PhD PROJECT PROPOSAL - Computational and Quantitative Biology

CHROMA

Chromatin-aware modelling and engineering of disorder in transcription factors.

Principal Investigator: Dr Davide Mercadante
Host Institute: Dipartimento di Farmacia, Universitá Federico II

Duration: 3 years (fully funded)

Background to the Proposed Research

Transcription factors (TFs) are central regulators of cellular identity, controlling gene expression by recognising specific DNA sequences and coordinating the assembly of regulatory complexes. However, TFs' function is not determined solely by their DNA-binding domains. Many TFs contain extensive intrinsically disordered regions (IDRs), which are flexible regulatory segments that lack a single defined structure but play essential roles in molecular recognition, chromatin interactions and transcriptional regulation. Increasing amount of data indicates that the dynamic properties of TFs and their IDRs are key determinants of TF ability to access and regulate genomic regions[5, 3, 2, 4]. This is particularly relevant for pioneer TFs, a specialised class of regulators capable of engaging compacted chromatin and influencing nucleosome accessibility[1]. Despite their biological importance, the physical principles governing how TF dynamics control chromatin recognition and remodelling remain poorly understood. Recent advances in molecular simulation, machine learning and computational protein design provide an unprecedented opportunity to investigate how sequence changes influence the dynamic behaviour of regulatory proteins.
By combining multiscale molecular modelling with data-driven approaches, this project aims to develop a computational framework to understand and engineer TF dynamics in the context of chromatin environments.

Proposed Research

The aim of this project is to develop a computational platform capable of predicting how sequence variation within TFs, particularly within intrinsically disordered regions, modulates their conformational ensembles, interactions with chromatin and regulatory function. The project will combine molecular simulations and computational design approaches to establish the relationship between TFs, their molecular dynamics-driven interactions and the conformational effects on nucleosomal arrays. The research will initially focus on large-scale molecular modelling and investigation of the IDR sequence space, to understand how changes in IDR sequence properties, such as charge distribution, flexibility and interaction propensity, influence TF behaviour in nucleosomal environments. The developed framework will then be used to explore whether TF dynamics can be rationally engineered to modulate chromatin accessibility and regulatory activity. Ultimately, this project will establish a new computational approach for studying and designing TF function beyond traditional sequence-based models, providing a foundation for predictive regulation of gene expression.

Research tasks and proposed timeline

Year 1: modelling transcription factor dynamics in nucleosomal environments.

Develop computational models to investigate the conformational behaviour of TFs and their interactions with nucleosomal substrates. Establish appropriate simulation approaches across multiple length scales, combining atomistic and coarse-grained methodologies where appropriate. Identify

physical descriptors linking dynamics to chromatin accessibility and molecular recognition.

Year 2: understanding how sequence properties of disordered regions regulate transcription factor function.

Explore how naturally occurring and designed variations in intrinsically disordered regions influence conformational ensembles and chromatin interactions. Use molecular simulations and machine learning approaches to identify sequence features that can control dynamic behaviour, including charge organisation, flexibility, interaction propensity and conformational adaptability.

Year 3: computational design of transcription factors with tailored dynamic properties.

Develop predictive models capable of designing variants with desired molecular behaviours. Investigate whether engineered changes in regulatory regions can alter chromatin engagement and nucleosome dynamics. Apply the developed framework to different systems and establish general principles governing dynamics-driven regulation.

Preferred Background and Skillset of the Candidate

This project lies at the interface between computational biophysics, molecular simulation, artificial intelligence and molecular biology. The successful candidate will develop expertise in multiscale molecular modelling, protein dynamics, machine learning, computational protein design and the analysis of complex biomolecular systems.
Suitable applicants may come from a range of quantitative disciplines, including chemistry, physics, computer science, mathematics, bioinformatics or computational biology.

Applicants with a stronger chemistry, physics or biology background should ideally have:

- strong quantitative and analytical skills.
- programming experience (preferably Python).
- an interest in molecular simulations and biomolecular dynamics.
- motivation to learn machine learning and computational design approaches.

Applicants with a stronger computer science or mathematical background should ideally have:

- strong programming (C++ and CUDA preferred) and computational skills.
- an interest in applying artificial intelligence to biological problems.
- curiosity about the molecular mechanics underlying gene regulation.
- willingness to work at the interface between disciplines.

Previous experience with molecular dynamics simulations, machine learning, structural biology or computational chemistry is not essential.

Project's wider relevance

Understanding how TFs control gene regulation remains one of the major challenges in molecular biology. Current approaches largely focus on DNA sequence recognition and static structural models, while the role of protein dynamics and conformational ensembles remains poorly understood.This project introduces a new perspective in which TF activity is viewed as an emergent property of molecular dynamics, where sequence changes alter conformational landscapes and

ultimately influence interactions with chromatin. By combining molecular simulation, artificial intelligence and computational design, the project will establish a predictive framework for understanding and engineering behaviour. Beyond individual TF, the methodology developed here has the potential to provide general principles for controlling regulatory proteins involved in development, cellular identity and disease. More broadly, this work contributes to the emerging field of dynamic biomolecular design, where protein function is engineered not only through structure, but through control of molecular flexibility and conformational landscapes.

References

[1] S. Bjarnason, J. A. P. McIvor, A. Prestel, K. S. Demény, J. T. Bullerjahn, B. B. Kragelund, D. Mercadante*, and P. O. Heidarsson*. "DNA binding redistributes activation domain ensemble and accessibility in pioneer factor Sox2". In: *Nature Communications* 15.1 (2024).

[2] C. M. MacCarthy, G. Wu, V. Malik, Y. Menuchin-Lasowski, T. Velychko, G. Keshet, R. Fan, I. Bedzhov, G. M. Church, R. Jauch, V. Cojocaru, H. R. Schöler, and S. Velychko. "Highly cooperative chimeric super-SOX induces naive pluripotency across species". In: *Cell Stem Cell* 31.1 (2024).

[3] B. T. Donovan, H. Chen, P. Eek, Z. Meng, C. Jipa, S. Tan, L. Bai, and M. G. Poirier. "Basic helix-loop-helix pioneer factors interact with the histone octamer to invade nucleosomes and generate nucleosome-depleted regions". In: *Molecular Cell* 83.8 (2023).

[4] M. Már, K. Nitsenko, and P. O. Heidarsson. "Multifunctional Intrinsically Disordered Regions in Transcription Factors". In: *Chemistry – A European Journal* 29.21 (2023).

[5] J. J. Ferrie, J. P. Karr, R. Tjian, and X. Darzacq. ""Structure"-function relationships in eukaryotic transcription factors: The role of intrinsically disordered regions in gene regulation". In: *Molecular Cell* 82.21 (2022).