Computational protein design with hybrid physical and data-driven approaches
PhD position

1 Project overview

Proteins are central to virtually all biological processes, and the ability to rationally design protein sequences and structures remains a major challenge in computational biology. Despite recent advances in protein structure prediction and sequence design, current computational methods still face limitations in terms of accuracy, generalization, and interpretability.

This PhD project aims to develop and apply novel computational and methodological approaches for protein design, combining traditional physics-based biomolecular modeling with modern data-driven techniques. The project will focus on improving the reliability and efficiency of computational protein design pipelines, with potential applications in biotechnology and molecular engineering.

2 Scientific context

Protein design aims to conceive new proteins or modify existing ones to obtain a given function. Computational approaches are a valuable help, to rationalize the predictions and guide experimental tests. Computational protein design relies on the accurate modeling of protein structure, dynamics, and energetics. Physics-inspired approaches [1–3], and data-driven methods [4, 5] have led to important breakthroughs, such as the creation of a protein with a new fold [6] or enzymes with new catalytic activities [7, 8]. Challenges remain in sampling sequence and conformational space, scoring designed sequences according to complex criteria, transferring methods across different protein systems and problems, and model interpretability.

3 Objectives

The main objective of this PhD is to contribute to the development of next-generation computational methods for robust and transferable protein design. Specific objectives include: 1) improve energy functions, scoring schemes, or sampling strategies used in protein design; 2) develop and implement computational strategies for hybrid physics and data-driven approaches; 3) benchmark and validate design methods on representative protein systems.

4 Methodology

The project will primarily rely on in silico approaches, including but not limited to: computational protein design frameworks (physics-based or deep learning approaches), molecular modeling and simulation methods, statistical analysis of protein sequence–structure relationships, high-performance computing and algorithmic optimization. The exact methodological focus may be adapted depending on the candidate's background and interests.

5 Research environment

The PhD will be carried out within the SyBioS team, a research group specialized in biomolecular modeling and computational structural biology, hosted at Ecole Polytechnique, a leading French academic institution. The doctoral candidate will benefit from a stimulating interdisciplinary environment, access to high-performance computing facilities, interactions with experimental collaborators, and training provided by the doctoral school.

6 Candidate profile

Required qualifications include a master's degree (or equivalent) in computational biology, bioinformatics, physics, chemistry, or applied mathematics, a strong background in quantitative or computational methods, programming skills, a strong motivation for methodological research in computational biology. Additional desired skills include experience in biomolecular modeling or protein structure analysis, familiarity with molecular mechanics or protein design tools, knowledge of statistical mechanics or machine learning, ability to work independently and collaboratively in an international environment.

7 Funding and practical information

The PhD position is subject to the candidate being awarded a doctoral school fellowship, following its competitive selection process. The position will start on 1 October 2026 and will last for 36 months. Salary is according to French doctoral school regulations. Location is Palaiseau (20 km from Paris), France. Working language is either English or French.

8 Application procedure

Interested candidates should submit the following documents: curriculum vitae, motivation letter describing research interests and relevant experience, academic transcripts (master's level), and contact information for 1–2 references. Applications should be sent to Thomas Gaillard (thomas.gaillard@polytechnique.edu) before 10 February 2026.

References

[1] Kuhlman, B.; Baker, D. Native Protein Sequences Are Close to Optimal for Their Structures. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10383–10388.

[2] Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P. et al. Macromolecular Modeling and Design in Rosetta: Recent Methods and Frameworks. *Nat. Methods* **2020**, *17*, 665–680.

[3] Mignon, D.; Druart, K.; Michael, E.; Opuu, V.; Polydorides, S.; Villa, F.; Gaillard, T.; Panel, N.; Archontis, G.; Simonson, T. Physics-Based Computational Protein Design: An Update. *J. Phys. Chem. A* **2020**, *124*, 10637–10648.

[4] Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N. et al. Robust Deep Learning-based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378*, 49–56.

[5] Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F. et al. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620*, 1089–1100.

[6] Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **2003**, *302*, 1364–1368.

[7] Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O. et al. Kemp Elimination Catalysts by Computational Enzyme Design. *Nature* **2008**, *453*, 190–195.

[8] Kim, D.; Woodbury, S. M.; Ahern, W.; Tischer, D.; Kang, A.; Joyce, E.; Bera, A. K.; Hanikel, N.; Salike, S.; Krishna, R. et al. Computational Design of Metallohydrolases. *Nature* **2026**, *649*, 246–253.