Supporting information - Tables

Table S1: *EG16* expression constructs screened for protein production

Construct	Production result
SmEG16	Strong production into inclusion bodies
PpEG16	Strong production into inclusion bodies
PpEG16-sfGFP	Weak production into inclusion bodies
PpEG16-SUMO	Strong production into inclusion bodies
BdEG16	Strong production into inclusion bodies
BdEG16-sfGFP	Soluble-aggregate production, precipitation after removal of sfGFP tag
BdEG16-SUMO	Soluble-aggregate production, precipitation after removal of SUMO tag
VvEG16	Strong soluble production
VvEG16-sfGFP	Slow soluble production
VvEG16-SUMO	Strong soluble production
VvEG16(ΔV152)	Strong soluble production
VvEG16(ΔV152)-sfGFP	Weak soluble production
VvEG16(ΔV152)-SUMO	Strong soluble production
VvEG16(ΔV152,E89A)	Strong soluble production
VvEG16(E89A)	Strong soluble production

Table S2: ^{18}O-labeling of each product from cellooligosaccharide hydrolysis.

Substrate Product	**% ^{18}O Labelled**[a]	**Error**
Cellotriose		
Glc	0.4	1.4
GG	35.2	2.8
Cellotetraose		
Glc	3.2	2.8
GG	51.2	0.9
GGG	89.4	1.6
Cellopentaose		
GG	17.6	2.6
GGG	79.0	0.8
Cellohexaose		
GG	15.6	2.6
GGG	50.7	0.8
GGGG	80.5	1.1

[a]Calculated as $\%L = ((I_{M+2}/(I_{M+2}+I_M))-(I_{N+2}/(I_{N+2}+I_N)))/x_{18O}$, where x_{18O} is the mole fraction of water which contains ^{18}O, I is the integration of an MS peak, $_M$ indicates base peak with ^{18}O water, $_N$ indicates base peak with ^{16}O water and $_{M+2}$ and $_{N+2}$ indicate the peak found 2 Da heavier than the each base peak. x_{18O} was estimated to be 0.85 for this experiment.

Table S3: X-ray diffraction data statistics.

	VvEG16(ΔV152,E89A): GGGG complex	VvEG16(ΔV152,E89A): bMLGO complex	VvEG16(ΔV152,E89A): tXyGO complex	VvEG16(ΔV152,C22S, C188S): tXyGO complex
PDB Code	**5DZE**	**5DZF**	**5DZG**	**5SV8**
		DATA COLLECTION		
Wavelength (Å)	0.9795	0.8266	1.0332	1.0332
Space group	$P2_12_12_1$	$P4_32_12$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c*, (Å)	45.15, 52.02, 83.06	74.33, 74.33, 149.28	41.3, 79.0, 132.6	46.2, 51.9, 82.6
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution, Å	44.09-0.97 (1.00-0.97)[b]	42.97-1.55 (1.69-1.65)	39.41-1.79 (1.83-1.79)	27.54-1.588 (1.62-1.59)
R_{merge}[a]	0.065 (0.901)	0.20 (1.25)	0.097 (0.317)	0.19 (0.69)
$I / \sigma(I)$	16.34 (1.25)	10.85 (1.43)	12.2 (3.6)	12.2 (2.8)
Completeness, %	95.6 (69.0)	100 (97.0)	98.2 (78.3)	99.5 (91.2)
Redundancy	6	12.1	5.9	6.2
		REFINEMENT		
No. of reflections:				
working, test	106175, 5588	48580, 2558	38822, 2056	27417, 1382
R-factor/free *R*-factor[c]	0.136/0.150 (0.350/0.366)	0.155/0.200 (0.304/0.376)	0.171/0.217 (0.225/0.285)	0.152/0.179 (0.220/0.259)
No. of refined atoms				
Protein	3361	3208	3092	3113
Ligand	114	146	215	63
Solvent	284	463	401	228
B-factors				
Protein	9.8	20.3	20.9	10.6
Ligand	13.0	21.0	28	18.4
Solvent	23.0	35.2	32.4	21.9
RMSD				
Bond lengths, Å	0.009	0.008	0.011	0.010
Bond angles, °	1.53	1.25	1.50	1.39
Ramachandran Plot:				
Favoured, %	96.6	97.3	97.5	98.0
Allowed, %	2.3	2.7	2.5	2.0
Disallowed, %	1.1	0	0	0

[a] $R_{merge} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)$, where $I_i(h)$ and $\langle I(h)\rangle$ are the *i*th and mean measurement of the intensity of reflection *h*.

[b] Figures in parentheses throughout the table indicate the values for the outer shells of the data.

[c] $R = \Sigma|F_p^{obs} - F_p^{calc}|/\Sigma F_p^{obs}$, where F_p^{obs} and F_p^{calc} are the observed and calculated structure factor amplitudes, respectively.

Table S4: Privateer validation results.

Residue Name	Conformation	Average B-factor	RSCC[1]
VvEG16(ΔV152,E89A): GGGG complex (5DZE)			
GLC-1	$^4C_1$	8.4	0.94
BGC-1	$^4C_1$	6.5	0.96
BGC-2	$^4C_1$	8.4	0.95
BGC-3	$^4C_1$	43.0	0.71
BGC-4	$^4C_1$	21.7	0.66
VvEG16(ΔV152,E89A): bMLGO complex (5DZF)			
GLC-1	$^4C_1$	16.4	0.95
GLC-2	$^4C_1$	16.8	0.93
BGC-1	$^4C_1$	13.8	0.96
BGC-2	$^4C_1$	16.9	0.94
BGC-3	$^4C_1$	24.5	0.92
BGC-4	$^4C_1$	24.8	0.92
BGC-5	$^4C_1$	25.8	0.90
BGC-6	$^4C_1$	21.7	0.95
BGC-7	$^4C_1$	16.6	0.95
BGC-8	$^4C_1$	15.0	0.94
BGC-9	$^4C_1$	19.4	0.94
BGC-10	$^4C_1$	32.7	0.87
BGC-11	$^4C_1$	28.6	0.77
VvEG16(ΔV152,E89A): tXyGO complex (5DZG)			
GLC-1	$^4C_1$	15.4	0.90
GLC-2	$^4C_1$	19.9	0.88
BGC-1	$^4C_1$	17.3	0.94
BGC-2	$^4C_1$	19.7	0.94
BGC-3	$^4C_1$	27.3	0.90
BGC-4	$^4C_1$	32.8	0.79
BGC-5	$^4C_1$	28.5	0.87
BGC-6	$^4C_1$	23.3	0.90
BGC-7	$^4C_1$	18.7	0.87
BGC-8	$^4C_1$	15.7	0.92
BGC-9	$^4C_1$	27.5	0.88
BGC-10	$^4C_1$	19.9	0.93
XYS-1	$^4C_1$	33.3	0.83
XYS-2	$^4C_1$	26.4	0.83
XYS-3	^2E	44.3	0.71
XYS-4	$^4C_1$	42.1	0.80
XYS-5	$^0S_2$	38.1	0.79
XYS-6	^2E	37.1	0.80
XYS-7	$^4C_1$	26.5	0.86
XYS-8	$^2S_0$	34.7	0.78
XYS-9	$^4C_1$	35.5	0.83
VvEG16(ΔV152,C22S,C188S): tXyGO complex (5SV8)			
BGC-1	$^4C_1$	7.9	0.93
BGC-2	$^4C_1$	9.1	0.89

BGC-3	$^4C_1$	13.9	0.87
BGC-4	$^4C_1$	19.7	0.84
XYS-1	$^4C_1$	33.6	0.66
XYS-2	$^4C_1$	31.7	0.75

[3]RSCC, short for Real Space Correlation Coefficient, measures the agreement between model and positive omit density. A RSCC below 0.8 is typically considered poor.

Table S5: Sequences used for generation of the GH16 phylogenetic tree.

Activity	GenBank Accession Codes
GH7 Cellulase	AAX28897.1, AAA65586.1, BAA09786.1, AEO58196.1, ABY56790.1
XTH	ABL75361.1, AAC09388.1, ACS13756.1, CAA48324.1, AAD31572.1
EG16	XM_002273939.3, XM_002301283.2, NM_001156874.1, AK356496.1, XM_002970606.1
Licheninase	CAA39426.1, AAB05759.1, CAA40547.1, AFX61538.1, AAA24896.1
Laminarinase	AAC38290.1, AAC44371.1, BAE54302.1, AAD35118.1, CAZ95067.1
Chitin Transglycosylase	AAB65002.1, EAL03096.1, EAL02960.1, EAL01616.1, CAA67525.1
Agarase	BAC99022.1, BAD29947.1, EDY95404.1, AAF21821.1, CAZ98378.1
Porphyranase	CBM41184.1, CBM41183.1, CBM41185.1, CBM41182.1, EDY95423.1
Kappa-Carrageenase	CAA50624.1, AGS43006.1, CAZ94309.1, BAJ61957.1, ADD92366.1

Table S6: Primer sequences used for cloning and mutagenesis.

Target	Primer Name	Primer Sequence
VvEG16(ΔV152)	Vv1-F Vv1-R	5'-TCGAAAACCTGTACTTCCAAGGCATGGCGGATCCGAGCTTG-3' 5'-CAGTGGTGGTGGTGGTGGTGCTCGAGTCACTCCACTGCGGTACCAAC-3'
VvEG16(ΔV152, E89A) and VvEG16(E89A)	Vv2-F Vv2-R	5'-AGAAATTCGAAATCGATTGCGTCCTGGCTTTTGTCCC-3' 5'-GGGACAAAAGCCAGGACGCAATCGATTTCGAATTTCT-3'
VvEG16((ΔV152, C22S)	Vv3-F Vv3-R	5'-GCGGTCGATTACTCTCCCGAATCTTGC -3' 5'-GCAAGATTCGGGAGAGTAATCGACCGC -3'
VvEG16((ΔV152, C22S,C188S)	Vv4-F Vv4-R	5'-GGTCCTTACGTTGGTTCCGATGCGCCGTATATC -3' 5'-GATATACGGCGCATCGGAACCAACGTAAGGACC -3'
pET24a(+)	pET24-F pET24-R	5'-GCCTTGGAAGTACAGGTTTTC-3' 5'-TGACTCGAGCACCACCAC-3'
P15TvLIC	p15Tv-F p15Tv-R	5'-CAAAGCCCGAAAGGAAGC-3' 5'-GCCCTGGAAATACAAGTTTTC-3'
PpXG5-pET21a (G323E)	PpXG-F PpXG-R	5'-GTTATCCGGTCGTGATTGGCGAATTTGGATCGATTGATAAAAC-3' 5'-GTTTTATCAATCGATCCAAATTCGCCAATCACGACCGGATAAC-3'