De novo generation of peptide binders using Protein Large Language Models: towards anti-obesity bioactive peptides

Bryan Renard[1,2], Jean-Yves Parfait[1], Catherine Michaux[2]

[1]*AI Department, Multitel, Pierre et Marie Curie street, 2, 7000 Mons, Belgium.*

[2]*Laboratoire de Chimie Physique des Biomolécules, UCPTS; Namur Institute of Structured Matter (NISM); Namur Research Institute for Life Sciences (NARILIS), University of Namur, 61 rue de Bruxelles, 5000 Namur, Belgium*

Bioactive peptides are often envisioned as the future of therapeutics. Indeed, peptides have many attractive properties such as a high specificity and few toxic effects. They are versatile molecules, with the potential to interact with a great variety of protein targets. However, it can be extremely costly to identify suitable bioactive peptides. In that context, numerical methods able to forecast the effect of given peptides, or identify interesting ones, are very useful. For this goal, exploring the possibilities of Artificial Intelligence and more specifically Large Language Models (LLMs) is a promising road as pretrained protein LLMs are available and becoming increasingly efficient [1]. These models are trained on huge protein datasets to learn meaningful representations of proteins and can be adapted (or finetuned) on many protein-related tasks.

Our goal is to use these LLMs to *de novo* generate peptide binders to a given protein using only its sequence. Two algorithms exploiting LLMs for the task at hand already exist, PepPrCLIP and PepMLM [2]. Other approaches inspired by language translation are also explored. To validate the generated peptides, a combination of docking and molecular dynamics tools is exploited with the objective to understand the interactions between the peptides and the protein, and estimate their binding affinity.

As a test target, the sequence of the MD2 protein is used to generate peptides that are subsequently analyzed into the TLR4/MD2 complex. This complex is involved in the recognition of bacterial lipopolysaccharides and highly implicated in obesity-associated inflammation [3], To prevent this, a competitive peptide antagonist of the receptor is particularly interesting to identify.

[1] G. Valentini *et al.*, Frontiers in Bioinformatics (2023), 3
[2] S. Bhat *et al.*, biorxiv (2023) doi:10.1101/2023.06.26.546591 ; T. Chen *et al.*, arxiv (2023). arXiv:2310.03842v2
[3] M.M. Rogero and P.C. Calder, Nutrients (2018), 10(4): 432.