Deciphering Protein-carbohydrate Recognition in Biodegradation of Plant Cell Wall using Carbohydrate Microarrays and X-ray Crystallography

D. Ribeiro[a*], E. J. Cabrita[a], C. A. Fontes[b], M. J. Romão[a], A. S. Palma[a] & A. L. Carvalho[a]

[a]UCIBIO-REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[b]CIISA/FMV-UL, Faculdade de Medicina Veterinária, Universidade de Lisboa, Portugal

Campus da Caparica, Quinta da Torre, 2928-516 Caparica, Portugal
www.ucibio.pt

*Corresponding author: d.ribeiro@campus.fct.unl.pt

CBMs in plant cell wall biodegradation

Some anaerobic microbial organisms assemble a multi-protein complex of Carbohydrate Active enZYmes (CAZYmes) highly efficient for plant cell-wall polysaccharide biodegradation - the cellulosome. In this assembly, the enzymes are often appended to non-catalytic modules – Carbohydrate Binding Modules (CBMs), which bring the adjoining catalytic modules in contact with target polysaccharides, highly-potentiating the enzymes' catalytic efficiency and the cellulolytic capability of the bacteria[1].

Bacterial genome sequencing revealed numerous putative CBM sequences, which are deposited in the CAZy database (www.cazy.org) and await elucidation. Combining the high-throughput screening feature of the microarray technology with X-ray crystallography, we are uncovering oligosaccharide ligands for CBMs of biotechnologically relevant cellulolytic bacteria, such as *Ruminococcus flavefaciens* FD-1 and *Clostridium thermocellum*.

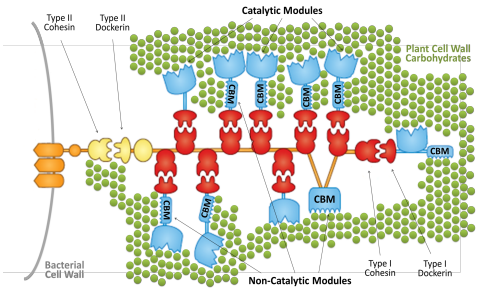
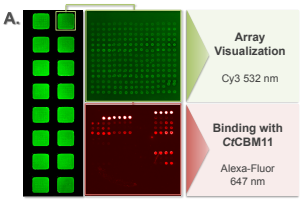


Figure 1. Representation of plant cell wall degradation by the *Clostridium thermocellum* cellulosome (adapted from Fontes & Gilbert, 2010)

Our most recent results on the carbohydrate-recognition of a *C. thermocellum* CBM assigned to CAZY family 11 (*Ct*CBM11)[2,3], reveal the structure of this module in complex with its natural ligands (manuscript in preparation), as indicated in the carbohydrate microarrays[4], ITC[2] and STD-NMR[3] analysis.

Carbohydrate Microarray analysis of *Ct*CBM11



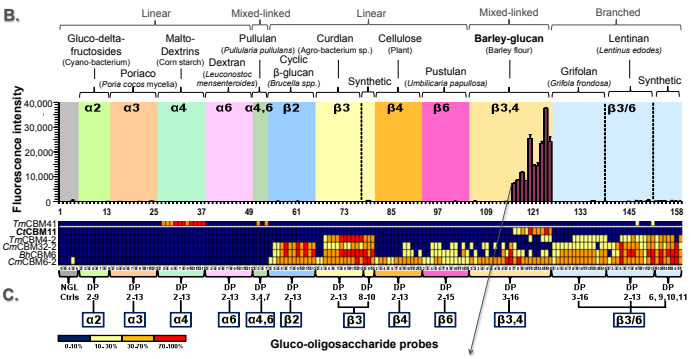


Figure 2. Analysis of the carbohydrate binding specificity of *Ct*CBM11. (**A**) 'Designer' oligosaccharide probes from plant, fungal and bacterial glucan polysaccharides are robotically arrayed, in the form of neoglycolipids (carbohydrate probes)[5], onto nitrocellulose-coated glass slides (16 pad). Cy3 dye is included as a marker (green emission). Binding spots are visualised with Alexa Fluor 647 (red emission)[6]. Binding of up to 16 CBMs to 64 different probes can be analyzed in the same slide. (**B**) A glucome microarray with gluco-oligosaccharide NGL probes, using ligand-bearing glucans as sources of gluco-oligosaccharides(REF), was used to study *Ct*CBM11. The fluorescence intensity signals are quantified and compared as absolute values or as relative binding intensities to reveal different patterns of binding, such as linkage specificity and chain length requirement, and these can be analyzed in the form of graphic or (**C**) matrix, where the binding patterns of different proteins can be compared[4]. (**D**) Chemical structure of the barley heptasaccharide 1,3:1,4-β-glucotetraose identified for *Ct*CBM11.

- *Ct*CBM11 binds exclusively to barley-derived oligosaccharides (Fig. 2B, C), confirming the previously described high-affinity to barley-derived tetrasaccharide (Fig. 2D) with a reducing terminal 3-linkage[2].
- Binding was observed only to DP-7 (Fig. 2D) and longer chain probes (Fig. 2B, C).
- The lack of binding to the tetrasaccharide may be due to steric hindrance of the reducing 3-linked Glc, which is derivatized to the lipid. This may prevent correct access by the protein for binding.

Concluding remarks

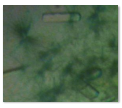
◆The use of carbohydrate microarrays is enabling the screening of a wide range of protein-carbohydrate interactions in a high-throughput manner using only minute amounts of sample.

◆X-ray crystallography allows to characterize at molecular and atomic level those interactions.

◆In this work, the combined approach has been successfully achieved, as *Ct*CBM11 structures in complex with its previously identified natural ligands, provide the atomic information on what has been observed and proposed by carbohydrate microarrays, as well as by ITC and STD-NMR studies.

Structural determinants of the specificity of *Ct*CBM11

The complex with barley 1,3:1,4-β-glucotetraose



- Co-crystallization setups of *Ct*CBM11 with barley 1,3:1,4-β-glucotetraose were done after overnight incubation at 4 °C.
- Single crystals grew within three to four weeks among earlier grown sea urchin-like crystals. Data was collected to 1.45 Å resolution at Diamond Light Source (I02) and crystals belong to the trigonal space group *R*3 (or *R*3:*H*), with unit cell dimensions *a*=*b*=103.2 Å and *c*=39.6 Å.

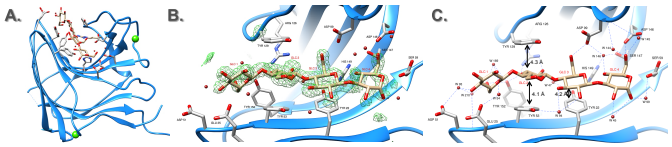


Figure 3. (**A**) Ribbon representation of the overall three-dimensional structure of *Ct*CBM11 complex with barley 1,3:1,4-β-glucotetraose (**PDB 5afc**) (manuscript in preparation) exhibiting the distorted β-barrel composed of two six-stranded antiparallel β-sheets, which form a convex side and a concave side[2]. The two calcium ions are indicated as green spheres. The concave side of *Ct*CBM11 forms the binding cleft where the ligand is accommodated. Residues inside binding cleft that interact with the ligand are shown as ball-and-stick models. (**B**) The mF_o-DF_c electron density map around the ligand molecule is shown in green and contoured at 2.5σ. (**C**) Direct and water-mediated hidrogen bonds between the protein residues and the ligand are represented by blue dashed lines. Black arrows evidence the ligand π-π stacking between Tyr 129, Tyr 53 and Tyr 22[3]. The structure of the complex is deposited in the PDB with accession number 5afc.

- As previously proposed, the ligand interacts with the protein mainly by hydrogen bonds, directly with residues Tyr152, Arg126, Asp99 and 146, as well as water-mediated contacts with residues Asp51, Glu25, Tyr22 and 53, His149, Ser147 and 59.
- The π-π stacking with Tyr22, 53 and 129 confirmed these residues to play a key role in the carbohydrate recognition, guiding and packing the ligand chain, where the β1,3-linked Glc position also appears to have an important role in the chain positioning in the binding cleft.

The complex with barley 1,3:1,4-β-glucohexaose

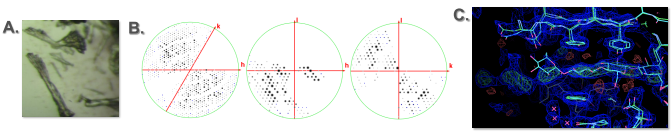


Figure 4. (**A**) Crystals of *Ct*CBM11 in complex with barley 1,3:1,4-β-glucohexaose grown within one to two weeks, in the same previous conditions. X-ray diffraction data on one of these crystals was collected to 2.5 Å resolution at ESRF (ID23-EH2) and crystals belong to the trigonal spacegroup *R*3 (or *R*3:*H*), with unit cell dimensions *a*=*b*= 104.9 Å and *c*= 39.5 Å. (**B**) Calculated precession images of HKL zones *hk*0, *h*0*l* and 0*kl* of an unmerged data set. The apparent hexagonal symmetry is visible in the precession images. (**C**) View of the binding cleft of the unrefined structure of *Ct*CBM11 complex with barley 1,3:1,4-β-glucohexaose, represented as stick in green, superposed with the 5afc structure, represented as stick in light blue. The 2mF_o-DF_c electron density map is shown in dark blue, contoured at 1σ, and the mF_o-DF_c electron density map is shown in green and red and contoured at 2.5σ. The density of the hexasaccharide can already be perceived.

- Although these crystals appeared to be quite multiple, data was possible to collect and integrate.
- Immediately after phasing, the ligand electron density could be observed in the binding cleft, and by superimposing with the tetrasaccharide, the density of the fifth and sixth sugar rings could be perceived.
- As structure refinement is not yet complete, position of the β1,3-linked Glc is not clear, but these results are in agreement with the binding specificities observed in the microarrays analysis.

References

1. Fontes, C.M., Gilbert, H.J. 2010. *Annu. Rev. Biochem.*, 79, 655-681.
2. Carvalho, A.L., *et al.* 2004. *J. Biol. Chem.* 279, 34785-34793.
3. Viegas, A. *et al.* 2008. *FEBS Journal.* 275, 2524–2535.
4. Palma, A.S., *et al.* 2015. *Mol. Cell. Proteomics.* 10 Feb. (Epub ahead of print)
5. Palma, A.S., *et al.* 2014. *Curr. Opin. Chem. Biol.*, 182, 87-94.
6. Palma, A.S., *et al.* 2010. *Methods Enzymol*, 478, 265-86.

Collaborations

- Ten Feizi group, Imperial College London
- Carlos Fontes group, CIISA, FMV-UTL

Funding

Supported by the Portuguese Science and Technology Foundation (FCT-MEC) through grants RECI/BBB-BEP/0124/2012, PTDC/QUI-QUI/112537/2009 and PTDC/BIA-PRO/118377/2010 and UID/Multi/04378/2013 (to UCIBIO).