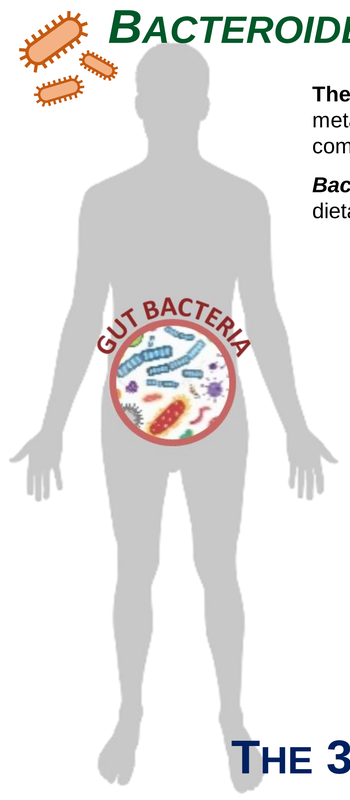
Bacteroides thetaiotaomicron, the human gut microbe that owns the required protein machinery to degrade complex pectic polysaccharides for us

Filipa Trovão[1,*], Viviana G. Correia[1], Frederico Lourenço[1], Joana L. A. Brás[2], Carlos M. G. A. Fontes[2,3], Ana Luísa Carvalho[1], Angelina S. Palma[1], Benedita A. Pinheiro[1]

1. UCIBIO-NOVA, Department of Chemistry, FCT-NOVA, Caparica, Portugal; 2. NYTech - genes & enzymes, Lisboa, Portugal; 3. CIISA, Faculdade de Medicina Veterinária, Universidade de Lisboa, Portugal

*f.trovao@campus.fct.unl.pt

Bacteroides thetaiotaomicron holds the machinery to degrade the most complex dietary carbohydrate



The human gut microbiota houses a densely microbial community of commensal and pathogenic bacteria, with a broad capacity to metabolize dietary and host-derived carbohydrates. The variation of the carbohydrate composition in the human diet is known to shape this community, promoting homeostasis or microbial imbalance (dysbiosis) associated with mechanisms of pathology and ultimately, disease [1].

Bacteroides thetaiotaomicron (*B. theta*, BT) is one of the most prominent bacteria found in the human gut, which can degrade both host and dietary carbohydrates.

- The breakdown of complex carbohydrates requires the action of **multi-specific modular protein assemblies** (enzymes and carbohydrate-targeting proteins) [1].
- ***B. theta* has the necessary protein assembly to efficiently degrade RG-II** (Fig. 1) [2, 3].
- Non-catalytic Carbohydrate Binding Modules (CBMs) are thought to mediate specific targeting and recognition of the carbohydrate substrate, enhancing degradation efficiency.

BT0996: involved in the degradation of RG-II is composed by:

- **2 catalytic glycoside hydrolase (GH) modules** at N-terminus - D-glucuronidase (GH2) (Fig. 2) and L-arabinofuranosidase (GH137) (Fig. 3) [3].
- A putative and yet uncharacterised **carbohydrate binding module (CBM)** at C-terminus - **BT0996-C** (Fig 1).

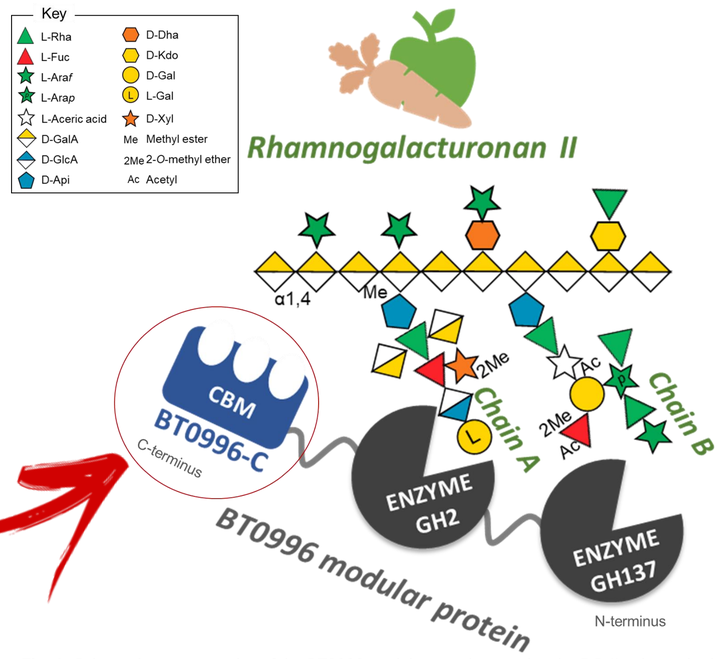
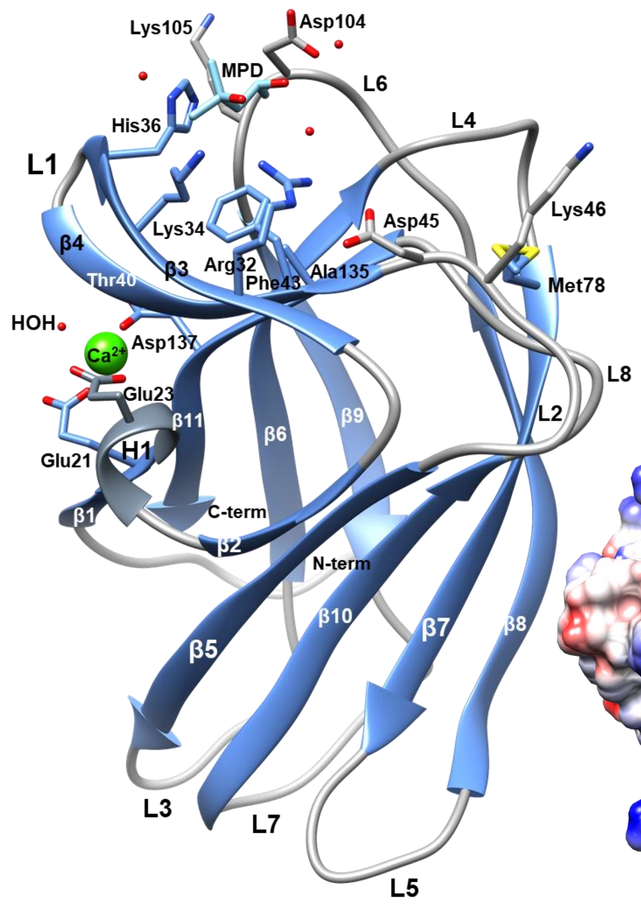


Fig. 1: Schematic representation of the BT0996 modular protein targeting the RG-II polysaccharide.

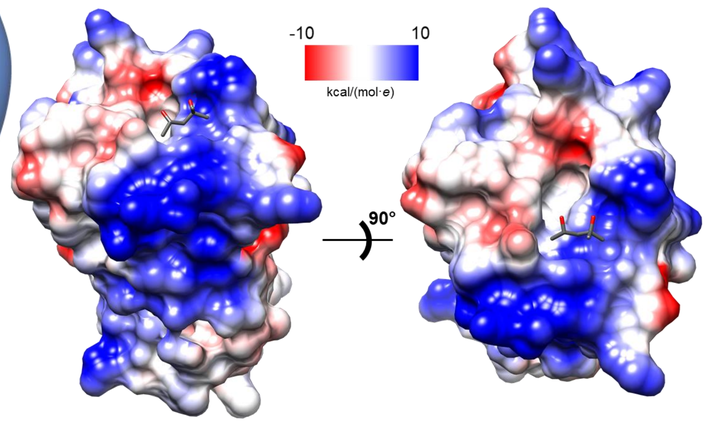
Aim

Structural characterization of BT0996-C CBM in order to understand its possible role in the bacterial degradation of complex pectic polysaccharides, such as RG-II.

The 3D structure of BT0996-C CBM



- Crystal Structure Resolution: 1.65 Å;
- Solved by Molecular Replacement from a preliminary Se-SAD 3D structure;
- Fold: distorted β-barrel-like structure, composed of two β-sheets;
- A structural calcium ion shows a pentagonal bipyramid coordination geometry;
- High content in positively charged residues (~20%, pI=9.43), **resulting in a prominent broad positive patch to the electrostatic surface.**



BT0996-C has homology with CBM35, but distinct features

The two closest structural homologues are **CBMs from family 35** (Fig. 4) [4], sharing a sequence identity of ~20%.

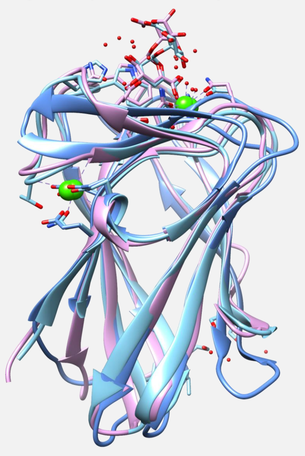


Fig. 2: Superposition of the two closest CBM35 structural homologues with BT0996-C CBM. Blue - BT0996-C; Cyan - CBM35 from *A. orientalis*, PDB ID 2VZQ; Pink - CBM35 from *C. thermocellum*, PDB ID 2W47.

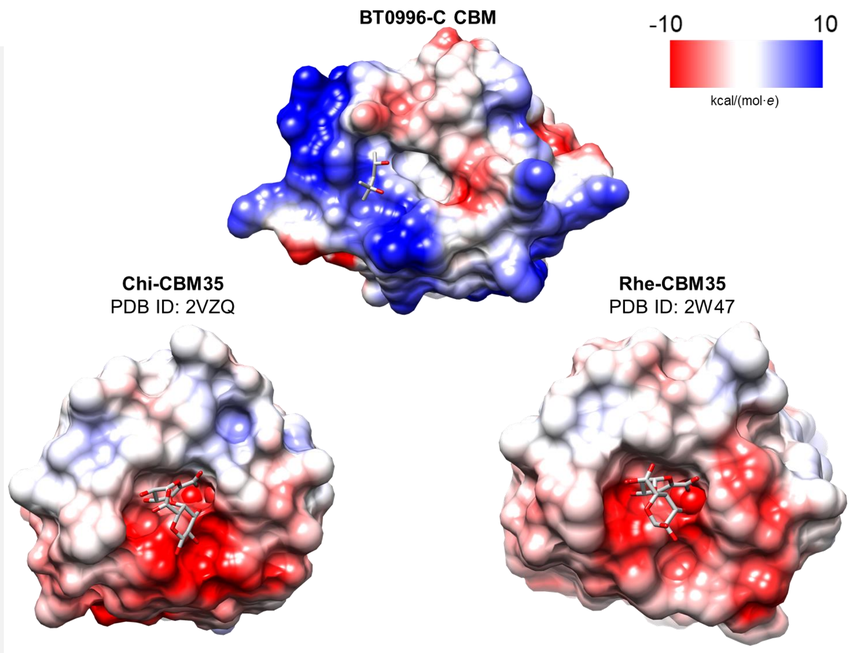


Fig. 3: Representation of the Coulombic surface charges distribution of the binding sites from BT0996-C and representative homologue CBMs. Chi-CBM35: exo-β-D-glucosaminidase from *A. orientalis;* Rhe-CBM35: rhamnogalacturonan acetyl esterase from *C. thermocellum*

The surface topology and electronegativity of BT0996-C are distinct from CBM35 structural homologues (Fig. 5):

- **Positively charged surface region in the vicinity of the putative binding site**.
- **Unusual small hydrophobic pocket** comprising Ala135 (Region C) and a Phe43 in the centre.

BT0996-C: Putative Binding Site inspection

The BT0996-C binding site lacks common carbohydrate-interacting aromatic residues. Instead has a more electron-poor aromatic residues (His36 and Phe43), then less likely to make CH−π interactions with sugar residues [5].

Five key regions (A-E) (Fig. 6), important for ligand binding and specificity, have been described in the binding sites of CBM families 35 [6], [7]:

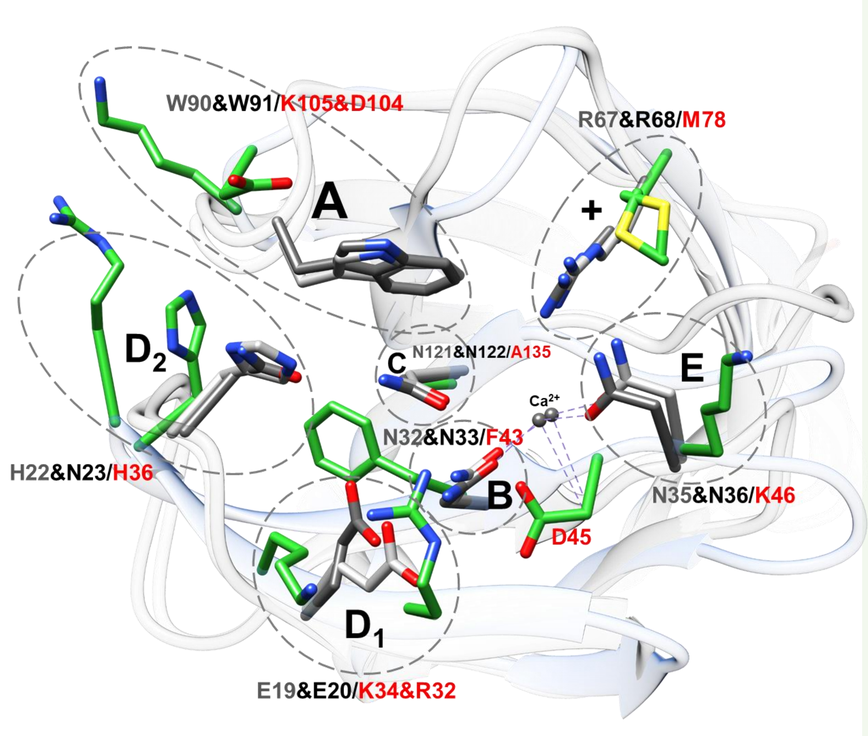


Fig. 4: Secondary structure matching superposition of BT0996-C CBM and the two family 35 CBMs of higher structural similarity. Amino acid residues are shown in green for BT0996-C, light grey for 2VZQ and dim grey for 2W47.

Region A – Conserved:
- CBM35 Family: Tryptophan (π-CH stacking interaction);
- BT0996-C: aromatic residue is absent.

Region B - Less conserved:
- CBM35: Asparagine – associated with the metal coordination site.
- BT0996-C: Phenylalanine – suggesting that this protein may not require calcium for ligand recognition.

Region C – Invariant:
- CBM35: Asparagine – critical hydrogen bond contacts with OH groups of the sugar residue.
- BT0996-C: Alanine – apolar residue (Ala135).

Region D_1 – Polar residues (make pocket-like binding site).
- BT0996-C: Lys34 and Arg35.

Region D_2 – Characteristic of uronate-specific CBM35.
- Conserved Histidine, also present in BT0996-C.

Region E – High degree of variation (contributes to ligand specificity):
- CBM35: Asparagine near a basic path (Arginine, +).
- BT0996-C: Lysine near a Methionine.

Exploring carbohydrate binding by BT0996-C

Carbohydrate Microarrays

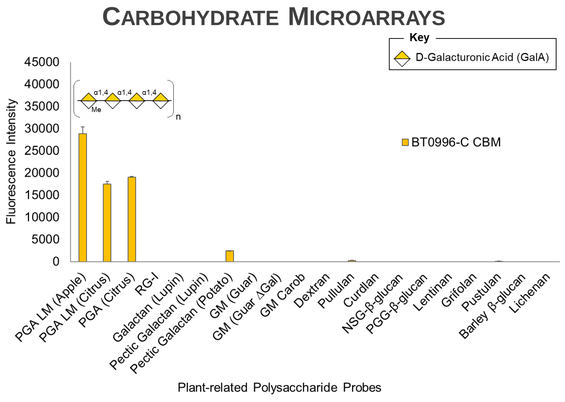


Fig. 5: Carbohydrate microarray analysis of BT0996-C to 20 plant-related polysaccharides. The major oligosaccharide sequence domains present in the bound polysaccharides are represented using the updated symbol nomenclature for glycans (SNFG). **PGA** – Polygalacturonic Acid; **LM** – Low Methoxyl; **RG-I** - Rhamnogalacturonan I; **GM** – Galactomannan)

MicroScale Thermophoresis (MST)

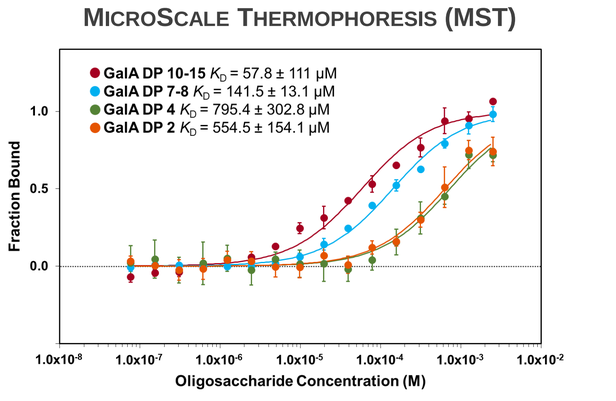


Fig. 6: Initial fluorescence analysis of BT0996-C with PGA derived oligosaccharides with degree of polymerization (DP) 2-15 α1,4-galacturonic acid units. Dose-response curves were fitted to a one-site binding model to obtain the K_D values. Error bars indicate the standard deviation of quadruplicate (n=4) or duplicate experiments (GalA DP 10-15, n=2).)

- **BT0996-C showed restricted binding** to the **pectic polysaccharides** that contain the anionic **α1,4-galacturonic acid chain**, resembling the backbone structure of RG-II (Fig. 7).
- **BT0996-C showed an increase of affinity with the increase of the oligosaccharide length** (Fig. 8).
- The **preference** shown by the positively charged BT0996-C **for anionic polysaccharides** suggests that **ionic interactions may be essential for recognition**.

Key findings for BT0996-C

- Despite the **overall fold conservation** with CBMs from CAZY family 35, BT0996-C presents **striking differences in its 3D structure, with a more positively charged putative binding site**, which may reflect its distinct carbohydrate-binding specificity detected in the microarray analysis.
- The binding and measured affinity to the negatively charged **α1,4-polygalacturonic acid** carbohydrates suggests that this **CBM could promote association of BT0996 to the complex structure of RG-II**, **potentiating its enzymatic degradation**.

References

1. Martens E. *et al*., PLoS Biol, 2011, 9, e1001221;
2. Pérez S. *et al*., Biochimie, 2003, 85, 109-121;
3. Ndeh D. *et al*., Nature, 2017, 544, 65-70;
4. Montanier C. *et al*., PNAS, 2008, 106, 3065-3070;
5. Hudson K. L. *et al*., JACS, 2015, 137, 15152-15160;
6. Sainz-Polo M. A. *et al*., JBC, 2014, 289, 31088-31101;
7. Abbott D. W. *et al*., COSB, 2014, 28, 32-40;

Funding

Project supported by the Portuguese Science and Technology Foundation (FCT-MCTES) through grants PTDC/BBB-BEP/0869/2014, PTDC/BIA-MIB/31730/2017, SFRH/BD/143494/2019 and DL-57/2016 Program Contract.
This work was also supported by the Applied Molecular Biosciences Unit - UCIBIO which is financed by national funds from FCT (UIDB/04378/2020).

Acknowledgments

The authors acknowledge the European Synchrotron Radiation Facility (Grenoble, France) and ALBA (Barcelona, Spain) for access to beamlines ID30B and BL-13 XALOC, respectively.