

The Research Group
Structural Biology Brussels

has the honor to invite you to the public defence of the PhD thesis of

Jose Gavaldá García

to obtain the degree of Doctor of Bioengineering Sciences

Title of the PhD thesis

**Probabilistic description of
protein backbone motions and conformations**

Promotor:
Prof. dr. Wim Vranken

The defence will take place on

**Friday, November 22, 2024 at 3:30 p.m.
in Auditorium D.0.05, VUB Campus
Etterbeek.**

The defence can also be followed through
a live stream: <https://shorturl.at/GwmQY>

Members of the jury

Prof. dr. Tom Lenaerts (VUB, chair)
Prof. dr. Sophie de Buyl (VUB, secretary)
Prof. dr. Joris Messens (VUB)
Prof. dr. Peter Tompa (VUB)
Prof. dr. Anastassia Vorobieva (VUB)
Prof. dr. Franca Fraternali (University College
London, UK)
Prof. dr. Rachid Tahzima (ULiège/ILVO)

Curriculum vitae

Jose Gavaldá-García earned his BSc in Biotechnology at Universitat Politècnica de València (Spain), followed by an MSc in Biomolecular Sciences at Vrije Universiteit Amsterdam and a joint MSc in Bioinformatics & Systems Biology from Vrije Universiteit Amsterdam and Universiteit van Amsterdam. In 2019, he began his PhD at Vrije Universiteit Brussel, where he developed Constava, a method to represent proteins in their dynamic form, and investigated AlphaFold's relation to protein flexibility and dynamics. Jose has co-authored six research articles, with one more under revision.

Abstract of the PhD research

Proteins in physiological conditions are dynamic and therefore better represented as an ensemble of conformations that reflects their complex thermodynamic landscape. Experimental and computational techniques such as Nuclear Magnetic Resonance (NMR) spectroscopy and Molecular Dynamics (MD) simulations provide information on such protein conformational ensembles and the range and timescale of protein backbone motions. Despite these advancements, characterising and interpreting protein dynamics and conformations remains challenging.

This thesis presents a novel probabilistic approach to describe protein local conformation and backbone dynamics based on probabilistically-defined conformational states, so acknowledging that amino acids in proteins can adopt multiple states. Six low-energy conformational states were defined from NMR data, thus capturing the behaviour of proteins in solution. A probability density function was defined for each conformational state by employing the backbone dihedrals extracted from NMR ensembles. A set of backbone dihedrals for a given residue, from large ensembles or MD data, can then be represented as a vector of conformational state propensities. In addition, a novel metric of conformational state variability was defined that can be also applied to regions with no defined stable structure.

The AlphaFold protein structure predictor, though capable of producing highly accurate structures from the amino acid sequence of proteins, lacks information on the backbone dynamics and flexibility of proteins. Consequently, this thesis also examines the capability of such predicted structures to capture backbone dynamics and flexibility. AlphaFold's pLDDT confidence metric inversely correlates with dynamics metrics predicted from sequence data. A further large-scale analysis with experimentally obtained backbone dynamics information further validates this trend. The study identified cases where AlphaFold's training data pre-processing resulted in the rigid prediction of flexible loops. Overall, our results highlight that while pLDDT distinguishes between rigid and dynamic regions, it fails to accurately represent subtle gradations in dynamics.

The work in this thesis therefore offers a quantitative probabilistic representation of protein conformational space and variability, and assesses the ability of AlphaFold models to capture backbone dynamics, suggesting inherent limitations in current predictive models in reflecting flexibility. In the spirit of open science, all tools and results are easily accessible to the research community, so offering new perspectives of quantifying the dynamic nature of proteins.