



Contribution ID: 5

Type: Talk

Unsupervised Machine Learning and Phase Space Reduction: A Robust and Generalisable Approach for Concurrently Solving the Protein Complex Conformation Classification and Quantification Problems.

Friday, 16 February 2024 10:30 (30 minutes)

The understanding of biochemical processes and the machinery of life hinges on comprehending the structural aspects of macromolecular interactions. This requires a systematic approach to analysing the vast manifold of macromolecular associations, or complexes, typically determined from crystallographic or electron microscopy experiments. Specific points of interest include similarity measurements, multiple structural alignments and superpositions, conformational analysis, classification, and functional annotation.

While a considerable number of tools have been developed for analysing covalently linked structures, or single chains [1-5], no methods applicable to the analysis of complexes are known to us. This may be explained by the higher diversity and absence of canonical ordering of chains in quaternary structures, leading to higher ambiguity compared to secondary and tertiary structure analyses.

We present FunCLAN, a novel approach and software solution for the analysis of protein complex conformations. FunCLAN combines unsupervised machine learning with physically informed scoring to transform a practicably infinite, continuous, geometric phase space into a tractable and measurable discrete one. This enables the robust classification and quantification of a protein complex's in-sample conformational landscape, canonical chain ordering and optimal multiple complex alignment and superposition.

The approach has been used to classify protein complexes, such as the SARS-CoV-2 spike protein, into distinct conformations, and measure the degree of similarity among them. FunCLAN provides a uniform and consistent approach to the analysis of macromolecular structures. It can be generalised to the case of single chains by applying the algorithm to secondary structures, or highly conserved "rigid" regions.

References:

- [1] Krissinel, E. (2012). Enhanced fold recognition using efficient short fragment clustering. *Journal of molecular biochemistry*, 1(2), 76.
- [2] Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D: Biological Crystallography*, 60(12), 2256-2268.
- [3] Ellaway, J. I., Anyango, S., Nair, S., Zaki, H. A., Nadzirin, N., Powell, H. R., ... & Velankar, S. (2023). Identifying Protein Conformational States in the PDB and Comparison to AlphaFold2 Predictions. *bioRxiv*, 2023-07.
- [4] Ye, Y., & Godzik, A. (2004). FATCAT: a web server for flexible structure comparison and structure similarity searching. *Nucleic acids research*, 32(suppl_2), W582-W585.
- [5] Li, Z., Natarajan, P., Ye, Y., Hrabe, T., & Godzik, A. (2014). POSA: a user-driven, interactive multiple protein structure alignment server. *Nucleic acids research*, 42(W1), W240-W245.

Submitting to:

8th CAPRI assessment meeting

Primary author: Dr CELIS GARZA, Daniel (CCP4, Research Complex at Harwell, STFC Rutherford-Appleton Laboratory, UK)

Co-authors: Dr ELLAWAY, Joseph (PDBe, EMBL European Bioinformatics Institute, Genome Campus, UK); Dr VELANKAR, Sameer (PDBe, EMBL European Bioinformatics Institute, Genome Campus, UK); Dr KRISSEL, Eugene (CCP4, Research Complex at Harwell, STFC Rutherford-Appleton Laboratory, UK)

Presenter: Dr CELIS GARZA, Daniel (CCP4, Research Complex at Harwell, STFC Rutherford-Appleton Laboratory, UK)

Session Classification: CAPRI