Supplementary Information for:

The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases

Kristian E. H. Frandsen,[1] Thomas J. Simmons,[2] Paul Dupree,[2] Jens-Christian N. Poulsen,[1] Glyn R. Hemsworth,[3] Luisa Ciano,[3] Esther M. Johnston,[3] Morten Tovborg,[4] Katja S. Johansen,[4,5] Pernille von Freiesleben,[4] Laurence Marmuse,[6] Sébastien Fort,[6] Sylvain Cottaz,[6] Hugues Driguez,[6] Bernard Henrissat,[7,8,9] Nicolas Lenfant,[7,8] Floriana Tuna,[10] Amgalanbaatar Baldansuren,[10] Gideon J. Davies,[3] Leila Lo Leggio[1,*] and Paul H. Walton[3,*]

1. Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100, Copenhagen Ø, Denmark.

2. Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

3. Department of Chemistry, University of York, York YO10 5DD, U.K.

4. Novozymes A/S, Krogshoejvej 36, 2880 Bagsvaerd, Denmark.

5. Current address: Division of Industrial Biotechnology, Chalmers University of Technology, Kemivägen 10, SE-412 96 Göteborg, Sweden.

6. Université de Grenoble Alpes, CERMAV, F-38000 Grenoble, France. CNRS, CERMAV, F-38000 Grenoble, France.

7. Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, 13288 Marseille, France.

8. INRA, USC 1408 AFMB, 13288 Marseille, France.

9. Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

10. EPSRC National EPR Facility, School of Chemistry and Photon Science Institute, University of Manchester, Oxford Road, Manchester M13 9PL, U.K.

Supplementary Results

Supplementary Table 1: Selected distances around the Cu sites in *Ls*(AA9)A and *Ta*(AA9)A for comparison. All distances are in Å.

	Unbound		G3 bound		G6 bound	Unbound
	Cu(I) *Ls*(AA9)A_highres	Cu(II) *Ls*(AA9)A_lowdose	Cu(I) *Ls*(AA9)A-G3	Cu(II) *Ls*(AA9)A-G3_lowdose	Cu(I) *Ls*(AA9)A-G6	PDB code 3ZUD *Ta*(AA9)A
Cu—NH_2(His1)	2.2	2.2	2.3	2.2	2.3	2.2
Cu—N_δ(His1)	1.9	1.9	1.8	1.9	1.9	1.9
Cu—N_ε(His78)	2.0	2.1	2.0	2.1	2.0	2.0
Cu—O(Tyr164)	2.8	2.7	2.7	2.5	2.6	2.9
Cu—X	3.9 and 2.1	2.2	2.8 and 4.1	2.3	2.4 and 4.0	2.1
Cu…$OH_{2,(ax)}$	3.3	2.8				2.9
$OH_{2,(ax)}$…$OH_{2,pocket}$	3.0	3.2				
$OH_{2,pocket}$…O=C(His1)	2.8	2.8	2.7	2.7	2.8	
$OH_{2,pocket}$ …NH_2(His1)	3.1	3.0	3.1	2.9	3.0	
$OH_{2,pocket}$ …H-N-CO(Ala75)	2.8	2.9	2.8	2.9	3.0	
OH_2…HO-$CH_{2(glucose)}$			2.8	2.8	2.7	

Supplementary Table 2: Potential hydrogen bonding distances (within 3.2 Å distance) in *Ls*(AA9)A-cellooligosaccharides complex structures

		***Ls*(AA9)A-G6**		***Ls*(AA9)A-G3**		***Ls*(AA9)A-G3_lowdose**	
Subsite	Glycosidic atom	Residue (atom)	Distances (Å)	Residue (atom)	Distances (Å)	Residue (atom)	Distances (Å)
GLC+2	O(2)	Asn28 ($N\delta_2$)	2.8	Asn28 ($N\delta_2$)	2.7	Asn28 ($N\delta_2$)	2.8
	O(2)	Asn67 ($O\delta_1$)	2.7	Asn67 ($O\delta_1$)	2.7	Asn67 ($O\delta_1$)	2.5
	O(3)	His66 ($O\varepsilon_2$)	2.7	His66 ($O\varepsilon_2$)	2.8	His66 ($O\varepsilon_2$)	2.7
	O(6)	-	-	H_2O (Asn29($O\delta_1$)	3.1	H_2O (Asn29($O\delta_1$)	3.0
GLC+1	O(6)	H_2O_{pocket}	2.7	H_2O_{pocket}	2.8	H_2O_{pocket}	2.8
GLC–1	O(2)	Ser77 ($O\gamma$)	2.7	Ser77 ($O\gamma$)	2.8	Ser77 ($O\gamma$)	2.6
	O(4)	-		H_2O (Glu148($O\varepsilon_1$))	2.3	H_2O (Glu148($O\varepsilon_1$))	2.5
	O(6)	H_2O_{352} (Glu148($O\varepsilon_1$))	3.1				
GLC–2	O(2)	Glu148 ($O\varepsilon_1$)	2.4	-	-	-	-
	O(3)	Arg159 ($N\omega_2$)	3.1	-	-	-	-
GLC–3	-	-	-	-	-	-	-
GLC–4	-	-	-	-	-	-	-

Supplementary Table 3: Torsion angles in bound cellooligosaccharides (model values from [1])

Glycosidic torsion angles		***Ls*(AA9)A-G6**	***Ls*(AA9)A-G3**	***Ls*(AA9)A-G3_lowdose**
Subsite +1/+2 (°)	ϕ	-91.7	-92.1	-91.6
	ψ	95.9	86.2	89.8
Subsite -1/+1 (°)	ϕ	-90.4	-88.9	-87.0
	ψ	101.7	98.8	88.7
Subsite -2/-1 (°)	ϕ	-77.3	-	-
	ψ	105.6	-	-
Subsite -3/-2 (°)	ϕ	-111.3	-	-
	ψ	130.3	-	-
Subsite -4/-3 (°)	ϕ	-96.9	-	-
	ψ	111.4	-	-
Values from model [1] (°)	ϕ	-88.9		
	ψ	95.0		
Definitions	ϕ	$O_{5'} - C_{1'} - O_4 - C_4$		
	ψ	$C_{1'} - O_4 - C_4 - C_3$		

Supplementary Table 4: Crystallization and soaking conditions

Crystallization Conditions	***Ls*(AA9)A_highres**	***Ls*(AA9)A_lowdose**	***Ls*(AA9)A-G6**	***Ls*(AA9)A-G3**	***Ls*(AA9)A-G3_lowdose**
Protein concentration	8.5 mg/mL	19.2 mg/mL	19.2 mg/mL	19.2 mg/mL	19.2 mg/mL
Protein buffer	20 mM acetate pH 5.5	20 mM acetate pH 5.5	20 mM acetate pH 5.5	20 mM acetate pH 5.5	20 mM acetate pH 5.5
Pre-incubation [Cu(II)acetate] Time	1.0 mM 1 hour	1.4 mM 30 min	1.4 mM 30 min	1.4 mM 30 min	1.0 mM 45 min
Precipitant concentration	3.0 M NaCl	3.9 M NaCl	3.6 M NaCl	3.0 M NaCl	3.6 M NaCl
Reservoir buffer	100mM citric acid pH3.5	100mM citric acid pH 4.0	100mM citric acid pH4.0	100mM citric acid pH3.5	100mM citric acid pH4.0
Drop volume and ratio (Prot:Res:H_2O)	0.5 µL 3:1:1	0.4 µL 3:1:0	0.4 µL 3:1:0	0.5 µL 3:1:1	0.4 µL 3:1:0
Soaking Conditions	***Ls*(AA9)A_highres**	***Ls*(AA9)A_lowdose**	***Ls*(AA9)A-G6**	***Ls*(AA9)A-G3**	***Ls*(AA9)A-G3_lowdose**
pH equilibration	Crystal transferred to 4.0 M NaCl 100 mM Citric acid pH 5.5 for 2 hours	0.5 µL reservoir added (3.9 M NaCl 100 mM Citric acid pH5.5) – 1 hour equilibration	Crystal transferred to 3.8 M NaCl 100 mM Citric acid pH 5.5 for 1 hour	Crystal transferred to 3.6 M NaCl 100 mM Citric acid pH 5.5 for 1 hour	Crystal transferred to 3.9 M NaCl 100 mM Citric acid pH 5.5 for 10 min
Soaking Conditions			Saturated G6 added to the drop. Soaked for 20 min	0.3 M G3added to final conc. of 150 mM. Soaked for 10 min	0.3 M G3 added to final conc. of 100 mM. Soaked for 10 min

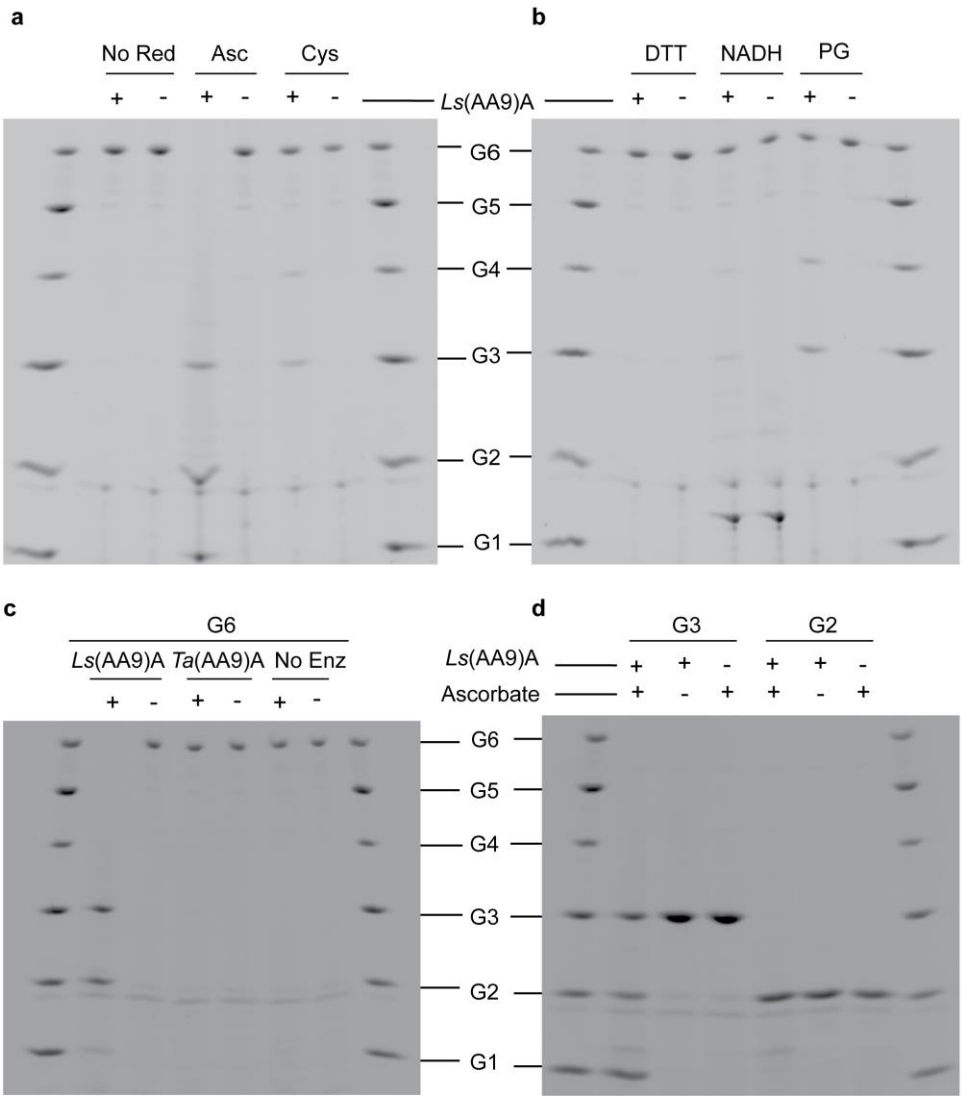
Supplementary Table 5: Crystallographic data collection and refinement statistics (each determination is from a single crystal of the species under study).

	***Ls*(AA9)A_highres**	***Ls*(AA9)A_lowdose**	***Ls*(AA9)A-G6**	***Ls*(AA9)A-G3**	***Ls*(AA9)A-G3_lowdose**
Data collection					
Space group	$P4_132$	$P4_132$	$P4_132$	$P4_132$	$P4_132$
Cell dimensions					
a, *b*, *c* (Å)	124.70	125.23	125.46	125.44	124.71
α, β, γ (°)	90	90	90	90	90
Resolution (Å)	41.57 – 1.28 (1.32 – 1.28)	20.00 – 1.91 (1.96-1.91)	20.00 – 1.75 (1.80 – 1.75)	30.00 – 1.70 (1.75 – 1.70)	30.00 – 1.80 (1.85 – 1.80)
R_{meas}(%)*	7.8 (108.4)	22.1 (105.1)	20.0 (184.6)	5.7 (44.1)	15.7 (91.2)
$I/\sigma I$	13.9 (1.4)	7.38 (1.64)	21.72 (2.59)	36.67 (6.16)	7.27 (1.38)
Completeness (%)	99.8 (99.3)	99.2 (98.2)	99.9 (99.9)	93.4 (98.1)	81.1 (80.5)
Redundancy	7.3 (4.4)	5.37 (5.23)	35.5 (31.4)	34.6 (19.4)	2.63 (2.68)
Refinement					
Resolution (Å)	41.57 – 1.28	20.00 – 1.91	20.00 – 1.75	30.00 – 1.70	30.00 – 1.80
No. reflections	80784	25178	32810	33655	24019
R_{work}/ R_{free}	11.75 / 14.71	17.76 / 20.85	16.47 / 19.55	19.29 / 22.69	17.06 / 21.25
No. atoms					
Protein§	1927	1812	1824	1826	1810
Ligand/ion	- / 16	- / 8	67 / 7	34 / 6	34 / 11
Water	548	366	416	412	361
B-factors					
Protein§	16.501	14.911	16.514	24.411	14.006
Ligand/ion	- / 29.358	- / 22.971	40.005 / 25.139	32.909 / 32.575	18.861 / 20.051
Water	34.972	24.863	28.177	35.294	22.570
R.m.s deviations					
Bond lengths (Å)	0.0207	0.0134	0.0139	0.0164	0.0162
Bond angles (°)	1.971	1.554	1.667	1.759	1.761

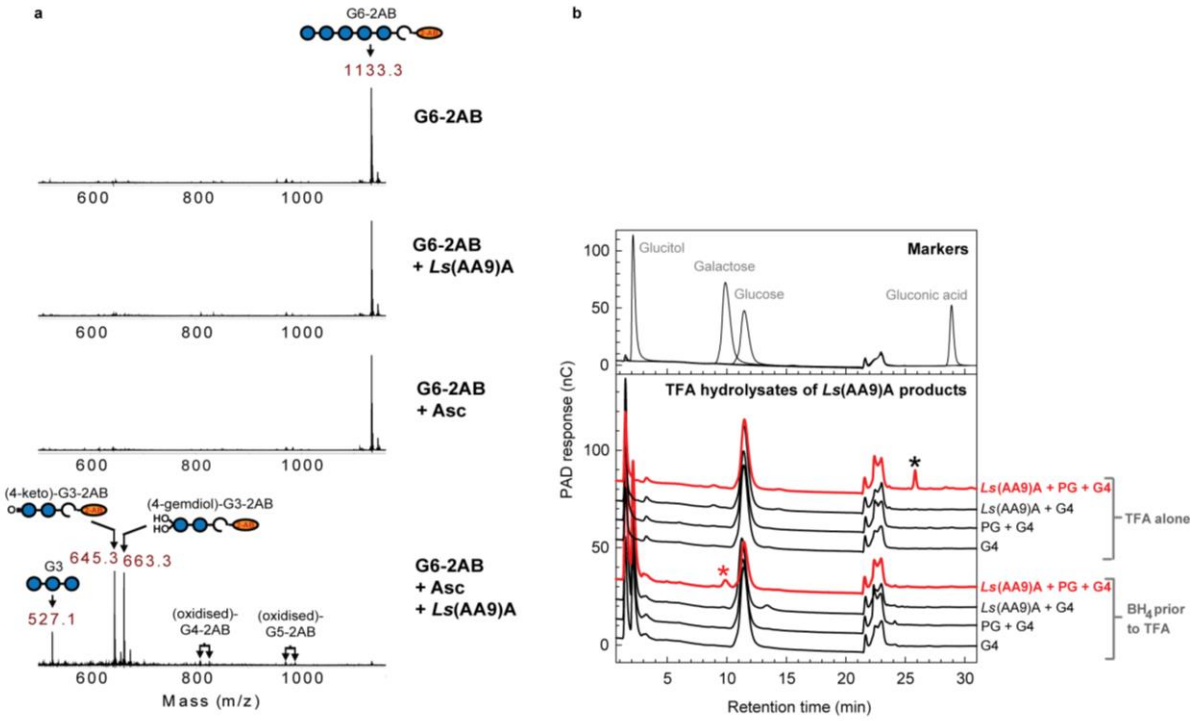
*Highest resolution shell is shown in parenthesis.
§Glycosylation (a single N-acetylglucosamine unit) and the active site copper are included in "Protein"
*R_{meas} is used instead of R_{merge} or R_{sym} as it is an improved, redundancy-independent *R*-factor.[2]

Supplementary Table 6: EPR spin Hamiltonian parameters from simulations of X Band cw spectra for *Ls*(AA9)A and *Ls*(AA9)A-G6.

		***Ls*(AA9)A-H_2O**	***Ls*(AA9)A-Cl**	***Ls*(AA9)A-G6-Cl**	***Ls*(AA9)A-G6-H_2O**
	g_x	2.044	2.01	2.032	2.041
***g* values**	g_y	2.085	2.09	2.063	2.069
	g_z	2.279	2.25	2.234	2.273
	A_x	58	20	10	20
A_{Cu} (/MHz)	A_y	78	40	69	38
	A_z	458	455	517	515
SHF A_N (isotropic) (/MHz)		37, 34		36, 31, 19	36, 30, 19
SHF A_{Cl} (/MHz)		---	---	$A_\perp$ = 43 $A_\parallel$ = 40	---
A_{Cu} strains (/MHz)		56, 90, 90	40, 40, 80	15, 35, 70	25, 38, 60
Line widths		0.2, 0.7	0.5, 0.4	0.1, 0.1	0.2, 0.2



Supplementary Figure 1: *Ls*(AA9)A specificity analysis by PACE. a,b, Small molecule reductant-specificity of *Ls*(AA9)A on G6. No Red, incubation in absence of added reducing agent; Asc, ascorbate; Cys, cysteine; DTT, dithiothreitol; NADH, nicotinamide adenine dinucleotide; PG, pyrogallol; +, incubation with *Ls*(AA9)A; –, incubation without *Ls*(AA9)A. Ascorbate was the best reductant tested, with pyrogallol and cysteine also functional. **c**, Comparison of *Ls*(AA9)A and *Ta*(AA9)A activities on G6. **d**, Cleavage of G3 but not G2 (right) by *Ls*(AA9)A. +, incubation with ascorbate (or *Ls*(AA9)A for right picture); –, incubation without ascorbate (or *Ls*(AA9)A for right picture).



Supplementary Figure 2: *Ls*(AA9)A product analysis. a, Cleavage of G6-2-aminobenzoic acid (reducing terminus)**.** MALDI spectra showing products of incubation of G6-2AB derivative ± *Ls*(AA9)A and ± ascorbate. *Ls*(AA9)A cleavage of this substrate yielded oxidised G3-2AB products and as well as non-oxidised G3, indicating that oxidation occurs on the non-reducing termini of these products. **b**, *Ls*(AA9)A produces C4 oxidation products**.** Products of *Ls*(AA9)A digestion of G4 were either TFA-hydrolysed or borohydride (BH_4)-reduced then TFA hydrolysed. Products were analysed by HPAEC. Positive reactions (± borohydride-reduction) are in red; negative controls are in black. Asterisked peaks: red, galactose from borohydride-reduced '*Ls*(AA9)A, pyrogallol (PG) and cellotetraose' reaction only – indicative of C4-oxidation; black, probably 4-keto glucose or derivative. Peaks between 21 and 24 min are caused by a change in the mobile phase. The presence of galactose in TFA hydrolysates of borohydride-reduced *Ls*(AA9)A products confirmed C4-oxidation. We did not observe gluconic acid in the hydrolysates, further supporting the absence of reducing terminal C1 oxidation products.

Supplementary Note 1: Synthesis and chemical characterization of intermediate compounds and final product during synthesis of FRET substrate

Chemicals were purchased from Sigma-Aldrich Chimie (Saint Quentin-Fallavier, France). Reactions were monitored by thin layer chromatography (TLC) using Silica Gel 60 F254 pre-coated plates (E. Merck, Darmstadt). Detection of carbohydrates was achieved by charring with the sulfuric acid/methanol/H_2O (1:1:0.5 *v/v*).

2,3,4-Tri-*O*-acetyl-6-*O*-levulinyl-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-α-D-glucopyranosyl fluoride (**2**): Hepta-*O*-acetyl α-lactosyl fluoride **1** [3] (955 mg, 1.05 mmol) was suspended in 2-methyl-2-butanol (36 mL) in presence of *Candida antarctica* lipase (Novozyme 435) (1g) and trifluoroethyl levulinate [4] (3.1 mL) was added. The reaction was placed on a rotative shaker for 3 d at 45~50 °C. The reaction was filtered, the filtrate evaporated and the residue was purified by flash column chromatography (dichloromethane/methanol, 10:0→9:1 v/v). The expected product was acetylated (acetic anhydride/pyridine, 1:10 v/v, 9 mL) in the presence of catalytic amount of 4-dimethylaminopyridine (DMAP). After 3 h of stirring at room temperature, the reaction mixture was evaporated *in vacuo* and co-evaporated with toluene. The crude product was purified by flash column chromatography (toluene/ethyl acetate, 1:1→2:3 *v/v*) to give fluoride **2** as a white solid (350 mg, 48%).

2,3,4-Tri-*O*-acetyl-6-*O*-methanesulfonyl-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-α-D-glucopyranosyl fluoride (**3**): Fluoride **2** (500 mg, 0.72 mmol) was suspended in ethanol (20 mL) and hydrazine acetate (140 mg, 1.55 mmol) was added. The reaction was stirred at room temperature for 2 h and neutralized with Et_3N. The reaction was concentrated *in vacuo* and the residue was diluted with CH_2Cl_2 and washed with a sat. aq NaCl solution and H_2O. The crude product was dissolved in pyridine (12 mL) in presence of catalytic amount of DMAP. Methanesulfonyl chloride (240 µL, 1.44 mmol) was added the mixture was stirred for 2 h at room temperature. The solution was evaporated *in vacuo* and co-evaporated twice with toluene. The residue was purified by flash column chromatography (toluene/ethyl acetate, 7:3→3:2 *v/v*) to give mesylate **3** as white foam (440 mg, 90%).

2,3,4-Tri-*O*-acetyl-6-azido-6-deoxy-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-α-D-glucopyranosyl fluoride (**4**): Compound **3** (380 mg, 0.56 mmol) was dissolved in dimethylformamide (DMF) (8 mL), NaN_3 (610 mg, 7.84 mmol) and 18-crown-6 (65 mg) were added. The reaction was heated at 80 ^oC for 4 d and concentrated *in vacuo*. The residue was diluted with CH_2Cl_2 and washed with a sat. aq. NaCl solution then H_2O. The crude product was purified by flash column chromatography (toluene/ethyl acetate, 1:1→2:3 *v/v*) to give compound **4** as white foam (206 mg, 59%).

6-Azido-6-deoxy-β-D-galactopyranosyl-(1→4)-α-D-glucopyranosyl fluoride (5): The peracetylated fluoride **4** (183 mg, 0.29 mmol) was treated with NaOMe (300 µL, 1M) in MeOH (7 mL) at 0 °C. The reaction mixture was stirred for 2 h and neutralized with Amberlite IR 120 H^+. After filtration of the resin and evaporation under reduced pressure the residue was freeze-dried to give the fluoride **5** (103 mg, 96%). This compound was used immediately without any further characterization.

Sodium *N*-[2-*N*[(S-(6-azido-6-deoxy-β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-β-D-glucopyranosyl)-(1→4)-β-D-glucopyranosyl)-2-thioacetyl]aminoethyl]-1-

naphthylamine-5-sulfonate (**7**)**:** Cel7B Glu197Ala glycosynthase[5] (0.62 mg) was added to a solution of fluoride **5** (31 mg, 0.083 mmol) and sodium *N*-[2-*N*[(S-(β-D-glucopyranosyl)-(1→4)-β-D-glucopyranosyl)-2-thioacetyl]aminoethyl]-1-naphthylamine-5-sulfonate[6] **6** (51 mg, 0.076 mmol) in sodium phosphate buffer (1.5 mL, 0.1M, pH 7). The solution was placed in a rotative shaker for 4 h at 37°C. Then the reaction was evaporated and the crude product was purified on octadecyl reversed phase silica cartridge (H_2O/MeOH, 1:0→19:1 *v/v*) to give tetrasaccharide **7** as an amorphous white solid (61 mg, 79%).

Sodium N-[2-N[(S-(6-deoxy-6-(4-((4-(dimethylamino)phenyl)azo)phenylthioureido-β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-β-D-glucopyranosyl)-(1→4)-β-D-glucopyranosyl)-2-thioacetyl]aminoethyl]-1-naphthylamine-5-sulfonate (**8**) (F*-G4-F): A solution of azido tetrasaccharide **7** (50 mg, 49 μmol) in pyridine and H_2O (1:1 v/v, 10 mL) was saturated with H_2S. The reaction was stirred at room temperature overnight. After evaporation under reduced pressure, the residue was dissolved in DMF (10 mL) and aq sodium hydrogenocarbonate solution (6 mL, 0.35M). Then 4-(4-isothiocyanatophenylazo)-N,N-dimethylaniline (DABITC) (22 mg, 78 μmol) was added. The reaction was stirred for 12 h at 40°C. The solution was evaporated under reduced pressure and the residue was purified on octadecyl reversed phase silica cartridge (H_2O/MeOH, 1:0→19:1 *v/v*). Appropriate fractions were pooled, concentrated *in vacuo* and freeze dried to give the title compound (F*-G4-F) as amorphous orange solid (45 mg, 69%).

Characterizations

The mass measurements were performed using Nermag R10-10C (Desorption-Chemical Ionisation-DCI MS, or Waters Micromass ZQ (Electrospay-ESI-MS) spectrometers (PSM facility, PCN-ICMG, Grenoble). NMR spectra were recorded on a Bruker AV400 spectrometer at 350 K. Chemical shifts (in ppm) were determined relative to deuterated solvent as reference. Coupling constant(s) in hertz (Hz) were measured from one-dimensional spectra and multiplicities were abbreviated as following: s(singlet), d (doublet), t (triplet), m (multiplet). In NMR data, H-1^{I} and C-1^{I} refers respectively to anomeric proton and carbon at the reducing unit. Optical rotations were measured with a Perkin-Elmer 341 polarimeter. Synthetic scheme is presented in **Supplementary Figure 3**.

2,3,4-Tri-*O*-acetyl-6-*O*-levulinyl-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-α-D-glucopyranosyl fluoride (2)

$[\alpha]_D$ (deg cm^3 g^{-1} dm^{-1}) = + 40 (*c* = 1.1 g cm^{-3} in $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ 5.62 (dd, $J_{1,2}$ = 2.7 Hz, $J_{1,F}$ = 53 Hz, H-1, 1H), 5.44 (t, $J_{3,4}$ = 9.6 Hz, H-3, 1H), 5.30 (d, $J_3{}^{II}{}_{,4}{}^{II}$ = 3.4 Hz, H-4^{II}, 1H), 5.06 (dd, $J_1{}^{II}{}_{,2}{}^{II}$ = 7.9 Hz, $J_2{}^{II}{}_{,3}{}^{II}$ = 10.4 Hz, H-2^{II}, 1H), 4.91 (dd, $J_3{}^{II}{}_{,4}{}^{II}$ = 3.5 Hz, $J_2{}^{II}{}_{,3}{}^{II}$ = 10.3 Hz, H-3^{II}, 1H), 4.79 (ddd, $J_{1,2}$ = 2.7 Hz, $J_{2,F}$ = 10.2 Hz, $J_{2,3}$ = 13.0 Hz, H-2, 1H), 4.51-4.43 (m, H-1^{II}, H-6a/H-6^{II}a, 2H), 4.14-4.03 (m, H-5, H-6a/H-6^{II}a, H-6^{II}b, 4H), 3.88-3.67 (m, H-4, H-5^{II}, 2H), 2.68 (t, *J* = 13.0 Hz, CH_2, 2H), 2.52 (m, CH_2, 2H), 2.15, 2.10, 2.05, 2.02; 2.00, 1.91 (s, 15H); ^{13}C-NMR (75 MHz, $CDCl_3$): δ 172.5, 170.5, 170.4, 170.3, 169.7, 169.2 (CH_3*C*=O), 103.9 (d, $J_{C,F}$ = 230 Hz, C-1^{I}), 101.1 (C-1^{II}), 77.6, 77.3, 77.0, 75.4, 71.2, 71.0, 70.9, 70.8, 70.6, 69.3, 69.2, 67.0, 61.5, 61.3 (C-$2^{I\text{-}II}$, C-$3^{I\text{-}II}$, C-$4^{I\text{-}II}$, C-$5^{I\text{-}II}$, C-$6^{I\text{-}II}$), 38.0, 29.9, 28.0 (CH_2, CH_3), 21.1, 21.0, 20.9, 20.8, 20.7 (C*H*$_3$C=O); (FAB-MS positive mode): *m/z* 717 $[M+Na]^+$, 695 $[M+H]^+$.

2,3,4-Tri-*O*-acetyl-6-*O*-methanesulfonyl-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-α-D-glucopyranosyl fluoride (3)

$[\alpha]_D$ (deg cm^3 g^{-1} dm^{-1}) = +19.2 (*c* = 1.0 g cm^{-3} in $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ 5.67 (dd, $J_{1,2}$ = 2.7 Hz, $J_{1,F}$ = 53 Hz, H-1^I, 1H), 5.40 (t, $J_{2,3}$ = 9.8 Hz, H-3^I, 1H), 5.15 (d, $J_3{}^{II},_4{}^{II}$ = 3.2 Hz, H-4^{II}, 1H), 4.99 (dd, $J_3{}^{II},_4{}^{II}$ = 3.4 Hz, $J_2{}^{II},_3{}^{II}$ = 10.3 Hz, H-3^{II}, 1H), 4.88 (ddd, $J_{1,2}$ = 2.8 Hz, $J_{2,F}$ = 10.2 Hz, $J_{2,3}$ = 13.0 Hz, H-2^I, 1H), 4.57-4.26 (m, H-1^{II}, H-6a^I/H-6^{II}a, 2H), 4.26-4.09 (m, H-5^I, H-6^{I}a/H-6^{II}a, H-6^{I}b, H-6^{II}b, 4H), 3.98 (m, H-5^I, 1H), 3.06 (s, OCH_3S, 3H), 2.17, 2.14, 2.10, 2.07, 2.05, 1.97 (s, C*H*$_3$C=O, 18H); ^{13}C-NMR (100 MHz, $CDCl_3$): δ 170.5, 170.4, 170.35, 170.2, 169.7, 169.2 (CH_3*C*=O), 103.9 (d, $J_{C,F}$ = 230 Hz, C-1^I), 100.9 (C-1^{II}), 75.1, 71.2, 71.0, 70.9, 70.7, 70.5, 69.4, 69.2, 66.8, 64.9, 61.4 (C-2^{I-II}, C-3^{I-II}, C-4^{I-II}, C-5^{I-II}, C-6^{I-II}), 38.01 (OSO_2CH_3), 21.1, 21.05, 20.85, 20.8, 20.7 (*C*H_3C=O);); (DCI-MS positive mode, ammoniac-isobutane): *m/z* 692 $[M+NH_4]^+$.

2,3,4-Tri-*O*-acetyl-6-azido-6-deoxy-β-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-acetyl-α-D-glucopyranosyl fluoride (4)

$[\alpha]_D$ (deg cm^3 g^{-1} dm^{-1}) = + 21.5 (*c* = 1.1 g cm^{-3} in $CHCl_3$); ^{1}H-NMR (400 MHz, $CDCl_3$): δ 5.65 (dd, $J_{1,2}$ = 2.6 Hz, $J_{1,F}$ = 53 Hz, H-1^I, 1H), 5.45 (t, $J_{2,3}$ = 9.7 Hz, H-3^I, 1H), 5.31 (d, $J_3{}^{II},_4{}^{II}$ = 3.4 Hz, H-4^{II}, 1H), 5.07 (dd, $J_1{}^{II},_2{}^{II}$ = 7.9 Hz, $J_2{}^{II},_3{}^{II}$ = 10.3 Hz, H-2^{II}, 1H), 4.93 (dd, $J_3{}^{II},_4{}^{II}$ = 3.4 Hz, $J_2{}^{II},_3{}^{II}$ = 10.3 Hz, H-3^{II}, 1H), 4.82 (ddd, $J_{1,2}$ = 2.6 Hz, $J_{2,F}$ = 7.7 Hz, $J_{2,3}$ = 10.3 Hz, H-2^I, 1H), 4.52-4.48 (m, H-1^{II}, H-6^{I}b, 2H), 4.13-4.05 (m, H-5, H-6^{I}a, 2H), 3.87 (t, $J_{3,4}$ = 9.8 Hz, H-4^I, 1H), 3.72 (m, H-5^{II}, 1H), 3.43 (m, H-6^{II}b, 1H), 3.20 (dd, $J_5{}^{II},_6{}^{II}$ = 5.0 Hz, $J_6{}^{II}{}_{a,6}{}^{II}{}_b$ = 13.0 Hz, H-6^{II}a, 1H), 2.14, 2.11, 2.03, 2.01, 1.94, (5 s, C*H*$_3$C=O, 12H); ^{13}C-NMR (75 MHz, $CDCl_3$): δ 170.4, 170.3, 170.1, 169.5 169.05 (CH_3*C*=O), 103.8 (d, $J_{C,F}$ = 230 Hz, C-1^I),100.7 (C-1^{II}), 74.4, 72.6, 71.0, 70.9, 70.7, 70.8, 70.5, 69.3, 69.2, 67.8, 61.3 (C-2^{I-II}, C-3^{I-II}, C-4^{I-II}, C-5^{I-II}, C-6^I), 50.5 (C-6^{II}), 21.0, 20.9, 20.8, 20.6, (*C*H_3C=O); (ESI-MS positive mode): *m/z* 644 $[M+Na]^+$, 660 $[M+K]^+$.

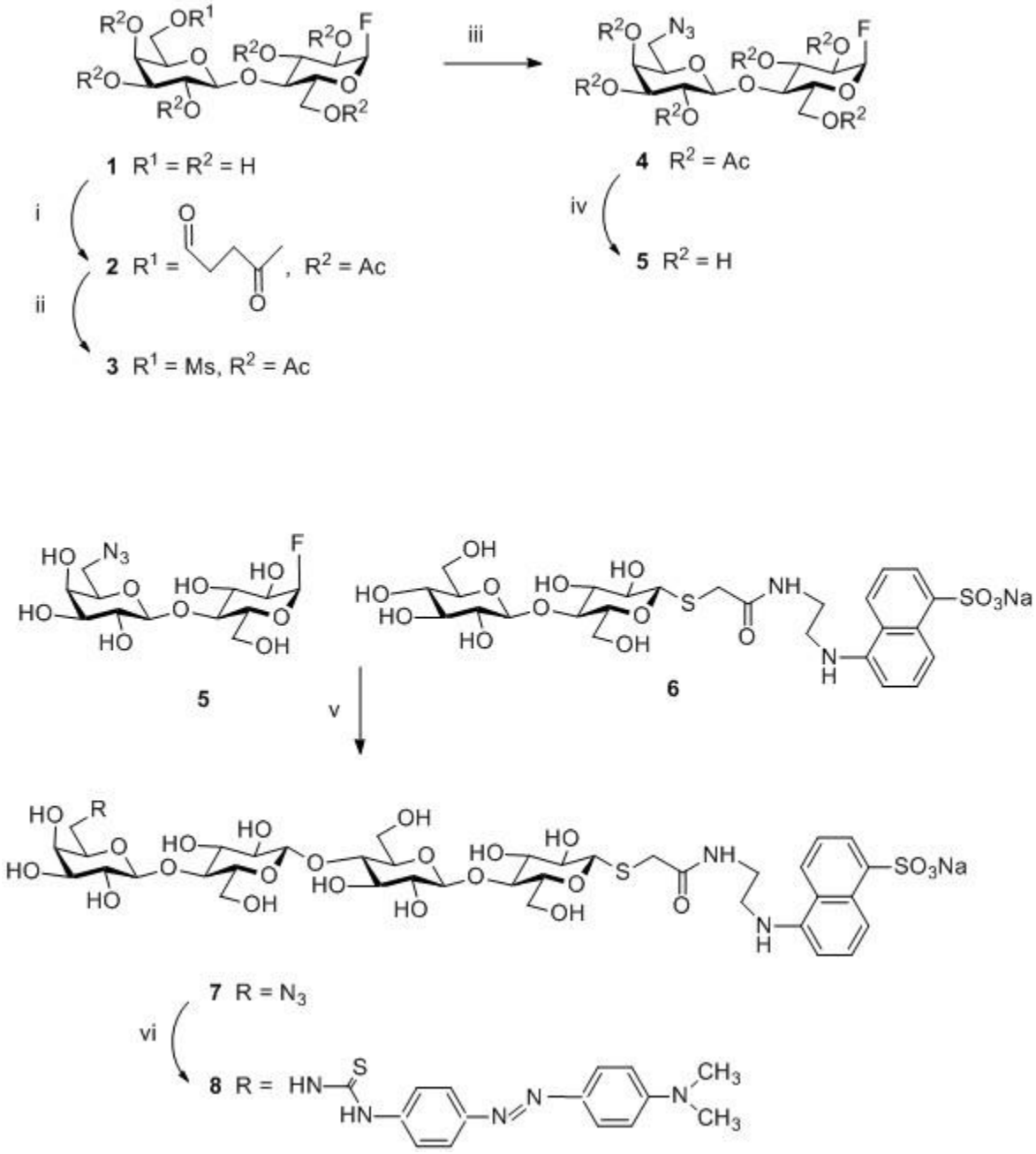
Sodium N-[2-N[(S-(6-azido-6-deoxy-β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-β-D-glucopyranosyl)-(1→4)-β-D-glucopyranosyl)-2-thioacetyl]aminoethyl]-1-naphthylamine-5-sulfonate (7)

$[\alpha]_D$ (deg cm^3 g^{-1} dm^{-1}) = - 0.59 (*c* = 0.6 g cm^{-3} in H_2O); ^{1}H NMR (400 MHz, D_2O): δ 8.16 (d, *J* = 7 Hz, arom H, 1H), 8.07 (d, *J* = 8.7 Hz, arom H, 1H), 7.59 (m, arom H, 2H), 6.89 (d, *J* = 7.8 Hz, arom H, 1H), 4.54 (d, *J* = 7.8 Hz, H-1, 1H), 4.48 (d, *J* = 7.8 Hz, H-1, 1H), 4.39 (d, *J* = 9.3 Hz, H-1^I, 1H), 4.24 (d, *J* = 8 Hz, H-1, 1H), 4.01-3.43 (m, 30H); ^{13}C NMR (100 MHz, D_2O): δ 173.9 (CONH), 144.4, 139.1, 129.8, 129.1, 126.9, 125.8, 125.0, 124.3, 115.6, (arom Cs), 106.8 (C-1^{IV}) 102.9 (C-1^{II,III}), 85.6 (C-1^I), 79.2, 79.1, 78.8, 75.9, 75.4, 75.3, 74.7, 74.5, 73.9, 73.5, 73.4, 73.0, 72.6, 71.4, 69.5 (C-2^{I-IV}, C-3^{I-IV}, C4^{I-IV}, C-5^{I-IV}), 60.65, 60.6, 60.4, (C-6^{I-III}), 51.5 (C-6^{IV}), 43.3 (S*C*H_2CONH), 39.1, 34.1 (CH_2NH); (ESI-MS positive mode): *m/z* 1058.21 $[M+Na]^+$, 1074.2 $[M+K]^+$.

Sodium N-[2-N[(S-(6-deoxy-6-(4-((4-(dimethylamino)phenyl)azo)phenylthioureido-β-D-galactopyranosyl-(1→4)-β-D-glucopyranosyl-(1→4)-β-D-glucopyranosyl)-(1→4)-β-D-glucopyranosyl)-2-thioacetyl]aminoethyl]-1-naphthylamine-5-sulfonate (8)

^{1}H NMR (400 MHz, [D7]DMF): δ 8.44 (d, *J* = 8.6 Hz, arom H, 1H), 8.33 (t, *J* = 5.7 Hz, CH_2N*H*, 1H), 8.16-8.11 (m, arom H, 2H), 7.89-7.76 (m, arom H, 6H), 7.34 (dd, *J* = 8.4 Hz, *J* = 7.2 Hz, arom H, 1H), 7.30 (t, *J* = 8.2 Hz, arom H, 1H), 6.90 (m, arom H, 2H), 6.62 (d, *J* = 7.3 Hz, arom, 1H), 6.03 (t, *J* = 5.4 Hz, CH_2N*H*, 1H), 5.63-5.56 (br s, OHs), 5.34 (br s, OHs), 4.90-4.86 (br s, OHs), 4.75 (t, *J* = 6.0 Hz, 1H), 4.55-4.42 (m, H-1, 4H), 3.11 (s, CH_3, 6H); ^{13}C NMR (100 MHz, [D7]DMF): δ 182.4 (NHCSNH), 171.7 (CONH), 153.7, 150.1, 145.0, 144.3, 142.3, 131.8, 127.2,

126.4, 125.7, 125.6, 125.2, 123.8, 123.7, 123.5, 123.4, 117.5, 112.7 (arom Cs), 104.9-104.0 (C-1^{II-IV}, arom C), 86.2 (C-1^{I}), 81.3, 81.0, 80.7, 80.5, 77.8, 76.5, 76.2, 76.2, 76.1, 74.8, 74.7, 74.6, 74.4, 74.1, 72.2, 70.6 (C-2^{I-IV}, C-3^{I-IV}, C4^{I-IV}, C-5^{I-IV}), 62.2, 61.9, 61.8 (C-6^{I-III}), 46.5 (C-6^{IV}), 44.5 (SCH_2CONH), 40.6 (CH_3), 39.5, 33.9 (CH_2NH); (MALDI-TOF-MS negative mode): *m/z* 1267.44 $[M]^-$, 1135.15 $[M\text{-}C_8H_8N_2]^-$.



Supplementary Figure 3: Synthetic scheme for Förster-resonance-energy-transfer (FRET) substrate, F*-G4-F.

Supplementary Note 2: Additional structural discussion

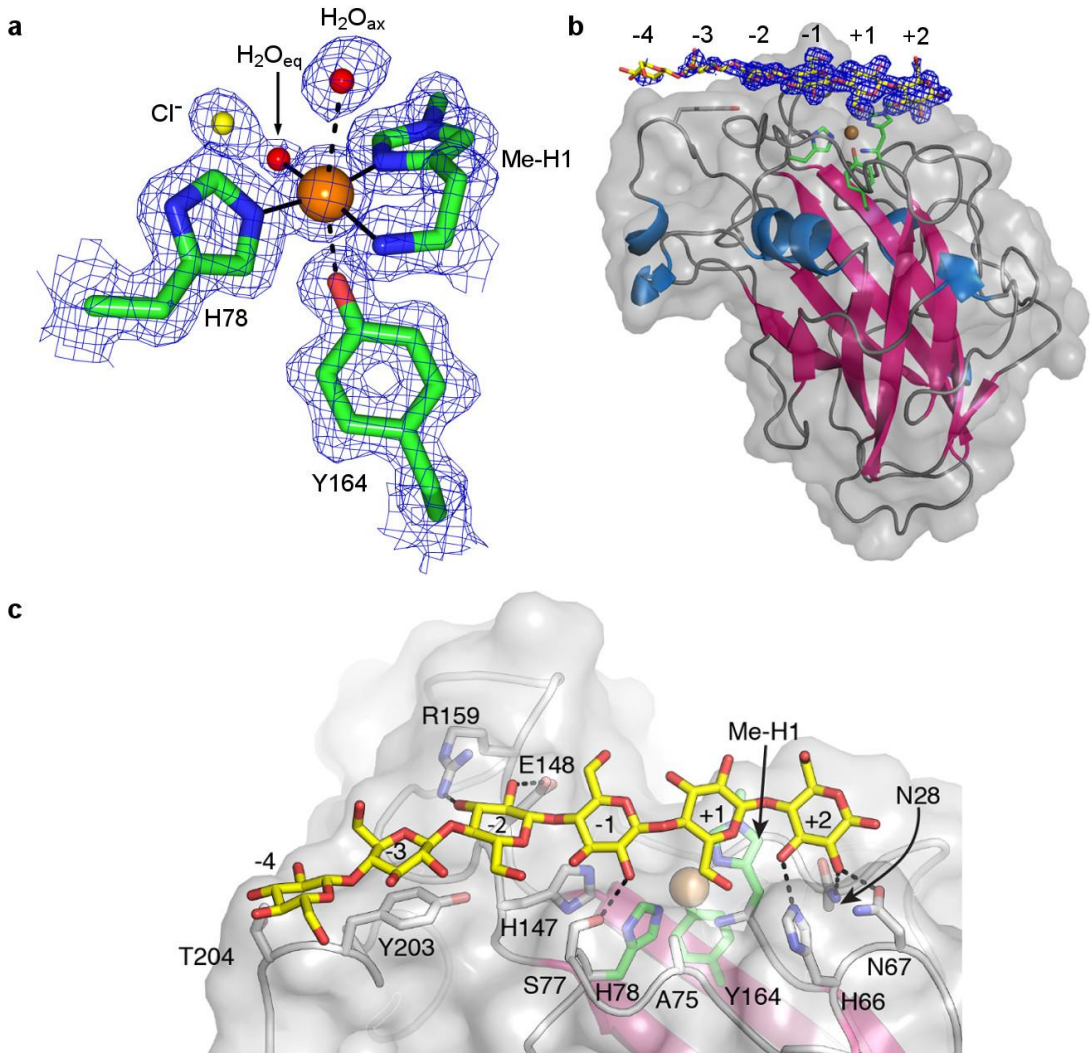
Copper coordination in structures determined at different X-ray doses

In the highest resolution structure obtained (*Ls*(AA9)A_highres, resolution 1.28 Å) and generally all the conventionally collected structures, the density around the active site metal is not always clear and the active site metal shows signs of photoreduction, as in many of the LPMO structures reported before. In *Ls*(AA9)A_highres structure (**Supplementary Fig. 4a**) this manifests itself with partial occupancy of the equatorial water, long distance of the axial water (3.3 Å) and high B-factors for both. Crystallographic refinement against the *Ls*(AA9)A_lowdose data set showed instead a typical Cu(II) AA9 LPMO active site, with His1, His78 and Tyr164 as coordinating ligands on the enzyme, and two additional water ligands (distances indicated in **Supplementary Table 1**, structure and density shown in **Fig. 3b**). The distances are very close to the ones reported for other AA9 LPMOs believed to be mostly in Cu(II) state, exemplified here by the crystal structure of *Ta*(AA9)A (**Supplementary Table 1**). Correspondingly the rmsd for 33 atoms in the Cu(II) liganding enzyme residues in *Ta*(AA9)A (3ZUD) and the low dose *Ls*(AA9)A_lowdose structure is low 0.898 Å. Both coordinating waters to the fully occupied Cu(II) are modelled at full occupancy, and have reasonable B factors compared to the surrounding residues (9.2 $Å^2$ for the apical water, 12.3 $Å^2$ for the equatorial water and 13.0 $Å^2$ for Cu(II) for comparison). In the *Ls*(AA9)_G3_lowdose structure, the axial ligand is displaced by G3, and the equatorial ligand was modelled as a full occupancy Cl^- with a B factor of 16.8 $Å^2$ (modelling of the equatorial ligand as water resulted in a very low B-factor of 2 $Å^2$). For comparison, the fully occupied Cu(II) in this structure has a B factor of 10.9 $Å^2$.

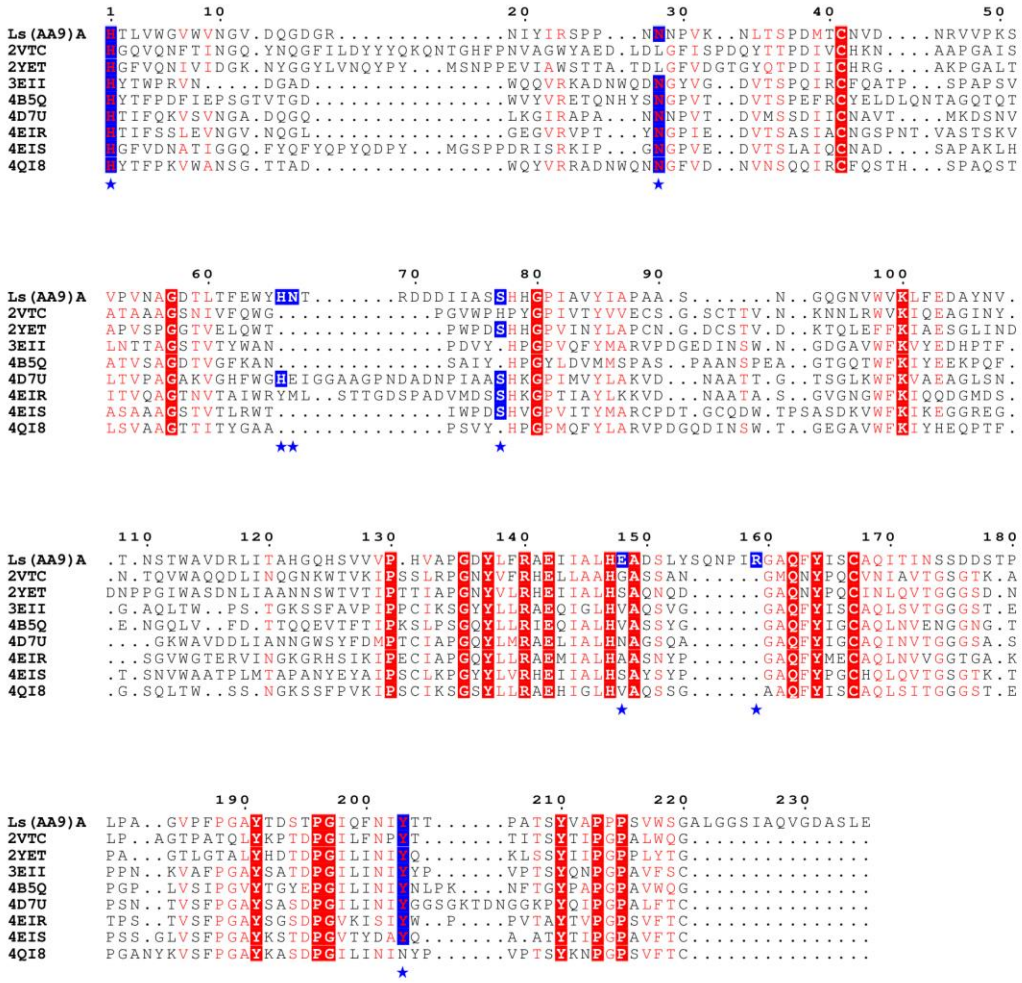
Additional discussion on cellooligosaccharide complexes

Interactions of *Ls*(AA9)A with cellooligosaccharides in the crystal structures are shown in **Figures 3c-e** and **Supplementary Figures 4b,c** and potential hydrogen bond interactions listed in **Supplementary Table 2**. The φ and ψ torsion angles for the glycosidic linkages of the bound cellooligosaccharides are shown in **Supplementary Table 3**, and show that while G3 binds in a flat conformation very close to the one found in cellulose, G6 shows some deviation from a cellulose-like structure, especially the glycosidic linkage between glucosyl units at subsites –3 and –2. From the current structure, it cannot be excluded that this deviation is due to crystal packing, since a cellulose-like conformation would result in clashes with a symmetry-related molecule. Inability of the oligosaccharide chains to assume the most preferred conformation for binding due to unfavorable crystal contacts could account for the poor definition of the terminal sugars in the G6 complexes.

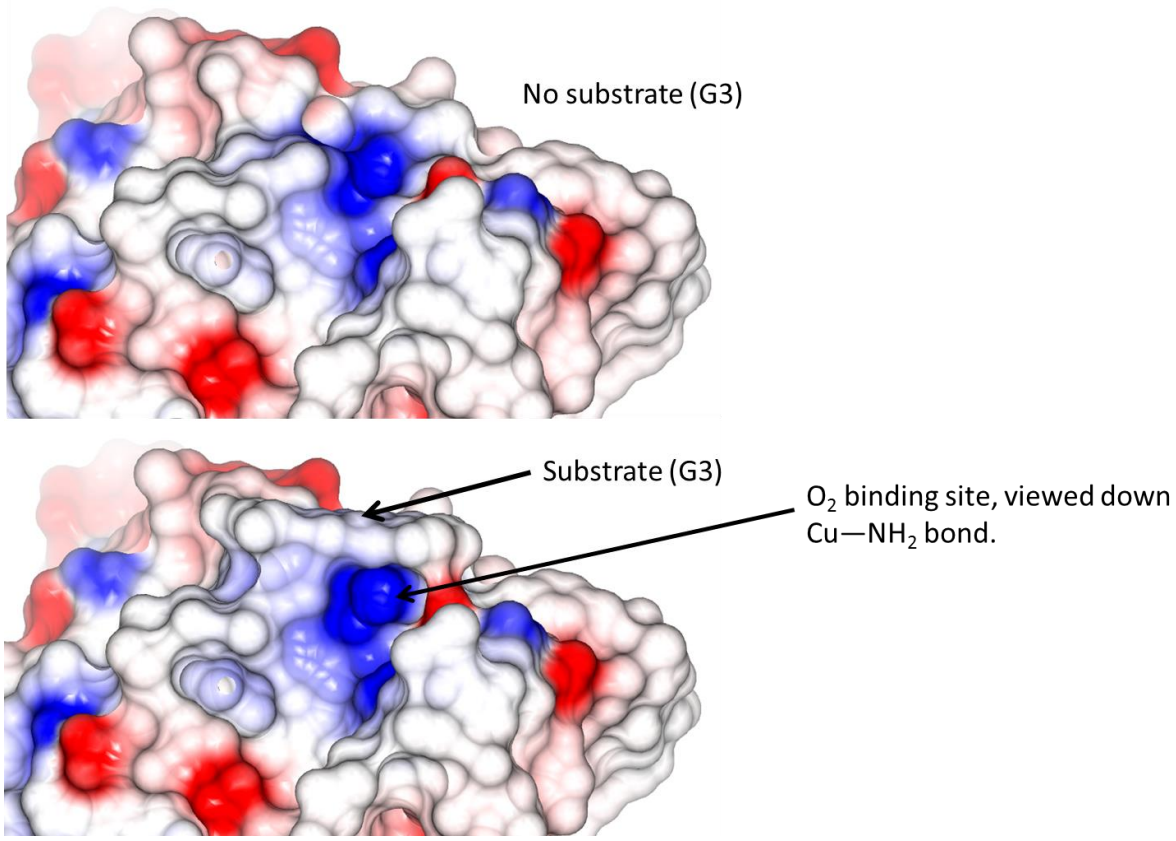
Despite the similarity of φ and ψ angles, one difference seen in the G3 complexes is that the O2…O6’ bonds generally formed in cellulose I are not formed, but rather the O…O distances suggest intrachain O3…O6’hydrogen bonds.



Supplementary Figure 4: Crystal structures of *Ls*(AA9)A. **a**, Active site density for photoreduced high resolution structure *Ls*(AA9)A_highres**.** The structure is shown in same orientation as **Figure 3b**. **b,** Overview of G6 binding and $2F_o$-F_c density of *Ls*(AA9)A_G6 structure. **c,** Close-up of G6 binding and interacting surface in the *Ls*(AA9)A_G6 structure.



Supplementary Figure 5: Structure-based alignment of AA9 sequences. The multiple alignment was computed using the 3DCoffee mode of TCoffee.[7] Residues printed white on a red background are strictly conserved in the alignment. Residues printed in red show 80% conservation. Residues that interact with G6 in *Ls*(AA9)A-G6 are marked with an asterisk and printed on a blue background. When conserved, the corresponding residues in other sequences are also shown on a blue background.

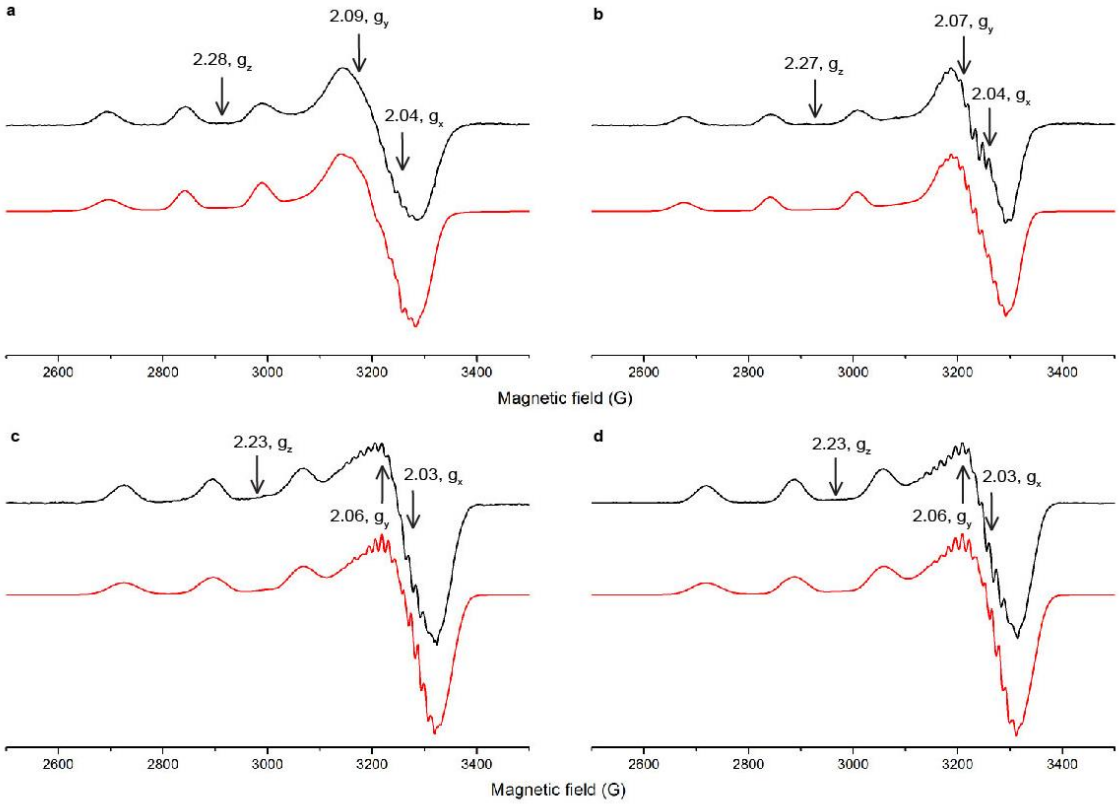


Supplementary Figure 6: O_2 binding site in *Ls*(AA9)A. Space-filling representation of *Ls*(AA9)A without substrate (top) and in the presence of G3 substrate (bottom).

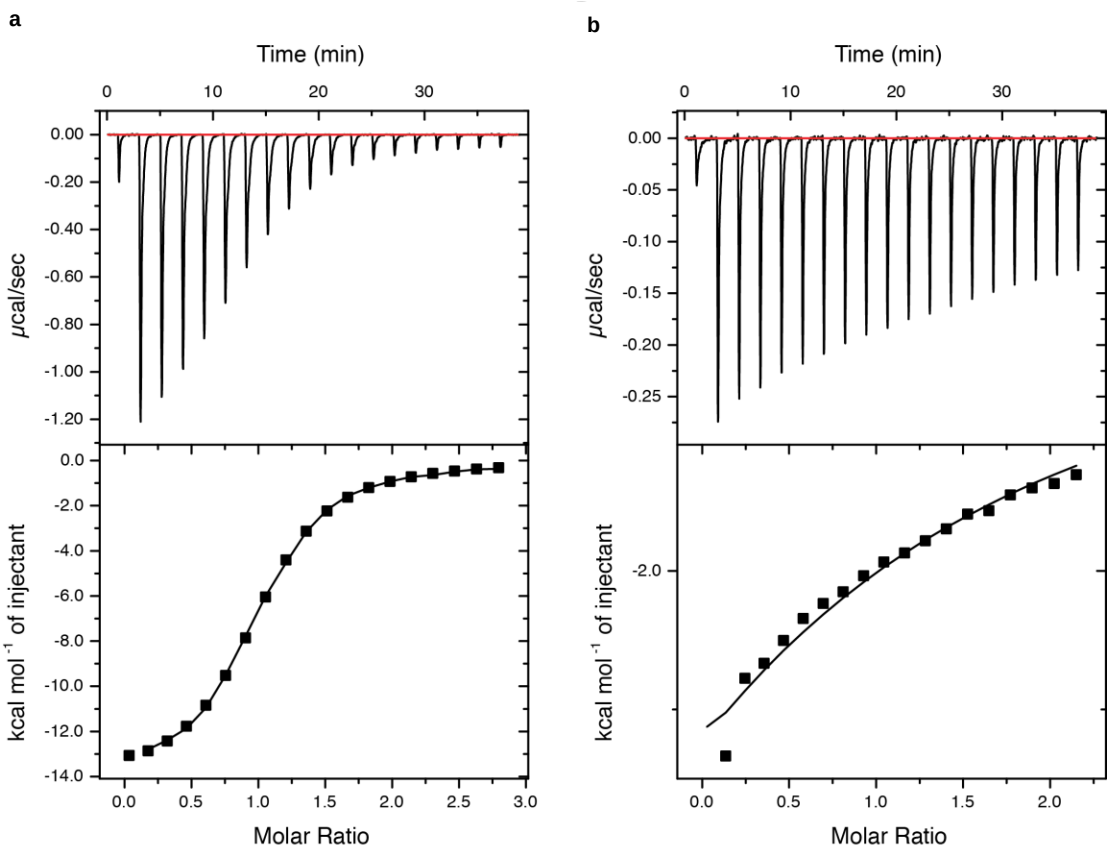
Supplementary Note 3: Pulsed EPR Analysis

Ls(AA9)A, X-band ^{1}H-HYSCORE and Davies ^{14}N-ENDOR spectra of *Ls*(AA9)A were obtained under high chloride concentrations (**Fig. 5c**, **Supplementary Fig. 9**). ^{14}N-ENDOR spectra in the presence of G6 substrate show hyperfine interactions between 9–25 MHz, typical of ^{14}N atoms strongly coupled to Cu(II) (**Supplementary Fig. 9f**).

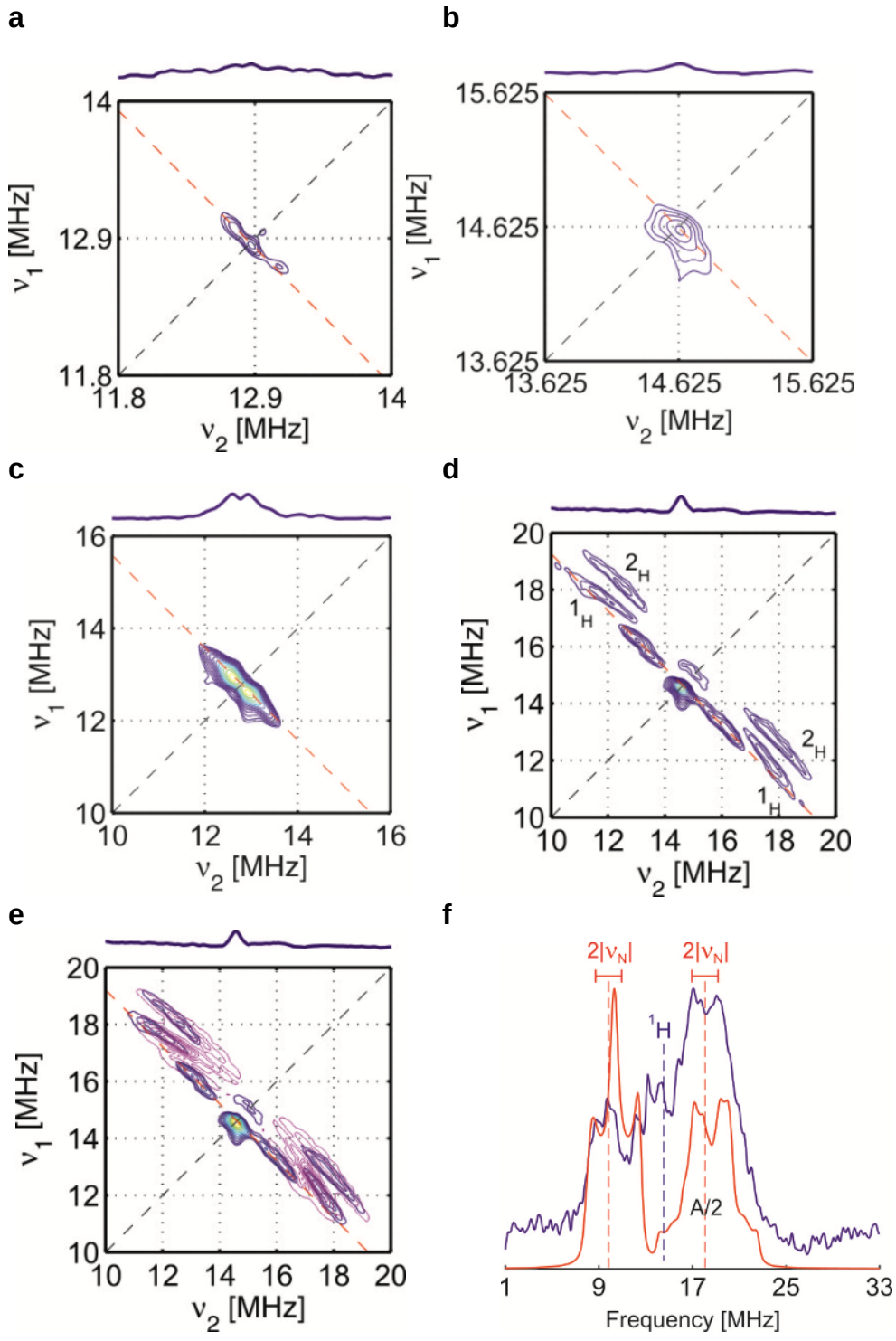
Spectra were simulated using EasySpin 5.0.0 [8] integrated into MATLAB R2014a software.[9]



Supplementary Figure 7: X-band cw EPR spectra and simulations of *Ls*(AA9)A. Simulations in red (bottom). **a,** *Ls*(AA9)A. **b,** *Ls*(AA9)-G6 after subtraction of the chloride complex (~15%). **c,** *Ls*(AA9)-PASC in the presence of 200 mM NaCl. **d,** *Ls*(AA9)-G6 in the presence of 200 mM NaCl



Supplementary Figure 8: Isothermal titration calorimetry (ITC) showing chloride enhancement of G6 binding. **a,** Titration of G6 into *Ls*(AA9)A in the presence of 200 mM NaCl and **b,** in the absence of chloride (ionic strength kept constant by addition of Na_2SO_4). K_d for G6 binding in the presence of chloride was 3.7 ± 0.1 μM. Binding in the absence of chloride was so weak that a curve could not be reliably fitted to the data. Experiments performed in triplicate.



Supplementary Figure 9: ^{1}H HYSCORE spectra of *Ls*(AA9)A and *Ls*(AA9)A-G6 and X-band Davies ^{14}N ENDOR spectrum of *Ls*(AA9)A-G6. Contour presentations of the ^{1}H HYSCORE in panels **a-e**, ^{14}N ENDOR spectrum panel **f**. **a,** *Ls*(AA9)A near $g_{\|}$ with τ = 136 ns at 304 mT. **b,** *Ls*(AA9)A near $g_{\perp}$ with τ = 136 ns at 345 mT. **c,** *Ls*(AA9)A-G6 near $g_{\|}$ with τ = 200 ns at 301.4 mT. The cross-peak labeled 1_H splits

symmetrically along the anti-diagonal. **d,** *Ls*(AA9)A near $g_{\perp}$ with τ = 136 ns at 343.6 mT. The cross-peaks labeled with 1_H and 2_H shift asymmetrically from the anti-diagonal. **e,** Numerical simulations (pink) of **d** which include two anisotropic protons with anisotropic constants $T \approx 4.7$ and $T \approx 6.7$ MHz, respectively. **f,** ^{14}N ENDOR spectrum of Ls(AA9)A-G6 (blue), recorded with a hard pulse $\pi/2$ = 32 ns and radiofrequency pulse π_{RF} = 10 µs at 343 mT near $g_{\perp}$ and ~ 9.8 GHz; (red) numerical simulation using hyperfine values $A_{eff} \approx$ 36.1 MHz and $A_{eff} \approx$ 19.5 MHz for two nitrogen atoms.

Supplementary References

1. Busse-Wicher, M. *et al*. The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana. *Plant J.* **79**, 492-506 (2014).
2. Diederichs, K. & Karplus, P.A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Mol. Biol.* **4**, 269-275 (1997).
3. Jünnemann, J., Thiem, J. & Pedersen, C. Facile synthesis of acetylated glycosyl fluorides derived from di- and tri-saccharides. *Carbohydr. Res.* **249**, 91-94 (1993).
4. Rencurosi, A., Poletti, L., Panza, L. & Lay, L. Improvement on lipase catalysed regioselective O-acylation of lactose:a convenient route to 2′-O-fucosyllactose. *J. Carbohydr. Chem.* **20**, 761-765 (2001).
5. Fort, S. *et al*. Highly Efficient Synthesis of β(1 → 4)-Oligo- and -Polysaccharides Using a Mutant Cellulase. *J. Am. Chem. Soc.* **122**, 5429-5437 (2000).
6. Boyer, V. *et al*. Chemoenzymatic synthesis of a bifunctionalized cellohexaoside as a specific substrate for the sensitive assay of cellulase by fluorescence quenching. *Chem. Eur. J.* **8**, 1389-1394 (2002).
7. O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D.G. & Notredame, C. 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *J. Mol. Biol.* **340**, 385-395 (2004).
8. Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J. Magn. Reson.* **178**, 42-55 (2006).
9. Inc., T.M. MATLAB and Statistics Toolbox Release 2014a. (The MathWorks, Inc., Natick, Massachusetts, United States).