An integrative strategy to unravel carbohydrate:protein interactions in the Human Gut Microbiome

Viviana G. Correia[1, §], Joana L.A. Brás[2], Benedita A. Pinheiro[1], Maria João Romão[1], Lisete Silva[3], Ten Feizi[3], Ana Luísa Carvalho[1], Carlos M.G.A. Fontes[2,4], Angelina S. Palma[1,3]

[1] UCIBIO@REQUIMTE, Department of Chemistry, Faculty of Science and Technology, NOVA University of Lisbon, Portugal; [2] NZYTech, Lda - Genes & Enzymes, Lisbon, Portugal; [3] Glycosciences Laboratory, Department of Medicine, Faculty of Medicine, Imperial College London, United Kingdom; [4] CIISA/FMV-UL, Faculty of Veterinary Medicine, University of Lisbon, Portugal; [§]*Presenting author (e-mail: vivianacorreia@campus.fct.unl.pt)*

Gut Microbiome

The human gut microbiota is a **highly carbohydrate-active microbial community** with a broad capacity to **metabolize dietary and host-derived glycans**, which is essential to human nutritional balance and immune system modulation. Thus, understanding carbohydrate recognition in the gut is of utmost importance for human health and nutrition.

Typically isolated gut strains, *e.g. Bacteroides spp*, exhibit many **substrate-specific polysaccharide-utilization loci (PULs)** that allow bacteria to cope with nutrient fluctuation (Fig.1). Each PUL orchestrates the recognition and degradation of a specific glycan, using a cohort of starch-utilization system (Sus)-like proteins and modular carbohydrate-active enzymes (CAZymes) with associated carbohydrate-binding modules (CBMs) [1]. **Architectural proteins such as non-catalytic CBMs and SusD-like proteins are thought to enhance and mediate specific glycan targeting and recognition**. The commonly isolated strain, *Bacteroides thetaiotaomicron*, carries over 31 CBMs in a total of 386 putative CAZymes [1].

As the microbiome (full collection of all microbiota genes) data piles up, there's an urgent need to develop and apply **high-throughput approaches** to study these recognition systems [1].

Here we present an **integrative strategy** combining high-throughput protein production with carbohydrate microarray technology [2,3] and protein X-Ray crystallography [4-6] to elucidate binding specificities of PULs architectural proteins from representative strains of the gut microbiome.

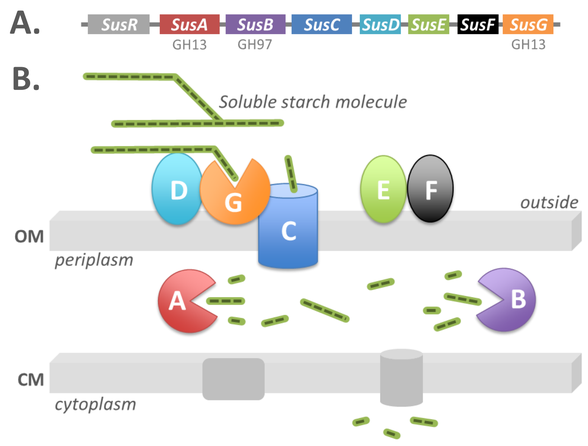
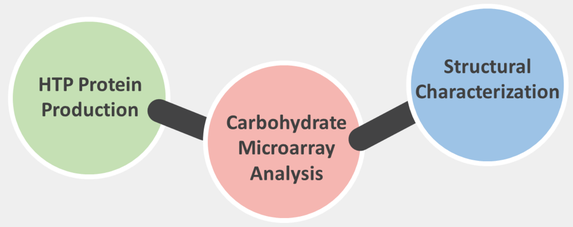


Figure 1 – Starch Utilization System (Sus) from *B. thetaiotaomicron*, the PUL paradigm, adapted from [1]. (A) Gene cluster or operon; (B) Putative system organization, hydrolyzing starch near the bacterial cell surface (OM outer membrane, CM cytoplasmic membrane).

AIM

- To elucidate **carbohydrate-binding specificities** of novel CBMs and SusD-like proteins of PULs representative of the **human microbiota metabolic diversity**

Applying an **integrative and combined approach**:



1. Recombinant Protein Library — Sequence Selection & High-throughput Protein Production

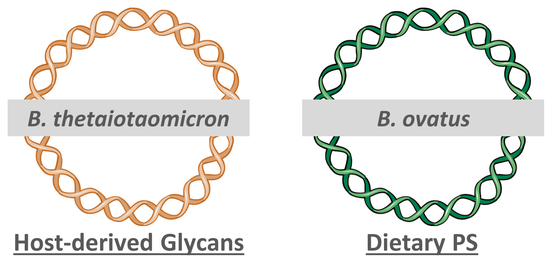
Protein Selection

- Sequenced and available strains representative of human microbiota metabolic diversity
- Putative **carbohydrate-binding activity** (SusDs & CBMs)
- **PULs databases**

Protein Production

- **High-throughput** cloning, expression and purification (Fig.2)
- **>100** different **recombinant proteins**
- Fully automated platform

- *Bacteroides* **selected strains**



Selection criteria
- CBMs and SusD from PULs active with given carbohydrates
- Cover broad diversity
- Potential target present in the arrays
- Novel structural derived information

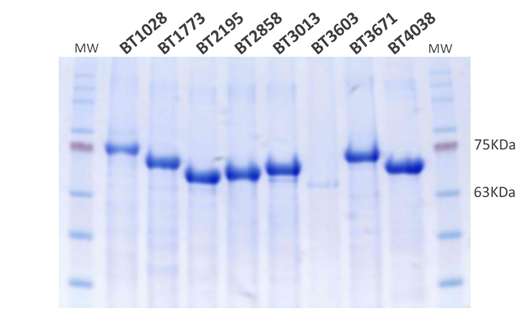


Figure 2 – Constructed protein library overview. (A) SDS-PAGE analysis of some recombinant proteins from *B. thetaiotaomicron*. Molecular weight (MW) ranging from 63 to 75 KDa. (B) High-throughput cloning and expression report showing a 75% success rate - 84/112 proteins were produced.

2. Protein-Glycan Interactions — Carbohydrate Microarray Analysis

Glycan Probes

- Polysaccharides and sequence-defined lipid linked oligosaccharides (neoglycolipids) [2]
- Plant, fungal, bacterial and mammalian type **glycan sequences** [2,3]

Design/Arraying

- **Robotic printing**
- Nitrocellulose-coated glass-slides
- Cy3 dye for spot location

Protein Binding

- Microarray probing (Fig.3)
- Fluorescence readout/scan
- **Antibody-biotin-streptavidin-AF$_{647}$**

Data Analysis

- Data analyses, storage and presentation *e.g.* ***heat-map*** (Fig.4)
- Dedicated Software [7]

- The use of **Carbohydrate Microarrays** enables the screening of a diverse and wide range of protein:glycan interactions on the same chip using only minute amounts of sample
- **Preliminary binding results** for 5 recombinant CBMs/SusD proteins are shown (Fig. 4)

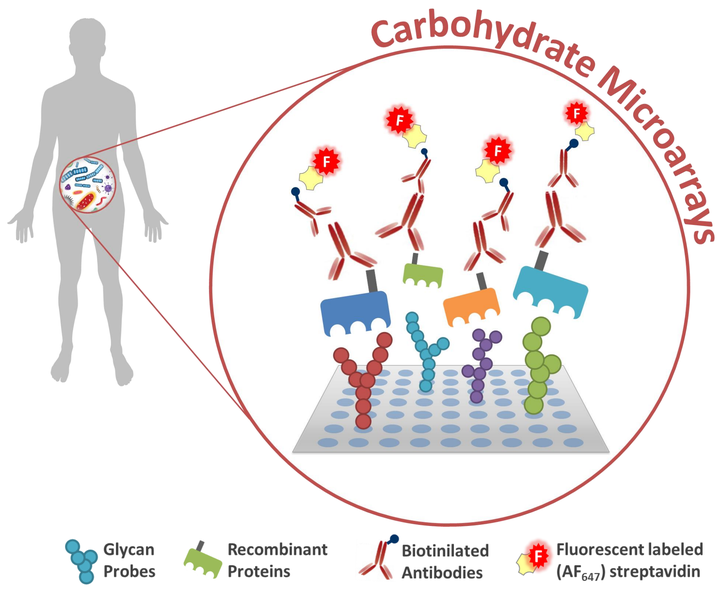


Figure 3 – Carbohydrate microarrays binding assay. Immobilized glycan probes, recombinant CBMs/SusD binding and detection system are depicted.

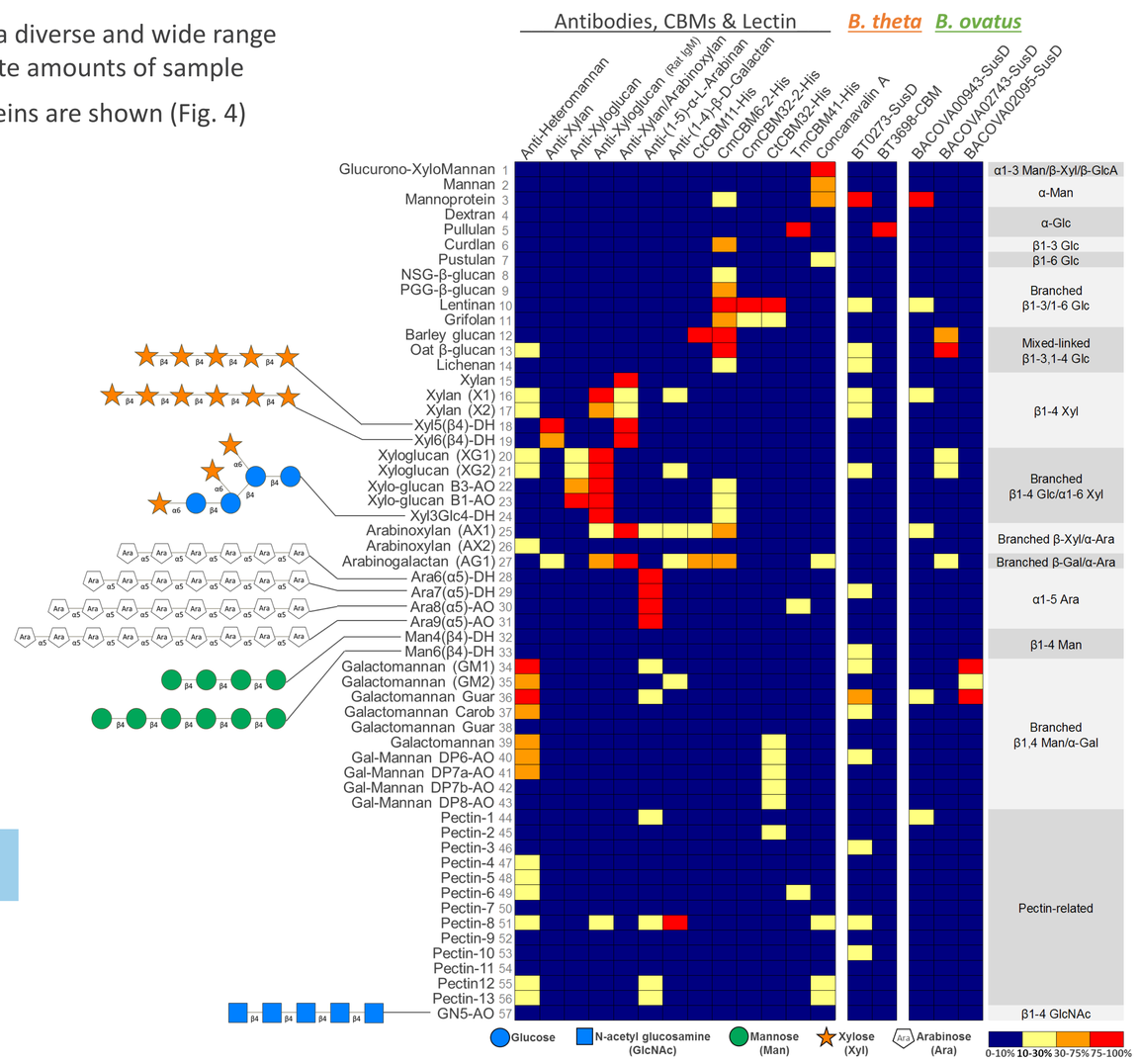


Figure 4 – Carbohydrate microarray screening analysis of *B. thetaiotaomicron* and *B. ovatus* CBMs and SusD proteins. The heat map represents the relative binding intensities calculated as the percentage of the fluorescence signal intensity given by the probe most strongly bound by each protein (normalized as 100%). Proteins used for validation of the microarray set include antibodies, CBMs and a lectin for which carbohydrate-binding specificity is known.

3. Structural Characterization — X-Ray Crystallography

Crystallization

- Diffraction-quality **protein crystals**
- Apo-protein and ligand-bound
- Automated or manual methods

X-Ray Diffraction

- X-ray diffraction **data collection**
- Radiation sources (in-house or a synchrotron facility)

3D Structure

- 3D structure solution - MR, Se-MAD or MIR(AS)
- **Model building** and refinement

- Structural analysis of **new protein-oligosaccharide complexes**

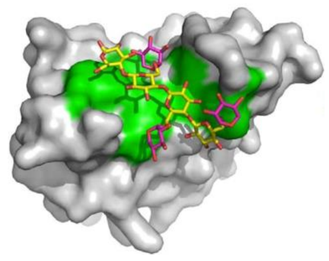


Figure 5 – Example of a crystal structure of a CBM-oligosaccharide complex, PDB 2YPJ [6].

Other complementary Biophysical Techniques
- Small Angle X-ray Scattering (**SAXS**)
- Saturation-Transfer Difference NMR (**STD-NMR**)
- Isothermal Titration Calorimetry (**ITC**)
- Surface Plasmon Resonance (**SPR**)

Conclusions & Futures perspectives

- **Preliminary data** on CBMs and SusD proteins reveal specific and promising binding patterns (Pullulan, Oat-β-glucan or α-Mannans)
- Specific interactions can be further addressed with sequence defined microarray sets
- This **combined approach** can be applied to other microbial strains and can contribute to a broad understanding of the **human microbiome metabolic capabilities**.

Find us!

www.ucibio.pt

References

1. Flint H.J., *et al.* (2012) GutMicrobes 3(4):289-306;
2. Palma A.S., *et al.* (2014) CurrOpinChemBiol 18C:87-94;
3. Palma A.S., *et al.* (2015) MolCellProteomics 14(4):974-88;
4. Carvalho A.L., *et al.* (2004) JBiolChem 279:34785-93;
5. Brás J.L., *et al.* (2011) ProcNatlAcadSciUSA 108(13):5237-42;
6. Luís A.S., *et al* (2013) J Biol Chem 288(7):4799-809;
7. Stoll M.S. & Feizi T. (2009) Proceeding Beilstein Glyco-Bioinformatics 123-40

Acknowledgements

We would like to thank Dr. Robert Childs from The Glycosciences Laboratory, Imperial College London for his assistance in the robotic microarray printing. And Professor Manuel Coimbra, from University of Aveiro, for kindly providing polysaccharide samples.

Funding

Project supported by the Portuguese Science and Technology Foundation (FCT-MEC) through grants PTDC/BBB-BEP/0869/2014; PTDC/QUI-QUI/112537/2009; RECI/BBB-BEP/0124/2012; SFRH/BPD/68563/2010; PD/BD/105727/2014. And by Unidade de Ciências Biomoleculares Aplicadas – UCIBIO, which is financed by national funds from FCT/MEC (UID/Multi/04378/2013) and co-financed by the ERDF under the PT2020 Partnership Agreement (POCI-01-0145-FEDER-007728).

Check out: **Diana Ribeiro's poster D4** *for more on Carbohydrate Microarrays and X-ray crystallography*