



STRUCTURAL  
BIOLOGY

Volume 72 (2016)

**Supporting information for article:**

**Conservation in the mechanism of glucuronoxylan hydrolysis revealed by the structure of glucuronoxylan xylanohydrolase (CtXyn30A) from *Clostridium thermocellum***

**Filipe Freire, Anil Verma, Pedro Bule, Victor D. Alves, Carlos M. G. A. Fontes, Arun Goyal and Shabir Najmudin**

**Table S1** Data collection and structure refinement statistics.

Values in parentheses are for the highest resolution shells.

PDB code	<b>4UQA</b>	<b>4UQB</b>	<b>4UQC</b>	<b>4UQE</b>	<b>4UQ9</b>	<b>5A6L</b>
<b>Beamline</b>	IO3-4 DIAMOND	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL	PROXIMA-1, SOLEIL
<b>Wavelength (Å)</b>	0.9795	0.9537	0.9537	0.9537	0.9537	0.953
<b>Space group</b>	P1	P1	P1	P1	P1	P2 <sub>1</sub>
<b>Number of molecules in A.U.</b>	1	1	1	1	1	1
<b>Unit cell parameters (Å, °)</b>	$a = 46.73$ $b = 50.37$ $c = 58.87$ $\alpha = 65.00$ $\beta = 67.28$ $\gamma = 76.84$	$a = 46.98,$ $b = 50.59,$ $c = 59.31,$ $\alpha = 114.77,$ $\beta = 101.43,$ $\gamma = 103.85$	$a = 50.39$ $b = 50.44$ $c = 58.19$ $\alpha = 111.90$ $\beta = 110.80$ $\gamma = 97.67$	$a = 50.39$ $b = 50.44$ $c = 58.19$ $\alpha = 111.90$ $\beta = 110.80$ $\gamma = 97.67$	$a = 46.62$ $b = 50.33$ $c = 58.71$ $\alpha = 65.16$ $\beta = 67.56$ $\gamma = 76.99$	$a = 50.13$ $b = 87.25$ $c = 60.03$ $\alpha = \gamma = 90.0$ $\beta = 114.68$
<b>Resolution limits (Å)</b>	50.39-1.52 (1.57-1.52)	50.68-1.68 (1.74-1.68)	44.84-1.30 (1.32-1.30)	44.84-1.30 (1.32-1.30)	45.53-1.77 (1.84-1.77)	46.25-1.80 (1.84-1.80)
<b>Total number of reflections</b>	182666 (11567)	214085 (18208)	463345 (19528)	463345 (19528)	173410 (1554)	164832 (10007)
<b>Number of unique reflections</b>	66458 (2618)	47651 (4525)	109732 (4770)	109732 (4770)	40534 (3836)	43517 (2642)

<b>Multiplicity</b>	4.3 (4.4)	4.3 (4.0)	4.2 (4.1)	4.2 (4.1)	4.1 (4.0)	3.8 (3.8)
<b>Completeness (%)</b>	96.51(94.56)	92.29 (77.25)	94.4 (82.80)	94.4 (82.80)	93.44 (79.54)	100.0 (100.0)
<b>&lt;I/σ(I)&gt;</b>	19.48 (10.22)	13.63 (1.77)	20.8 (7.9)	20.8 (7.9)	14.15 (1.91)	17.5 (9.8)
<b>CC<sub>1/2</sub></b>	0.999 (0.996)	0.997 (0.837)	0.994 (0.973)	0.994 (0.973)	0.991 (0.764)	0.972 (0.568)
<b>R<sub>merge</sub><sup>a</sup></b>	0.050 (0.142)	0.085 (0.827)	0.055 (0.177)	0.055 (0.177)	0.132 (0.952)	0.280 (1.494)
<b>R<sub>p.i.m.</sub><sup>b</sup></b>	0.063 (0.778)	0.101 (1.438)	0.031 (0.098)	0.031 (0.098)	0.175 (2.080)	0.152 (0.795)
<b>Refinement</b>						
<b>Reflections used</b>	63184	45706	107084	93795	38598	41363
<b>Resolution used in refinement, Å</b>	50.39-1.52	45.49-1.68	44.84-1.30	40.18-1.28	45.53-1.77	54.54-1.80
<b>R<sub>work</sub> (%)</b>	13.5	17.9	11.2	9.4	20.7	13.6
<b>R<sub>free</sub> (%)</b>	17.1	20.6	14.1	11.1	23.7	16.2
<b>Nº protein atoms</b>	3158	3165	3266	3256	3149	3186
<b>Nº waters</b>	320	415	582	602	221	404
<b>Ligands*</b>	-	2 EPE 2 SO <sub>4</sub> <sup>2-</sup>	1 TAR, 1TLA 1 GLC	2 GOL, 1PGO, 1 TRS, 2PEG	-	4 XYP
<b>r.m.s. bond lengths</b> (Å)	0.010	0.007	0.011	0.009	0.019	0.018
<b>r.m.s. bond angles</b> (degrees)	1.392	1.212	1.443	1.504	1.866	1.763
<b>Mean B factors (Å<sup>2</sup>)</b>						

<b>Protein main chain atoms</b>	19.958	26.580	10.614	7.321	20.441	8.425
<b>Protein side chain atoms</b>	22.934	27.816	13.079	9.547	21.681	9.499

<sup>a</sup>,  $R_{merge} = \sum_{hkl} \sum_i (I_i(hkl) - \langle I(hkl) \rangle) / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the  $i^{\text{th}}$  intensity measurement of reflection  $hkl$ , including symmetry-related reflections, and  $\langle I(hkl) \rangle$  is its average.

<sup>b</sup>,  $R_{p.i.m.} = \left( \sum_{hkl} \sqrt{\frac{1}{n-1}} \sum_{j=1}^n |I_{hkl,j} - \langle I_{hkl} \rangle| \right) / (\sum_{hkl} \sum_j I_{hkl,j})$ , where  $\langle I_{hkl} \rangle$  is the average of symmetry-related observations of a unique reflection. CC<sub>1/2</sub> is the half-data-set correlation coefficient (Diederichs and Karplus, 2013)

$R_{work} = \left( \sum_{hkl} |F_{hkl}^{obs} - F_{hkl}^{calc}| \right) / \left( \sum_{hkl} F_{hkl}^{obs} \right) \times 100$  where  $F^{calc}$  and  $F^{obs}$  are the calculated and observed structure factor amplitudes, respectively.  $R_{free}$  is calculated for a randomly chosen 5% of the reflections.

\* EPE - 4-(2-hydroxyethyl)-1-piperazine ethane sulfonic acid; TAR - (D) - tartaric acid; TLA - L(+) – tartaric acid; GLC - D-glucose; GOL - glycerol ; PGO - 1,2-propanediol; TRS - 2-amino-2-hydroxymethyl-propane-1,3-diol; PEG - polyethylene glycol; XYP –  $\beta$ -D-xylofuranose.

**Table S2** Substrate specificity of *CtXyn30A* from *C. thermocellum*.

All assays were performed in triplicate at 70°C using 20 mM sodium phosphate buffer (pH 6.0).

Substrates	Specific Activity (U.mg <sup>-1</sup> )
Xylan (Beechwood)	31±0.8
Xylan (Birchwood)	28±0.7
4-O-methyl Glucuronoxyran	27±0.7
Oat spelt xylan	10±0.6
Arabinoxylan (wheat, insoluble)	9±0.5
Arabinoxylan (wheat, soluble)	7±0.4
Arabinoxylan (Rye)	6±0.03
Arabinogalactan	4.5±0.3
Arabinan (sugar beet)	4.0±0.3
Glucomannan (locus bean)	2.7±0.2
Xyloglucan	2.2±0.2
Glucomannan (Konjac)	1.4±0.2
Other Substrates*	ND

ND= No activity detected.

\* Carob galactomannan, Galactan (Lupin), β-D-Glucan (Barley),Carboxy methyl cellulose, Pecticgalactan (Apple), Pecticgalactan (Citrus), Pecticgalactan (Lupin), Rhamnogalactouronan, Curdlan, Pullulan and Pustulan

**Table S3** Conservation of key residues in the catalytic binding pocket of *CtXyn30A* and homologous structures.

<i>CtXyn30A</i>	<i>BsXynC</i>	<i>EcXynA</i>	<i>CpC71</i>	<i>PbXyn30D</i>
Trp23	Trp27	Trp55	Trp25	Trp26
Arg49	Arg53	Arg81	Arg47	Arg52
Trp81	Trp85	Trp113	Trp88	Trp84
Asn135	Asn139	Asn164	Asn142	Asn137
Tyr139	Tyr143	Tyr168	Tyr146	Tyr141
Trp143	Trp147	Tyr172	Trp150	Trp145
Tyr200	Tyr204	Tyr232	Tyr207	Tyr203
Tyr227	Tyr231	Tyr255	Tyr234	Tyr230
Trp264	Trp268	Trp289	Trp265	Tyr267
Tyr265	Tyr269	Tyr290	Trp266	Tyr268
Arg268	Arg272	Arg293	Trp269	Arg271
Tyr270	Tyr274	Tyr295	Asn271	Tyr273