1

Oral presentation

Guiding structural model predictions with experimental information towards dynamics and assemblies with VAIRO

Josep Triviño¹, Elisabet Jiménez¹, Iracema Caballero¹ and Isabel Usón^{1,2}

Instituto de Biología Molecular de Barcelona (IBMB-CSIC), Barcelona Science Park, Baldiri Reixach 15, Barcelona, 08028, Spain 2ICREA: Institució Catalana de Recerca i Estudis Avançats. Pg. Lluis Companys 23, Barcelona, 08010, Spain. jtvcri@ibmb.csic.es

To solve the folding problem, the attention algorithm, encoded in the transformer succeeded in relating pairwise evolutionary conservation¹ and structural data in the PDB². AlphaFold³ succeeded in harnessing this knowledge, in many cases matching the accuracy of close homologs in its predictions, as assessed in CASP14-15⁴ experiments. AlphaFold derives information from three sources: learned parameters capturing intrinsic amino acid secondary structure and environment propensity; models of related proteins providing structural templates; and aligned sequences⁵ encoding profiles and concerted evolutionary changes of residues involved in contacts. Our method uses these three channels to include prior knowledge, through site-specific variants, predefined alignments and templates.

We have implemented a way to analyze and modify the information underlying AlphaFold predictions. Our program, VAIRO (Fig. 1), allows to selectively constrain structure conferring targeted degrees of freedom and setting boundary conditions. I will demonstrate how to use it in order to integrate predictions and biochemical and structural data as a route to broadly relate knowledge across experimental cases.

VAIRO: guiding structural predictions with prior knowledge		
INPUT	MODES TO STEER AF PREDICTION	RESULTS
Analyze input and set up predictions	Compose features.pkl for AF prediction	Filter and process models, analyze and recombine
Mandatory input: • Query sequence(s) • Define multimer	Naïve • Cluster and divide AF search	Score results to compose dynamics
Optional input: • Templates • Sequences	Guided Compare to template • Templates references • Sequence Alignments (GESAMT/LSQKAB) • Tragment libraries • Conformations (HINC CCANALYSIS, ALEPH) • Modify query sequence and revert • Interfaces (PISA)	Compare to template / references (GESAMT/LSQKAB) Conformations (HINGES)
Fragment librariesfeatures.pkl		CCANALYSIS, ALEPH) • Interfaces (PISA) • Energy (OpenMM) • Graphical Pymol session
Reference structures Modify: add, mask, weight	Mosaic Divide and assemble large prediction 	

Figure 1. Scheme of VAIRO. In addition to AlphaFold and HH-suite, CCANALYSIS⁶, ALEPH⁷ and CCP4 software⁸ are used.

- [1] Marks, D. S. et al. (2011). Protein 3D structure computed from evolutionary sequence variation. PLoS One 6, e28766.
- [2] Berman, H. M. et al. (2000). The Protein Data Bank. Nucleic Acids Research 28, 235-242.
- [3] Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589.
- [4] Kryshtafovych, A. et al. (2019). Critical assessment of methods of protein structure prediction (CASP)-Round XIII. Proteins 87, 1011-1020.
- [5] Steinegger, M. et al. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics 20, 473.
- [6] Diederichs, K. (2017). Dissecting random and systematic differences between noisy composite data sets. *Acta Crystallographica* D73, 286-293.
- [7] Medina, A. *et al.* (2019). ALEPH: a network-oriented approach for the generation of fragment-based libraries and for structure interpretation. *Acta Crystallographica* **D76**.
- [8] Agirre, J. *et al.* (2023). The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallographica* **D79**, 449-461.