Computational Protein Design with Artificial Intelligence

Marianne Defresne, Thomas Schiex, Sophie Barbe

Computational Protein Design with Artificial Intelligence

Marianne Defresne[1,2], Thomas Schiex[2], Sophie Barbe[1]

[1] TBI, Université de Toulouse, CNRS, INRAE, INSA, ANITI, Toulouse, France

[2] Université Fédérale de Toulouse, ANITI, INRAE, UR 875, Toulouse, France

Computational Protein Design (CPD) aims to design proteins with enhanced or new functions and/or properties. Structure-based methods target a function via a 3D input backbone, and aim to find sequence(s) folding onto this backbone. Traditional methods minimize a score function combining the energy of the sequence on the input backbone [1] with specific design constraints related to the precise design objective. However, various generally desirable properties such as expressivity, solubility or hydrophilicity are extremely difficult to capture mathematically in such score functions. Recent deep learning-based methods can implicitly capture these general properties by training on natural proteins. But then, specific design constraints become difficult to enforce with deep learning.

In this talk, we aim to take the best of both worlds by combining two artificial intelligence technologies: automated reasoning, augmented with deep learning. We keep the formulation CPD as the optimization of a pairwise decomposable energy function, but we learn the energy on known protein structures. This energy is then optimized exactly for each protein to design using automated reasoning [2]. We *in silico* tested the quality of the learned energy function, and found it to outperform state-of-the art hybrid and statistical functions such as those available in Rosetta [1] or KORP [3] in terms of Native Sequence Recovery, a commonly used metric for CPD. We are currently testing it on applied projects that we will briefly present.

[1] Park, Hahnbeom et al. (2016). “Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules”. In: Journal of Chemical Theory and Computation 12.12, pp. 6201–6212.

[2] Traoré, Seydou et al. (2013). “A new framework for computational protein design through cost function network optimization”. In: Bioinformatics 29.17, pp. 2129–2136. issn:1367-4803.

[3] Lopez-Blanco, José Ramon and Pablo Chacon (2019). “KORP: knowledge-based 6D potential for fast protein and loop modeling”. In: Bioinformatics 35.17, pp. 3013–3019. issn: 1367-4803.