Addressing the challenge of oligomerization in computational protein design

Diogo Silva [1], Pedro Moreira [1,2], Rita I. Teixeira [1], Bárbara Fernandes [1], Mariana Parada [1], João B. Vicente [1], Diana Lousa [1], Cláudio M. Soares [1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
2. Centro de Engenharia Biológica, Escola de Engenharia da Universidade do Minho, Braga, Portugal

Computational protein design is a field of research with potential to greatly impact areas such as drug development or enzyme technology, by combining knowledge-based and physics-based methods to design tailor-made proteins. This concept was introduced by David Baker with the development of Rosetta Commons. This is an extensive framework containing tools that enable protein design using physics-based scoring functions. Recently, the field of computational protein design has had major breakthroughs with the introduction of artificial intelligence (AI)-driven tools, namely RFDiffusion1 and ProteinMPNN2, which, when combined with AI-based 3D structure prediction tools (AlphaFold23), provide more efficient protein design pipelines.

This project aims to combine AI-driven methods with physics-based methods to re-design a protein binder and increase its affinity for a given target, considering potential oligomeric states of the design candidates. For this purpose, the receptor binding domain (RBD) of the Sars-CoV-2 spike protein was utilized as proof of concept, by starting with a known binder, that was predicted to dimerize, and improving its binding affinity to the RBD. The protocol starts with a molecular dynamics (MD)-based analysis of the original protein binder in solution (in monomer and dimer forms), as well as of the Target:Bindermonomer and Target:Binderdimer complexes. This guides the subsequent design steps, where ProteinMPNN2 is used to improve the binder's affinity for the target, considering both the monomer and dimer states of the binder in parallel re-design runs. The best candidates obtained in each re-design run will then be produced in bacterial systems, and their binding affinity and stability evaluated through binding assays and biophysical characterization, respectively. Hopefully, this project will contribute to the validation of computational AI-driven protein design as an approach that holds promise for biopharmaceutical applications and show the importance of combining MD simulations with protein design methodologies to obtain robust protein structures.