

[English version below]

Proposition de sujet de thèse de doctorat sur le thème de l'analyse topologique de données en biologie moléculaire et biologie évolutive

Titre du sujet de thèse. Analyse topologique des structures tridimensionnelles de protéines et réseaux de neurones topologiques pour l'étude de traits d'histoire de vie des organismes.

Directeur. Philippe Malbos, Professeur, Université Claude Bernard Lyon 1, Institut Camille Jordan (UMR CNRS 5208).

Co-encadrement. Philippe Oger, Directeur de recherches CNRS, Laboratoire de Microbiologie, Adaptation et Pathogénie (UMR CNRS 5240), Céline Brochier-Armanet, Professeur, Université Claude Bernard Lyon 1, Laboratoire de Biométrie et Biologie Évolutive (UMR CNRS 5558).

Période de la thèse. Octobre 2025 – Septembre 2028.

Contexte scientifique

L'Institut Camille Jordan (ICJ, UMR CNRS 5208, CNRS, Université Claude Bernard Lyon 1) propose une thèse de doctorat à temps plein sur le thème « *Analyse topologique des structures tridimensionnelles de protéines et réseaux de neurones topologiques pour l'étude des traits d'histoire de vie des organismes* ».

Cette thèse vise à développer des méthodes à l'interface de la biotopologie moléculaire et de l'apprentissage automatique, afin d'identifier les déterminants moléculaires et structuraux de la piézo-résistance.

La biotopologie moléculaire consiste à étudier les macromolécules biologiques, essentiellement les protéines, et leurs propriétés fonctionnelles à l'aide de techniques issues de l'analyse topologique des données (TDA) et de la géométrie spectrale. Dans cette thèse, il s'agira de proposer de nouvelles représentations multi-échelles de ces macromolécules et d'en déterminer les caractéristiques géométriques (marqueurs de courbure) et spectrales (marqueurs spectraux). Ces marqueurs, construits à partir de courbures discrètes (Forman-Ricci et Ollivier-Ricci), du Laplacien persistant et de facteurs structuraux, génomiques et environnementaux, constituent des représentations vectorielles des macromolécules.

Des métriques adaptées seront définies afin d'entraîner des modèles prédictifs sur ces représentations, en combinant apprentissage profond et régression statistique, pour mieux comprendre les mécanismes de la piézo-résistance.

La thèse portera également sur la modélisation de l'évolution de ces marqueurs, en prenant en compte les traits d'histoire de vie des organismes, par des méthodes combinant phylogénie moléculaire et reconstruction de séquences ancestrales. L'objectif est de retracer l'évolution conjointe des structures, des séquences protéiques et des marqueurs topologiques le long des arbres phylogénétiques afin de tester les modèles proposés.

Cette thèse combine ainsi modélisation mathématique, biologie évolutive et apprentissage automatique pour explorer les mécanismes de l'adaptation aux hautes pressions.

Les résultats attendus incluent :

- Le développement de modèles d'apprentissage pour prédire les traits d'histoire de vie à partir de la structure des protéines ;
- L'identification de signatures moléculaires associées à la piézo-résistance ;
- L'intégration des avancées théoriques et logicielles dans l'environnement DeltaFold.

Environnement de travail

Le projet de thèse s'inscrit dans le cadre d'une approche interdisciplinaire permettant de modéliser et prédire les traits d'histoire de vie des organismes à partir des structures tridimensionnelles des protéines.

Il est financé dans le cadre d'un projet de la Mission pour les Initiatives Transverses et Interdisciplinaires (MITI) du CNRS.

La thèse se déroulera au sein d'une équipe pluridisciplinaire sur le campus scientifique LyonTech-la Doua, à Lyon. Cette équipe regroupe des mathématiciens et des biologistes issus de trois laboratoires : l'Institut Camille Jordan (ICJ, UMR CNRS 5208, CNRS, Université Claude Bernard Lyon 1), le Laboratoire de Biométrie et Biologie Évolutive (LBBE, UMR CNRS 5558, CNRS, Université Claude Bernard Lyon 1, VetAgro Sup), et le Laboratoire de Microbiologie, Adaptation et Pathogénie (MAP, UMR CNRS 5240, CNRS, Université Claude Bernard Lyon 1, INSA Lyon).

La thèse s'inscrira dans le cadre du projet DeltaFold, qui vise à développer de nouveaux outils mathématiques, algorithmiques et logiciels combinant topologie, apprentissage automatique et biologie moléculaire, afin d'identifier les signatures structurales des protéines associées aux traits d'histoire de vie.

Les méthodes développées dans le cadre de la thèse seront implémentées dans la bibliothèque DeltaFold, développée en Python par une équipe interdisciplinaire de l'ICJ et du LBBE. Cette bibliothèque constitue un environnement de développement pour la programmation de modèles prédictifs en biologie évolutive, fondés sur des vectorisations des structures tridimensionnelles de protéines, via l'homologie persistante.

Les modèles prédictifs piézométriques seront validés par comparaison avec des résultats expérimentaux produits au sein du laboratoire MAP. Ces validations s'appuieront sur deux modèles procaryotes : les Thermococcales (81 espèces, chacune correspondant à environ 2 000 protéines, soit un jeu de données d'environ 162 000 protéines) et les Alteromonadales (> 1 000 espèces, chacune correspondant à environ 6 000 protéines) pour lesquels il conviendra de sélectionner un sous-ensemble réaliste centré sur les genres présentant des organismes piézophiles, *Shewanella* et *Colwellia*.

Candidatures

Nous recherchons un candidat (H/F), disposant d'un solide dossier académique et titulaire d'un Master en mathématiques ou en bioinformatique. Une excellente formation théorique est indispensable. Une expérience préalable en analyse topologique des données (TDA), statistiques ou sciences des données constituera un atout apprécié.

Compétences et qualifications :

- Master ou équivalent en mathématiques ou bioinformatique.
- Bonne maîtrise de l'anglais écrit et oral.
- Capacités organisationnelles et aptitude à s'investir dans un travail interdisciplinaire et collaboratif.
- Volonté de participer à des séminaires, ateliers et réunions scientifiques.

Les candidats (H/F) seront intégrés dans une équipe pluridisciplinaire. Le candidat (H/F) doit donc posséder d'excellentes compétences personnelles et être capable de travailler en équipe.

Une présélection sera réalisée sur la base du CV, de l'expérience, des compétences et de la lettre de motivation. Les candidats présélectionnés seront convoqués pour une audition devant un comité de sélection.

Les candidatures doivent comprendre les documents suivants :

- Un Curriculum Vitae complet ;
- Une lettre de motivation (1 page), exposant les intérêts de recherche du candidat en lien avec le projet ;
- Une ou deux lettres de recommandation ;
- Les relevés de notes de Master (M1 et M2) ou équivalent, ainsi que tout certificat pertinent.

Pour postuler : <https://emploi.cnrs.fr/Offres/Doctorant/UMR5208-PHIMAL-002/Default.aspx>

Contact : malbos (-at-) univ-lyon1.fr

Proposal for a PhD thesis on topological data analysis in molecular biology and evolutionary biology

Thesis title. Topological analysis of three-dimensional protein structures and topological neural networks for studying life-history traits in organisms.

Director. Philippe Malbos, Professor, Université Claude Bernard Lyon 1, Institut Camille Jordan (UMR CNRS 5208).

Co-supervision. Philippe Oger, CNRS Research Director, Laboratoire de Microbiologie, Adaptation et Pathogénie (UMR CNRS 5240), Céline Brochier-Armanet, Professor, Université Claude Bernard Lyon 1, Laboratoire de Biométrie et Biologie Évolutive (UMR CNRS 5558).

Thesis period. October 2025 - September 2028.

Scientific background

The Institut Camille Jordan (ICJ, UMR CNRS 5208, CNRS, Université Claude Bernard Lyon 1) is offering a full-time PhD thesis on "*Topological analysis of three-dimensional protein structures and topological neural networks for studying life-history traits in organisms*".

This thesis aims to develop methods at the interface of molecular biotopology and machine learning, in order to identify the molecular and structural determinants of piezoresistance.

Molecular biotopology is the study of biological macromolecules, mainly proteins, and their functional properties, using techniques derived from topological data analysis (TDA) and spectral geometry. The aim of this thesis is to propose new multiscale representations of these macromolecules and to determine their geometric (curvature markers) and spectral (spectral markers) characteristics. These markers, based on discrete curvatures (Forman-Ricci and Ollivier-Ricci), the persistent Laplacian and structural, genomic and environmental factors, constitute vector representations of macromolecules.

Relevant metrics will be defined to train predictive models on these representations, combining deep learning and statistical regression, to better understand the mechanisms of piezo-resistance.

The thesis will also focus on modeling the evolution of these markers, taking into account the life-history traits of organisms, using methods combining molecular phylogeny and ancestral sequence reconstruction. The aim is to trace the joint evolution of structures, protein sequences and topological markers along phylogenetic trees, in order to test the proposed models.

This thesis combines mathematical modeling, evolutionary biology and machine learning to explore the mechanisms of adaptation to high pressure.

Expected results include:

- The development of learning models to predict life-history traits from protein structure;
- Identification of molecular signatures associated with piezoresistance ;
- Integration of theoretical and software advances in the DeltaFold environment.

Working environment

The thesis project is part of an interdisciplinary approach to modeling and predicting the life-history traits of organisms based on the three-dimensional structures of proteins. It is funded as part of a CNRS Mission pour les Initiatives Transverses et Interdisciplinaires (MITI) project.

The thesis will be carried out as part of a multidisciplinary team on the LyonTech-la Doua science campus in Lyon. The team includes mathematicians and biologists from three laboratories: Institut Camille Jordan (ICJ, UMR CNRS 5208, CNRS, Université Claude Bernard Lyon 1), Laboratoire de Biométrie et Biologie Évolutive (LBBE, UMR CNRS 5558, CNRS, Université Claude Bernard Lyon 1, VetAgro Sup), and Laboratoire de Microbiologie, Adaptation et Pathogénie (MAP, UMR CNRS 5240, CNRS, Université Claude Bernard Lyon 1, INSA Lyon).

The thesis will be part of the DeltaFold project, which aims to develop new mathematical, algorithmic and software tools combining topology, machine learning and molecular biology, in order to identify the structural signatures of proteins associated with life-history traits.

The methods developed in this thesis will be implemented in the DeltaFold library, developed in Python by an interdisciplinary team from ICJ and LBBE. This library provides a development environment for programming predictive models in evolutionary biology, based on vectorizations of three-dimensional protein structures, via persistent homology.

The piezometric predictive models will be validated by comparison with experimental results produced within the MAP laboratory. These validations will be based on two prokaryotic models: Thermococcales (81 species, each corresponding to around 2,000 proteins, i.e. a dataset of around 162,000 proteins) and Alteromonadales (> 1,000 species, each corresponding to around 6,000 proteins), for which a realistic subset centered on the genera presenting piezophilic organisms, *Shewanella* and *Colwellia*, will be selected.

Applications

We are looking for a candidate (M/F) with a solid academic record and a Master's degree in mathematics or bioinformatics. An excellent theoretical background is essential. Previous experience in topological data analysis (TDA), statistics or data science will be a welcome asset.

Skills and qualifications:

- Master's degree or equivalent in mathematics or bioinformatics.
- Good command of written and spoken English.
- Organizational skills and ability to engage in interdisciplinary and collaborative work.
- Willingness to participate in seminars, workshops and scientific meetings.

Candidates will be integrated into a multidisciplinary team. Candidates must have excellent personal skills and be able to work as part of a team.

Pre-selection will be based on CV, experience, skills and cover letter. Shortlisted candidates will be invited to audition before a selection committee.

Applications must include the following documents:

- A complete Curriculum Vitae ;
- A cover letter (1 page) outlining the applicant's research interests in relation to the project;
- One or two letters of recommendation;
- Transcripts of your Master's degree (M1 and M2) or equivalent, as well as any relevant certificates.

To apply: <https://emploi.cnrs.fr/Offres/Doctorant/UMR5208-PHIMAL-002/Default.aspx>

Contact: malbos (-at-) univ-lyon1.fr