SEBASTIAN PUJALTE OJEDA

GitHub | sp2120@cam.ac.uk | LinkedIn | Google Scholar

EDUCATION

University of Cambridge *Cambridge, UK*
PhD Chemistry Apr 2024-Present
   Research: Drug discovery targeting disordered proteins associated with neurodegenerative disease
   Supervisor: Prof. Michele Vendruscolo
   Award: Emmanuel College Finlay Scholarship, Full International Tuition & Maintenance

MPhil Computational Biology Sep 2022-Aug 2023
   Grade: 74/100

National Autonomous University of Mexico (UNAM) *Mexico City, Mexico*
BSc Mathematics Aug 2016-Jan 2021
   Grade: 88/100

EXPERIENCE

Department of Chemistry, University of Cambridge *Cambridge, UK*
Doctoral Researcher Apr 2024-Present
- Developed DynamicMPNN in shared codebase (increased yield of multi-state protein designs by 25%) and curated the largest ML-ready multi-state protein structure dataset.
- Developed automated pipeline to leverage residual HPC resources to perform large-scale (>1ms) enhanced sampling MD simulations of $A\beta42$ with small molecules, increasing simulation throughput on GPU by 300%.
- Evaluated performance of generative diffusion protein structure prediction models (i.e. Boltz, AF3, BioEmu, DiG) in generating structural ensembles of ordered and disordered proteins.
- Serve as the liaison between computational and experimental teams: communicating ML methods to wet-lab collaborators and coordinating experimental validation of model predictions.

Research Assistant Sep 2023-Apr 2024
- Developed the first method (LigandTransformer) for reliably predicting small-molecule binders to disordered proteins (20% binding assay hit rate).

Institute of Cellular Physiology, UNAM *Mexico City, Mexico*
Research Assistant Feb 2020-Aug 2022
- Developed Toposort, a manifold learning algorithm for spike-sorting that automatically processed and clustered decades of previously unanalyzed primate neural recordings, enabling novel scientific insights and multiple publications.

PUBLICATIONS

Multi-state Protein Design with DynamicMPNN. A. Abrudan*, **S. Pujalte***, C.K. Joshi, M. Greenig, F. Engelberger, A. Khmelinskaia, J. Meiler, M. Vendruscolo, T.P.J. Knowles (Accepted ICLR 2026; ICML 2025 GenBio Workshop). Link

Sequence-based virtual screening using transformers. S. Zhang, D. Huo, R.I. Horne, Y. Qi, **S. Pujalte**, A. Yan, M. Vendruscolo (Nature Communications 2025). Link

Coordination of nuclear BuGZ and cytoplasmic ATGL phase separation drives lipolysis and reverses hepatic steatosis. T. Jia, S. Zhang, X. Wang, W. Meng, L. Li, Z. Hou, L. Tang, Q. Zhang, **S. Pujalte**, H. Chen, X. Huang, M. Vendruscolo, L. Yang, H. Jiang (Under Review 2025).

Efficient and reliable spike sorting from neural recordings with UMAP-based unsupervised nonlinear dimensionality reduction. D. Suarez-Barrera, L. Bayones, S. Parra, V. Monroy, **S. Pujalte**, B. Andrade-Ortega, H. Díaz, M. Alvarez, A. Zainos, A. Franci, R. Rossi-Pool (PLOS Biology 2025). Link

An abstract categorical decision code in dorsal premotor cortex. G. Diaz-deLeon, M. Alvarez, L. Bayones, A. Zainos, J. Zizumbo, S. Parra, **S. Pujalte**, R. Romo, R. Rossi-Pool, V. De Lafuente (PNAS 2022). Link

Hierarchical unimodal processing within the primary somatosensory cortex during a bimodal detection task. S. Parra, H. Diaz, A. Zainos, M. Alvarez, J. Zizumbo, **S. Pujalte**, L. Bayones, R. Romo, R. Rossi-Pool (PNAS 2022). Link

*Asterisk denotes equal contribution

INVITED TALKS

Multi-State Protein Design-IDP2 Biomed, University of Padua, Italy   Nov 2025
Introduction to Transformers-European Union Bio2Brain Network (Online)   Jun 2024
Automated ML Driven Spike Sorting-Institute of Cellular Physiology, UNAM   Feb 2021

TEACHING

Systems Biology-DAMTP, Cambridge University   Apr 2024-Jun 2024
Computational Methods in Neuroscience-Faculty of Medicine, UNAM   Feb 2021-Aug 2022

OPEN SOURCE CONTRIBUTIONS

umap-learn maintainer and contributor; created new features, fixed bugs and wrote documentation.
scikit-learn contributor; expanded functionality for creating synthetic datasets and dimensionality reduction documentation.
BioEmu contributor; added support for custom MSA servers to enable large-scale inference.

SKILLS

Programming Languages: Python (experienced), **Bash** (experienced), **R** (proficient), **Julia** (proficient), **JavaScript** (familiar), **C** (familiar)

Machine Learning & Data Science: PyTorch (experienced), **scikit-learn** (experienced), **NumPy** (experienced), **pandas** (experienced), **PyTorch Lightning** (experienced), **JAX** (proficient)

Cheminformatics: RDKit (experienced), **GROMACS** (experienced), **PLUMED** (experienced), **OpenMM** (proficient), **OpenBabel** (proficient),

Computational Tools and Frameworks: Slurm (experienced), **HPC environment** (experienced), **distributed training** (experienced), **Git** (experienced), **Linux** (experienced), **LaTeX** (experienced)

Spoken Languages: English (native)**, Spanish** (native)

Citizenship: **Spanish** and **Mexican**