Are biological molecules nanoscale machines, and can AI-based tools be used to understand and even "break" them?

Leverhulme Centre for Algorithmic Life funded PhD studentship (3.5 years), Durham University, in collaboration with the Centre for Programmable Biological Matter

Start date: 1 October 2027

In this PhD project, we will address these questions through the investigation of the enzyme DNA gyrase.

DNA gyrase is an enzyme found ubiquitously in bacteria but absent from humans. Its unique biological role is the introduction of negative supercoils into DNA. What is currently known about its mechanism strongly suggests a machine-like mode of operation. The enzyme must first wrap its DNA substrate into a precise geometry and handedness. It then cleaves the DNA while simultaneously capturing another segment of the same DNA molecule within a protein clamp. A separate part of the enzyme subsequently pulls the cleaved strands apart, allowing the captured DNA segment to be transported through the break. The cleaved DNA strands are then religated, and a third and final gate opens to allow the transported DNA to exit the enzyme.

Because the release of broken DNA would be catastrophic for the cell, gyrase must operate with extraordinary precision and fidelity. Each gate must open and close in the correct sequence, and failures in coordination can be lethal. In many respects, DNA gyrase functions as a sophisticated molecular machine.

One of the grand challenges in protein design is the creation of artificial proteins with bespoke functions. In recent years, AI-based methods have revolutionised our ability to design proteins de novo with almost arbitrary structures. However, because these tools largely learn from existing structural data, they remain relatively poor at designing dynamic proteins that undergo coordinated conformational changes. Understanding how naturally evolved molecular machines such as gyrase operate may therefore provide critical insights for the next generation of protein design technologies.

Although numerous structural studies of gyrase have been performed, the available structures represent only snapshots of the enzyme frozen in time. Most correspond to relatively stable conformational states, providing an incomplete picture of the full reaction cycle. Our current "movie" of gyrase action therefore contains many missing frames.

To obtain these missing frames, we aim to trap gyrase in transient conformations that are only briefly populated during its functional cycle. By definition, this is a challenging task. Our approach will combine computational modelling and AI-based structure prediction to

identify plausible intermediate conformations. We will then use modern AI-driven peptide and protein design methods to generate binders predicted to stabilise and lock gyrase in these otherwise transient states. These trapped complexes can subsequently be isolated and characterised using structural biology techniques, particularly cryo-electron microscopy (cryo-EM).

DNA gyrase has an additional attraction as a model system. It is one of the most important antibacterial drug targets and is inhibited by clinically important antibiotics such as the fluoroquinolones. These compounds stabilise the cleaved-DNA intermediate and prevent religation, eventually leading to lethal DNA damage. However, antibiotic resistance is becoming increasingly widespread. The ability to access and understand previously unseen gyrase conformations could therefore not only transform our understanding of how this remarkable molecular machine functions but could also reveal entirely new opportunities for antibiotic development against targets for which resistance mechanisms do not yet exist.

LCAL Integration and Interdisciplinary Themes

This project is funded by the Leverhulme Centre for Algorithmic Life and is supervised collaboratively, led by Prof. Jonathan Heddle at the Centre for Programmable Biological Matter (CPBM). The supervisory team will be interdisciplinary and will support the student in the twinned aspects of the project: whether biological molecules can be understood as information-processing systems; and whether algorithmic methods can be used not only to model biological function but also to intervene in and redesign it.

Modern AI systems are increasingly capable of proposing protein sequences, predicting structures and generating experimental hypotheses. However, these systems do not possess a mechanistic understanding of biology in the way a human scientist might claim to. DNA gyrase therefore provides a unique test case through which to explore the relationship between algorithmic prediction, mechanistic understanding and experimental validation.

The project raises a number of questions central to the LCAL thematic programme "Synthetic Worlds":

- Can machine-learning systems discover biological mechanisms that humans cannot readily infer from experimental observations?
- What distinguishes prediction from understanding in biological science, and do algorithmic systems introduce new epistemologies of prediction?
- To what extent can biological molecular machines be described as computational or algorithmic systems?
- Can AI systems move beyond reproducing patterns in existing data and begin to generate genuinely novel biological designs?

- How should societal responsibility and scientific credit be assigned when discoveries emerge from increasingly autonomous design pipelines?

The student would therefore engage not only with cutting-edge structural biology and protein design, but also with broader questions concerning the nature of biological information, machine intelligence and scientific discovery.

Specific LCAL Contributions and Training Opportunities

Overall, the project investigates leveraging AI to discover missing states of a molecular machine that have never been experimentally observed. The interdisciplinary research at LCAL focuses on the contextual questions of what discover, observation, and synthesis mean for science and society. Through LCAL, the student will therefore have access to additional training and experience in:

AI for Scientific Discovery: Understanding what AI generativity and benchmarking mean for the life sciences, and for protein and peptide design specifically.

Explainable AI: Investigating how AI-derived hypotheses can be interpreted biologically rather than simply accepted as predictions.

Biological Information Processing: Exploring whether molecular machines such as gyrase can be formally described in terms of computation, state transitions and information flow.

Human-AI Collaboration: Studying how experimental scientists interact with AI design tools and what forms of expertise remain uniquely human.

Ethics and Governance of AI-Driven Biotechnology: Considering how conventional ethics frameworks for experimental biotechnology may not "fit" for algorithmic processes, and assessing the opportunities and risks associated with increasingly autonomous biological design systems.

Leverhulme Centre for Algorithmic Life, Durham University, and postgraduate study Details of Award

- Duration: 3.5 years full-time or part-time equivalent.
- Fees at the UKRI indicative home fee level; stipend at UKRI standard rate; tax-free.
- Additional resource for training, fieldwork, career development, and conferences from the Centre.
- Open to both Home and International applicants; **a limited contribution to international fees may be available.** We particularly welcome applications from people of Global Majority backgrounds and candidates with disabilities, who are underrepresented at this level.