SI Appendix

Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide degrading enzymes

Meiling Zhang[1,2,3], Jonathan R. Chekan[4], Dylan Dodd[1,2,5], Pei-Ying Hong[3], Lauren Radlinski[1,2,5], Vanessa Revindran[1,2], Satish K. Nair[2,4*], Roderick I. Mackie[1,2,3*], and Isaac Cann[1,2,3,5*]

Energy Biosciences Institute[1], Institute for Genomic Biology[2], Department of Animal Sciences[3], Department of Biochemistry[4], Department of Microbiology[5], University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

SI Materials and Methods

Circular dichroism (CD) spectroscopy. The circular dichroism spectra for the wild type and mutant proteins were recorded to evaluate their secondary structures. The CD analyses were carried out using a J-815 CD spectropolarimeter (Jasco, Japan) equipped with a constant temperature cell holder. The proteins were equilibrated in a 1 mm path-length quartz cuvette at 25 °C for 5 min before scanning. The initial wavelength of the measurement was 260 nm and the final wavelength was 190 nm with a wavelength step of 0.1 nm. Secondary structure analysis was performed on the DichroWeb (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) using an algorithm of CDSSTR with reference set 4.

SI Results

Circular dichroism spectra of Bixyn10A, CBM1, CBM2 and their mutants. The CD spectra of BiXyn10A, CBM1, CBM2 and their mutants were assessed to determine whether mutations introduced into their wild-type form led to gross changes in secondary structural elements. For the mutations that impacted the binding activity of CBM1 (i.e., W176A and W249A), neither yielded α-helix and β-turn compositions which were different compared to the wild-type protein (Table S8). In the case of CBM2, W363A abolished binding without impacting the secondary structural elements compared to the wild-type protein (Table S9). Importantly, when W176A, W249A and W363A were introduced into BiXyn10A as individual mutations or to create a triple mutant, the secondary structural elements of the mutants were not statistically different from the wild-type protein. A similar observation (no statistical significant changes in secondary structural elements) was made for BiXyn10A-E634A, the mutant made to abolish catalytic activity (Table S10).

SI Tables

Table S1. Anaerobic medium for culturing *Bacteroides intestinalis* DSM 17393 and *Bacteroides ovatus* ATCC 8483

Ingredients	Concentration in Media (mg/L)
Xylose[a]	5000
Wheat arabinoxylan (medium viscosity, 22 cSt)[a]	5000
Clarified rumen fluid	200 mL
Yeast extract	500
Tryptone	1000
Resazurin (0.1% w/v)	1.0 mL
Cysteine-HCl (2.5%, w/v)	10 mL
Hemin (0.1% w/v)	1.0 mL
Sodium sulfide (2.5%, w/v)	5.0 mL
Sodium carbonate (8%, w/v)	30 mL
Mineral solution 1 (mL/L)	75 mL
KH_2PO_4 (6 g/L)	
Mineral solution 2 (mL/L)	75 mL
KH_2PO_4 (6 g/L)	
$(NH_4)_2SO_4$ (12 g/L)	
NaCl (12 g/L)	
$MgSO_4 \cdot 7H_2O$ (1.2 g/L)	
$CaCl_2 \cdot 2H_2O$ (1.2 g/L)	
Trace element solution (mL/L) [b]	10 mL
Nitrilotriacetic acid (1.5 g/L)	
$MgSO_4 \cdot 7H_2O$ (3 g/L)	
$MnSO_4 \cdot 2H_2O$ (0.5 g/L)	
NaCl (1 g/L)	
$FeSO_4 \cdot 7H_2O$ (0.1 g/L)	
$CoCl_2$ (0.1 g/L)	
$CaCl_2 \cdot 2H_2O$ (0.1 g/L)	
$ZnSO_4$ (0.1 g/L)	
$CuSO_4 \cdot 5H_2O$ (0.01 g/L)	
$AlK(SO_4)_2$ (0.01 g/L)	
H_3BO_3 (0.01 g/L)	
$Na_2MoO_4 \cdot 2H_2O$ (0.01 g/L)	
Vitamin solution (mL/L)	5.0 mL
Thiamine-HCl (200 mg/L)	
Riboflavin (200 mg/L)	
Pyridoxine-HCl (200 mg/L)	
Para-aminobenzoic acid (10 mg/L)	
Biotin (2.5 mg/L)	
Folic acid (2.5 mg/L)	
Calcium pantothenate (200 mg/L)	
Nicotinamide (200 mg/L)	
Vitamin B_{12} (1 mg/L)	
VFA solution (mL/L)	10 mL
Acetic acid (16.8% v/v)	
Propionic acid (6.68% v/v)	
n-butyric acid (3.96% v/v)	
Isobutyric acid (0.8% v/v)	
n-valeric acid (0.92% v/v)	
Isovaleric acid (0.92% v/v)	
D,L-2-methylbutyric acid (0.8% v/v)	

[a] Either xylose or soluble wheat arabinoxylan were used as the sole carbohydrate source.

[b] Trace element solution was prepared as described by Balch *et al.* (1). Nitrilotriacetic acid was first dissolved by adjusting the pH to 6.5 with KOH. Next, the remaining trace elements were added, the pH was adjusted to 7.0 with KOH, and then the volume was brought to 1 L with milliQ H_2O.

1. Balch WE, Fox GE, Magrum LJ, Woese CR, & Wolfe RS (1979) Methanogens: reevaluation of a unique biological group. *Microbiol Rev* 43(2):260-296.

Table S2. RNAseq mapping results

RNAseq Sample ID	Total reads (avg. length)	Reads after Trimming[a] (avg. length)	Uniquely mapped reads[b] (%)	Non-specifically mapped reads (%)	Unmapped reads (%)
BACINT_WAX1	10,101,023 (100 nt)	10,100,976 (94.7 nt)	8,597,177 (85.1%)	8,786 (0.09%)	1,495,013 (14.8%)
BACINT_WAX2	22,325,130 (100 nt)	22,325,047 (94.8 nt)	18,839,356 (84.4%)	20,881 (0.09%)	3,464,810 (15.5%)
BACINT_Xylose1	17,076,556 (100 nt)	17,076,484 (94.7 nt)	14,281,134 (83.6%)	14,737 (0.09%)	2,780,613 (16.3%)
BACINT_Xylose2	11,009,018 (100 nt)	11,008,962 (94.9 nt)	9,142,140 (83.0%)	9,677 (0.09%)	1,857,145 (16.9%)
BACOVA_WAX1	21,824,448 (100 nt)	21,824,352 (95.1 nt)	16,738,339 (76.7%)	125,379 (0.6%)	4,960,634 (22.7%)
BACOVA_WAX2	27,234,520 (100 nt)	27,234,391 (95.3 nt)	20,244,307 (74.3%)	105,564 (0.4%)	6,884,520 (25.3%)
BACOVA_Xylose1	11,265,889 (100 nt)	11,265,845 (95.1 nt)	9,240,294 (82.0%)	58,302 (0.5%)	2,025,551 (18.0%)
BACOVA_Xylose2	15,992,242 (100 nt)	15,992,175 (95.0 nt)	12,609,816 (78.8%)	60,094 (0.4%)	3,322,265 (20.8%)

[a] Reads were trimmed using CLC Genomics Workbench v5.0 with a quality score limit of 0.05 and maximum number of ambiguities of 2.
[b] Reads were mapped to the *B. intestinalis* DSM 17393 or *B. ovatus* ATCC 8483 genome using CLC Genomics Workbench v5.0 with a minimum length fraction of 0.9, a minimum similarity fraction of 0.8, and maximum number of hits for a read of 10.

Table S3. Primers used in this study.

Gene name	Primer name	Orientation	Sequence (5' → 3')
Cloning[a]			
BACINT_04215	P1	Forward	GACGACGACAAGATGGTGGACGACAAACCGTTGGCATTCGAAG
BACINT_04215	P2	Forward	GACGACGACAAGATGAATAACAAGTTTTTGAATTCGTTGATT
BACINT_04215	P3	Reverse	GAGGAGAAGCCCGG*TTA*AGCCTCAGAATGAGTAATCTTAACATA
BACINT_04215	P4	Forward	GACGACGACAAGATGGCTCCGGTTATGGAAATCTTTACCGAC
BACINT_04215	P5	Reverse	GAGGAGAAGCCCGG*TTA*GAGAGGAATAGAATTGCCCTTAGTTAC
BACINT_04215	P6	Reverse	GAGGAGAAGCCCGGTTATTTTCCAGCCAATCCATCAGC
BACINT_02810	P1	Forward	GACGACGACAAGATGGCTTCGGGAGGCGCACAATATTTTC
BACINT_02810	P2	Reverse	GAGGAGAAGCCCGG*TTA*GTAAGCATGCCACTC
BACOVA_00247	P1	Forward	GACGACGACAAGATGAATAATGGGACATTAGTG
BACOVA_00247	P2	Reverse	GAGGAGAAGCCCGG*TTA*CTTAGTTGCAGCAAC
BACINT_04197	P1	Forward	GACGACGACAAGATGCAGCAGAAACAAATGTAC
BACINT_04197	P2	Reverse	GAGGAGAAGCCCGG*TTA*GTCGGCAAACGAAG
PBR_0377	P1	Forward	GACGACGACAAGATGTCTCAGCAGCAAGTAGC
PBR_0377	P2	Reverse	GAGGAGAAGCCCGG*TTA*CTTTACTTCTTCGGTC
BACOVA_04390	P1	Forward	GACGACGACAAGATGACTAATCAGAATGCGAG
BACOVA_04390	P2	Reverse	GAGGAGAAGCCCGG*TTA*AAAATCTCCGTTAGAAATC
BACEGG_01298	P1	Forward	GACGACGACAAGATGGTAGGCAAACATACTGGTG
BACEGG_01298	P2	Reverse	GAGGAGAAGCCCGG*TTA*GCAGGCTTCCATC
PRU_2739	P1	Forward	GACGACGACAAGATGAGCTATGACGGTACTCTTG
PRU_2739	P2	Reverse	GAGGAGAAGCCCGG*TTA*CTCTGTCAGAGGCAG
Mutagenesis[b]			
BACINT_04215	Y173A	Forward	GAAATAGACTATTCTACTGCGGGCGCTAGTTTTTGGAACGAAGTAAATGAA
BACINT_04215	F175A	Forward	GACTATTCTACTGCGGGCTATAGTGCTTGGAACGAAGTAAATGAAGAAG
BACINT_04215	W176A	Forward	CTATTCTACTGCGGGCTATAGTTTTGCGAACGAAGTAAATGAAGAAG
BACINT_04215	W249A	Forward	GAATTTTGGTGTAGGACCTGCGAGTGGTAGAGCAGAAGGT
BACINT_04215	W263A	Forward	AAGGTTCTTTTTCTTTCAACACAGAAGCGAAAGAGTATGAGTTTTCATTCAAAG
BACINT_04215	W363A	Forward	CAGCAGGTGGTGGTGTAAATAGTGCGGATACTCAGTTCTT
BACINT_04215	W421A	Forward	GGATGTGCTGTTGATTTTACTACAGAAGCGCAACATTTTGAAAAAACTATAAC
BACINT_04215	W445A	Forward	GAAAACATGCAGACTTTTGCAGCGAATCTTGATGTAGGTGTGCC
BACINT_04215	E654A	Forward	TGGTAAACTGGTAAAGATATCAGCACTTGATATGGGTTATATAGATG

[a] Underlined sequences indicate the incorporated T4 exonuclease digestion sites. Italicized nucleotides indicate translational stop codons inserted at the end of the construct. The primer names correspond to those mentioned in the text and shown in Figure S1.
[b] Primers were designed using the Agilent Technologies QuikChange primer design tool. Underlined sequences indicate the substituted codon.

Table S4. Genes up-regulated greater than five-fold during growth of *B. intestinalis* DSM 17393 with WAX compared to xylose assessed by RNAseq and listed by magnitude of induction.[a]

ORF	Annotation[b]	Fold change	CAZy Designation
BACINT_04218	SusC-like	214	-
BACINT_04220	SusC-like	182	-
BACINT_04215	endo-xylanase	170	GH10/CBM
BACINT_04216	hypothetical protein	167	-
BACINT_04219	SusD-like	142	-
BACINT_00570	beta-xylosidase	135	GH43
BACINT_00569	arabinofuranosidase/endo-xylanase	132	GH43/GH10
BACINT_04217	SusD-like	98.1	-
BACINT_04201	Na/sodium co-transporter	97.0	-
BACINT_04202	endo-xylanase/arabinofuranosidase	90.0	GH10/GH43
BACINT_04199	beta-galactosidase	87.5	GH35
BACINT_04200	esterase	85.1	GH2/CE6
BACINT_04198	hypothetical protein	78.4	-
BACINT_04203	beta-xylosidase	71.0	GH43
BACINT_04213	endo-xylanase	70.1	GH5
BACINT_04204	oxidoreductase	65.6	-
BACINT_04205	alpha-glucuronidase	62.5	GH67
BACINT_04214	hypothetical protein	60.7	-
BACINT_04196	hypothetical protein	41.9	-
BACINT_01180	hypothetical protein	41.7	-
BACINT_01179	hypothetical protein	41.4	-
BACINT_04197	endo-xylanase	39.2	GH10/CBM
BACINT_04195	SusD-like	36.4	-
BACINT_02768	beta-xylosidase	34.2	GH43
BACINT_04194	SusC-like	32.1	-
BACINT_00076	xylanase-esterase	28.8	GH10/CE1
BACINT_02774	SusC-like	28.2	-
BACINT_02773	SusD-like	26.3	-
BACINT_02772	SusC-like	25.8	-
BACINT_02776	hypothetical protein	25.2	-
BACINT_01178	hypothetical protein	23.7	-
BACINT_02767	arabinofuranosidase	22.9	CBM4/GH51
BACINT_02775	SusD-like	22.5	-
BACINT_02771	hypothetical protein	21.8	-
BACINT_00908	SusC-like	21.4	-
BACINT_00927	exo-xylanase	19.7	GH8
BACINT_02777	beta-xylosidase	19.7	GH43
BACINT_00926	beta-xylosidase	19.2	GH3
BACINT_02766	beta-xylosidase	19.1	GH43
BACINT_00907	SusD-like	18.7	-
BACINT_01181	hypothetical protein	17.5	-
BACINT_02782	L-arabinose isomerase	17.5	-
BACINT_04210	exo-xylanase	17.2	GH8
BACINT_04212	esterase	16.8	CE1
BACINT_02783	xylulose kinase	15.3	-
BACINT_04706	hypothetical protein	15.1	-
BACINT_01183	hypothetical protein	14.4	-

BACINT_02784	hypothetical protein	14.4	-
BACINT_04211	alpha-rhamnosidase	14.3	CE6/GH95
BACINT_02786	arabinofuranosidase	14.3	GH51
BACINT_02781	L-ribulose-5-phosphate 4-epimerase	14.2	-
BACINT_02785	alpha-glucosidase	13.8	GH97
BACINT_01095	beta-xylosidase	13.5	GH43/CBM6
BACINT_01182	hypothetical protein	13.4	-
BACINT_01094	esterase	13.1	CE1
BACINT_01184	hypothetical protein	12.8	-
BACINT_02779	Na/sodium cotransporter	10.8	-
BACINT_04223	hypothetical protein	10.7	GH115
BACINT_02780	ADP-ribose pyrophosphorylase	10.7	-
BACINT_02778	aldose 1-epimerase	9.92	-
BACINT_00772	hypothetical protein	9.14	-
BACINT_04222	esterase	7.91	GH2/CE6
BACINT_01185	hypothetical protein	7.47	-
BACINT_03137	hypothetical protein	7.16	-
BACINT_01190	hypothetical protein	6.78	-
BACINT_01096	hypothetical protein	6.17	-
BACINT_01186	hypothetical protein	5.89	-

[a] *B. intestinalis* DSM 17393 was cultured in a synthetic medium with either soluble wheat arabinoxylan (WAX) or xylose as the sole carbohydrate source. RNA was then extracted and RNAseq experiments were performed as described in experimental procedures.
[b] Open reading frame assignments and gene annotations were from the Joint Genome Institute's Integrated Microbial Genomes database (http://img.jgi.doe.gov/cgi-bin/w/main.cgi).

Table S5. Genes up-regulated greater than five-fold during growth of *B. ovatus* ATCC 8483 with WAX compared to xylose assessed by RNAseq and listed by magnitude of induction.[a]

ORF	Annotation[b]	Fold change	CAZy Designation
BACOVA_04393	SusC-like	341	-
BACOVA_04390	endo-xylanase	335	GH10/CBM
BACOVA_04389	Esterase	319	GH2/CE6
BACOVA_04384	hypothetical protein	312	-
BACOVA_03427	SusD-like	300	-
BACOVA_03422	alpha-glucosidase	294	GH31
BACOVA_03425	beta-xylosidase	286	GH43
BACOVA_04386	beta-xylosidase	286	GH43
BACOVA_04385	alpha-glucuronidase	256	GH67
BACOVA_04387	endo-xylanase	244	GH10
BACOVA_03426	SusC-like	241	GH35
BACOVA_03423	alpha-glucosidase	234	GH97
BACOVA_03428	SusC-like	217	-
BACOVA_03424	beta-xylosidase	201	GH43/CBM6
BACOVA_03431	hypothetical protein	201	-
BACOVA_04391	hypothetical protein	194	-
BACOVA_03433	endo-beta-galactosidase	171	GH98/CBM35
BACOVA_03429	SusD-like	166	-
BACOVA_03432	endo-xylanase	164	GH30
BACOVA_03421	beta-xylosidase	163	GH43
BACOVA_03417	beta-xylosidase	158	GH43
BACOVA_04392	SusD-like	145	-
BACOVA_04388	Na/sugar cotransporter	112	-
BACOVA_03434	alpha-glucuronidase	105	GH115
BACOVA_03436	beta-xylosidase	94.0	GH43/CBM6
BACOVA_01707	hypothetical protein	71.2	-
BACOVA_03430	hypothetical protein	58.6	-
BACOVA_03435	Esterase	49.2	CE6/CE1
BACOVA_03449	alpha-glucuronidase	43.6	GH115
BACOVA_01708	alpha-arabinofuranosidase	42.1	GH51
BACOVA_03450	Esterase	41.3	GH2/CE6
BACOVA_01709	hypothetical protein	36.6	-
BACOVA_01715	aldose 1-epimerase	29.6	-
BACOVA_01714	Na/sugar cotransporter	21.8	-
BACOVA_01710	xylulose kinase	20.0	-
BACOVA_03419	beta-xylosidase	19.8	GH3
BACOVA_01713	ADP-ribose pyrophosphatase	19.5	-
BACOVA_01711	L-arabinose isomerase	19.2	-
BACOVA_01712	L-ribulose-5-phosphate 4-epimerase	16.9	-
BACOVA_03637	hypothetical protein	15.7	-

BACOVA_04183	hypothetical protein	11.8	-
BACOVA_03438	hypothetical protein	10.2	-
BACOVA_01199	SusD-like	9.63	-
BACOVA_03483	alpha-arabinofuranosidase	9.00	GH51
BACOVA_03440	hypothetical protein	8.17	-
BACOVA_01866	hypothetical protein	8.12	-
BACOVA_03443	SusC-like	8.00	-
BACOVA_03447	hypothetical protein	7.98	-
BACOVA_03437	HTCS regulator	7.83	-
BACOVA_01865	hypothetical protein	7.69	-
BACOVA_03444	SusD-like	7.59	-
BACOVA_03446	hypothetical protein	7.38	-
BACOVA_01864	hypothetical protein	7.06	-
BACOVA_03445	hypothetical protein	7.03	-
BACOVA_03439	hypothetical protein	6.87	-
BACOVA_03484	alpha-glucosidase	6.59	GH97
BACOVA_01868	hypothetical protein	6.41	-
BACOVA_01867	hypothetical protein	6.28	-
BACOVA_01869	hypothetical protein	5.83	-
BACOVA_04621	site-specific recombinase	5.75	-
BACOVA_04960	hypothetical protein	5.62	-
BACOVA_01497	hypothetical protein	5.23	-
BACOVA_05005	pectate lyase	5.02	PL1

[a] *B. ovatus* ATCC 8483 was cultured in a synthetic medium with either soluble wheat arabinoxylan (WAX) or xylose as the sole carbohydrate source. RNA was then extracted and RNAseq experiments were performed as described in experimental procedures.
[b] Open reading frame assignments and gene annotations were from the Joint Genome Institute's Integrated Microbial Genomes database (http://img.jgi.doe.gov/cgi-bin/w/main.cgi).

Table S6. Kinetic properties of BiXyn10A and mutants with OSX[a]

Protein	k_{cat} (s^{-1})	K_M (ml mg^{-1})	k_{cat}/K_M (ml mg^{-1} s^{-1})
WT	54 ± 6	7.4 ± 2	7.4 ± 2
W176A	32 ± 7	11 ± 5	2.9 ± 1
W249A	39 ± 6	11 ± 4	3.4 ± 1
W363A	27 ± 2	8.8 ± 2	3.1 ± 0.7
W176A/W249A/W363A	15 ± 1	6.7 ± 1	2.2 ± 0.4

[a] BiXyn10A wild type and mutants were incubated with OSX and initial rates of reducing sugar release were monitored at different substrate concentrations. The rate of reducing sugar release was plotted against the substrate concentration and non-linear regressions were used to estimate the kinetic parameters. Experiments were performed in triplicate and the data are represented as means ± standard deviations.

Table S7. Kinetic properties of BiXyn10A and mutants with WAX[a]

Protein	k_{cat} (s^{-1})	K_M (ml mg^{-1})	k_{cat}/K_M (ml mg^{-1} s^{-1})
WT	93 ± 6	1.7 ± 0.4	57 ± 10
W176A	70 ± 4	2.7 ± 0.5	26 ± 5
W249A	57 ± 2	1.5 ± 0.2	37 ± 5
W363A	57 ± 3	2.0 ± 0.3	29 ± 5
W176A/W249A/W363A	69 ± 3	2.1 ± 0.3	33 ± 5

[a] BiXyn10A wild type and mutants were incubated with WAX and initial rates of reducing sugar release were monitored at different substrate concentrations. The rate of reducing sugar release was plotted against the substrate concentration and non-linear regressions were used to estimate the kinetic parameters. Experiments were performed in triplicate and the data are represented as means ± standard deviations.

Table S8. CD spectroscopy analysis of CBM1 and its mutants with abolished substrate binding [a]

Protein	α-helices (%)	β-strands (%)	β-turns (%)	unordered (%)
CBM1	1 ± 1	49 ± 1	21 ± 0	26 ± 0
CBM1 W176A	2 ± 1	47 ± 1	22 ± 1	26 ± 1
CBM1 W249A	3 ± 1	46 ± 2	22 ± 1	27 ± 3

[a] CD spectra were recorded in the far-UV range utilizing a J-815 CD spectropolarimeter. The spectra were recorded from 190 to 260 nm at a scan rate of 50 nm/s and a 1-nm wavelength step with five accumulations. Each reaction was performed three times, and the data are represented as means ± standard deviations of the means. The spectra were uploaded onto the DICHROWEB on-line server and analyzed as described in supplemental materials and methods.

Table S9. CD spectroscopy analysis of CBM2 and its mutant with abolished substrate binding[a]

Protein	α-helices (%)	β-strands (%)	β-turns (%)	unordered (%)
CBM2	2 ± 1	38 ± 0	23 ± 6	34 ± 0
CBM2 W363A	2 ± 1	39 ± 1	22 ± 1	35 ± 1

[a] CD spectra were recorded in the far-UV range utilizing a J-815 CD spectropolarimeter. The spectra were recorded from 190 to 260 nm at a scan rate of 50 nm/s and a 1-nm wavelength step with five accumulations. Each reaction was performed three times, and the data are represented as means ± standard deviations of the means. The spectra were uploaded onto the DICHROWEB on-line server and analyzed as described in supplemental materials and methods.

Table S10. CD spectroscopy analysis of BiXyn10A and its mutants [a]

Protein	α-helices (%)	β-strands (%)	β-turns (%)	unordered (%)
WT	10 ± 0	37 ± 1	22 ± 0	31 ± 1
E634A	10 ± 2	37 ± 2	22 ± 1	30 ± 1
W176A	10 ± 2	35 ± 2	22 ± 1	30 ± 0
W249A	11 ± 1	35 ± 1	22 ± 1	31 ± 1
W363A	10 ± 2	37 ± 2	22 ± 1	31 ± 1
W176A/W249A/W363A	10 ± 2	37 ± 2	22 ± 0	30 ± 1

[a] CD spectra were recorded in the far-UV range utilizing a J-815 CD spectropolarimeter. The spectra were recorded from 190 to 260 nm at a scan rate of 50 nm/s and a 1-nm wavelength step with five accumulations. Each reaction was performed three times, and the data are represented as means ± standard deviations of the means. The spectra were uploaded onto the DICHROWEB on-line server and analyzed as described in supplemental materials and methods.

Table S11. Crystallographic Statistics for BiXyn10A CBM1 and PbXyn10C CBM

	BiXyn10A CBM1	BiXyn10A CBM1-X6	PBXyn10C CBM
Cell Dimensions			
Space group	$P2_12_12$	$P2_12_12_1$	$P3_12_1$
a, b, c (Å)	53.7, 90.7, 30.9	36.9, 57.0, 85.3	68.7, 68.7, 111.5
α, β, γ, (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution	22.79-1.80 (1.83-1.80)	57.049-1.144 (1.148-1.144)	15.83-1.68 (1.71-1.68)
Total Reflections	112,403	513,642	161,095
Unique reflections	14,401 (695)	65,859 (4,183)	35,308 (1,720)
R_{sym} (%)	9.0 (59.5)	6.4 (74.7)	5.0 (74.4)
I/σ	21.5 (2.4)	16.2 (2.3)	26.5 (2.0)
Completeness (%)	100.0 (100.0)	100.0 (99.8)	99.9 (99.8)
Redundancy	7.8 (6.5)	7.8 (6.4)	4.6 (4.6)
Phasing			
Figure of Merit	0.414		0.260
Refinement			
Resolution (Å)	22.79-1.801	47.44-1.14	15.83-1.68
No. reflections	14,363	62,558	33,795
R_{work}/R_{free}	0.159/0.186	0.160/0.178	0.186/0.208
No. of atoms			
Protein	1125	1127	1268
Water	190	311	418
B-factors			
Protein	16.52	13.64	14.51
Oligosaccharide		21.74	
Water	29.14	33.56	30.31
Ramachandran analysis			
Favored	98.58	97.9	98.09
Allowed	1.42	2.1	1.91
Outliers	0.0	0.0	0.0
RMSD			
Bond lengths (Å)	0.006	0.0313	0.006
Bond angels (°)	1.024	2.362	1.168
PDB Code	4MGS	4QPW	4MGQ

SI Figures

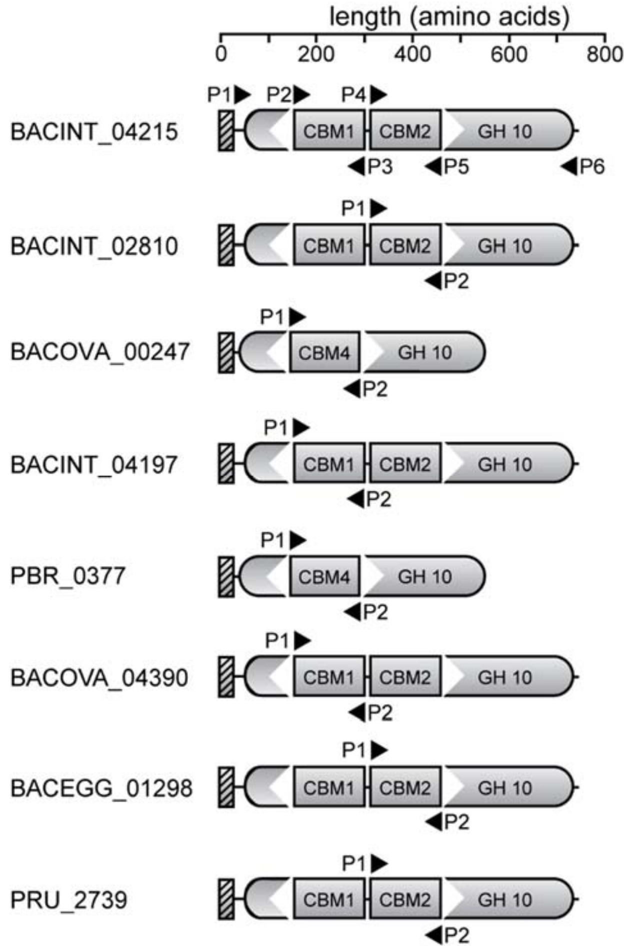


Figure S1. Primers used to amplify designated truncational mutants in this study. The nucleotide sequences for the individual primers are provided in Table S3. The BACINT_04215 truncational mutants, the primers used, and the amino acid regions constructed are as follows: WT (P1-P6, 33-746); TM1 (P2-P5, 138-477); TM2 (P2-P3, 138-301); TM3 (P4-P5, 302-477). The amino acid regions amplified for the remaining genes are as follows: BACINT_02810 CBM 2 (305-484); BACOVA_00247 CBM, 123-358; BACINT_04197 CBM1, 140-329; PBR_0377 CBM, 133-306; BACOVA_04390 CBM1, 138-330; BACEGG_01298 CBM2, 265-471; PRU_2739 CBM2, 278-442.

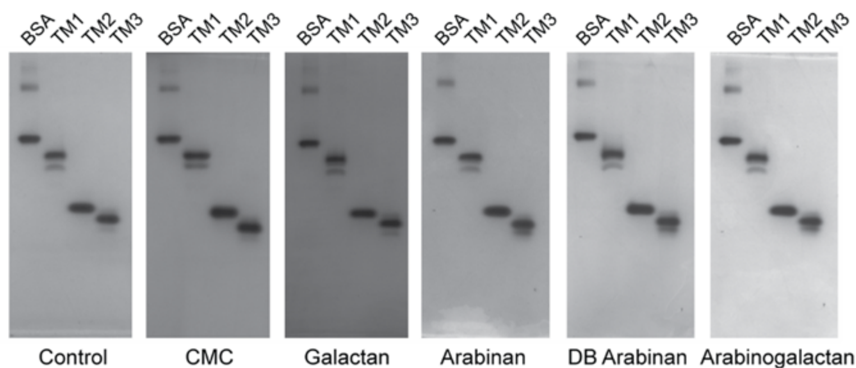


Figure S2. Affinity Gel Electrophoresis. Native polyacrylamide gels were prepared with or without 0.1% w/v polysaccharide, and 2 μg of each truncational mutant or 1 μg of bovine serum albumin (BSA) was loaded and electrophoresed at 100 V for 6 h at 4 °C. The proteins were then visualized by staining with Coomassie Brilliant Blue G-250. Polysaccharides used were carboxymethyl cellulose (CMC), potato galactan, sugar beet arabinan, debranched (DB) arabinan, and larch arabinogalactan.

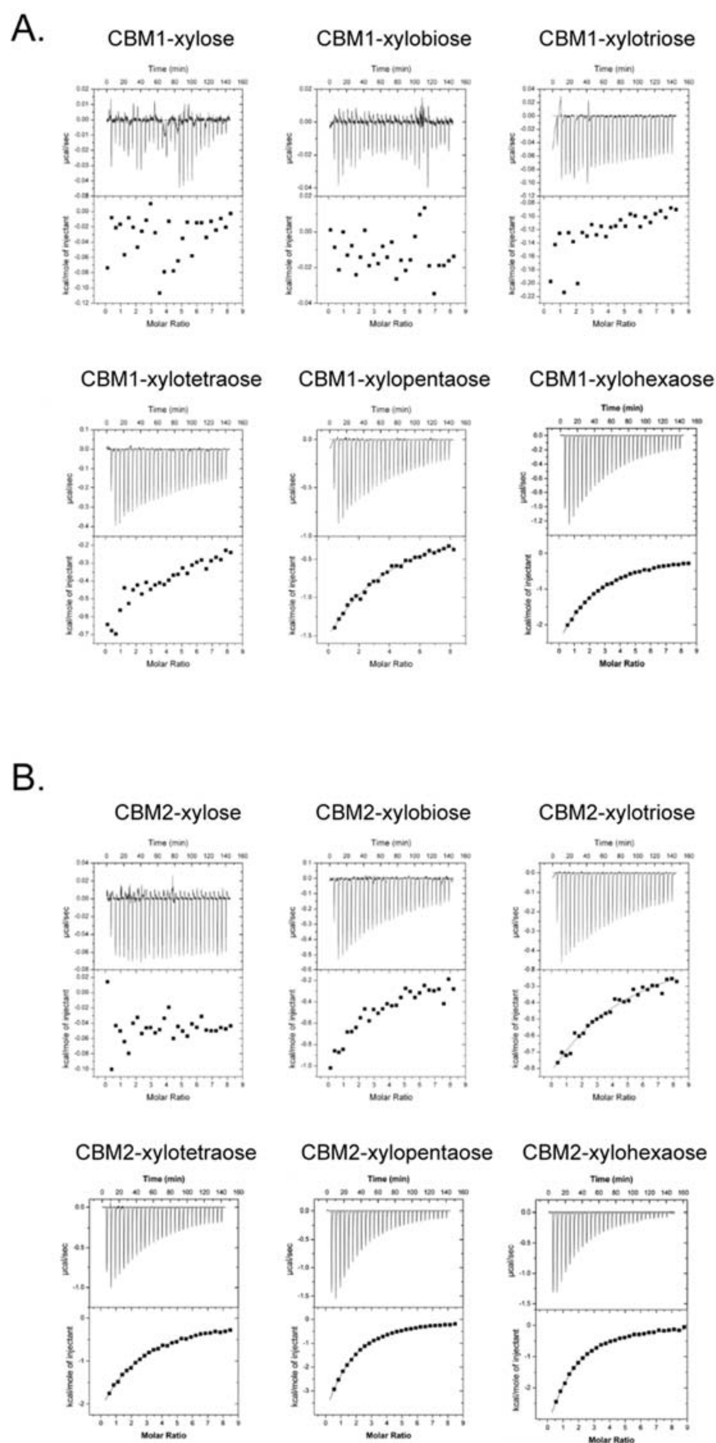


Figure S3. Representative ITC data for binding of (A) CBM1 and (B) CBM2 with xylo-oligosaccharides as the ligand. All experiments were carried out in 50 mM sodium phosphate buffer, pH 7.0 at 25 °C.

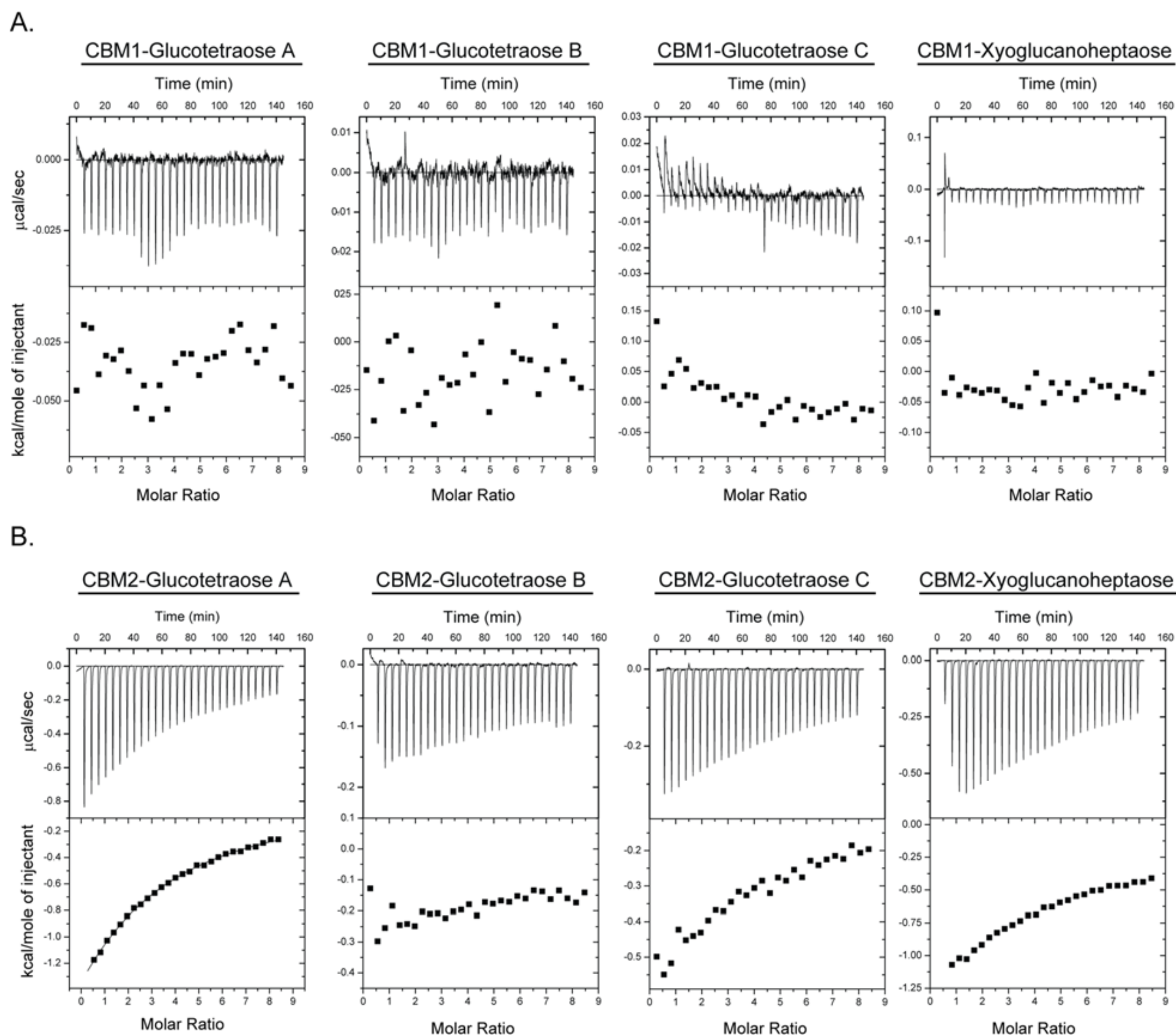


Figure S4. Representative ITC data for binding of (A) CBM1 and (B) CBM2 with lichenan and xyloglucan derived oligosaccharides as the ligand. Linkages within the ligands are as follows: Glucotetraose A, Glcβ1-3Glcβ1-4Glcβ1-4Glc; Glucotetraose B, Glcβ1-4Glcβ1-4Glcβ1-3Glc, Glucotetraose C, Glcβ1-4Glcβ1-3Glcβ1-4Glc; Xyloglucanoheptaose, X_3Glc_4. All experiments were carried out in 50 mM sodium phosphate buffer, pH 7.0 at 25 °C.

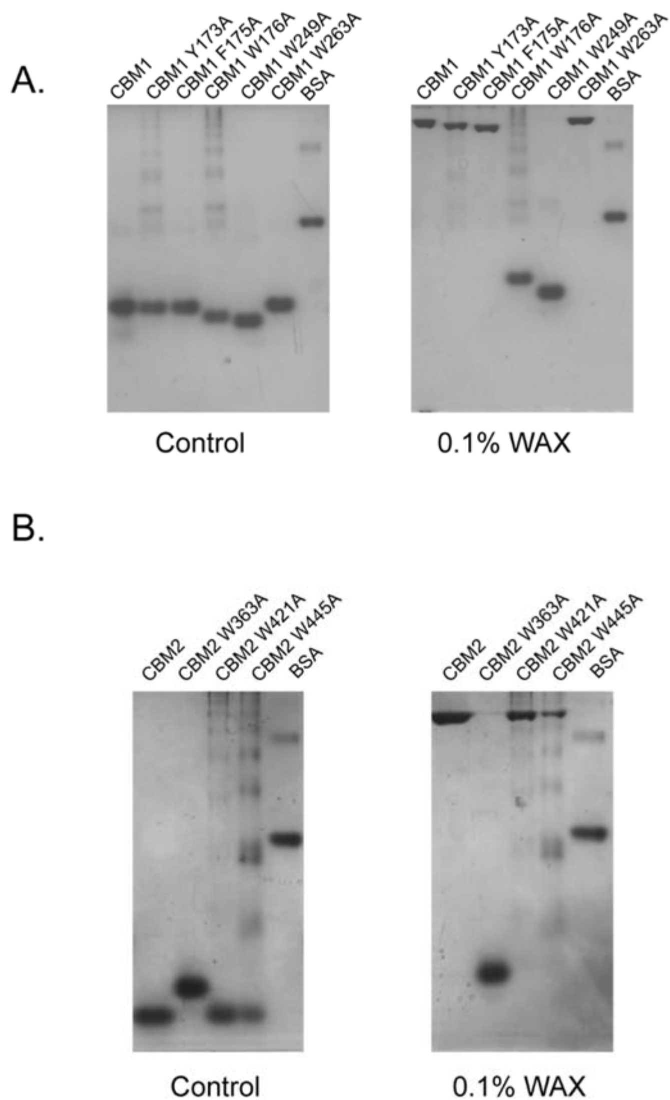
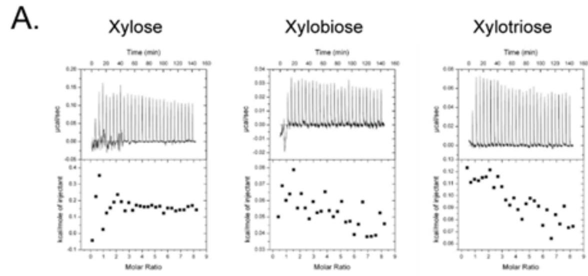
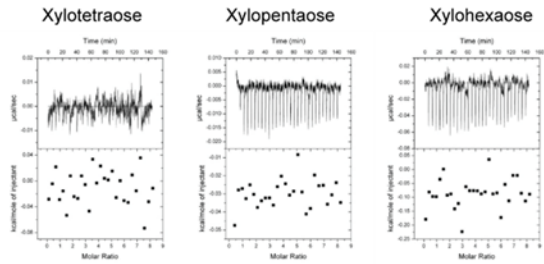
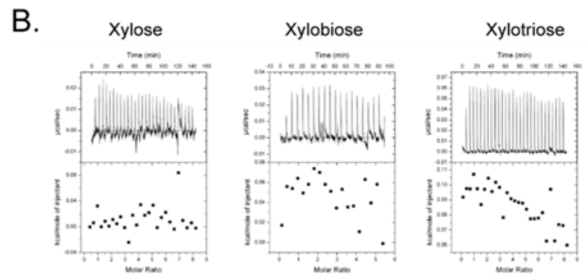
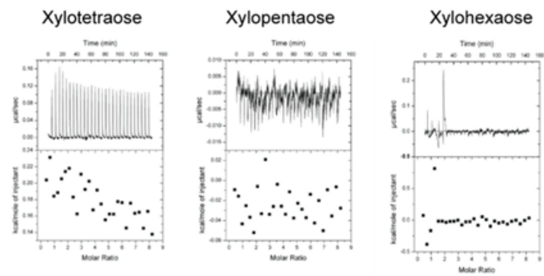


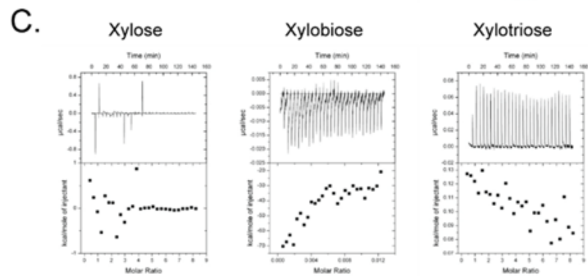
Figure S5. Affinity gel electrophoresis. Native polyacrylamide gels were prepared with or without 0.1% w/v wheat arabinoxylan (WAX). Two µg of each protein or 1 µg of bovine serum albumin (BSA) was loaded and electrophoresed at 100 V for 6 h at 4 °C. (A) CBM1 and its mutants. (B) CBM2 and its mutants.


A.
Xylose
Xylobiose
Xylotriose


Xylotetraose
Xylopentaose
Xylohexaose


B.
Xylose
Xylobiose
Xylotriose


Xylotetraose
Xylopentaose
Xylohexaose


C.
Xylose
Xylobiose
Xylotriose

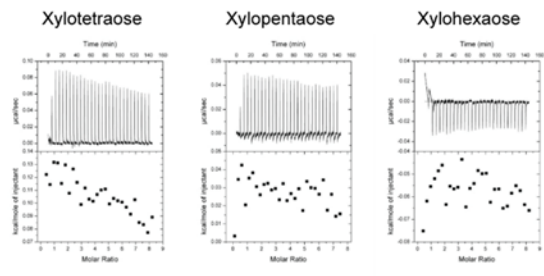
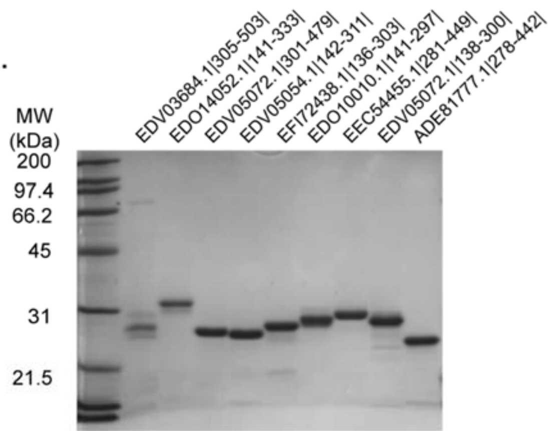
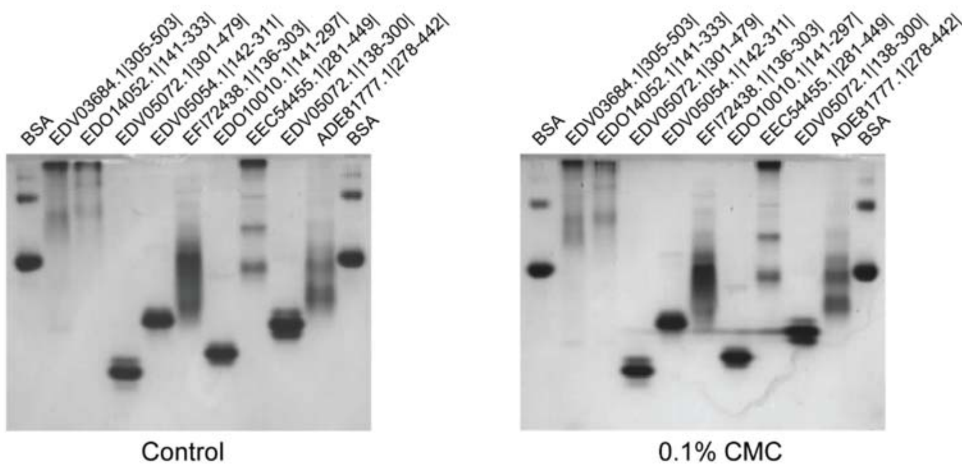

Xylotetraose
Xylopentaose
Xylohexaose

Figure S6. Representative ITC data for binding of (A) CBM1 W176A, (B) CBM1 W249A and (C) CBM2 W363A, with xylo-oligosaccharides as the ligand. All experiments were carried out in 50 mM sodium phosphate buffer, pH 7.0 at 25 °C.


A.
MW (kDa)
200
97.4
66.2
45
31
21.5
EDV03684.1|305-503|
EDO14052.1|141-333|
EDV05072.1|301-479|
EDV05054.1|142-311|
EFI72438.1|136-303|
EDO10010.1|141-297|
EEC54455.1|281-449|
EDV05072.1|138-300|
ADE81777.1|278-442|


B.
BSA
EDV03684.1|305-503|
EDO14052.1|141-333|
EDV05072.1|301-479|
EDV05054.1|142-311|
EFI72438.1|136-303|
EDO10010.1|141-297|
EEC54455.1|281-449|
EDV05072.1|138-300|
ADE81777.1|278-442|
BSA
Control
0.1% CMC

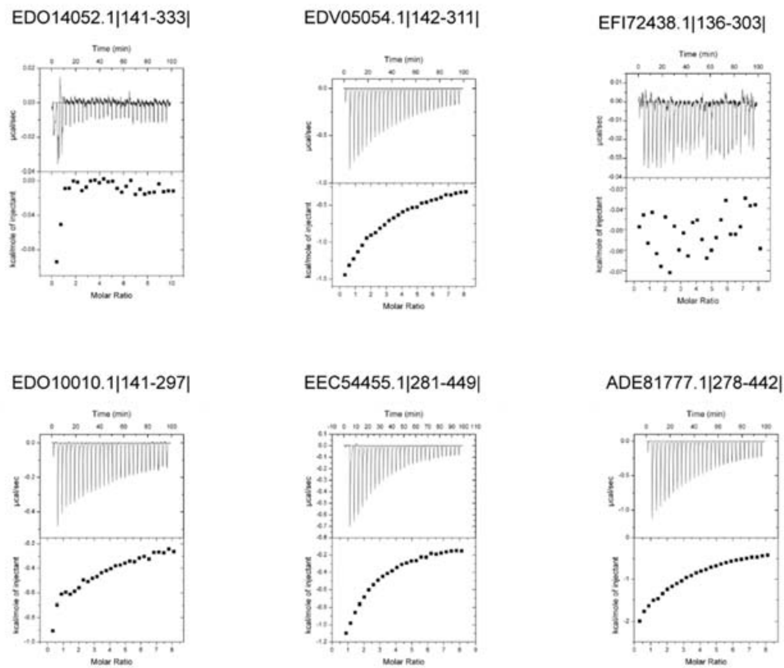

C.
EDO14052.1|141-333|
EDV05054.1|142-311|
EFI72438.1|136-303|
EDO10010.1|141-297|
EEC54455.1|281-449|
ADE81777.1|278-442|
Time (min)
µcal/sec
kcal/mole of injectant
Molar Ratio

Figure S7. The detection of binding activity of selected proteins with CMC and xylo-oligosaccharides. (A) SDS-PAGE showing the purified proteins. The proteins were purified as described in materials and methods. Two µg of protein was loaded on the 12% w/v polyacrylamide gel and stained with Coomassie brilliant blue G-250. (B) Affinity gel electrophoresis. Native polyacrylamide gels were prepared with or without 0.1% w/v carboxymethyl cellulose (CMC). Two µg of each protein or 1 µg of bovine serum albumin (BSA) was loaded and electrophoresed at 100 V for 6 h at 4 °C. (C) ITC data for binding activity of selected proteins with xylopentaose as the ligand. All experiments were carried out in 50 mM sodium phosphate buffer, pH 7.0 at 25 °C.

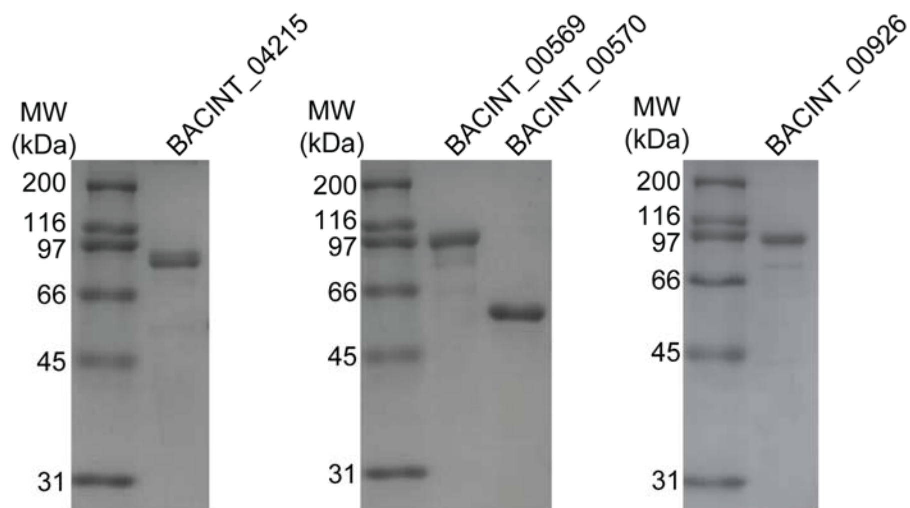


Figure S8. SDS-PAGE analysis of the purified recombinant proteins BACINT_04215 (BiXyn10A), BACINT_00569, BACINT_00570, and BACINT_00926 (BiXyl3A). Two µg of each of the proteins was resolved on a 12% w/v SDS-polyacrylamide gel and stained with Coomassie brilliant blue G-250.

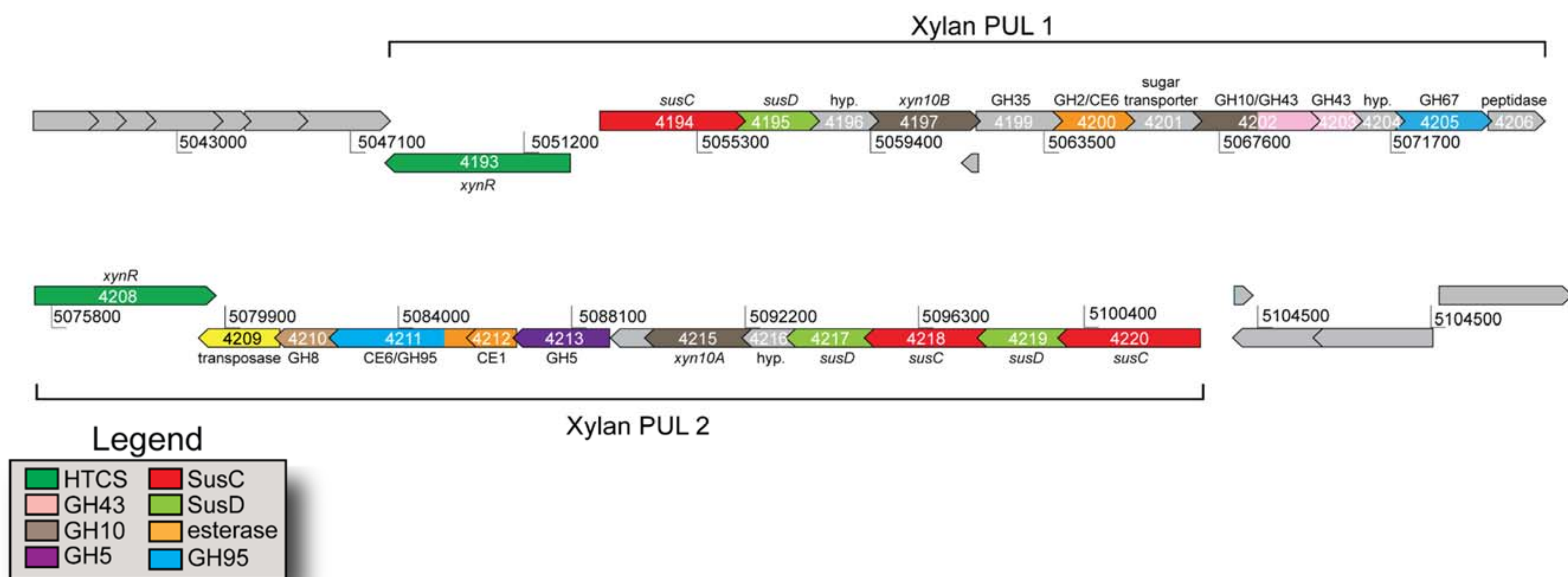


Figure S9. Genomic organization of the major xylan utilization locus in *Bacteroides intestinalis* DSM 17393. ORF numbers indicated within each of the genes correspond to those listed in the GenBank database. Numbering indicates the nucleotide position along the genome. Annotations were made using a hybrid approach with dbCAN, RAST, and Pfam as described in the text.

```
BACINT_4202  MKLKYLALSVCAATFMSCNSDKPVAADPTLKDILGNKFLVGVAINSEQAVGRDTSGVDVVRRHFNSIVAENCMKSEVIHPEEDRYDFSLADEFVKFGEDN
CCY87565.1   MKLKYLALSVCAATFMSCNSDKPVAADPTLKDILGNKFLVGVAINSEQAVGRDTSGVDVVRRHFNSIVAENCMKSEVIHPEEDRYDFSLADEFVKFGEDN

BACINT_4202  GMFIIGHCLVWHSQLSPWFCVDADGKNVSPEVLKERLRSHIHTIVGRYKGRIKGWDVVNEAIEGDGSYRKSKFYEILGEEYIPLAFQYAHEADPEAELYY
CCY87565.1   GMFIIGHCLVWHSQLSPWFCVDADGKNVSPEVLKERLRSHIHTIVGRYKGRIKGWDVVNEAIEGDGSYRRSKFYEILGEEYIPLAFQYAHEADPEAELYY

BACINT_4202  NDYGMHEPGRRDAVVRMVNSLKEKGLRIDAVGMQGHMGLDYPSIGEYETSIFAFASTGAKVMITEWDMSALPTVNRGANIADKVAFEKALNPYPEALPDS
CCY87565.1   NDYGMHEPGRRDAVVRMVNSLKEKGLRIDAVGMQGHMGLDYPSIGEYETSLLAFASTGAKVMITEWDMSALPTVNRGANIADKVAFEKALNPYPEALPDS

BACINT_4202  VSNLWNARMKSFMELFIKHSDVITRVTAWGVSDGDSWKNDWPVPGRREYPLLFDRNYQPKPFLKEILEPRVAKFYEFSYSIDNSQLTIDNSMQRDSAANC
CCY87565.1   VSNLWNARMKSFMELFIKHSDVITRVTAWGVSDGDSWKNDWPVPGRREYPLLFDRNYQPKPFLKEILEPKKAVFDEFTYTVAP...........KDTDKA

BACINT_4202  QLSTVNCQLQNPILPGCYPDPSICRVGNDYYMVNSSFAFYPGVPIWHSTNLTNWEQLGYVLNRPSQLPMYDGLRISGGIYAPDIKYNPHNGLFYLITTAV
CCY87565.1   TDQLTTPGTLNPVLPGCYPDPSICRVGNDYYMVNSSFAFYPGVPIWHSTDLTNWEQLGYVLNRPSQLPMYDGLRISGGIYAPDIKYNPHNGLFYLITTAV

BACINT_4202  DGGGNFFVTTDDPKKGSWSDPTFLPEVGGIDPGFLFDEDGKAYIVNNDGPAGKPEYDGHRAIWIREFDWKNGCTVGKQKMIVDGGVDKTRHPIWIEGPHL
CCY87565.1   DGGGNFFVTTDDPKKGSWSDPTFLPEVGGIDPGFLFDEDGKAYIVNNDGPAGKPEYDGHRAIWIREFDWKNGCTVGKQKMIVDGGVDKTRHPIWIEGPHL

BACINT_4202  YHINGTYYLMAAEGGTGPNHSEVIFTSASPFGPFKPCAINPILTQRGLPGDRPNPVTCVGHADLVETPAGDWYAVFLGVRPYRDGHDVMGRETFMLPVTW
CCY87565.1   YHINGTYYLMAAEGGTGPNHSEVIFTSASPFGPFKPCAINPILTQRGLPGDRPNPVTCVGHADLVETPAGDWYAVFLGVRPYRDGHDVMGRETFMLPVTW

BACINT_4202  KENQPIILPEGDVITYTADRSYGPAPLWTANGLAKEAFFIRTPLVPCYSINDKGQLEMTASSTDLNQKRQPAAIGRWINNWTFTAQTGLDFVPQQPKDFA
CCY87565.1   KENQPIILPEGDVITYTADRSYGPAPLWTANGLAKEAFFIRTPLVPCYSINDKGQLEMTASSTDLNQKRQPAAIGRWINNWTFTAQTGLDFVPQQPKDFA

BACINT_4202  GIICFHDDNCYIRFGKTLDKDGKPVMLLETYSHGRLCSQAGSPLTWTDGKVYLKVEGDNAVNYTFCYSTSPKGSWTQVGEPVSADLISTQTAGGFTGTMV
CCY87565.1   GIICFHDDNCYIRFGKTLDKDGKPVMLLETYSHGRLCSQAGSPLTWTDGKVYLKVEGDNAVNYTFCYSTSPKGSWTQVGEPVSADLISTQTAGGFTGTMV

BACINT_4202  GIYATGNYTN
CCY87565.1   GIYATGNYTN
```

Figure S10. Amino acid sequence alignment of BACINT_04202 and the most similar protein (CCY87565.1) from *B. intestinalis* CAG:564. The amino acid sequences were obtained from GenBank, aligned with ClustalW, and then visualized using ESPript 3.0.

```
Inverted Repeat 1

 385            390            395            400            405
 |              |              |              |              |
 Q  L  T  I  D  N  S  M  Q  R  D  S  A  A  N  C  Q  L  S  T  V  N  C  Q  L
CAATTGACAATTGACAATTCTATGCAGCGTGACAGCGCAGCCAATTGTCAATTGTCAACTGTCAATTGTCAATTG
       |         |         |         |         |         |         |
      1160      1170      1180      1190      1200      1210      1220

Inverted Repeat 2

 385            390            395            400            405
 |              |              |              |              |
 Q  L  T  I  D  N  S  M  Q  R  D  S  A  A  N  C  Q  L  S  T  V  N  C  Q  L
CAATTGACAATTGACAATTCTATGCAGCGTGACAGCGCAGCCAATTGTCAATTGTCAACTGTCAATTGTCAATTG
       |         |         |         |         |         |         |
      1160      1170      1180      1190      1200      1210      1220
```

Figure S11. Analysis of the inverted repeats in BACINT_04202. The nucleotide sequence for BACINT_04202 was analyzed for inverted and direct repeats using UGENE (1). Two pairs of inverted repeats with 23 (IR1) and 43 (IR2) nucleotide spacers were identified with partial overlap at the 5′ end. Shaded regions represent the inverted repeats. Top numbers indicate amino acid position and the bottom numbers indicate nucleotide positions.

1. Okonechnikov K, Golosova O, & Fursov M (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28(8):1166-1167.