

REBONTO HAQUE

+447404265414, rebonto.haque@magd.ox.ac.uk

SUMMARY

PhD researcher working on interpretable deep learning for protein sequence–structure–function relationships utilising SAE-driven mechanistic interpretability studies. Experienced in structure-conditioned generative models, namely ESM-IF and finetuned variants, high-dimensional biological data, and HPC-based ML workflows.

EDUCATION

University of Oxford – Magdalen College **2025-2029**

PhD, Statistics: Interpretable machine learning for protein sequence–structure–function relationships in therapeutic antibodies.

University of Oxford – St Peter’s College **2021-2025**

MBiochem, Biochemistry (Molecular and Cellular): First Class Honours (rank: 10/110)

S.F.X. Greenherald International School, Bangladesh **2009-2021**

A levels: 5 A* (Biology, Chemistry, Physics, Mathematics, Computer Science); O levels: 8 A*

- Awarded Cambridge Highest in the World Academic Achievement Award in O Level Physics
- Awarded Cambridge Highest in the Country Academic Achievement Award in O Level Bangladesh Studies

RESEARCH EXPERIENCE

Oxford Protein Informatics Group – University of Oxford : Master’s + PhD Student **09/2024 – Present**

- Working under Prof. Charlotte M. Deane on **interpretable deep learning models for protein sequence–structure–function relationships**, with applications to antibody–antigen interactions and cross-reactivity
- Currently investigating how **sequence likelihood of structure-conditioned inverse folding models correlates with cross-reactivity**
- Investigating methods to condition **ESM-IF** to generate sequences which maintain bi-specificity
- Trained **Sparse Autoencoders (SAEs)** to develop **mechanistic interpretability methods for inference-time model steering** of antibody-specific autoregressive language models (IgGen, ProGen2, p-IgGen)
- Identified latents associated with key antibody-specific concepts such as **V/J germline identity, CDR identity**
- Used SAE latents to steer model generation to generate germline-specific sequences whilst maintaining **diversity, novelty, and antibody-likeness** of sequences
- SAE paper was accepted to the **MLSB workshop, co-located at NeurIPS 2025**

TraitSeq – Norwich, UK : ML Intern/ ML Software Developer **07/2024 – Present**

- Continuing to work part-time for **TraitSeq**, an agritech using machine learning to enhance breeding strategies
- Analysed **high dimensional metabolomics, transcriptomics, and genomics data** to inform model predictions
- Implemented differential expression methods (**edgeR**) for feature selection to increase predictive accuracies
- Integrated **statistical differential expression (edgeR), ML feature importance, and gene regulatory network connectivity** to identify biologically meaningful biomarkers
- Explored biological questions to help identify future services the company could offer, namely, investigating whether stressed phenotypes can be accurately predicted using non-stressed expression data
- Developed robust ensemble learning strategies improving model generalisation by **~15%**

Earlham Institute – Norwich, UK : Research Intern **06/2023 – 08/2023**

- Selected amongst the top 14 students from over 500 applicants
- Developed an **end-to-end image analysis and ML pipeline** for biological specimens (wildflower seeds), integrating segmentation, feature extraction, and classification
- Designed ImageJ plugins for **automated microscopy-style image processing** and feature extraction
- **32,915** wildflower seeds across 20 species analysed
- Final pipeline showed a balanced accuracy of **98%** (an improvement from the original **90%**)
- Optimised computational pipelines, **reducing runtime from ~20 hours to ~2.5 hours**
- The results from the internship have been submitted for a publication which is currently under review

Winner of the “OxBioHackathon 2023” organised by the Oxford University Biotechnology Society **2023**

- Competed among **40+ students** across Oxford, including undergraduates, postgraduates, and PhD candidates
- Pitched a project exploring the therapeutic potential of novel piRNA technology, addressing key limitations in existing gene therapy approaches
- Awarded first place by an adjudication committee of industry experts for delivering the most compelling pitch and project. I specifically focused on the feasibility and efficacy of our proposal based on the existing literature

International Centre for Diarrhoeal Disease Research, Bangladesh – Bangladesh : Intern **08/2022 - 09/2022**

- Received hands-on training on biological sample handling, including **labelling, processing, and storage** following the appropriate bio-safety rules and regulations
- Set up the necessary equipment for toxicology experiments and prepared samples for **Atomic Absorption Spectrometry** and **Ficoll-Paque density-based PBMC separation** experiments
- Shadowed researchers, learning various experimental techniques and principles, including ELISA, Ficoll-Paque density-based PBMC separation process, immune cell profiling through flow cytometry, heavy metal and trace element measurement by Atomic Absorption Spectrometry, and different types of microbiological analysis such as serum-based bactericidal assays

PUBLICATIONS

- Mechanistic Interpretability of Antibody Language Models Using SAEs (pre-print; first author): <https://doi.org/10.48550/arXiv.2512.05794>
- Integrating Machine Learning, Deep Learning, and Image Analysis for Seed Species Classification (in review; co-second author)

EXTRACURRICULARS

Oxford University Chemistry and Biochemistry Society : President/Secretary **03/2023 – 04/2025**

- Elected to be President in March 2024. Previously carried out the role of Secretary
- Strategically managed society funds in partnership with the Treasurer, ensuring sustainable growth and responsible budgeting to expand event offerings
- Enhanced communication with members by overseeing a mailing list of **1,000+**, delivering regular updates that drive event participation and member engagement
- Fostered inter-society collaborations, forging partnerships with other student groups to expand our reach and deliver large-scale events with **~150** attendees, boosting cross-disciplinary networking

The Oxford Strategy Challenge : Student Consultant **03/2023 – 04/2023**

- Developed the strategy action plan for a biotech startup (TrialeX)
- TrialeX aimed to increase the participation of under-represented groups in clinical trials
- Developed targeted marketing strategies to boost engagement with key populations, leveraging data-driven insights to tailor outreach efforts for maximum impact
- Developed novel strategies for increasing participation of specifically homeless populations in clinical trials
- Presented findings to the founders of TrialeX alongside other teams participating in the project

SKILLS

- **Computational & ML:** PyTorch, TensorFlow; deep generative models, representation learning, interpretability; statistical modelling, differential expression (edgeR)
- **Biological Data:** Protein sequence & structural data; antibody annotation (ANARCI, OAS, PLAbDab); transcriptomics, genomics, metabolomics
- **Systems:** Linux, HPC (SLURM), GPU computing; Python, Java