LAUREA MAGISTRALE IN BIOINFORMATICS
INTERNATIONAL BOLOGNA MASTER IN BIOINFORMATICS
ALMA MATER STUDIORUM ▪ UNIVERSITÀ DI BOLOGNA

Application Form

Surname: Zianni **Name:** Domenico

Beginning date of the internship: 15/02/2026

Host laboratory: Acibadem University

Scientific supervisor: prof. Ugur Sezerman

Proposed thesis title: Graph-Based Deep Learning for Predicting and Interpreting Peptide–Molybdenum Binding from Experimental Elution Data

Enclosed documents:

- short summary of the thesis project (page 2).
- acceptance letter of the scientific supervisor (by email or signed PDF) .

Bologna, 08/01/2026

Requirements:

- Methods (Minimum length 400 words)
- 1st Report: after two months from the beginning date (Minimum length 400 words), including Methods (Minimum length 400 words).
- 2nd Report: after four months from the beginning date (Minimum length 400 words)
- 3rd Report: a month before delivering the thesis to the Bologna student secretary (Minimum length 400 words)

Notes:

- The written final version should be given to the "Commissione Didattica" of the CdL in Bioinformatics one month before the official presentation to the student's secretary for the final approval by the tutor.
- The tutor is the thesis coordinator when the work is performed in one of the laboratories of the Bologna University supporting the Laurea Magistrale in Bioinformatics; the tutor is one of the faculty of the CdL in Bioinformatics, assigned to the student by the "Commissione Didattica" upon presentation for the approval of the internship-thesis work; the local tutor is supposed to support the student and his work through the whole internship-thesis period performed in any laboratory outside the Bologna University and is approving all the reports.

LAUREA MAGISTRALE IN BIOINFORMATICS
INTERNATIONAL BOLOGNA MASTER IN BIOINFORMATICS
ALMA MATER STUDIORUM ▪ UNIVERSITÀ DI BOLOGNA

Thesis project summary

Student (Name, Surname): Domenico Zianni

Proposed thesis title: Graph-Based Deep Learning for Predicting and Interpreting Peptide–Molybdenum Binding from Experimental Elution Data

Background (100-200 words): Peptide–metal interactions play a crucial role in biological catalysis, materials science, and biotechnology, where short peptides are often engineered to selectively bind metal ions or metal-containing surfaces. Molybdenum-binding peptides are of interest due to the central role of molybdenum in enzymatic cofactors and catalytic processes. Experimental techniques such as elution-based binding assays enable the high-throughput measurement of peptide binding strength, generating quantitative proxies for binding affinity. However, translating these experimental measurements into predictive rules that link peptide sequence to binding behavior remains challenging due to the combinatorial complexity of sequence space and the subtle physicochemical determinants of metal coordination. Recent advances in deep learning, especially graph neural networks (GNNs) and protein language models, provide powerful tools for learning structured representations of biological molecules and their interactions. By integrating experimental elution data with modern representation learning techniques, it is possible to develop predictive and interpretable models that capture the underlying sequence–interaction relationships governing peptide–metal binding.

Data (origin, type 100-200 words): The primary dataset used in this project consists of experimentally measured elution data derived from molybdenum-binding assays provided by the hosting laboratory. Each data point corresponds to a peptide sequence associated with a quantitative elution or enrichment score, which serves as a proxy for its binding affinity to molybdenum. The dataset may include replicate measurements, enabling normalization and noise assessment. Peptide sequences are short and experimentally validated, making the dataset well suited for supervised learning. In addition to experimental data, complementary computational data will be incorporated to enrich the model, including physicochemical residue properties and embeddings generated by pretrained protein language models. Where applicable, predicted structural or coordination features may be derived computationally to support model generalization. All data will be curated, pre-processed, and split into training, validation, and test sets to ensure robust statistical evaluation.

Methods (proposed analysis 100-200 words): Peptides will be represented as graphs, where nodes correspond to amino acid residues and edges represent peptide bonds and, optionally, potential coordination interactions with a molybdenum node. Node features will include physicochemical descriptors and contextual embeddings extracted from pretrained protein

language models, enabling the integration of evolutionary and biochemical information. Graph neural network architectures (e.g., message-passing or attention-based GNNs) will be employed to learn peptide-level representations that capture both local residue properties and global interaction patterns. The learned graph embeddings will be used to predict experimental elution scores via a regression framework. Model performance will be evaluated using appropriate statistical metrics, including correlation coefficients and error measures, and compared against baseline sequence-based models. Model interpretability techniques, such as attention analysis or feature attribution, will be applied to identify residues and motifs contributing most strongly to predicted binding, enabling biological insight alongside predictive accuracy

Aims (expected results 100-200 words): The project seeks to provide a validated computational framework that bridges experimental binding data and modern deep learning, supporting both predictive modelling and rational peptide design. This primary aim is carried out developing a graph-based deep learning model capable of accurately predicting peptide–molybdenum binding behavior from experimental elution data. A key objective is to assess whether incorporating graph representations and protein language model embeddings improves predictive performance compared to traditional sequence-based approaches. Beyond prediction, the project aims to interpret the learned models to identify sequence features and residue patterns associated with strong molybdenum binding, thereby contributing to a mechanistic understanding of peptide–metal interactions.