

A bioinformatic investigation of linkers in multi-modular enzymes

Project supervisors:

Juan CORTÉS

LAAS-CNRS, Toulouse

juan.cortes@laas.fr

Nicolas TERRAPON

AFMB, Marseille

nicolas.terraupon@univ-amu.fr



Context and objectives:

The majority of proteins, in prokaryotes and eukaryotes, are composed of several domains connected by linkers, domain-linker-domain (DLD) being the most common architecture. In most cases, linkers are flexible and lack permanent secondary structure, but they are not mere spacers between domains. Indeed, their length and sequence have been evolutionarily tailored to play key functional roles, being frequently involved in allosteric mechanisms. Nevertheless, in general, the relationships between sequence, structural properties and function of flexible linkers are still misunderstood.

This project is focused on the investigation of multi-modular enzymes dedicated to the deconstruction of plant cell wall (PCW), which are composed of a glycosyl hydrolase (GH) catalytic module and a carbohydrate binding module (CBM) separated by a linker [1]. Data will be extracted from the CAZy database (www.cazy.org/), which contains hundreds of thousands GH and CBM sequences. The analysis will be focused on linkers connecting xylanase catalytic domains of the GH11 family to CBM domains (>2000 sequences), but linkers of other GH families involved in PCW deconstruction will also be analyzed in order to investigate possible specificities of each family of enzymes. The linker flexibility has a significant effect on the binding to carbohydrates and on the GH activity through various mechanisms, such as processivity or stability [2]. Importantly, sequence-based analyses performed for flexible linkers and tails attached to catalytic domains have highlighted conserved characteristics potentially related to function [3,4].

Analyses will be performed to investigate several features of the linkers: length, amino-acid composition, conservation, charge distribution, etc. For this, we will apply a combination of standard statistical and bioinformatics tools, such as MUSCLE (<https://www.ebi.ac.uk/Tools/msa/muscle/>) or JACKHMMER (<https://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer>) for multiple sequence alignment (MSA), and specific scripts (in Python) that will be developed during the project. We will also investigate structural properties using sequence-based predictors. More precisely, we will apply predictors such as DisEMBL (<http://dis.embl.de/>) or SPOT-Disorder2 (<https://sparks-lab.org/server/spot-disorder2/>) to assess the level of disorder within the linkers. In addition, we plan to apply our predictor of local structural propensities, LS2P [5] (<https://moma.laas.fr/applications/LS2P/>), to analyze the possible presence of structural patterns. We will investigate possible correlations between conservation of structural propensities and sequences. We will classify linkers with respect to the aforementioned sequence and structure characteristics, and we will select representative members for subsequent investigation using experimental and computational methods in the framework of a collaboration with other laboratories. In fact, this work is part of a larger project that aims to design optimized multi-modular enzymes for the degradation of PCW. PCW biomass has a vast potential as a renewable resource, principally for the manufacture of advanced fuels, chemical intermediates and products. Thus, with this project, we aim to contribute to the development of an environmentally sustainable economy.

The project will be co-supervised between Juan Cortés (LAAS-CNRS, Toulouse) and Nicolas Terrapon (AFMB, Marseille). The student will have the choice to be hosted among these two laboratories. Frequent meetings and travels to the other laboratory will be organized during the project. The M2 internship may continue with a PhD thesis on the computational design of flexible linkers.

References:

- [1] Gilbert HJ. The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol.* 2010; 153(2):444-455. DOI: 10.1104/pp.110.156646
- [2] Ruiz DM, Turowski VR, Murakami MT. Effects of the linker region on the structure and function of modular GH5 cellulases. *Sci Rep.* 2016;6:28504. DOI: 10.1038/srep28504
- [3] Sammond DW, Payne CM, Brunecky R, Himmel ME, Crowley MF, Beckham GT. Cellulase linkers are optimized based on domain type and function. *PLoS One.* 2012;7:e48615. DOI: 10.1371/journal.pone.0048615
- [4] Tamburrini KC, Terrapon N, Lombard V, Bissaro B, Longhi S, Berrin J-G. Bioinformatic analysis of lytic polysaccharide monooxygenases reveals the pan-families occurrence of intrinsically disordered C-terminal extensions. *Biomolecules.* 2021; 11(11):1632. DOI: 10.3390/biom11111632
- [5] Estaña A, Barozet A, Mouhand A, Vaisset M, Zanon C, Fauret P, Sibille N, Bernadó P, Cortés J. Predicting secondary structure propensities in IDPs using simple statistics from three-residue fragments. *J Mol Biol.* 2020;432(19):5447-5459. HAL: hal-02920302v1

Expected skills:

Strong background in bioinformatics, as well as good programming skills (Python, R).
Background in structural biology is not necessary, but it would be a plus.

Possibility of funding:

The student will be provided with a monthly stipend of around 550 euros during up to six months.
There is a possibility to continue with a funded PhD thesis.

Applications:

Please send an email containing your CV to the project supervisors, indicating in the subject "Candidate bioinfo linkers project".