Gabriel CRETIN

Ph.D in AI for computational biology • Linux system administrator

Email
gabriel.cretin@u-paris.fr

Personal Website
https://gabrielcretin.fr

DSIMB Team Website
(that I developed)
https://dsimb.inserm.fr

Phone
(+33) 0781703515

Address
Paris, France

Social networks

Github
@gabrielctn

LinkedIn
gabriel-cretin

Education

- 2021 - 2025 Université Paris Cité, France
  Ph.D
  Computational Biology
  Deep learning and computational approaches for protein structure analysis, prediction, and generation
- 2017 - 2019 Université Paris Cité, France
  Master of Science (MSc)
  Biology, Computer Science, Bioinformatics (BIB)
  With honors • Rang 2/23
- 2016 - 2017 Université Paris Cité, France
  Bachelor of Science (BSc)
  Biology, Computer Science, Bioinformatics (BIB)
  Rank 6/21

Languages

French
Native speaker

English
Scientific
Cambridge English Advanced
TOEIC (Score 875)
Erasmus+ OLS (C1)

Experience

Sep 2019 - Present · Paris, France

Linux System Administrator Engineer
Université Paris Cité
Team DSIMB • U1134 unit
Team Leader
alexandre.debrevern@inserm.fr

Installation, management, and maintenance of a Linux IT infrastructure:
Web server and dedicated computing blades:
- 14 web services, 2 APIs, 10 databases (~18K users/year)
- Self-hosted Free and Open-Source Software (FOSS): GitLab, Mattermost, WikiJS, JupyterHub, GLPI

Shared computing cluster: 5 GPU servers (deep learning, molecular dynamics), 300+ CPUs and 800+ Gb of RAM
SGI HPC cluster: 48 nodes, 768 cores
Linux scientific workstations: 25 PCs, 16 GPUs, 700 CPUs, 1+ To RAM
Storage: 1+ Pb of storage (3 dedicated servers)

Sep 2021 - Oct 2025 · Paris, France

Ph.D in Computational Biology
Université Paris Cité
Team DSIMB • U1134 unit
Ph.D Supervisor
jean-christophe.gelly@u-paris.fr

Thesis: "Deep learning and computational approaches for protein structure analysis, prediction, and generation"

Multi-scale structural analysis: developed a comprehensive suite of tools for protein study, ranging from local conformation to global structure analysis:
- **PYTHIA:** Created a deep learning predictor for local protein conformations based on the Protein Blocks structural alphabet.
- **SWORD2:** Developed an interactive web server for the hierarchical decomposition of 3D structures into domains and Protein Units.
- **ICARUS:** Designed a non-sequential flexible structural alignment method utilizing Protein Peeling to identify rigid sub-units.
- **PEGASUS:** Built a sequence-based predictor of molecular dynamics-derived flexibility (RMSF, LDDT) leveraging Protein Language Models (pLMs).

Representation learning & Generative modeling:
- Designed and optimized an Adversarial Autoencoder (AAE) architecture to compress high-dimensional pLM embeddings (Ankh, ESM-2) into a fixed-size, continuous latent space.
- Implemented contrastive triplet learning to reconfigure latent spaces, significantly improving structural fold recognition and surpassing state-of-the-art structure-based methods.
- Explored de novo protein design through latent space interpolation, validating the generation of physically plausible protein sequences.

HPC & Technical implementation:
- Managed large-scale training workflows (40k+ A100/H100 GPU hours) on national supercomputers (IDRIS/CNRS).
- Engineered high-performance APIs for real-time pLM embedding generation and asynchronous web-server processing

---

Jan - Jun 2019 · Paris, France

Internship Master 2
Team DSIMB • BIGR U1134 unit

Development of a new strategy for the prediction of the 3D protein structure using Protein Units and Deep Learning.

Mar - May 2018 · Halle, Allemagne

Internship Master 1
Leibniz Institute Of Plant Biochemistry

Optimization of the metabolomics data processing parameters of XCMS with the R/Bioconductor package IPO as part of theEuropean project PhenoMeNal, a standardized e-infrastructure for large-scale computing of medical metabolomics data.

Apr - Jun 2016 · Clermont-Ferrand, France

Bioinformatics Internship
French National Institute for Agricultural Research

As part of the French project Workflow4Metabolomics, development of a tool for the annotation of GC-MS data exploited with Golm Metabolome Database, for its integration into the Galaxy platform.

Skills

Frequently used · Occasionnaly

Bioinformatics

Deep Learning
PyTorch · Lightning Fabric · HPC distributed GPU training

Protein sequence/structure analysis
HH-Suite · MMseqs2 · TM-align · AlphaFold · pLMs · PyMOL · Foldseek · GROMACS

Softwares - DevOps
CI/CD (Travis CI/Jenkins) · Containerization (Docker / Singularity) · Anaconda · VSCode · JupyterHub/Lab · GitLab

Data Analyses
Biostatistics (models, visualization, tests) with R

Databanks
PDB · SCOPe · CATH · EMBL-EBI · SwissProt · Pfam · NCBI · GenBank · Interpro · KEGG · UniProtKB

System administration

OS
GNU-Linux/Unix (Ubuntu, Debian) · macOS

Programming Languages
Python · Bash/Shell · R · C · Java · Perl · SQL (PostgreSQL, MySQL)

Web technologies
HTML5 · CSS3 · Bootstrap · JavaScript · PHP

Infrastructure as Code (IaC) - Virtualization
Ansible · Docker · Docker swarm · Kubernetes

Networking
EfficientIP (DHCP, IPAM)

Software - Configuration
JupyterHub · SLURM · Apache2 · Samba · NFS · OpenLDAP · Environment Modules · Spack · GLPI

Soft skills

Teaching & Mentoring
Co-supervision of Master level interns · Technical guidance for peers

Scientific Communication
Poster sessions · Oral presentations · Technical / Documentation writing

Critical Problem-Solving
Diagnosing infrastructure incidents (HPC/Linux) in production environments

Interdisciplinary Collaboration
Working effectively at the interface of AI, IT, and Biology

Hobbies

Coffee Enthusiast
Tech & Science Monitoring
Travel

Selected Publications

2025

medRXiv, 2025. · Scholar

PATHOS: Predicting Variant Pathogenicity by Combining Protein Language Models and Biological Features
Radjasandirane Ragousandirane, **Cretin Gabriel**, Diharce Julien, de Brevern Alexandre G, Gelly Jean-Christophe

Protein Science, 2025. · Scholar

PEGASUS: Prediction of MD-derived protein flexibility from sequence
Vander Meersche Yann*, Duval Gabriel*, **Cretin Gabriel***, Gheeraert Aria, Gelly Jean-Christophe, Galochkina Tatiana
**authors contributed equally*

2024

Nucleic Acids Research, 2024. · Scholar

ATLAS: protein flexibility description from atomistic molecular dynamics simulations
Vander Meersche Yann , **Cretin Gabriel** , Gheeraert Aria , Gelly Jean-Christophe and Galochkina Tatiana

2023

Bioinformatics, 2023. · Scholar

ICARUS: flexible protein structural alignment based on Protein Units
Cretin Gabriel , Périn Charlotte , Zimmermann Nicolas , Galochkina Tatiana , and Gelly Jean-Christophe

2022

Nucleic acids research, 2022. · Scholar

SWORD2: hierarchical analysis of protein 3D structures
Cretin Gabriel , Galochkina Tatiana , Vander Meersche Yann , de Brevern Alexandre G , Postic Guillaume , and Gelly Jean-Christophe

2021

International journal of molecular sciences, MDPI, 2021. · Scholar

PYTHIA: Deep learning approach for local protein conformation prediction
Cretin Gabriel , Galochkina Tatiana , de Brevern Alexandre G , and Gelly Jean-Christophe