug discovery**: performed large-scale virtual screening campaign, created robust accelerated virtual screening [approach](#), communicated with CROs, oversaw functional tests.

🤖**machine learning**: did many ad-hoc ML applications in computer vision (background removal with NMF decomposition), clustering, supervised learning. Always try linear regression first, have a [paper](#) about it.

👾**deep learning**: know about protein language models and their properties, ~~Alpha~~Openfold-like models and their applications, wrote toy discrete diffusion models, adapted open-source discrete diffusion models for other domains.

Education

University of Groningen, 2019–2023 PhD, diploma on "On the methods of studying protein-ligand interaction dynamics"

Computer Science Center, 2020–2022 Full-time extracurricular education in computer science: 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, summa cum laude, with specialization in biophysics and structural biology

Moscow Institute of Physics and Technology, 2013–2017 BSc in applied mathematics and physics, magna cum laude

Last updated: September 2025.

2025 Egor Marin | [Archie Theme](#) | Built with [Hugo](#)