Molecular insight into a new low affinity xylan binding module from the xylanolytic gut symbiont *Roseburia intestinalis*

Maria Louise Leth[1*], Morten Ejby[1*], Eva Madland[2], Yoshihito Kitaoku[2], Dirk Jan Slotboom[3], Albert Guskov[3], Finn Lillelund Aachmann[2], Maher Abou Hachem[1]

1. Dept. of Biotechnology and Biomedicine, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark. 2. NOBIPOL, Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, N-7491 Trondheim, Norway. 3. Membrane Enzymology, Institute for Biomolecular Sciences & Biotechnology, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. * These authors contributed equally to this work.

***Correspondence to Maher Abou Hachem*:** maha@bio.dtu.dk, Søltofts plads 224, 2800 Kgs. Lyngby, Denmark

Running title

New xylan binding module from *R. intestinalis*

Key words

Carbohydrate binding modules, xylanase, butyrate, human gut microbiota, prebiotics.

Abbreviations

GH, glycoside hydrolase; CBM, carbohydrate binding modules; X6, xylohexaose; X4, xylotetraose; HGM, human gut microbiota; SCFA, short chain fatty acid, AX, arabinoxylan; GX, glucuronoxylan; WAX, wheat arabinoxylan; BGX, birch glucuronoxylan; XOS, xylo-oligosaccharide; XAXXX, 3^3-α-L- and 2^3-α-L-arabinofuranosyl-xylotetraose; XUXXX, 2^3-(4-*O*-methyl-α-D-glucuronyl)-xylotetraose; SAD, single-wavelength anomalous diffraction; ^{15}N-HSQC, heteronuclear single quantum coherence spectroscopy.

Databases

Structural data are available in the PDB database under the accession number 6SGF. Sequence data are available in the GenBank database under the accession number EEV01588.1. The assignment of

the *R. intestinalis* xylan binding module into the CBM86 new family is available in the CAZy database (http://www.cazy.org/CBM86.html).

Abstract

Efficient capture of glycans, the prime metabolic resources in the human gut, confers a key competitive advantage for gut microbiota members equipped with extracellular glycoside hydrolases (GHs) to target these substrates. The association of glycans to the bacterial cell surface is typically mediated by carbohydrate binding modules (CBMs). Here we report the structure of *Ri*CBM86 appended to a GH10 xylanase from *Roseburia intestinalis*. This CBM represents a new family of xylan binding CBMs present in xylanases from abundant and prevalent healthy human gut Clostridiales. *Ri*CBM86 adopts a canonical β-sandwich fold, but shows structural divergence from known CBMs. The structure of *Ri*CBM86 has been determined with a bound xylohexaose, which revealed an open and shallow binding site. *Ri*CBM86 recognizes only a single xylosyl ring with direct hydrogen bonds. This mode of recognition is unprecedented amongst previously reported xylan-binding type-B CBMs that display more extensive hydrogen-bonding patterns to their ligands or employ Ca^{2+} to mediate ligand binding. The architecture of *Ri*CBM86 is consistent with an atypically low binding affinity (K_D≈0.5 mM for xylohexaose) compared to most xylan binding CBMs. Analyses using NMR spectroscopy corroborated the observations from the complex structure and the preference of *Ri*CBM86 to arabinoxylan over glucuronoxylan, consistent with the largely negatively charged surface flanking the binding site. Mutational analysis and affinity electrophoresis established the importance of key binding residues, which are conserved in the family. This study provides novel insight into the structural features that shape low-affinity CBMs that mediate extended bacterial glycan capture in the human gut niche.

Introduction

The human gut microbiota (HGM) consists of trillions of microorganisms that exert a profound impact on human health, especially via modulation of host immune- and metabolic homeostasis[1,2]. The molecular dialogue of the microbiota with the host is typically communicated via microbial metabolites, whereby short chain fatty acids (SCFAs) produced from fiber fermentation play a key role[3]. The most common SCFAs are acetate, propionate and butyrate, all of which are considered beneficial to human health[4]. Notably, SCFA profiles generated from fiber fermentation are specific to distinct taxonomic groups, e.g. members of the dominant genus *Bacteroides* produce mainly acetate (and lower amounts of propionate), whereas members from *Clostridium* group XIVa group[5,6] are key butyrate producers[7]. Bacterially produced butyrate has received increasing attention due to its role in enforcing the gut barrier by increasing the proliferation rate of colonocytes and strengthening tight junctions. Moreover, butyrate down-regulates the expression of inflammatory cytokines and increases colonic regulatory T cells by inhibition of host histone deacetylases[8,9]. Thus, butyrate producers are considered an indicator of a healthy HGM and make a marked contribution to maintaining a balanced and healthy community in the human gut[10]. Despite these pronounced physiological roles, little attention has been dedicated to understating the interactions of butyrate producing members of the HGM with dietary glycans, as opposed to other taxonomic groups that are ascribed a probiotic status, e.g. bifidobacteria[11–13] and lactobacilli[14,15].

Roseburia intestinalis from the *Clostridium* cluster XIVa is an abundant (up to 5 % of the total microbiota) and prevalent butyrate producing Firmicute[7,16]. The abundance of *R. intestinalis* is reduced in type 2 diabetes[17], Chron's disease[18–20], and colorectal cancer[21] patients, which is consistent with the association of this species to a balanced microbiota in healthy humans. *R. intestinalis* has also been shown to adhere to mucin[22], reflecting intimate association with the host and production of butyrate close to the surface of the enterocytes. *R. intestinalis* is atypical amongst human gut Firmicutes by encoding a considerable repertoire (>130) of glycoside hydrolases (GHs) and

74 polysaccharide lyases[23] indicative of extensive saccharolytic potential. Accordingly, *R. intestinalis* is
75 an appropriate model to investigate the strategy of complex glycan utilisation by butyrate producing
76 *Clostridium* XIVa members.

77 *R. intestinalis* and *Eubacterium rectale*, both affiliated to the *Clostridium* XIVa, have been proposed to
78 be key primary degraders of the prime dietary fiber xylan based on enrichment from faecal samples
79 and *in vitro* growth experiments[24,25]. Xylan comprises a β-(1→4)-xylosyl backbone with a variety of
80 side chain substitutions that vary considerably according to botanic origin and tissue. Arabinoxylan
81 (AX), the dominant structural component in the cereal cell wall[26], is substituted with L-arabinosyl
82 residues at C2, C3 or both positions of backbone xylosyl units. Xylan is also present in lower amounts
83 in vegetables and fruits as glucuronoxylan[27] (GX), which is decorated with (4-*O*-methyl)glucuronic
84 acid at the C2 position of xylosyl units. Both AX and GX are further acetylated at C2, C3 or both
85 positions. The molecular apparatus of xylan utilisation by *R. intestinalis* has been recently
86 described[5]. Extracellular capture and break down of xylan is mediated by a modular xylanase of
87 GH10 (*Ri*Xyn10A). This enzyme, which is conserved within the species, comprises an N-terminal
88 carbohydrate binding module (CBM) from a previously unknown family (henceforth designated as
89 *Ri*CBM86) followed by a CBM22, a GH10 catalytic module, a tandem repeat of CBM9 and two C-
90 terminal putative cell-attachment domains. Curiously, *Ri*CBM86 was specific to xylan, but it displayed
91 relatively low affinity (K_D≈0.5 mM for xylohexaose (X6) as opposed to about a 7-fold higher average
92 affinity of the truncated enzyme lacking this CBM for the same ligand[5]. Interestingly, *Ri*CBM86
93 prefers the nutritionally more abundant arabinoxylan as compared to glucuronoxylan judged by
94 retardation in affinity electrophoresis gels.

95 Association to complex glycans, such as xylan, offers a competitive advantage for bacteria in the
96 densely populated milieu of the gut. Firmicutes from *Clostridium* XIVa group frequently have large
97 modular cell-attached glycoside hydrolase (GHs) containing multiple carbohydrate binding modules
98 (CBMs) for capture and hydrolysis of polysaccharides[5,6,28,29]. To examine the mode of recognition

and discrimination of *Ri*CBM86 to different xylans, we have determined the structure of this module and performed binding analyses to glucurono- and arabinoxylan and oligosaccharides thereof using NMR spectroscopy. *Ri*CBM86 displays an open and shallow binding site with only direct hydrogen bonds to the C2-OH and C3-OH of a single xylosyl moiety, which rationalises the low affinity recognition of xylan. These finding highlight the diversity of CBMs associated with xylan catabolism in the human gut and merit further work to bring insight into the role of low-affinity glycan recognition in enzymes from this ecological niche.

Results

Crystal structure

We determined the structure of *Ri*CBM86 in complex with X6. The structure was solved in the hexagonal space group $P6_5$ (6 molecules in the asymmetric unit) using single-wavelength anomalous diffraction (SAD) with the experimental phase information obtained from the Tb anomalous scattering for data collected on crystals soaked with Tb-Xo4[30]. The data collection and refinement statistics are in Table 1. The structure of *Ri*CBM86 was solved to a maximum resolution of 1.8 Å revealing a β-sandwich fold, consisting of two sheets formed by 11 antiparallel β-strands and 2 helical turns (right handed 3_{10}-helices) connected by loops (Fig. 1A). β-Sheet 1 forms the concave face of the β-sandwich and consists of the strands β2(K39-G43), β5(Y62-T68), β7(I92-Y97), β8(T108-L112) and β10(D129-I135). β-Sheet 2 is formed by β1(V29-T34), β3(D46-A50), β4(G53-F58), β6(N79-A86), β9(E117-I120) and β11(A143-L154). The chemical shifts obtained from the NMR assignment are in good agreement with the secondary structure in the X-ray structure[31]. A striking feature of the CBM is the open solvent accessible ligand-binding site that runs almost orthogonal to the β-strands of sheet 1 (Fig. 1A). A DALI server search against the protein data bank (PDB) identified the closest structural relative of *Ri*CBM86 to be CBM29.2 from the fungus *Piromyces equi*[32] (1W9F, *Z*-score=12.8, primary structure identity 12%), which shows specificity for both β-manno- and β-gluco-oligosaccharides[33]. The second closest structural hit is the CBM84 from xanthan lyase family 8 of *Paenibacillus nanensis*[34]

(6F2P, *Z*-score=11.9). Although the overall structural fold is shared between these modules, the low shared sequence identity (<12%) and the divergence of the binding sites (especially key residues mediating aromatic stacking onto ligands) justify the assignment of *Ri*CBM86 as a representative of a new CBM family.

Ligand binding site

The crystal structure of *Ri*CBM86 in complex with X6 shows clear density for four xylosyl units. The ligand-binding site features an open and shallow surface with the ligand bound in a relaxed helical conformation[35]. The ligand-binding site is defined by Y110, which stacks onto the terminal reducing end moiety of the xylosyl that defines position 1 (Fig. 1B). Xylo-oligosaccharide (XOS) ligands can, however, be accommodated in the opposite directionality with equivalent direct hydrogen bonds (non-reducing end xylosyl stacking onto Y110), but this seems to be less likely as it places the endocylic oxygen at close proximity to the indole ring of W42. Our description will focus on the former orientation for clarity. The second aromatic ridge is provided by Y62 that stacks onto the xylosyl unit at position 3. A third potential stacking residue is W42 (Fig. 1C). The indole solvent accessible face of this residue, however, is largely blocked by a methionine side chain from a neighboring molecule in the crystals. Nonetheless, the terminal non-reducing xylosyl at position 4 stacks onto the edge of the indole ring (Fig. 1C). The recognition of the helical conformation of the XOS is facilitated by the planes of the aromatic rings of Y62 and Y110 being almost orthogonal (≈100˚) to each other (Fig. 1B). The only direct potential hydrogen bonds are observed at position 3 between the C2-OH and K95 N^{ζ}, C3-OH and Q64 $N^{\varepsilon 2}$, K95 N^{ζ} or D102 $O^{\delta 2}$ (Only two of these three potential H-bonds are possible). Additional water mediated potential hydrogen bonds may also contribute to the recognition. Dynamic analysis by NMR characterized *Ri*CBM86 as being predominantly rigid, with limited flexibility in two loop regions, E71-I73 and G124-A127 as well as the termini (Fig. 2).

Ligand binding analysis using NMR spectroscopy

The changes in ^{15}N-HSQC (heteronuclear single quantum coherence spectroscopy) spectra of *Ri*CBM86 were monitored and the change in chemical shifts for both the N and H atoms upon titration with undecorated xylotetraose (X4), a 1:1 mixture of 3^3-α-L- and 2^3-α-L-arabinofuranosyl-xylotetraose (XAXXX) and 2^3-(4-*O*-methyl-α-D-glucuronyl)-xylotetraose (XUXXX) was followed. The affinity of the *Ri*CBM86 was lowest for XUXXX, while the higher affinity for XAXXX and X4 resulted in a chemical shift difference in the same order for the two latter ligands (Fig. 3, Table 2). This binding profile and the range of affinity for X4 are in excellent agreement with the previously reported data[5]. The change in chemical shift occurred mainly at the binding site and the flanking area (Fig. 3). The amino acids Y62, Q64, K95, D102 and Y110, which are observed to interact with the ligand in the crystal structure, showed a significant chemical shift difference after titration with the three ligands, except for Q64 with XUXXX. An interesting observation is that G111 undergoes a change in chemical shift in the ^{1}H dimension only for the decorated substrates, which is suggestive this region of *Ri*CBM86 may be involved in the accommodation of side chains substituted at the C2-OH of the xylosyl at position 1. Neighboring G111, is Y110 which provides aromatic stacking interactions for the xylan back bone of substrates.

The interactions between *Ri*CBM86 and birch glucuronoxylan (BGX) as well as wheat arabinoxylan (WAX) were also analyzed by monitoring the ^{15}N-HSQC spectra upon titration (Fig. 4). Due to the strong interaction between *Ri*CBM86 and WAX, some of the signals were broadened beyond detection. The signals for only the WAX ligand expanded to the backside of the protein. The chemical shift difference was lower for BGX, indicating weaker binding affinity to *Ri*CBM86 than WAX. This, in addition to the observations made with oligomeric substrates, provides evidence for the preference of *Ri*CBM86 for arabinosyl substitutions compared to glucuronosyl substitutions both on XOS and xylan.

Mutational analysis of binding residues

The crystal structure and the NMR binding analyses suggested that Y62 and Y110 likely provide aromatic stacking interactions to two xylosyl units of bound xylan. The edge of the indole ring of W42

makes van der Waals contacts with the xylosyl at position 4, which may contribute to restricting the ligand confirmation at this site. An alanine scanning mutagenesis approach was used to investigate functional significance of the aromatic residues together with the invariant lysine (K95), which recognizes the xylosyl at position 3 with a potential bidendate polar interaction (Fig. 1B). The wild type *Ri*CBM86 was thermostable with an unfolding temperature T_m= 74.1˚C, which was only modestly affected by the mutations based on the identical thermograms under 55˚C (Fig. 5A). This suggests that the overall protein structure was retained by the mutants, despite local rearrangements. The binding of the Y62A, K95A and Y110A to xylan was abolished based on affinity electrophoresis, whereas the affinity of the mutant W42A was markedly reduced, especially on WAX (Fig. 5B). The side chain of K95 is crucial for binding as it provides the only charged hydrogen bond to the xylosyl ring that is stacked onto Y62. Similarly, each of the two aromatic stacking tyrosines Y62 and Y110 is also essential for xylan binding, whereas W42 contributes to the xylan affinity, albeit to a less extent. This latter residue possibly stabilizes the xylosyl at position 4 as observed in the crystal structure. The chemical shift changes of W42 are just above significance threshold for XAXXX and X4 and below that for the lower affinity XUXXX, consistent with the observed limited contacts of the indole side chain with the XOS ligand. Notably, W42 is conserved in all but two homologues of *Ri*CBM86 (see sequence analysis in the next section), which is in agreement with the observed impact on the function of the CBM.

RiCBM86 represent a new family of CBMs from xylanases observed in a taxonomically related Clostridiales

*Ri*CBM86 confers affinity to xylan and XOS but lacks homologues with an assigned function[5]. A blast search against the non-redundant database identified 19 homologs from different butyrate producing strains from the Clostridiales order of gut Firmicutes. An analysis of these sequences revealed that several structural residues, e.g. glycines and prolines, in addition to residues involved in xylan binding are conserved. Members of CBM86 are exclusively located at the N-termini of GH10 xylanases (Fig. 6),

198 which together with the narrow distribution among related gut bacteria points to a highly specialized
199 nature of these binding modules.

200 **Discussion**

201 *Architecture of the ligand-binding site of Ri*CBM86 *is consistent with low affinity ligand binding.*
202 The ligand-binding site of *Ri*CBM86 features a shallow and open binding surface that accommodates
203 four xylosyl units. Only about a 4-fold increase in affinity for X6 was previously observed as compared
204 with X4[5], consistent with the presence of only minor additional contacts that stabilise the binding
205 beyond the observed X4 ligand similar to other xylan binding modules, e.g. of CBM6[36] and
206 CBM15[37]. The increase in affinity could also be due to entropic factors, i.e. more stable helical
207 structure of the longer xylan or oligomers thereof as compared to a tetraose.

208 The architecture of the binding site of *Ri*CBM86 is different from most type-B xylan specific CBMs[38]
209 e.g. from CBM4[39], CBM6[40], CBM15[37] and CBM22[41] (Fig. 7A-E). The deeper and more occluded
210 binding site in these latter CBM families is defined by loops connecting the sandwich β-strands and
211 pointing into the binding site. By contrast, the equivalent loops in *Ri*CBM86 are pointing downwards
212 and away from the ligand, which creates a relatively flat open binding surface topology (Fig. 1). To our
213 knowledge, only a few characterized type-B xylan specific CBMs, have similar open binding sites
214 reminiscent of *Ri*CBM86, e.g. CBM36[42] (Fig. 7F) and CBM60[43] that are structurally similar to each
215 other. Similarly, to *Ri*CBM86, a single xylosyl-binding site dominates ligand recognition in the shallow
216 cleft of these CBMs. A key difference, however, between *Ri*CBM86 and CBM36 is that a Ca^{2+} ion
217 mediates the binding in the latter CBM, which appears to yield an affinity about 6-fold higher toward
218 xylohexoase as compared to *Ri*CBM86 [5,42]. Indeed, the affinity of *Ri*CBM86 to X6 (K_D=0.48 mM) is
219 at least 10-fold lower than typical type-B xylan-specific CBMs[40,41,44]. While most xylan-binding
220 counterparts from other families typically recognize 2−3 xylosyl rings along the binding sites with
221 direct hydrogen bonds[37,44], *Ri*CBM86 has a focused recognition of a single xylosyl unit by three
222 direct hydrogen bonds (Fig. 1B). The surface of *Ri*CBM86 flanking the active site is mainly negatively

charged or apolar, which may explain the preferential affinity to arabinoxylan as compared to glucuronic acid substituted xylan (Fig. 1C). Arabinosyl decorations are either tolerated or recognized, based on the similar affinities for the undecorated and decorated ligand X4 and the markedly higher affinity for WAX as compared to BGX (Fig. 3, 4).

Despite the typical β-sandwich fold observed in CBMs, *Ri*CBM86 does not display high structural similarity to any CBM families or other characterized proteins. The closest structural homologues were CBMs with affinity to polysaccharides with a different structural symmetry than xylan, such as β-mannan or xanthan. Indeed the closest structural homologue is from CBM29, which shares a shallow binding site that prefers cello-oligosaccharides (K_D =31.4 μM) [32]. The lack of conserved ligand binding residues between *Ri*CBM86 and distant functionally described orthologues, is consistent with the functional divergence of the new CBM family represented by *Ri*CBM86. To date, 19 non-redundant sequences with high similarity to *Ri*CBM86 are retrieved from the NCBI database. Both the aromatic and the polar residues that interact with the bound ligand in *Ri*CBM86 are highly conserved in these sequences (Fig. 1B). Additionally structurally important amino acid residues such as glycines and a proline are either invariant or highly conserved in this new CBM family.

Rationale for having lower affinity xylan binding in modular xylanase?

Having large extracellular enzymes with a variety of CBMs seems to be common in Clostridiales from the human gut. *R. intestinalis* has a large modular GH26 mannanase with two CBMs[6] and both *Eubacterium rectale* and *Butyrivibrio fibrisolvens* possess large modular α-amylases with 5 and 2 CBMs, respectively for capturing starch[28,45]. *Ri*CBM86 is followed by a CBM22, a GH10 catalytic module and a tandem repeat of CBM9 (Fig. 6). Notably the architecture of characterized CBM22 and CBM9 are different from each other and from the *Ri*CBM86. Members of CBM9 are type-C CBMs that possess a binding slot able to accommodate two terminal xylosyl units in xylan [46], whereas CBM22 possess

a deep extended binding cleft[47,48] for the accommodation of a single xylan chain. Thus, the three different families of CBMs in *Ri*Xyn10A orchestrate the binding of substrate by being able to capture either the terminal reducing ends or internal regions of xylan by the CBM9 (assuming similar binding mode to known members) or CBM22/CBM86, respectively. These CBMs also appear to have variable affinities as judged from average affinities for X6 of *Ri*CBM86, the full-length enzyme and a truncated variant lacking *Ri*CBM86, which have affinities of 479 μM, 128 μM and 65 μM, respectively[5]. Thus, the enzyme construct lacking *Ri*CBM86 possesses an affinity about 7 fold higher than *Ri*CBM86, which suggests that at least one or more of the three remaining CBMs in *Ri*Xyn10A possess markedly higher affinities for X6. This variable affinity and multiplicity of CBMs may confer a dynamic binding where the substrate is anchored to the enzyme surface in between consecutive catalytic cycles to minimize diffusional loss. Notably, similar low affinity CBMs in the α-amylase that confers the capture and breakdown of starch by the related gut symbiont *E. rectale* have been reported. Thus, the N-terminal CBM82 and the C-terminal CBM83 of this α-amylase displays affinities of ≈1 and 3 mM, respectively to maltoheptaose[45], which is substantially lower than the internal CBMs constructs. Another example of low-affinity (K_D≈0.58 mM for the full-length enzyme towards β-mannohexoase) CBM from the human gut niche is the mannan specific CBM10 connected to a GH5 β-mannanase from *Bifidoabcterium animalis* subsp. *lactis*. Interestingly the latter enzyme is one of the most efficient β-mannanases reported[13]. The evolution of low affinity CBMs may be an adaptation to increase the area of substrate binding with minimal reduction of turnover, i.e. maximizing k_{cat}/k_{off}. Additional experiments are required to evaluate the dynamics of substrate binding and translocation to *Ri*Xyn10A as a model to evaluate the contribution of multiple CBM binding.

Materials and Methods

Chemicals

All chemical were of analytical grade. Wheat arabinoxylan (WAX), xylohexaose (X6), xylotetraose (X4), 3^3-α-L- and-2^3-α-L-arabinofuranosyl-xylotetraose (XAXXX) in mixture of ≈1:1 were from Megazyme (Wicklow, Ireland). 2^3-(4-*O*-methyl-α-D-glucuronyl)-xylotetraose (XUXXX) was from Cambridge Glycoscience (Cambridge, United Kingdom). Birchwood glucuronoxylan (BGX) was from Carl Roth (Karlsruhe, Germany).

Cloning

The gene fragment encoding the *Ri*CBM86 from *Roseburia intestinalis* L1-82 was amplified from a plasmid encoding the full length xylanase *Ri*Xyn10A (EEVO1588.1, ROSINTL182_06494)[5] using a primer pair (TTTCAGGGCGCCATGGGGGTAAAAAAAGTTTTTACTGCAGAT, GACGGAGCTCGAATTTTAATCCCCCAATTTTGCA). The amplicon, encoding amino acids 28-165 in *Ri*Xyn10A, was cloned into the EcoRI and NcoI restriction site of a pETM-11 vector (kind gift from Dr. Gunter Stier, EMBL, Center for Biochemistry, Heidelberg, Germany)[49] using In-Fusion cloning (Takara). The construct was transformed into *Escherichia coli* DH5α and verified by full sequencing.

Site directed mutagenesis

Specific mutants of *Ri*CBM86 were generated by PCR *Ri*CBM86 as template. The primer pairs were; W42A (CAGCTGAAAGTGGCA**gcg**GGAGACGCGGATTATG, CATAATCCGCGTCTCC**cgc**TGCCACTTTCAGCTG), Y62A (GTCTTTTGCAAAACAG**gct**AATCAGGTGAAATGGACG, CGTCCATTTCACCTGATT**agc**CTGTTTTGCAAAAGAC), K95A (GTACCGATCAGTCTG**gca**GTATACAACGGTGGAGATG, CATCTCCACCGTTGTATAC**tgc**CAGACTGATCGGTAC) and Y118A (GATTAAGCGGACAGACGGAG**gct**ACGATAAATCCATC,

GATGGATTTATCGT**agc**CTCCGTCTGTCCGCTTAATC). The amplicons were incubated with DpnI restriction endonuclease (New England Biolabs) at 37°C for 30 min to remove the template DNA plasmid. The mutated constructs were then transformed into *E. coli* DH5α and each mutants were sequenced to ensure that only the desired mutations had been incorporated into the nucleic acids.

Expression and purification

Recombinant plasmids were transformed into BL21(DE3) (Novagen) for expression of unlabeled and $^{13}C/^{15}N$ double labeled protein and B834(DE3) (Novagen) expression selenomethionine labelled protein. Protein production was performed as previously described for unlabeled protein[5], selenomethionine labelled protein[11], and double labelled $^{13}C/^{15}N$ labelled protein used for the NMR studies[31]. Cell pellets were resuspended in buffer (20 mM HEPES pH 7.5, 0.5 M NaCl, 10% glycerol) and disrupted at 1000 bar by a single passage in a high pressure homogenizer (Standsted Fluid Power, Essex, UK). Recombinant proteins were purified from the supernatant by affinity chromatography using a 5 mL His-Trap HP column (GE Healthcare) and a standard protocol. Pure fractions were concentrated and loaded onto a Hiload 16/60 Superdex 75 pg size exclusion chromatography column (GE Healthcare) mounted on an ÄKTA-AVANT chromatograph (GE Healthcare). For crystallization the His-tag was removed using a TEV-protease. This was done by buffer exchange into buffer (50 mM Tris-HCL pH 8.0, 0.5 mM EDTA, 1 mM DTT) and next adding TEV-protease in a ratio of 1:100 (v/v). After incubation for 24 hours at room temperature, the mixture was passes through a His-Trap column, and the flow through containing the cleaved protein dialyzed into buffer (20 mM MES pH 6.5, 150 mM NaCl). Protein purity was determined by SDS-PAGE and protein concentration were measured spectrophotometrically and calculated from the theoretical molar extinction coefficient (ε_{280nm}= 26930 and 23950 M^{-1} cm^{-1}, for tagged and non-tagged proteins, respectively).

Crystallization and structure determination

Crystals were only obtained in the presence of 1 mM X6 by vapour diffusion in hanging or sitting drops, and grew for 2 days at 5°C with a 1:1 ratio of the protein (18 mg mL^{-1} in 10 mM MES pH 6.5 and 150 mM NaCl) and reservoir solution (0.2 M Cadmium chloride hemi(pentahydrate) 0.1 M Sodium acetate pH 4.8 and PEG 400 35% v/v). An initial crystallisation condition (0.1 M Cadmium chloride hemi(pentahydrate), 0.1 M Sodium acetate pH 4.6 and PEG400 30% v/v at 5 °C) was identified with the Structure Screen (Molecular Dimensions Ltd, UK), using a Mosquito® liquid handling robot (TTP Labtech, UK). The crystals were flash frozen in liquid nitrogen without cryo-protectant. Diffraction data were collected to a maximum resolution of 1.91 and 1.76 Å for derivatized and native crystals respectively, at the DESY beamlines, Hamburg, Germany. The dataset was processed with XDS[50]. The structure was solved in the hexagonal space group $P6_5$ using single-wavelength anomalous diffraction (SAD) with the experimental phase information obtained from data collected at 7.575 KeV for crystals soaked for 1 min with 100 mM Tb-Xo4[30] (Molecular Dimensions) using the Tb anomalous scatterer for phasing. Experimental phasing, initial model building and refinements were performed in the Phenix software suite[51]. Further corrections and model building using the program Coot[52]) resulted in a complete model, which was used in molecular replacement to solve the structure of *Ri*CBM86 in a slightly higher resolution dataset. Manual structure improvement was done in Coot[52]. Ligand molecules were included after the protein parts were build and water molecules were added with Coot, all refinements were performed in phenix_refine. The overall quality of all models was checked using MolProbity[53]. The data collection and refinement statistics are presented in Table 1. The PyMOL Molecular Graphics System, Version 2.0.6 Schrödinger, LLC was used to explore the models and for rendering.

NMR spectroscopy

NMR spectra of 0.1-0.2 mM *Ri*CBM86 in 50 mM sodium phosphate buffer pH 6.5 and 10% D_2O were recorded at 25°C on a Bruker Ascend 800 MHz spectrometer Avance III HD (Bruker Biospin) equipped with a 5 mm Z-gradient CP-TCI (H/C/N) cryoprobe at the NV-NMR-Centre/Norwegian NMR Platform

344 at NTNU (Trondheim, Norway). A single NMR titration was preformed with three oligomeric
345 substrates: X4, XAXXX or XUXXX. Titration points for X4 (mM): 0.5, 1.0, 2.5, 5 and 10 M; XAXXX (mM):
346 0.2, 0.5, 1.0, 2.5, 5.0 and 10; XUXXX (mM): same as for XAXXX with the addition of the following four
347 points of 12.5, 15.0, 20.0 and 25.0. In addition, NMR titrations were also carried out with two xylans:
348 BGX and WAX. The titration with BGX was performed with nine concentrations within 0.04−1.0 mg
349 mL^{-1} and a final point at 2.0 mg BGX. For WAX eight concentrations within 0.04−0.73 mg mL^{-1} and a
350 final point of 1.4 mg WAX. 1D and ^{15}N-HSQC spectra were recorded for each titration point and
351 processed with Topspin version 3.5 and CARA version 1.5 using backbone and side-chain assignments
352 of *Ri*CBM86 have been published elsewhere[31]. The chemical shift perturbation upon titration was
353 followed in ^{15}N-HSQC. Binding parameters were estimated by Gnuplot 5.2 (www.gnuplot.info) using
354 an average of the chemical shift difference ($\Delta\delta$) from the titration of three amino acids, K_D X4 (A59,
355 N63, N93), K_D XAXXX (N63, N93, G111) and K_D XUXXX (N63, N93, G111).

356 Relaxation measurements (T_1, T_2 and ^{1}H-^{15}N NOE) for amide ^{15}N labelled *Ri*CBM86 were recorded. The
357 nuclear spin relaxation times T_1 and T_2 were recorded as pseudo-3D spectra where the two frequency
358 dimensions corresponded to the amide ^{1}H and ^{15}N chemical shifts, respectively. The third dimension
359 was made up of the following variable relaxation time delays: T_1 time points: 0.1, 0.2, 0.5, 1.0, 1.5, 2.0,
360 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0 s and T_2 time points: 17, 34, 68, 136, 170, 204, 237 and 271 ms. The
361 heteronuclear ^{1}H-^{15}N NOE spectra composed of two 2D planes were recorded with and without
362 presaturation, respectively.
363
364 *Affinity electrophoresis*

365 Binding of *Ri*CBM86 and the mutants to WAX (0.1% w/v) and BGX (1% w/v) was assessed in 10%
366 polyamide gels as described in[5].

367 *Differential scanning calorimetry*

368 The thermal stability of the *Ri*CBM86 mutants (1 mg mL^{-1}) was assessed in 10 mM Sodium Phosphate
369 buffer, pH 6.5 using differential scanning calorimetry (DSC) between 20°C and 90°C, 1°C min^{-1} in a
370 Nano DSC instrument (TA Instruments, New Castle, DE, USA). Baseline scans, collected with buffer in
371 both reference and sample cells, were subtracted from sample scans, and NanoAnalyse (TA
372 Instruments) was used to model the reference cell and baseline-corrected thermograms using a two-
373 state model to determine T_m. *Ri*CBM86 was scanned with cooling to assess the reversibility of thermal
374 transitions.

375

376

References

1 Marchesi JR, Adams DH, Fava F, Hermes GD a, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, Thomas L V, Zoetendal EG & Hart A (2015) The gut microbiota and host health: a new clinical frontier. *Gut*, 1–10.

2 Sonnenburg JL & Bäckhed F (2016) Diet–microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64.

3 Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, De los Reyes-Gavilán CG & Salazar N (2016) Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* **7**, 1–9.

4 Cummings JH, Pomare EW, Branch WJ, Naylor CP & Macfarlane GT (1987) Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28**, 1221–1227.

5 Leth ML, Ejby M, Workman C, Ewald DA, Pedersen SS, Sternberg C, Bahl MI, Licht TR, Aachmann FL, Westereng B & Abou Hachem M (2018) Differential bacterial capture and transport preferences facilitate co-growth on dietary xylan in the human gut. *Nat. Microbiol.* **3**, 570–580.

6 La Rosa SL, Leth ML, Michalak L, Hansen ME, Pudlo NA, Glowacki R, Pereira G, Workman CT, Arntzen MØ, Pope PB, Martens EC, Abou Hachem M & Westereng B (2019) The human gut Firmicute Roseburia intestinalis is a primary degrader of dietary β-mannans. *Nat. Commun.* **10**, 905.

7 Louis P & Flint HJ (2009) Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol. Lett.* **294**, 1–8.

8 Chang P V, Hao L, Offermanns S & Medzhitov R (2014) The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl. Acad. Sci.* **111**, 2247 LP – 2252.

9 Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, Glickman JN & Garrett WS (2013) The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science (80-.).* **341**, 569 LP – 573.

10 Louis P, Hold GL & Flint HJ (2014) The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12**, 661–672.

11 Ejby M, Fredslund F, Andersen JM, Žagar¶ AV, Henriksen JR, Andersen TL, Svensson B, Slotboom DJ & Abou Hachem M (2016) An atp binding cassette transporter mediates the uptake of α-(1,6)-linked dietary oligosaccharides in bifidobacterium and correlates with competitive growth on these substrates. *J. Biol. Chem.* **291**, 20220–20231.

12 Viborg AH, Fredslund F, Katayama T, Nielsen SK, Svensson B, Kitaoka M, Lo Leggio L & Abou Hachem M (2014) A β1-6/β1-3 galactosidase from *Bifidobacterium animalis* subsp. lactis Bl-04 gives insight into sub-specificities of β-galactoside catabolism within Bifidobacterium. *Mol. Microbiol.* **94**, 1024–1040.

13 Morrill J, Kulcinskaja E, Sulewska AM, Lahtinen S, Stålbrand H, Svensson B & Abou Hachem M (2015) The GH5 1,4-β-mannanase from *Bifidobacterium animalis* subsp. lactis Bl-04 possesses a low-affinity mannan-binding module and

highlights the diversity of mannanolytic enzymes. *BMC Biochem.* **16**, 26.

14 Møller MS, Fredslund F, Majumder A, Nakai H, Poulsen J-CN, Lo Leggio L, Svensson B & Abou Hachem M (2012) Enzymology and Structure of the GH13_31 Glucan 1,6-α-Glucosidase That Confers Isomaltooligosaccharide Utilization in the Probiotic *Lactobacillus acidophilus* NCFM. *J. Bacteriol.* **194**, 4249 LP – 4259.

15 Møller MS, Goh YJ, Rasmussen KB, Cypryk W, Celebioglu HU, Klaenhammer TR, Svensson B & Abou Hachem M (2017) An Extracellular Cell-Attached Pullulanase Confers Branched α-Glucan Utilization in Human Gut *Lactobacillus acidophilus*. *Appl. Environ. Microbiol.* **83**, e00402-17.

16 Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Antolín M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD & Bork P (2011) Enterotypes of the human gut microbiome. *Nature* **473**, 174–180.

17 Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto J-M, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K & Wang J (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60.

18 Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, Pan C, Shah M, Halfvarson J, Tysk C, Henrissat B, Raes J, Verberkmoes NC, Fraser CM, Hettich RL & Jansson JK (2012) Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. *PLoS One* **7**.

19 Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, Järnerot G, Tysk C, Jansson JK & Engstrand L (2010) A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844-1854.e1.

20 Shen Z, Zhu C, Quan Y, Yang J, Yuan W, Yang Z, Wu S, Luo W, Tan B & Wang X (2018) Insights into *Roseburia intestinalis* which alleviates experimental colitis pathology by inducing anti-inflammatory responses. *J. Gastroenterol. Hepatol.* **33**, 1751–1760.

21 Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S & Zhao L (2012) Structural segregation of gut microbiota

between colorectal cancer patients and healthy volunteers. *ISME J.* **6**, 320–329.

22 Van den Abbeele P, Belzer C, Goossens M, Kleerebezem M, De Vos WM, Thas O, De Weirdt R, Kerckhof F-M & Van de Wiele T (2013) Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an *in vitro* gut model. *ISME J.* **7**, 949–61.

23 El Kaoutari A, Armougom F, Gordon JI, Raoult D & Henrissat B (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11**, 497–504.

24 Chassard C, Goumy V, Leclerc M, Del'homme C & Bernalier-Donadille A (2007) Characterization of the xylan-degrading microbial community from human faeces. *FEMS Microbiol. Ecol.* **61**, 121–131.

25 Sheridan PO, Martin JC, Lawley TD, Browne HP, Harris HMB, Bernalier-Donadille A, Duncan SH, O'Toole PW, Scott KP & Flint HJ (2016) Polysaccharide utilisation loci and nutritional specialisation in a dominant group of butyrate-producing human colonic Firmicutes. *Microb. Genomics* **2**.

26 Izydorczyk MS & Biliaderis CG (1995) Cereal arabinoxylans: advances in structure and physicochemical properties. *Carbohydr. Polym.* **28**, 33–48.

27 Selvendran RR (1987) Chemistry of plant cell walls and dietary fibre. *Scand. J. Gastroenterol.* **5521**, 33–41.

28 Ramsay AG, Scott KP, Martin JC, Rincon MT & Flint HJ (2006) Cell-associated α-amylases of butyrate-producing Firmicute bacteria from the human colon. *Microbiology* **152**, 3281–3290.

29 Cockburn DW, Orlovsky NI, Foley MH, Kwiatkowski KJ, Bahr CM, Maynard M, Demeler B & Koropatkin NM (2015) Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*. *Mol. Microbiol.* **95**, 209–230.

30 Engilberge S, Riobé F, Di Pietro S, Lassalle L, Coquelle N, Arnaud C-A, Pitrat D, Mulatier J-C, Madern D, Breyton C, Maury O & Girard E (2017) Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chem. Sci.* **8**, 5909–5917.

31 Madland E, Kitaoku Y, Sætrom GI, Leth ML, Ejby M, Abou Hachem M & Aachmann FL (2018) 1H, 13C and 15N backbone and side-chain assignment of a carbohydrate binding module from a xylanase from *Roseburia intestinalis*. *Biomol. NMR Assign.*

32 Flint J, Bolam DN, Nurizzo D, Taylor EJ, Williamson MP, Walters C, Davies GJ & Gilbert HJ (2005) Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules. *J. Biol. Chem.* **280**, 23718–26.

33 Freelove AC, Bolam DN, White P, Hazlewood GP & Gilbert HJ (2001) A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of *Piromyces equi*. *J. Biol. Chem.* **276**, 43010–7.

34 Jensen PF, Kadziola A, Comamala G, Segura DR, Anderson L, Poulsen J-CN, Rasmussen KK, Agarwal S, Sainathan RK, Monrad RN, Svendsen A, Nielsen JE, Lo Leggio L & Rand KD (2018) Structure and Dynamics of a Promiscuous Xanthan Lyase from *Paenibacillus nanensis* and the Design of Variants with Increased Stability and Activity. *Cell Chem. Biol.*

35 Lütteke T, Frank M & von der Lieth C-W (2005) Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Res.* **33**, D242–D246.

36 Pires VMR, Henshaw JL, Prates JAM, Bolam DN, Ferreira LMA, Fontes CMGA, Henrissat B, Planas A, Gilbert HJ & Czjzek M (2004) The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *J. Biol. Chem.* **279**, 21560–8.

37 Szabó L, Jamal S, Xie H, Charnock SJ, Bolam DN, Gilbert HJ & Davies GJ (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose: Evidence that xylan binds in an approximate 3-fold helical conformation. *J. Biol. Chem.* **276**, 49061–49065.

38 Gilbert HJ, Knox JP & Boraston AB (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol* **23**.

39 Simpson PJ, Jamieson SJ, Abou Hachem M, Karlsson EN, Gilbert HJ, Holst O & Williamson MP (2002) The Solution Structure of the CBM4-2 Carbohydrate Binding Module from a Thermostable *Rhodothermus marinus* Xylanase. *Biochemistry* **41**, 5712–5719.

40 Czjzek M, Bolam DN, Mosbah A, Allouch J, Fontes CMGA, Ferreira LMA, Bornet O, Zamboni V, Darbon H, Smith NL, Black GW, Henrissat B & Gilbert HJ (2001) The Location of the Ligand-binding Site of Carbohydrate-binding Modules That Have Evolved from a Common Sequence Is Not Conserved. *J. Biol. Chem.* **276**, 48580–48587.

41 Xie H, Gilbert HJ, Charnock SJ, Davies GJ, Williamson MP, Simpson PJ, Raghothama S, Fontes CMG a, Dias FM V, Ferreira LM a & Bolam DN (2001) *Clostridium thermocellum* Xyn10B carbohydrate-binding module 22-2: The role of conserved amino acids in ligand binding. *Biochemistry* **40**, 9167–9176.

42 Jamal-Talabani S, Boraston AB, Turkenburg JP, Tarbouriech N, Ducros VMA & Davies GJ (2004) Ab initio structure determination and functional characterization of CBM36: A new family of calcium-dependent carbohydrate binding modules. *Structure* **12**, 1177–1187.

43 Montanier C, Flint JE, Bolam DN, Xie H, Liu Z, Rogowski A, Weiner DP, Ratnaparkhe S, Nurizzo D, Roberts SM, Turkenburg JP, Davies GJ & Gilbert HJ (2010) Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules. *J. Biol. Chem.* **285**, 31742–31754.

44 Sainz-Polo MA, González B, Menéndez M, Pastor FIJ & Sanz-Aparicio J (2015) Exploring multimodularity in plant cell wall deconstruction: Structural and functional analysis of Xyn10C containing the CBM22-1-CBM22-2 tandem. *J. Biol. Chem.* **290**, 17116–17130.

45 Cockburn DW, Suh C, Medina KP, Duvall RM, Wawrzak Z, Henrissat B & Koropatkin NM (2018) Novel carbohydrate binding modules in the surface anchored α-amylase of *Eubacterium rectale* provide a molecular rationale for the range of starches used by this organism in the human gut. *Mol. Microbiol.* **107**, 249–264.

504 46 Notenboom V, Boraston AB, Kilburn DG & Rose DR (2001) Crystal structures of the family 9 carbohydrate-binding
505 module from *Thermotoga maritima* xylanase 10a in native and ligand-bound forms. *Biochemistry* **40**, 6248–6256.
506 47 Najmudin S, Pinheiro B a., Prates J a M, Gilbert HJ, Romão MJ & Fontes CMG a (2010) Putting an N-terminal end to the
507 *Clostridium thermocellum* xylanase Xyn10B story: Crystal structure of the CBM22-1-GH10 modules complexed with
508 xylohexaose. *J. Struct. Biol.* **172**, 353–362.
509 48 Charnock SJ, Bolam DN, Turkenburg JP, Gilbert HJ, Ferreira LM a, Davies GJ & Fontes CMG a (2000) The X6
510 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: Structure and biochemistry of the
511 *Clostridium thermocellum* X6b domain. *Biochemistry* **39**, 5013–5021.
512 49 Dümmler A, Lawrence A-M & de Marco A (2005) Simplified screening for the detection of soluble fusion constructs
513 expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.* **4**, 34.
514 50 Kabsch W (2010) *XDS*. *Acta Crystallogr. Sect. D* **66**, 125–132.
515 51 Adams PD, Afonine P V, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW,
516 McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) *PHENIX*:
517 a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213–221.
518 52 Emsley P, Lohkamp B, Scott WG & Cowtan K (2010) Features and development of *Coot*. *Acta Crystallogr. Sect. D* **66**,
519 486–501.
520 53 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM & Kapral GJ (2010) MolProbity: all-atom structure validation
521 for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**.
522

523 **Tables**

Table 1 Data collection and refinement statistics

	*Ri*CBM86 X6 Polyvalan Crystallophore No1	*Ri*CBM86 X6 Native
Beamline	PETRA III P13	PETRA III P13
PDB ID		6SGF
Wavelength (Å)	1.649	1.000
Resolution range (Å)	70.9 - 1.91 (1.98 - 1.91)	46.4 - 1.76 (1.82 - 1.76))
Space group	$P6_5$	$P6_5$
Unit cell	141.87 141.87 60.6 90 90 120	141.87 141.87 60.6 90 90 120
Unique reflections [a]	53405 (5006)	67325 (4933)
Multiplicity [a]	9.6 (6.8)	5.8 (1.9)
Completeness (%) [a]	99.30 (93.65)	96.74 (71.29)
CC½ [a]	0.997 (0.898)	0.998 (0.398)
Mean I/σ(I) [a]	14.48 (3.21)	15.83 (1.55)
Wilson B-factor	19.18	21.84
R-factor		0.1794

R-free	0.2237
Number of atoms	6891
Macromolecules	5960
Ligands	248
Water	683
Protein residues	786
RMS bonds (Å)	0.013
RMS angles (°)	1.68
Ramachandran favored (%)	98.19
Ramachandran outliers (%)	0.00
Clash score	6.97
Average B-factor	26.51
Macromolecules	25.48
Ligands	31.97
Water	33.51

[a] Values in the parenthesis are for the highest resolution shell.

Table 2 Binding parameters determined by NMR

	K_D (mM)	B_{max} (Δδ at saturation)
X4	1.09	0.19
XAXXX	1.23	0.17
XUXXX	22.89	0.15

Binding parameters are estimated from a single titration experiment.

Figure legends

Fig. 1. Crystal structure of *Ri*CBM86. (A) Cartoon model of β-sandwich structure of *Ri*CBM86 (PDB accession: 6SGF). The left panel is a top view of sheet 1 formed by five β-strands. The four visible rings of the soaked xylohexaose (X6) are shown in sticks. The view is rotated 180° in the right panel to show sheet 2 formed by six β-strands. (B) The left panel is a close-up of the ligand binding site with subsites numbered in Arabic numerals starting from the reducing end at position 1. The two aromatic residues Y110 and Y62 that stack onto xylosyl rings at positions 1 and 3, respectively. The aromatic side chain of W42 makes limited contacts with the xylosyl at position 4, but it is not positioned for aromatic stacking. The only direct hydrogen bonds that recognize the C2 and C3 hydroxyl groups of the xylosyl at position 3 are shown and the $2F_{obs}–DF_{calc}$ composite omit electron map for the bound ligand is shown at a contour level of 1σ (blue mesh). The right panel shows the binding site rotated about 90° along the axis of the ligand and a sequence logo that reflects the conservation of the binding residues is shown. (C) The electrostatic potential of *Ri*CBM86 (at pH=7) is shown to highlight the topology and the chemistry

of the ligand binding site. The two aromatic stacking residues Y62 and Y110 and W42 are labeled for clarity. The figure was generated with PyMOL.

Fig. 2 Dynamics of *Ri*CBM86 as evaluated by NMR relaxation analysis. ^{1}H-^{15}N NOEs and ^{15}N T_1 and T_2 relaxation times for *Ri*CBM86 were recorded at 800 MHz and 25 °C. Apart from two loops (E71-I73 and G124-A127) and the terminals (parts that normally can display flexibility), the data shows a well-folded and rigid protein structure. Data are with error bars calculated based on the signal-to-noise ratios.

Fig. 3 Interaction of *Ri*CBM86 with xylo-oligosaccharides using NMR chemical shift analysis. The chemical shift differences are after titration with xylo-oligosaccharides; (A) glucurono-xylotetraose (XUXXX), (B) α-L-arabinofuranosyl-xylotetraose (XAXXX) and (c) xylotetraose (X4). The figure was generated with PyMOL.

Fig. 4 Interaction of *Ri*CBM86 with xylans using NMR chemical shift analysis. The chemical shift differences are after titration with xylans; (A) birch glucuronoxylan (BGX) and (B) wheat arabinoxylan (WAX). The figure was generated with PyMOL.

Fig. 5 Analysis of thermal stability and binding to xylan for *Ri*CBM86 and mutants thereof. (A) Reference and baseline subtracted differential scanning calorimetry thermograms, which are normalized to protein concentration. The unfolding temperatures (T_m) were determined using a two state model, which is justified due to the partial reversibility of the traces as judged by partial area recovery following unfolding. (B) Binding of CBMs to a negative control gel (no polysaccharide), 0.1 (w/v) wheat arabinoxylan (WAX) or 1% (w/v) birch glucuronoxylan (BGX) is analyzed using affinity electrophoresis. Lane 1: native marker, lane 2: *Ri*CBM86, lane 3: W42A, lane 4: Y62A, lane 5: K95A, lane 6: Y110A.

Fig. 6 Modular organization of 19 *Ri*CBM86 homologous sequences. The modular organization was predicted using HMMR (http://hmmer.org/)[2] and dbCAN (http://bcb.unl.edu/dbCAN2/blast.php)[3]. Purple: novel carbohydrate binding module (CBM86), pink: carbohydrate binding module of family 22 (CBM22), yellow: catalytic module of glycoside hydrolase family 10 (GH10), green: carbohydrate binding module of family 9 (CBM9). The asterisk indicates that this putative CBM9 cannot be predicted with these tools, even though it is assigned as CBM9 is the CAZy database.

Fig. 7 Comparison of the binding site architecture of xylan-specific CBMs. (A) *Ri*CBM86 from *Roseburia intestinalis* (PDB ID 6SGF), (B) CBM4 from *Rhodothermus marinus* (PDB ID 2Y64), (C) CBM6 from *Clostridium stercorarium* (PDB ID 2UY4), (D) CBM15 from *Cellvibrio japonicus* (PDB ID 1GNY), (E) CBM22 from *Paenibacillus barcinonensis* (PDB ID 4XUR), (F) CBM36 from

Paenibacillus polymyxa (PDB ID 1UX7). A calcium ion is represented in brown in panel F. The figure was generated with PyMOL.

Acknowledgement

This project was funded by a Graduate School DTU Scholarship, Lyngby, Denmark. Additional funding were from the Independent Research Fund Denmark (DFF, FNU), Research Project 2 grant for MAH (ID: 4002-00297B) and the Norwegian NMR Platform, NNP from the Research Council of Norway for FLA (ID: 226244). We wish to thank Bernard Henrissat, architecture et fonction des macromolécules biologiques, CNRS, Aix-Marseille University, for discussion on the *Ri*CBM86 assignment.

Author Contributions

M.L.L., M.A.H, F.L.A. planned experiments; M.L.L., M.E., E.M. and Y.K. performed experiments and analyzed data; D.J.S, A.G., F.L.A, M.A.H supervised experiments and contributed reagents or other essential material; M.L.L. and M.A.H wrote the paper with contribution from all authors. All authors approved the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Figure 1

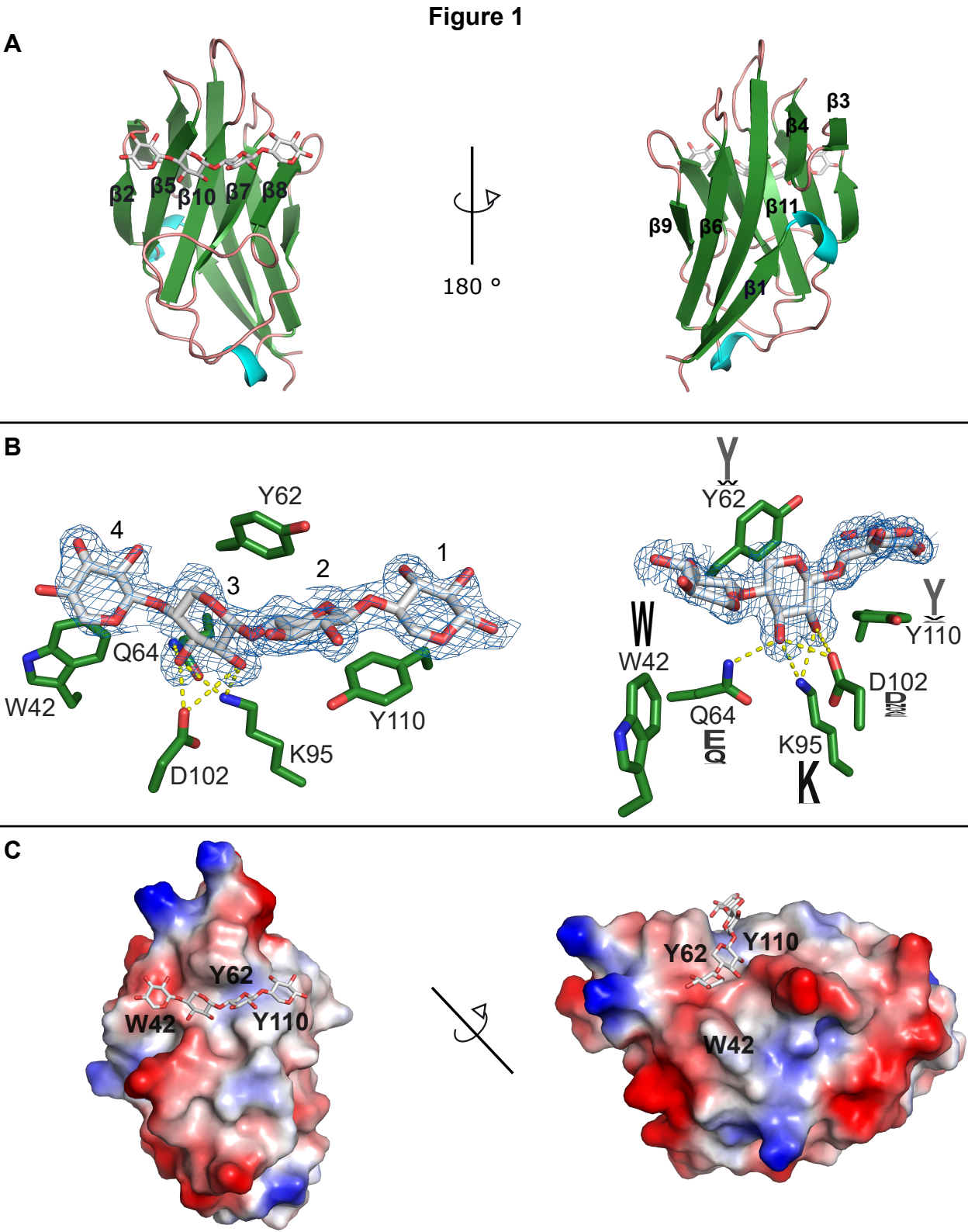

A
β2
β5
β10
β7
β8
180 °
β3
β4
β11
β9
β6
β1
B
Y62
4
3
2
1
Q64
W42
Y110
K95
D102
Y62
W42
Q64
K95
D102
Y110
C
Y62
W42
Y110
Y62
Y110
W42

Figure 2

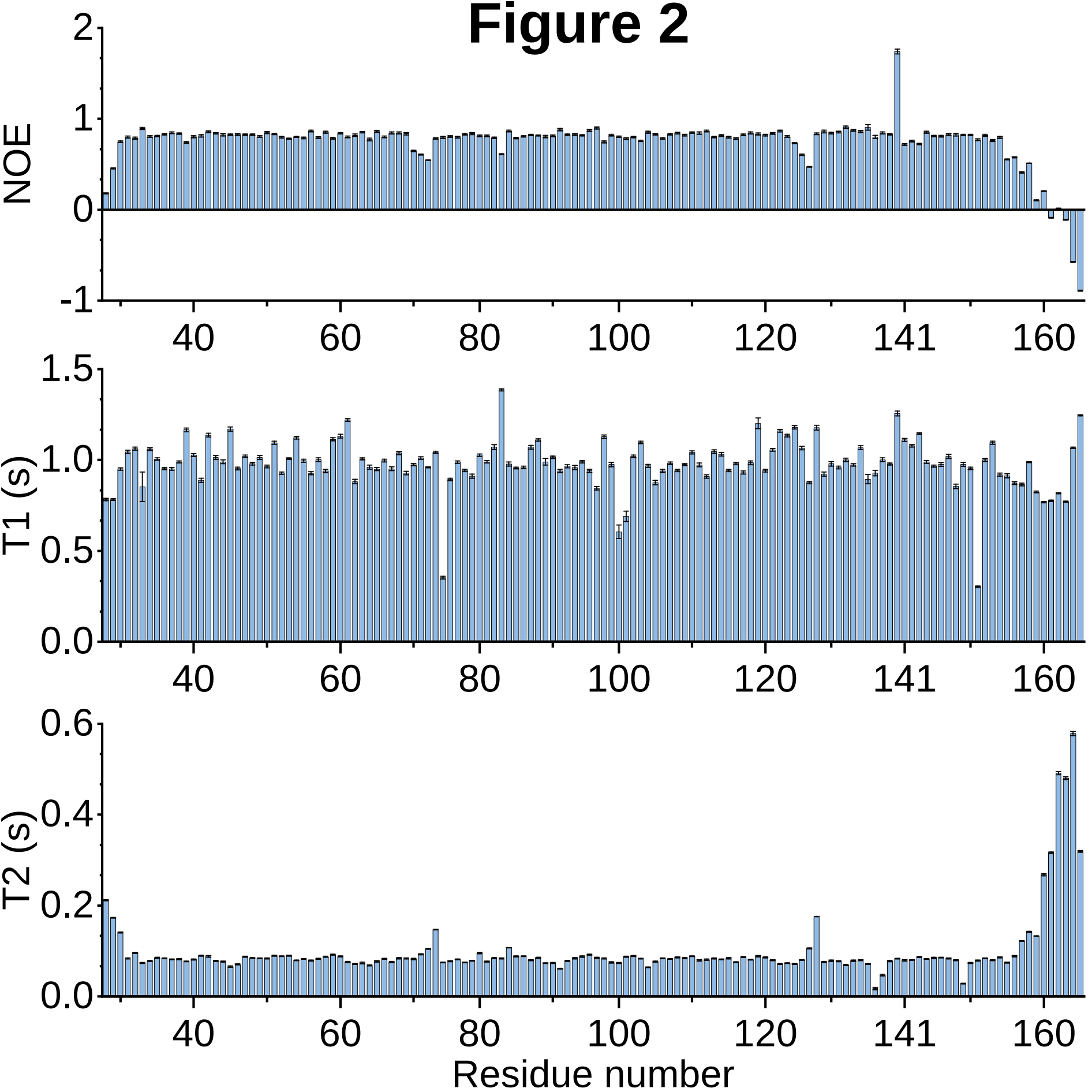

NOE
2
1
0
-1
40
60
80
100
120
141
160
T1 (s)
1.5
1.0
0.5
0.0
40
60
80
100
120
141
160
T2 (s)
0.6
0.4
0.2
0.0
40
60
80
100
120
141
160
Residue number

Figure 3

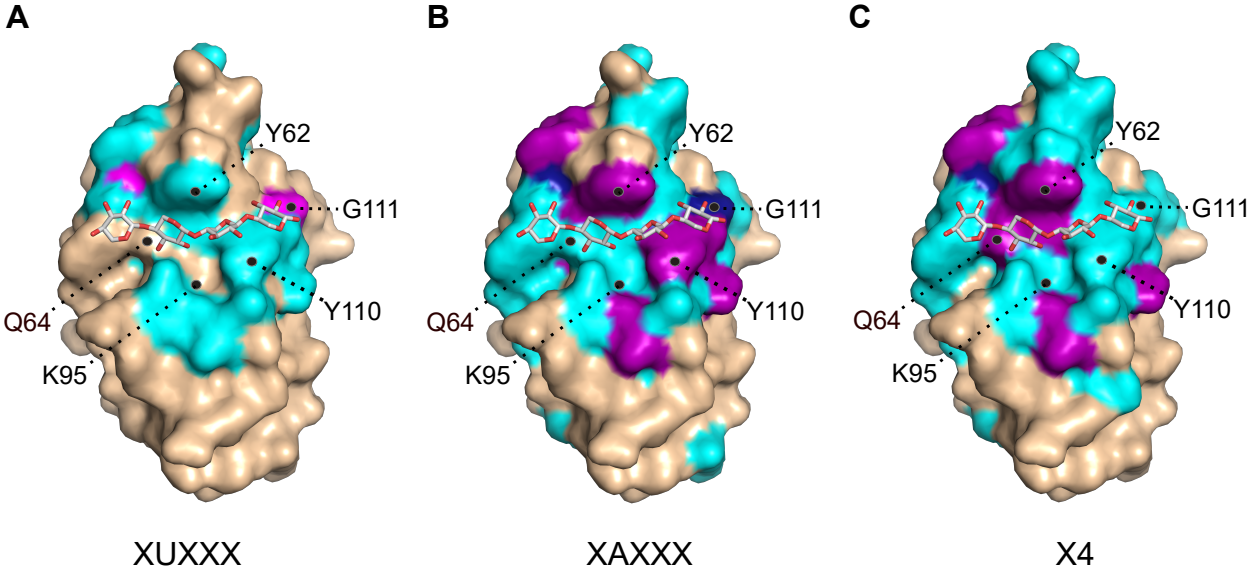

A
B
C
Y62
G111
Y110
Q64
K95
XUXXX
XAXXX
X4
0.025-0.075 ppm
0.075-0.125 ppm
>0.125 ppm

Figure 4

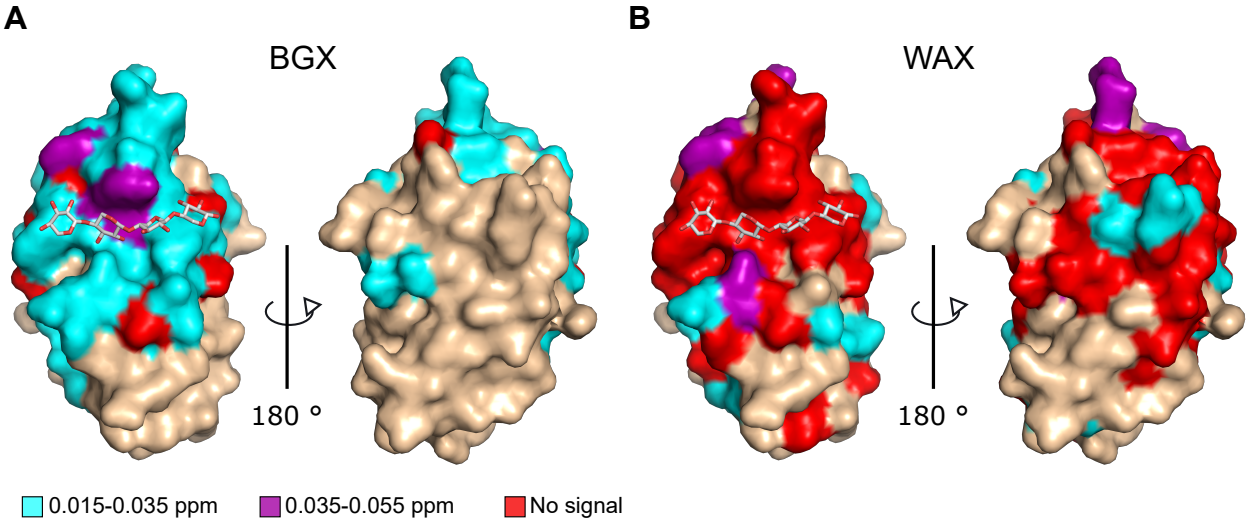
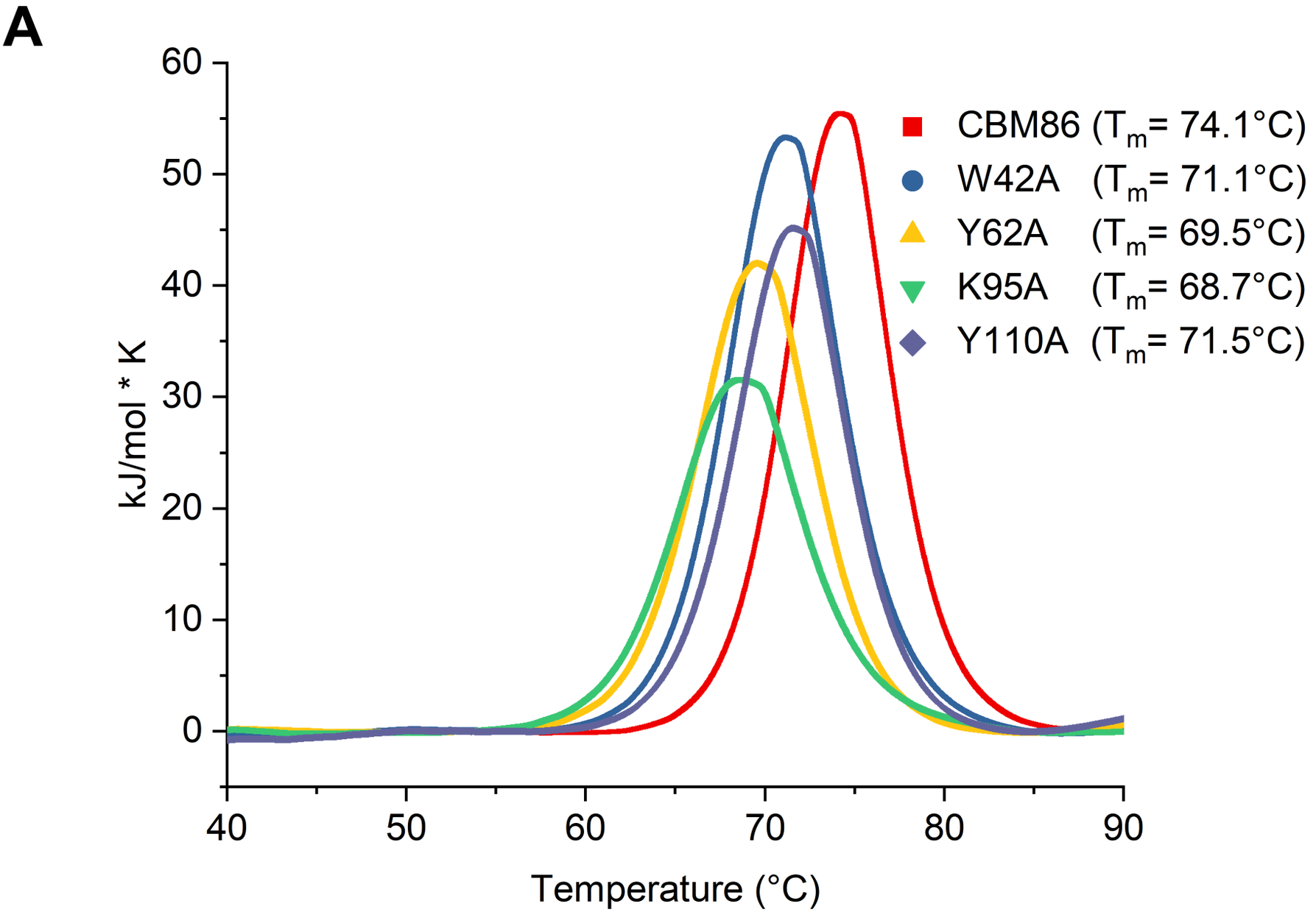


Figure 5


A
60
50
40
30
20
10
0
kJ/mol * K
40
50
60
70
80
90
Temperature (°C)
CBM86 (Tm= 74.1°C)
W42A (Tm= 71.1°C)
Y62A (Tm= 69.5°C)
K95A (Tm= 68.7°C)
Y110A (Tm= 71.5°C)

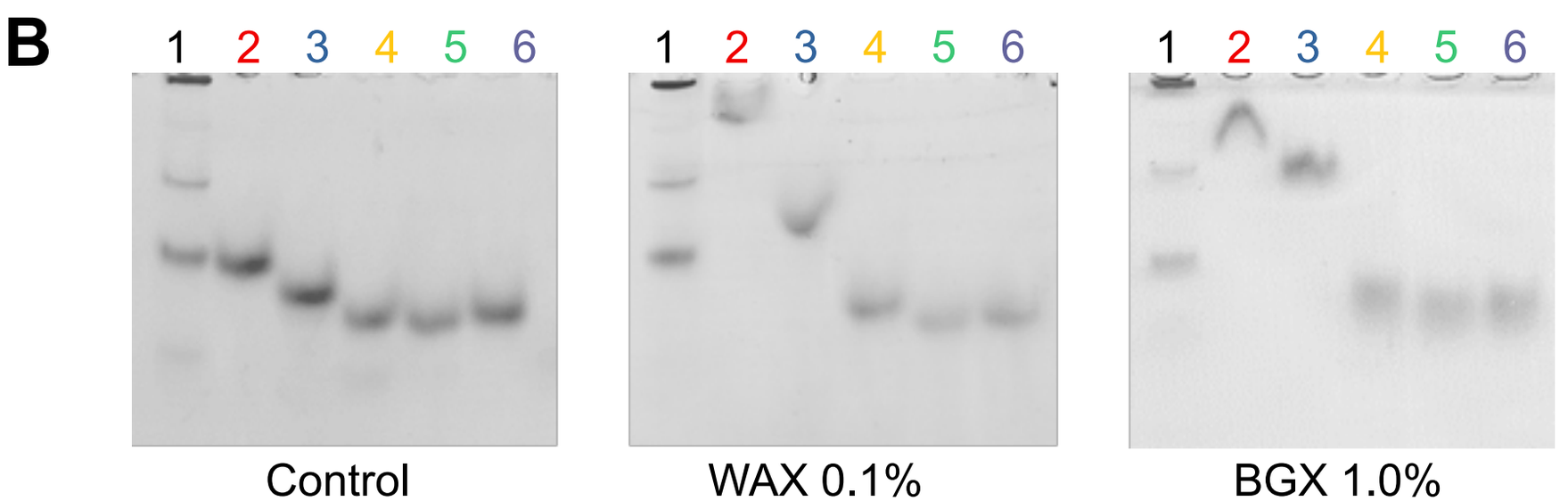
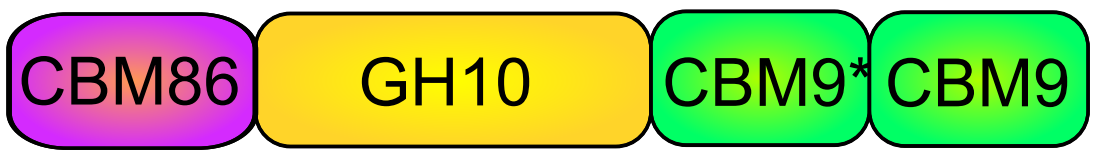

B
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
Control
WAX 0.1%
BGX 1.0%

Figure 6

EEV01588.1
CBL13458.1
CDA56980.1
WP_118209778.1
WP_117920611.1
CUN06914.1
CRL32809.1

CRL34489.1
WP_081671307.1
SCX91715.1
WP_090036222.1
WP_081646650.1
WP_044913615.1
WP_026526370.1
SDB09990.1
WP_026491692.1



AFU34339.1
ABX41884.1
WP_033165005.1

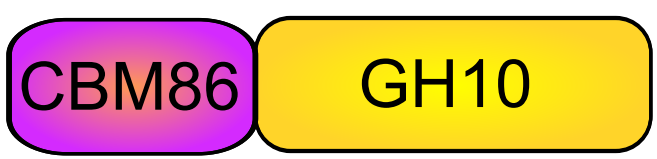


Figure 7