Deep learning for protein function discovery and in silico protein engineering

Project description:
New developments in laboratory technologies allow fast identification of DNA and protein sequences. Although ever increasing amounts of data are generated and stored in databases like Uniprot [1], functional annotation of proteins is lagging. Algorithms like EggNOG [4] use sequence homology to predict function, but for proteins without known homologs function prediction is still a challenge. Over the last few decades deep learning algorithms have made great advances in various fields, but have rarely been applied in biological studies. Recently, Schwartz et al. [3] showed with the development of D-Space that applying a deep learning model trained on 70 million protein sequences from Uniprot can make the connection between sequence and function [3].

With the starting point of this paper and various ideas on applying deep learning methods for our in-house experimental data, we are looking for a talented MSc student to work on this challenge. In this thesis/internship project, in collaboration with DSM, you will train and test deep convolutional neural networks for protein annotation. You will explore the resulting high-dimensional embeddings for protein discovery and in silico protein engineering. Machine learning knowledge, linear algebra skills, programming proficiency, and interest in deep learning and biotechnology are desirable.

References:

Contact:
Marcel Reinders
E: m.j.t.reinders@tudelft.nl

Stavros Makrodimitris
E: s.makrodimitris@tudelft.nl

Jurgen Nijkamp
E: jurgen.nijkamp@dsm.com