A β-mannannase with a lysozyme-like fold and a novel molecular catalytic mechanism

SUPPORTING INFORMATION

Yi Jin, Marija Petricevic, Alan John, Lluís Raich, Huw Jenkins, Leticia Portela De Souza, Fiona Cuskin, Harry J. Gilbert, Carme Rovira,* Ethan D. Goddard-Borger,* Spencer J. Williams,* Gideon J. Davies*

Contents

Supplementary Figures and Tables

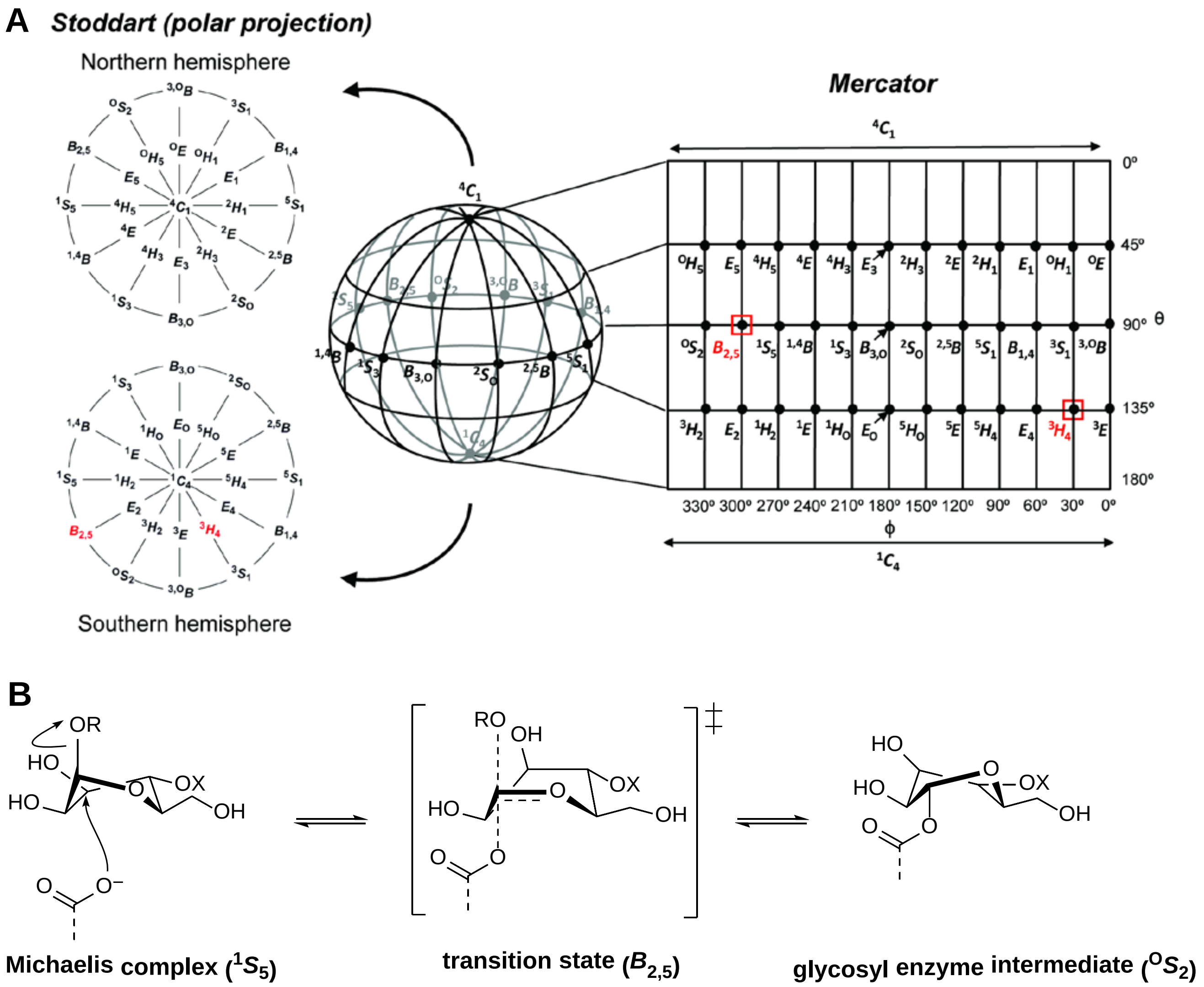


Figure S1: The conformational space of β-mannosidase catalysis

(**A**) The conformational surface for pyranosides shown as a sphere, and with Stoddart (polar) and Mercator projections. GH family 5, 26, and 113 β-mannanases (and GH family 2 β-mannosidases) are believed to operate through the '*Equatorial*' $B_{2,5}$ TS conformation, and in the present study GH family 134 β-mannanases operate through the '*Southern hemisphere*' $^{3}H_{4}$ TS conformation (shown in red).

(**B**) The conformational itinerary for the glycosylation half-reaction of the retaining β-mannosidases/β-mannanases of glycoside hydrolase families GH2, 26 and 113.

A

>gi|663231811|ref|WP_030268297.1| hypothetical protein [Streptomyces sp. NRRL B-24484]
MRRTASLLGSAVGTLAALTLALAPTAAAETAPNGYPYCANGSASDPDGDGWGWENNRSCVVRTGSGSGSG
SGSSACPSGATCGSYTVGGLGSRKQQVRNAGGSSLDLAVAML**E**TERMDTAYPYG**D**NKSGDAANFGIFKQN
WLMLRSACAQFGGQGAGQYDNGAALNSSLGQDVSCLHQSQSHYGLDAWFAGHRNGASGLSSPNTADIAAY
KAAVYWIKAQLDADSANLGNDTRFWVQVPAI

Yellow = signal peptide
Green = CBM10
Blue = GH134

B

>MBP-SsGH134
MKIHHHHHHEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAAT
GDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEAL
SLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYEN
GKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWS
NIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAV
NKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASG
RQTVDEALKDAQTNSSSNNNNNNNNNNLGLEVLFQGPLGSSACPSGATCGSYTVGGLGSR
KQQVRNAGGSSLDLAVAML**E**TERMDTAYPYG**D**NKSGDAANFGIFKQNWLMLRSACAQFGG
QGAGQYDNGAALNSSLGQDVSCLHQSQSHYGLDAWFAGHRNGASGLSSPNTADIAAYKAA
VYWIKAQLDADSANLGNDTRFWVQVPAI*

Yellow = hexahistidine tag
Green = MBP
Grey = HRV 3C protease site
Blue = GH134

Figure S2: Native and expression construct sequences for *Ss*GH134.

(**A**) Predicted protein sequence of *Ss*GH134. (**B**) Recombinant *Ss*GH134 protein sequence, consisting of an N-terminal hexahistidine tag, maltose binding protein (MBP) fusion partner and HRV 3C protease cleavage site. The general acid and base residues are underlined/bold.

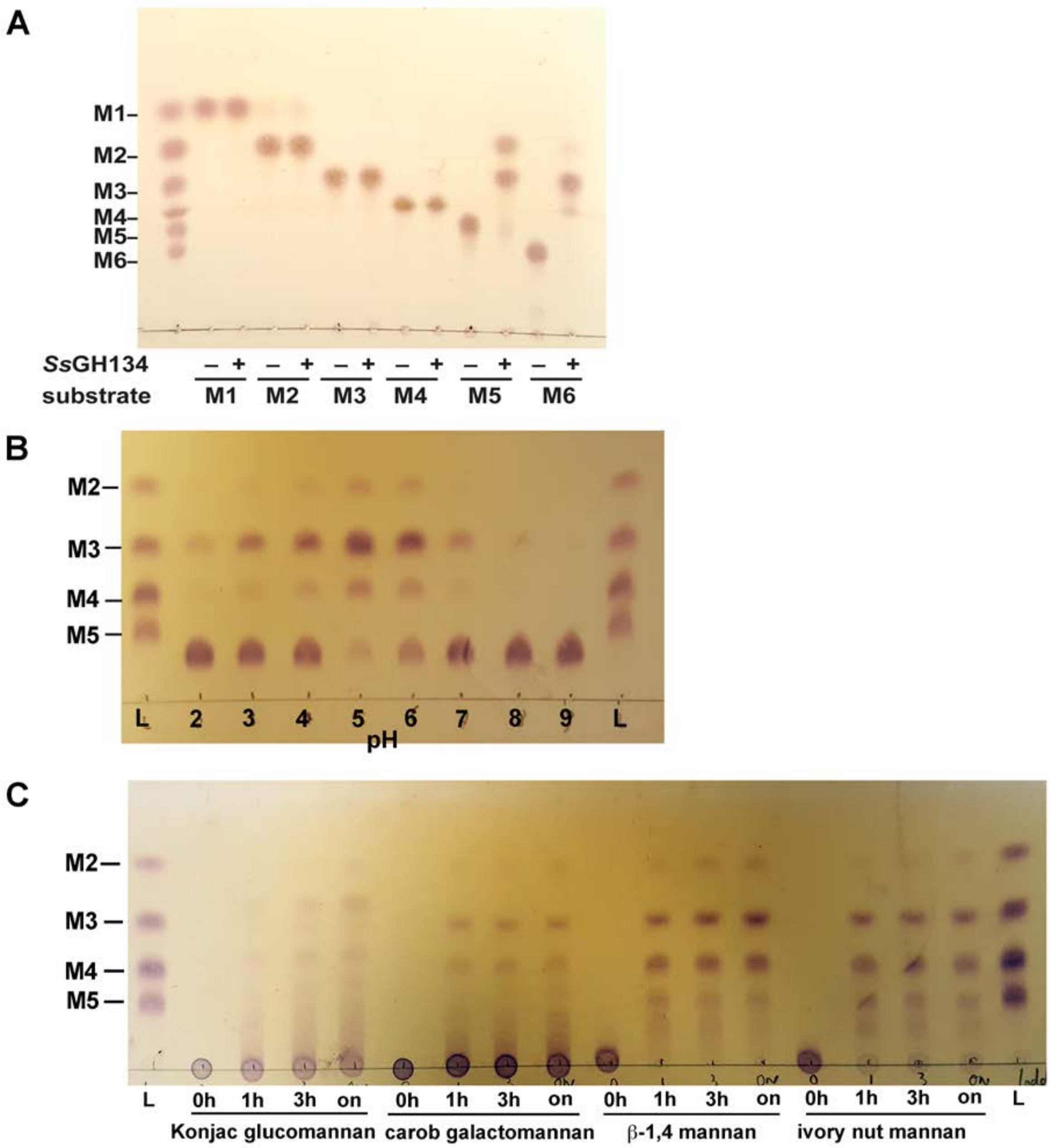


Figure S3: *Ss*GH134 digests of β-1,4-mannooligosaccharides and assorted mannans.

(**A**) β-1,4-Mannooligosaccharides (0.1 mg) in H_2O (25 µl) were digested with MBP-*Ss*GH134 at room temperature for 30 min. M1, mannose; M2, mannobiose; M3, mannotriose; M4, mannotetraose; M5, mannopentaose; M6, mannohexaose; -/+ indicate the absence/presence of enzyme. (**B**) The pH assay was performed using 1 mM M6 digested with 0.5 µM MBP-*Ss*GH134 at 37 ºC for 5 min, in 20 mM NaOAc buffer (pH 2.0 – 4.0), 20 mM MOPS buffer (pH 5.0), 20 mM MES buffer (pH 6.0), 20 mM phosphate buffer (pH 7.0), or 20 mM Tris (pH 8.0-9.0). (**C**) The substrate preference assay was performed for four different mannnans at 2 mg/mL, digested with 0.5 µM MBP-*Ss*GH134 for 0, 1 h, 3 h and overnight (on). β-1,4-mannan is a purified preparation from Megazyme.

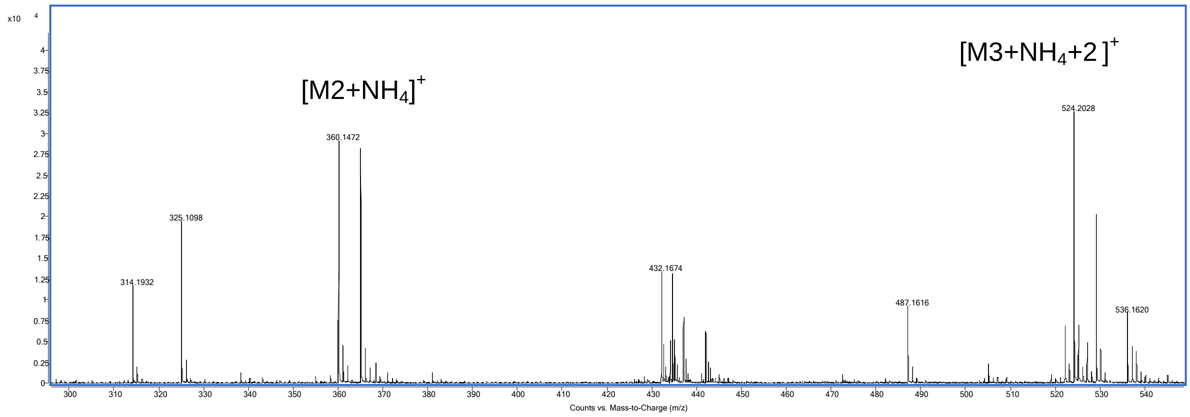
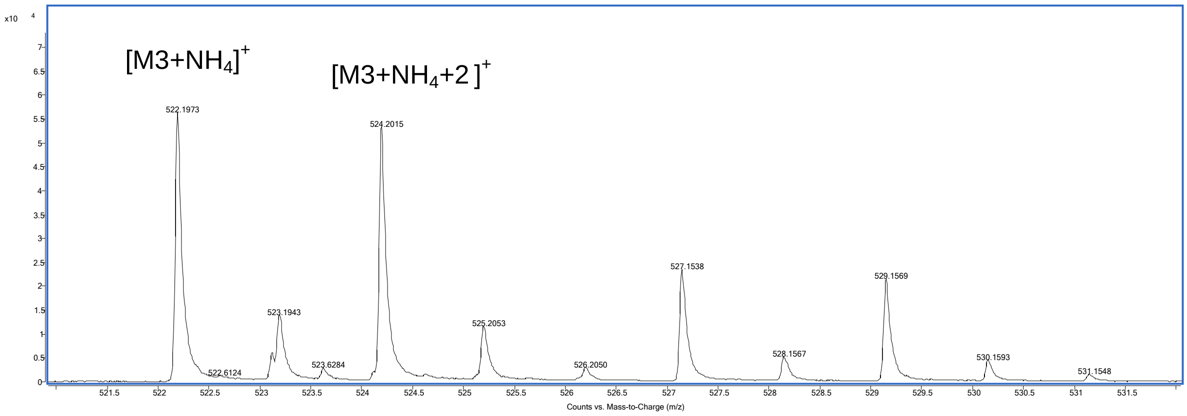


Figure S4: Isotope-mapping of *Ss*GH134 catalyzed cleavage of β-1,4-mannopentaose (M5).

β-1,4-Mannopentaose was dissolved in buffered ^{18}O-labelled water and *Ss*GH134 was added. The reaction mixture was analyzed by mass spectrometry. Mass spectrum showing M2 and M3+2 quasi-molecular ions (NH_4^+ adducts), consistent with ^{18}O incorporation into the latter fragment. These data are consistent with hydrolysis of M5 across the -3 → +2 subsites.

A



B

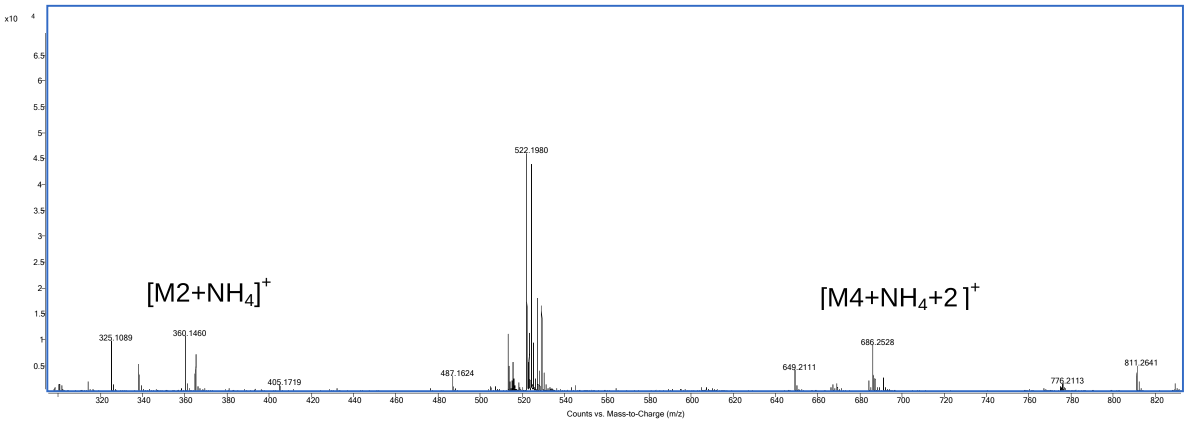


Figure S5: Isotope-mapping of *Ss*GH134 catalyzed cleavage of β-1,4-mannohexaose (M6).

β-1,4-mannohexaose was dissolved in buffered ^{18}O-labelled water and *Ss*GH134 was added. The reaction mixture was analyzed by mass spectrometry. (**A**) Mass spectrum showing M3 and M3+2 quasi-molecular ions (NH_4^+ adducts), consistent with ^{18}O incorporation into the latter fragment. (**B**) Mass spectrum showing M2 and M4+2 quasi-molecular ions (NH_4^+ adducts), consistent with ^{18}O incorporation into the latter. These data are consistent with hydrolysis of M6 across the -3 → +3 subsites as well as -4 → +2 subsites.

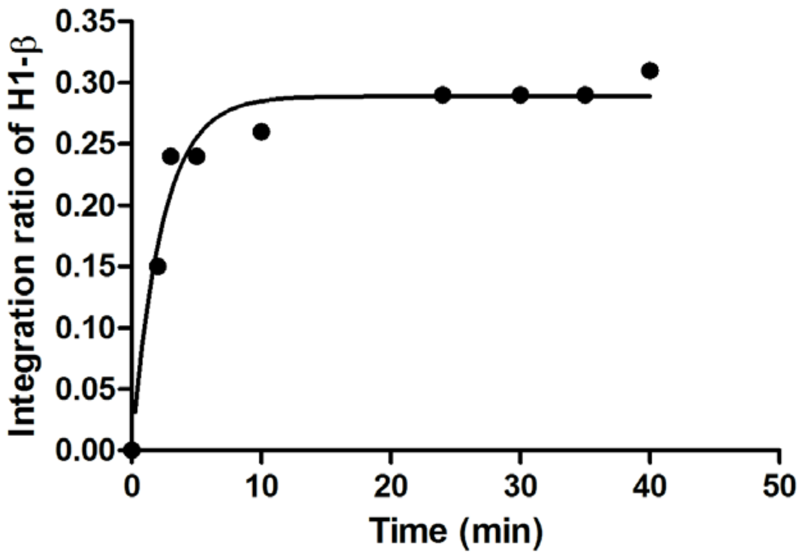
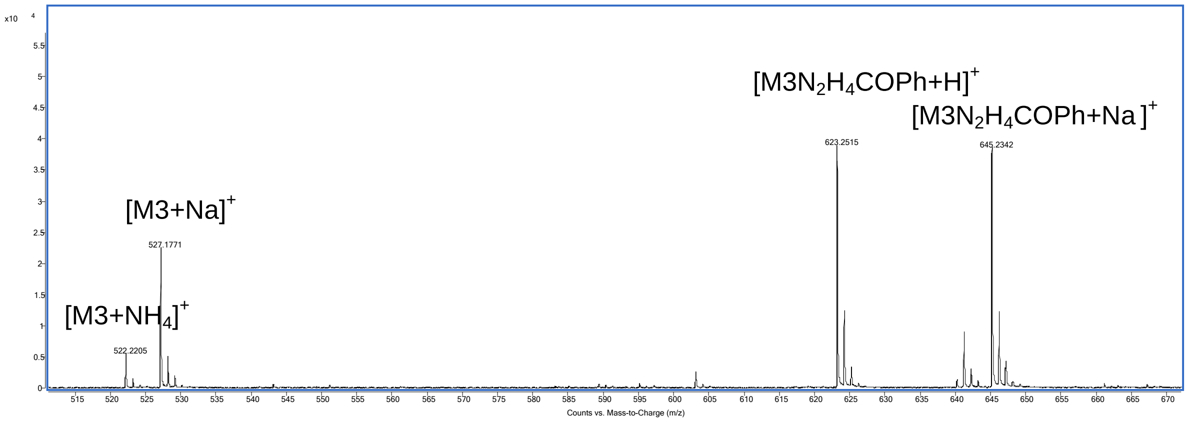


Figure S6: Plot showing mutarotation to the β-anomer, integration ratio of H1β = (integral H1β)/(sum of integrals of H1α and H1β).

A



B

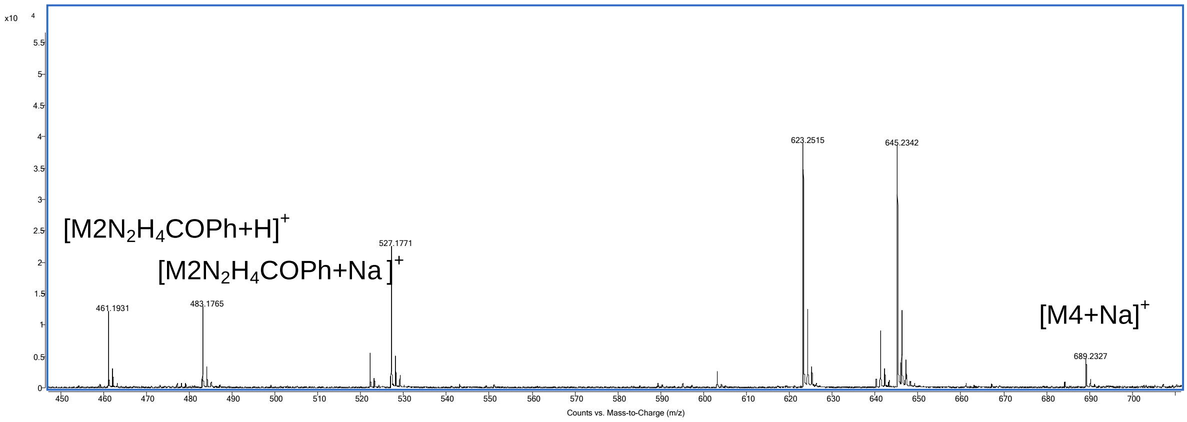


Figure S7: Product analysis of *Ss*GH134 catalyzed cleavage of mannohexaosyl benzoylhydrazide.

Mannohexaosyl benzoylhydrazide was diluted into water and analyzed by mass spectrometry. (**A**) Mass spectrum showing M3 and M3-N_2H_4Bz quasi-molecular ions (H^+ and Na^+ adducts). (**B**) Mass spectrum showing M2 and M4-N_2H_4Bz quasi-molecular ions (H^+ and Na^+ adducts). These data reveal hydrolysis of M6-N_2H_4Bz across the -3 → +3 subsites as well as -4 → +2 subsites.

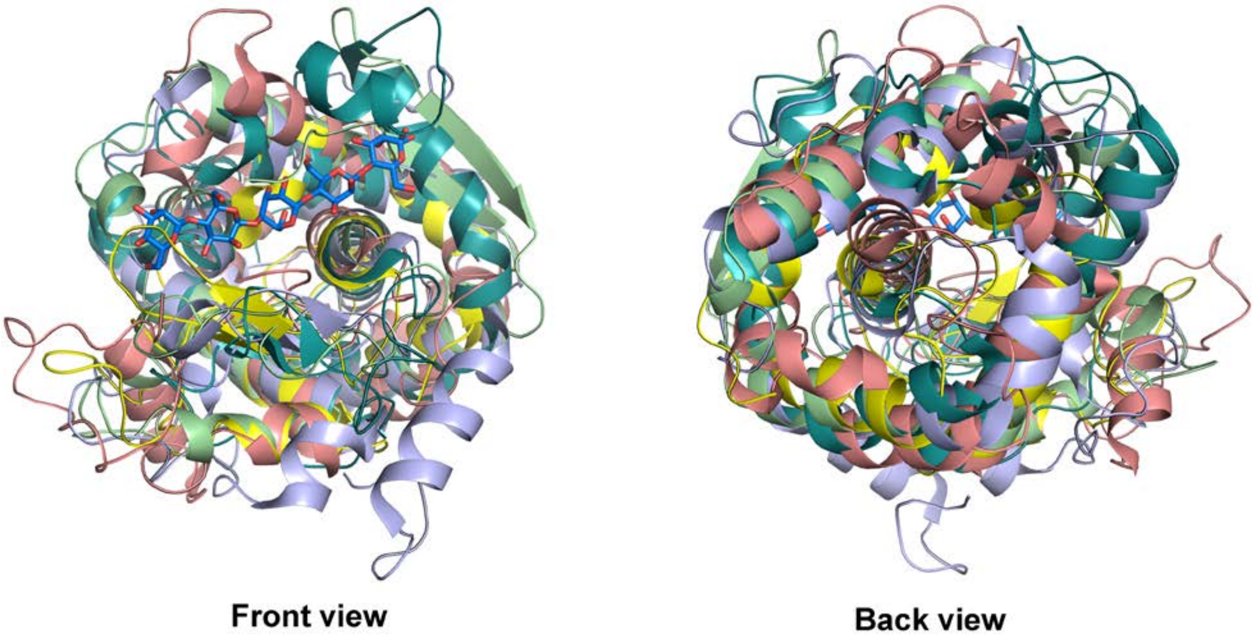
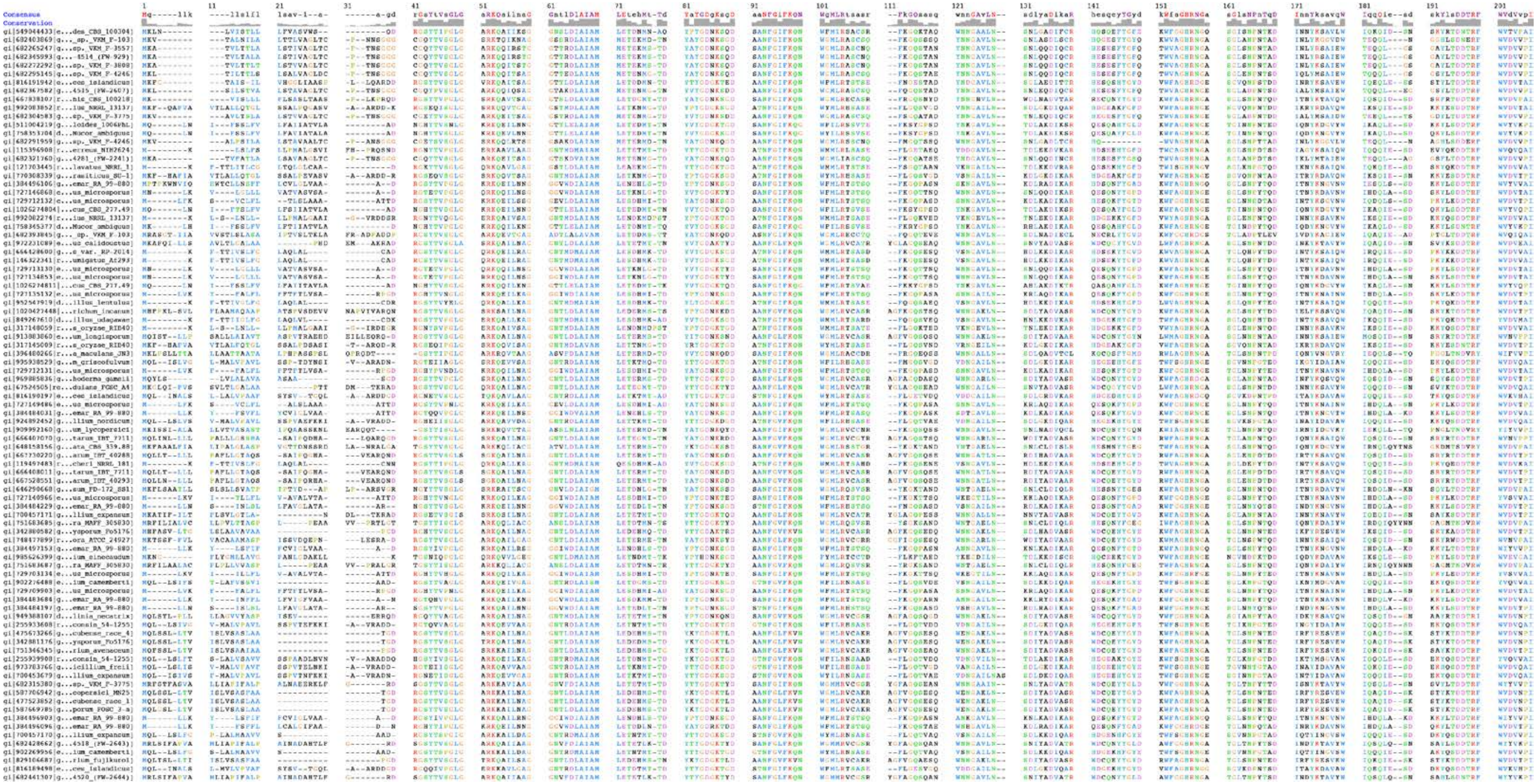


Figure S8: Overlay of structures of representatives of glycoside hydrolases with lysozyme-like folds.

Overlay of *Ss*GH134 (PDB code 5JUG, in light green) bound to mannopentaose, with the inverting chitinase A (GH19, PDB code 3WH1, in pink), retaining hen egg white lysozyme (HEWL; GH22, PDB code 2WAR, in yellow), inverting G-type lysozyme (GH23, PDB code 3GXR, in dark green), and inverting cellulase *Ct*Cel124 (GH124, PDB code 2XQO, in lavender). All polysaccharide ligands in the overlay were omitted for clarity.

Figure S9. Family GH134 sequence alignment



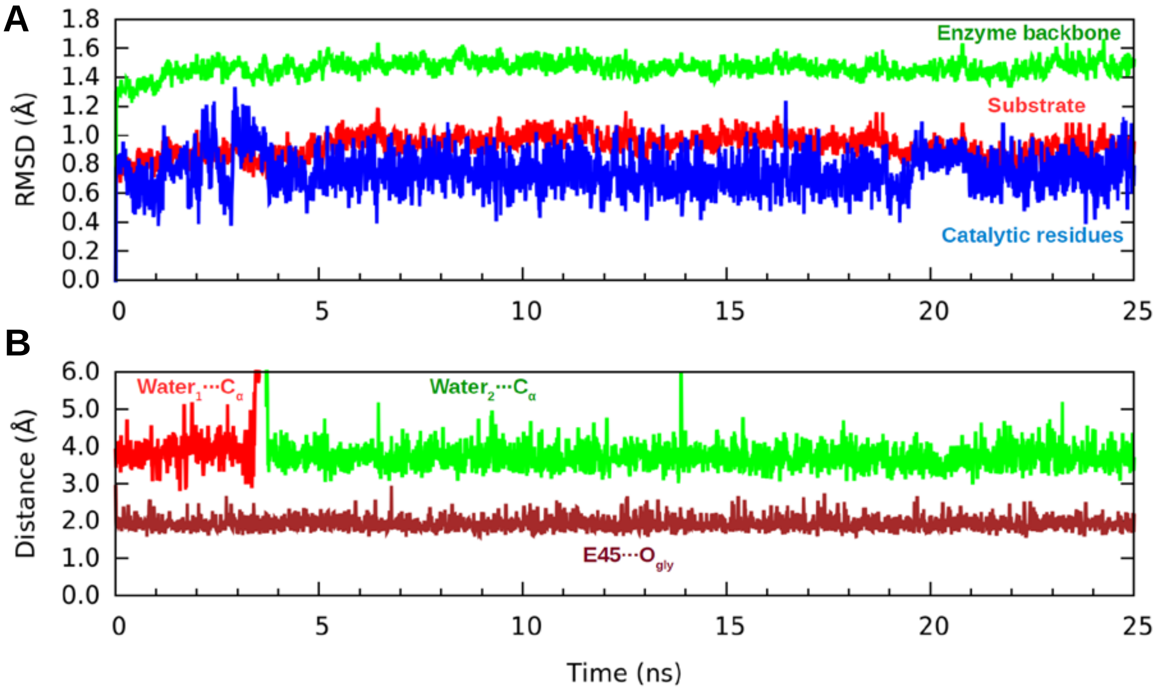


Figure S10: Time evolution of structural parameters during the classical MD simulation.

(**A**) Time evolution of the RMSD of the enzyme backbone (green), substrate (red) and catalytic residues (blue) during the classical MD simulation. (**B**) Distance between the catalytic water and the anomeric carbon (red and green). C_α refers to the C1 atom of the –1 subsite mannose. The water molecule is replaced by another one from the solvent at ≈ 3.5 ns. The time evolution of the hydrogen bond distance between the acid residue (E45) and the glycosidic oxygen is shown in brown.

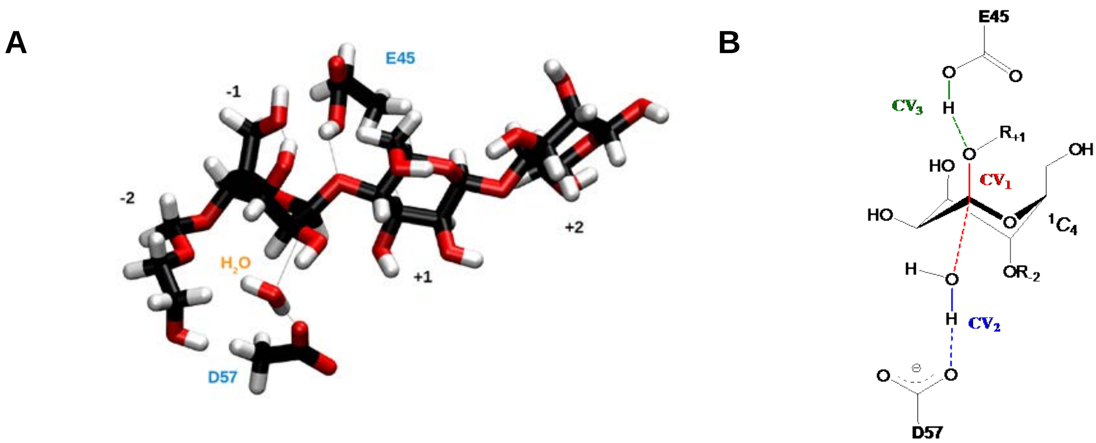


Figure S11: QM region and collective variables.

(**A**) The QM region (98 atoms) used in the QM/MM simulations. (**B**) Collective variables (CVs) used in the metadynamics simulations.

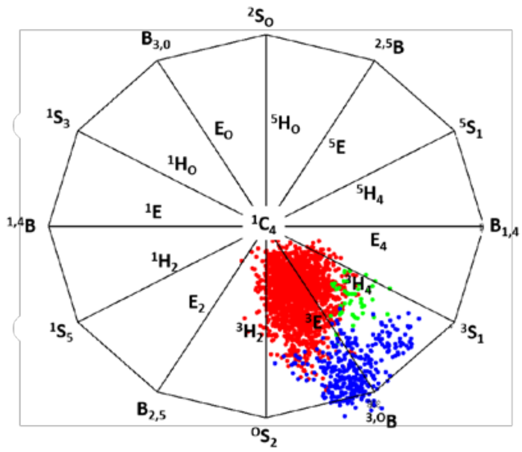


Figure S12: Conformational itinerary of the –1 subsite mannose.

Conformational itinerary of the –1 subsite mannosyl ring during the reaction, mapped onto the projection of the Cremer-Pople sphere on the Southern hemisphere. The conformations have been extracted from the reactants well (**R**), the products well (**P**) and a small region around the transition state (**TS**).

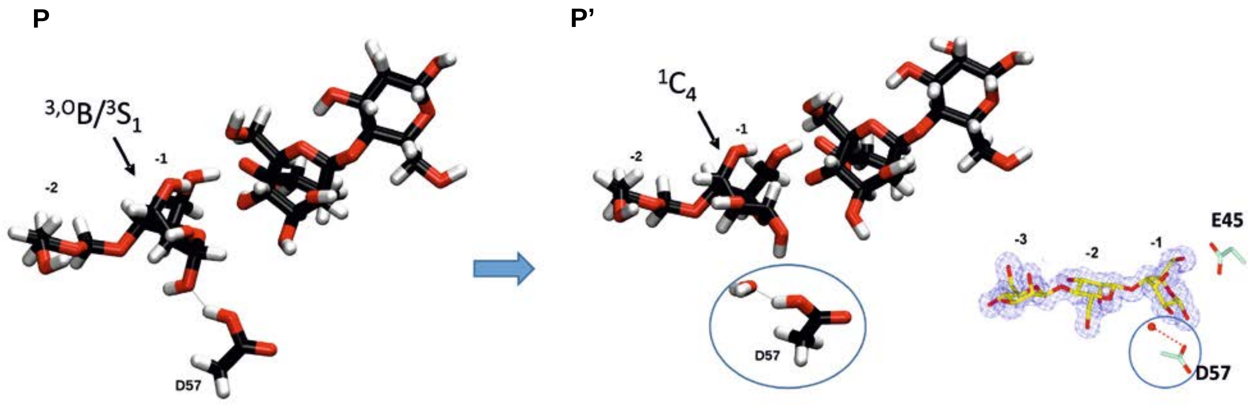


Figure S13: Evolution of the conformation of the –1 subsite mannose upon cleavage of the glycosidic bond

The –1 subsite mannose adopts a $^{3,O}B/^{3}S_1$ conformation at the **P** well on the free energy landscape (**Figure 5a**). Afterwards, D57 changes its hydrogen bond partner from the 1-OH to a water molecule (transition from **P** to **P'** in Figure 5A, $\Delta G^{\ddagger}$ = 2 kcal mol^{-1}) that fills the space previously occupied by the 1-OH mannose substituent. This results in the spontaneous change of the sugar ring to a $^{1}C_4$ conformation.

Table S1: *Ss*GH134 X-ray data collection, processing and refinement statistics.

	Native	*Ss*GH134-mannotriose	E45Q-mannopentaose
	5JTS	5JU9	5JUG
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, b, c (Å)	36.9, 53.7, 81.5	36.7, 53.6, 82.1	36.7, 53.8, 82.6
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	33.62 (1.09) *	44.91 (1.18)*	45.1 (0.96) *
R_{sym} or R_{merge}	0.066 (0.864)	0.056 (0.558)	0.054 (0.520)
$I/\sigma I$	12.2 (1.8)	16.6 (2.4)	12.4 (1.1)
Completeness (%)	99.9 (99.9)	97.5 (80.2)	84.1 (29.2)
Redundancy	6.1 (5.8)	7.0 (3.7)	5.9 (2.5)
Refinement			
Resolution (Å)	1.09	1.18	0.96
No. reflections	68171	52696	84331
$R_{work/}$ R_{free}	0.139/0.163	0.131/0.153	0.122/0.141
No. atoms			
Protein	1237	1276	1276
Ligand/ion	23	36	75
Water	150	162	230
B-factors			
Protein	13.66	13.31	8.65
Ligand/ion	18.8	9.97	8.12
Water	28.5	26.67	21.47
R.m.s deviations			
Bond lengths (Å)	0.014	0.012	0.012
Bond angles (º)	1.56	1.57	1.67

Table S2: Change of the main distances (in Angstrom) involving the active site residues of *Ss*GH134 for each characteristic point along the reaction coordinate.

	O_{wat}-C1	C1-O1'	O_{wat}-H_{wat}	H_{wat}-O_{D57}	O_{E45}-H_{E45}	H_{E45}-O1'
R	3.68±0.18	1.55±0.15	1.01±0.03	1.74±0.06	1.20±0.30	1.78±0.57
TS	1.99±0.09	3.33±0.14	1.09±0.02	1.50±0.01	1.53±0.06	1.04±0.00
P	1.51±0.07	3.87±0.05	1.50±0.07	1.06±0.03	1.93±0.73	1.12±0.20
P'	1.43±0.03	3.28±0.03	2.60±0.02	0.98±0.04	1.62±0.01	0.99±0.00

Cloning, gene expression and protein purification of a Streptomyces sp. GH134

Hypothetical protein WP_030268297.1 from *Streptomyces* sp. NRRL B-24484 possesses a putative signal peptide, CBM10 domain, GS repeat linker, and GH134 domain (**Figure S2A**). A dsDNA polynucleotide encoding the GH134 domain of WP_030268297.1 (residues 74–241) was synthesized (Integrated DNA Technologies) and cloned into the vector pHisMALP to give pHisMALP-*Ss*GH134, which encodes hexahistidine-tagged MBP fused through an HRV 3C protease site to *Streptomyces* sp. GH134 (**Figure S2B**). *Ss*GH134 coding sequence highlighted in yellow:

```
>pHisMALP-SsGH134
CCGACACCATCGAATGGTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACG
ATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGA
TGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGC
CGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGG
CGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTC
CGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGG
GTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGC
CGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCA
ACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATTTCGGTAGTGGGATACGACGATACCGAAGACAGCT
CATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGA
AGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGC
AGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACAATTCTCATGTTTGACAGCT
TATCATCGACTGCACGGTGCACCAATGCTTCTGGCGTCAGGCAGCCATCGGAAGCTGTGGTATGGCTGTGCAGGTCGTAAATCACTGCATAATTCGTGTCGC
TCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGT
ATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCCAGTCCGTTTAGGTGTTTTCACGAGCAATTGACCAACAAGGACCATAGATTATG
AAAATCCATCACCATCACCATCACGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAG
AAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGG
GCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCC
GTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAA
GAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGAC
GGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTG
ATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGG
TCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATT
AACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGT
GCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGTGAAAGATCCGCGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGCCGAACATCCCG
CAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCG
AGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGGCTGGAAGTTCTGTTCCAGGGGCCCCTGGGATCCAGTGCCTGCCCGTCCGGTGCGACGTGC
GGCTCCTATACCGTGGGTGGATTGGGCAGCCGCAAACAGCAAGTTCGCAATGCAGGTGGCTCCAGTCTGGATCTGGCGGTAGCGATGCTGGAAACGGAACGT
ATGGATACCGCGTATCCGTATGGGGATAACAAATCGGGAGACGCTGCAAATTTTGGCATTTTCAAACAAAATTGGTTAATGCTGCGCAGCGCTTGCGCGCAG
TTTGGCGGCCAGGGCGCCGGCCAATATGATAACGGTGCGGCCCTGAACTCCAGTCTGGGCCAGGACGTCAGCTGTCTGCACCAATCTCAGTCCCACTACGGC
TTAGATGCTTGGTTCGCGGGTCACCGCAATGGCGCTAGCGGCTTAAGTTCTCCCAATACTGCGGACATCGCCGCCTATAAAGCAGCGGTGTATTGGATCAAA
GCTCAGCTGGATGCGGATTCAGCGAACTTAGGAAATGATACCCGTTTCTGGGTTCAAGTTCCAGCTATTTGACTCGAGAAGCTTCAAATAAAACGAAAGGCT
CAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAA
CGGCCCGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACT
CTTTCGGTCCGTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGT
ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCT
GAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTCCCAATGATGAGC
ACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAG
TACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTT
CTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCC
ATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAA
TTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGT
GGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAAT
AGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTCCTTAGGACTGAGCGT
CAACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTG
TTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCA
CCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGA
CTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA
CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGA
GCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCT
ATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAA
CCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTA
TTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATAAGGTGCACTGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGC
CCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCG
CGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGCAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTA
ATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGA
TACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGC
GGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCC
GGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGC
AGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGA
TCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAAATT
```

Escherichia coli BL21 (DE3) transformed with pHisMALP-*Ss*GH134 were grown in 1000 ml LB media with shaking (200 rpm) at 37 °C (100 μg ml^{-1} ampicillin) until the culture reached an OD_{600} value of 0.8. The culture was cooled to room temperature, IPTG added to a final concentration of 200 µM, and then incubated with shaking (200 rpm) at 18 °C for 16 h. Cells were harvested by centrifugation (17,000 g, 20 min, 4 °C) and resuspended in 40 ml of binding buffer (50 mM NaP_i, 500 mM NaCl, 5 mM imidazole, pH 7.5) containing EDTA-free protease inhibitor cocktail and lysozyme (0.1 mg/ml) by incubating at 4 °C for 30 min. Benzonase (1 ul) was added to the mixture then lysis was effected by sonication. The lysate was centrifuged (17,000 g, 20 min, 4 °C) and the supernatant collected. The supernatant was filtered (0.45 µm) and then subjected to immobilized metal-ion affinity chromatography. Fractions containing product (as determined by SDS-PAGE) were combined and further purified by size exclusion chromatography (GE HiLoad *16/600 Superdex 200*) using 50 mM sodium phosphate, 150 mM NaCl, pH 7.5 buffer. The protein obtained was estimated to be >95% pure by Coomassie-stained SDS-PAGE. Protein concentration was determined by Bradford assay. The yield of MBP-*Ss*GH134 fusion protein was >50 mg l^{-1}.

For crystallography, the MBP fusion partner was removed by incubating MBP-*Ss*GH134 (5 mg ml^{-1}) with HRV 3C protease (50 µg ml^{-1}) in PBS with 1 mM EDTA and 1 mM DTT at 4 °C for 16 h. The liberated *Ss*GH134 was purified by size exclusion chromatography (GE HiLoad *16/600 Superdex 75*) using 50 mM sodium phosphate, 150 mM NaCl, pH 7.5 buffer.

Mutagenesis of *Ss*GH134

Site-directed mutagenesis of *Ss*GH134 to generate the general base and acid variants E45Q and D57N was accomplished using a PCR approach with the following oligonucleotide primers:

```
E45Q_f: GTAGCGATGCTGCAAACGGAACGTATG
E45Q_r: CATACGTTCCGTTTGCAGCATCGCTAC
D57N_f: CGTATCCGTATGGGAATAACAAATCGG
D57N_r: CCGATTTGTTATTCCCATACGGATACG
```

The mutagenized vectors were verified by sequencing, and the mutants were expressed and purified as for wild-type *Ss*GH134.

Preparation of β-1,4-mannohexaosyl benzoylhydrazine (M6-benzoylhydrazine)

Mannohexaose (4 mg, 0.004 mmol) was refluxed in H_2O (1 ml) with benzoyl hydrazide (0.8 mg, 0.006 mmol) and acetic acid (25 µl of a 0.5% solution in water) for 2.5 h. The solution was evaporated to dryness. Reverse phase chromatography (Alltech Prevail C18-silica), eluted with water gave M6-benzoylhydrazide (3 mg, 68%); ^{1}H NMR (500 MHz, $CDCl_3$): δ 3.46-3.42 (44H, m, 5 × H1, 6 × H2, 6 × H3, 6 × H4, 6 × H5, 12 × H6, 4 × NH), 4.44 (1H, s, H1β), 4.75 (1H, s, H1α), 7.54-7.89 (12H, m, 2 × Ph); HRMS $(ESI)^+$ *m/z* 1109.3818 [$C_{43}H_{68}N_2O_{31}$ $(M+H)^+$ requires 1109.3879].

Isotope-mapping of *Ss*GH134 catalyzed cleavage of β-1,4-mannopentaose

To determine the regiospecificity of mannooligosaccharide cleavage by *Ss*GH134, ^{18}O incorporation into the products through ^{18}O-water was determined by mass spectrometry. Solutions of 0.5 mg of mannotetraose (M4), mannopentose (M5), and mannohexose (M6) respectively, were each dissolved into 50 µL of $H_2{}^{18}O$. 1 µl of *Ss*GH134 (25 mg ml^{-1} in 50 mM NaPi and 150 mM NaCl at pH 7.5) was added to each solution. After 30 min the solution was analyzed using an Agilent ESI-TOF mass spectrometer using electrospray ionisation in positive ion mode.

X-ray data collection, processing and structure solution

Well-diffracting crystals of *Ss*GH134 were obtained by mixing 15 mg ml^{-1} protein stock with an equal volume of precipitant composed of 18-22% (v/v) PEG6000, 0.15-0.2 M $CaCl_2$, and HEPES-NaOH, pH 7.0 after 1-2 days at 20 °C using the sitting drop vapor diffusion method. The general base and acid variants E45Q and D57N were crystallized under the same conditions as for the native enzyme. The wt-M3 and E45Q-M5 complexes were produced using the soaking method. Drops containing crystals of wildtype or E45Q mutant were supplemented with 10 mM M3 or 100 mM M5 in the same precipitant solution for 1 h at 20 ºC before collecting and freezing crystals. 20% glycerol was used as cryoprotectant for the crystals before they were flash frozen in liquid nitrogen. Diffraction data were collected at 100 K on beamline I02 of the Diamond Light Source and were processed using the *xia2* implementation of XDS[1] and AIMLESS from CCP4 suite.[2] The structure was determined by molecular replacement with an 8-residue ideal alpha-helical fragment placed using Phaser.[3] Phases calculated from the placed fragment were improved by density modification with ACORN[4] using normalized structure factors extended to 1.0 Å resolution. The resulting phases were of excellent quality and a complete model was built with ARP/wARP[5], with maximum likelihood refinement of the protein model using numerous cycles of REFMAC[6,7] and manual correction using COOT.[8] The statistics of the data processing and structure refinement are listed in Supplementary Information **Table S1**.

Enzyme kinetics

Reactions were carried out in a total volume of 1 ml with 200 µM M6 at 37 °C in 20 mM MOPS buffer, pH 5.0. Various enzyme concentrations (1 µM, 100 nM and 10 nM) were used in order to choose the optimal enzyme concentration for the assay. Assays were carried out with 200 µM M6 and 100 nM enzyme, incubated for 25 min and four 100 µl aliquots were taken at the times indicated. Each aliquot was heat inactivated by boiling for 10 min. Following inactivation, the samples were diluted 2-fold and subjected to HPLC analysis. Each reaction contained an internal standard of fucose (50 µM) and all data obtained was normalized to this standard using a standard curve.

Thin Layer Chromatography (TLC)

TLC plates (Silica gel 60, 20 × 20, Merck) were cut to 10 cm in height. 2 µL of samples were spotted on the plate, separated by 10 mm. Solvent (50 ml) comprising freshly made 1-butanol/acetic acid/water (2:1:1, v/v) was poured into a glass chromatography tank (23 × 23 × 7.5) and covered tightly. Vapors were allowed to equilibrate for at least 2 h before use. The loaded TLC plate was placed into the tank and samples allowed to migrate until the running buffer reached approximately 1 cm from the top of the plate. The plate was dried gently using a hairdryer and put back in the tank and eluted a second time. The plate was dried again and immersed for a few seconds in orcinol/sulfuric acid reagent (sulfuric acid/ethanol/water 3:70:20 v/v, 1% orcinol), dried carefully and heated until sugars were developed, at 120 °C (5-10 min). Standards consisting of known monosaccharides and oligosaccharides were spotted on the TLC plate.

Computational Methods

Classical and QM/MM molecular dynamics simulations

The initial structure for the simulations was taken from the present reported structure of *Ss*GH134 in complex with mannopentaose (PDB 5JUG). To simulate the wild type enzyme, the mutation of the acid residue (E45Q) was manually reverted (changing atom N by O without modifying its orientation). The protonation states and hydrogen atom positions of all amino acid residues were taken according to protein environment. A total number of 12.102 water molecules were added to within a radius of 15 Å from the protein and one sodium ion was added to neutralize the enzyme charge. Molecular dynamics (MD) simulations were performed using Amber11 software.[9] The protein was modeled using the FF99SB force field.[10] The carbohydrate substrate and water molecules were described with the GLYCAM06[11] and TIP3P[12] force fields, respectively. The MD simulation was carried out in several steps. First, the system was minimized, holding the protein and substrate fixed, followed by energy minimization on the entire system. To gradually reach the desired temperature, weak spatial constraints were initially added to the protein and substrate, while water molecules and the sodium ion were allowed to move freely at 100 K. The constraints were then removed and the working temperature of 300K was reached after two more 100 K heating steps in the NVT ensemble. Afterwards, the density was converged up to water density at 300 K in the NPT ensemble and the simulation was extended to 25 ns in the NVT ensemble. The system reached equilibrium according to the root mean squared deviation of enzyme backbone (**Figure S10A**). The two catalytic residues (E45 and D57) kept their orientation during the simulation, with a water molecule properly oriented for nucleophilic attack (**Figure S10B**). Interestingly, the water molecule was replaced by another one at ≈ 3.5 ns, suggesting that the active site is highly dynamic. Analysis of the trajectory was carried out using standard tools of AMBER and VMD.[13]

QM/MM MD simulations were performed using the method developed by Laio *et al.*,[14] which combines Car–Parrinello MD,[15] based on Density Functional Theory (DFT), with force-field MD methodology. In this approach, the system is partitioned into quantum mechanics (QM) and molecular mechanics (MM) fragments. The dynamics of the atoms on the QM fragment depend on the electronic density, ρ(r), computed with Density

Functional Theory, whereas the dynamics of the atoms on the MM fragment are ruled by an empirical force field. The QM/MM interface is modeled by the use of link-atom that saturates the QM region. The electrostatic interactions between the QM and MM regions were handled via a fully Hamiltonian coupling scheme,[14] where the short-range electrostatic interactions between the QM and the MM regions are explicitly taken into account for all atoms. An appropriately modified Coulomb potential was used to ensure that no unphysical escape of the electronic density from the QM to the MM region occurs. The electrostatic interactions with the more distant MM atoms were treated via a multipole expansion. Bonded and van der Waals interactions between the QM and the MM regions were treated with the standard AMBER force-field. Long-range electrostatic interactions between MM atoms were described with the P3M implementation,[16] using a 64 x 64 x 64 mesh. A large QM region including the mannose rings at the –1, +1 and +2 subsites and half ring of the saccharide at the -2 subsites and the catalytic residues (E45 and D57), leading a total number of 98 QM atoms (including capping hydrogens; **Figure S11A**) and 38.693 MM atoms for the system. The QM region was enclosed in an isolated supercell of size 20.1 x 17.7 x 20.9 $Å^3$. Kohn–Sham orbitals were expanded in a planewave basis set with a kinetic energy cutoff of 70 Ry. Norm-conserving Troullier–Martins *ab initio* pseudopotentials[17] were used for all elements. The calculations were performed using the Perdew, Burke and Ernzerhoff generalized gradient-corrected approximation (PBE).[18] This functional form has been proven to give a good performance in the description of hydrogen bonds[19] and was already used with success in previous works on glycoside hydrolases and transferases.[20] A fictitious electronic mass of 700 au and a timestep of 5 au was used to ensure an adiabaticity of $4.12 \cdot 10^{-5}$ $a.u \cdot ps^{-1} \cdot atom^{-1}$ for the fictitious kinetic energy.

Metadynamics simulations

The free energy landscape (FEL) of the reaction was explored using the metadynamics approach with three collective variables (CVs). We used the metadynamics driver provided by the Plumed2 plugin.[21] The first collective variable (CV_1) was defined as the difference between the O_{wat}-C1 and the C1-O distances. This variable accounts for the nucleophilic attack of the water molecule and the cleavage of the glycosidic bond. The second collective variable (CV_2) was defined as the distance difference between the O_{wat}-H and H-O_{D57}. This variable accounts for proton transfer between D57 and the water molecule. Finally, CV_3 was defined as the distance difference of O_{E45}-H and H-$O_{glycosidic}$, which thus accounts for the transfer of the E45 proton to the glycosidic oxygen atom. A hill height of 1 kcal/mol, deposition time of 30 fs (250 MD steps). The three-dimensional FES was completed after 552 deposited Gaussians. Enlarging the simulation leads to a relaxation of the –1 subsite mannose to a $^1C_4$ conformation, as observed experimentally.

References

1. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **2010**, *43*, 186-190
2. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760-763
3. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658-674
4. Jia-xing, Y., Woolfson, M. M., Wilson, K. S., and Dodson, E. J. A modified ACORN to solve protein structures at resolutions of 1.7 A or better. *Acta Crystallogr. D, Biol. Crystallogr.* **2005**, *61*, 1465-1475
5. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **2008**, *3*, 1171-1179
6. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **2011**, *67*, 235-242
7. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D* **1997**, *53*, 240-255
8. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D* **2010**, *66*, 486-501
9. Case, D. A., Darden, T. A., Cheatham, T. E., and C. Simmerling, J. W., R. Duke, R. Luo, M. F. Crowley, R. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. E. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. Kollman. (2010) AMBER11. University of California, San Francisco
10. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712-725
11. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622-655
12. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174
13. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926-935
14. Humphrey, W., Dalke, A., and Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38
15. Laio, A., VandeVondele, J., and Rothlisberger, U. A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. *J. Chem. Phys.* **2002**, *116*, 6941-6947
16. Car, R., and Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55*, 2471-2474
17. Troullier, N., and Martins, J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **1991**, *43*, 1993-2006
18. Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868
19. Ireta, J., Neugebauer, J., and Scheffler, M. On the Accuracy of DFT for Describing Hydrogen Bonds: Dependence on the Bond Directionality. *J. Phys. Chem. A* **2004**, *108*, 5692-5698
20. Ardèvol, A., and Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 7528-7547
21. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604-613