Supporting information

Regioselective glucosylation of (+)-catechin using a new variant of the sucrose phosphorylase from *Bifidobacterium adolescentis*

Marie Demonceaux, Marine Goux, Johann Hendrickx, Claude Solleux, Frédéric Cadet, Émilie Lormeau, Bernard Offmann and Corinne André-Miral

Contents

1 Materials and methods

1.1 Chemicals

(+)-catechin was purchased from Extrasynthèse, Ni-NTA resin from Macheray-Nagel and PCR primers from Eurofins. All other chemicals were purchased from Sigma-Aldrich or VWR.

1.2 Vector construction

pCXP34h-*Ba*SP was kindly provided by Tom Desmet from Universiteit Gent.
pCXP34h-*Ba*SP Q345F and pCXP34h-*Ba*SP Q345F/P134D were obtained by quick-change mutagenesis with the support of D-Zyme (Capacités, Nantes).
Hexahistidine-tagged enzymes were constructed and amplified by PCR using the pairs of oligonucleotides 5'-TATACCATGGGCAAAAATAAAGTCCAACTGATTAC CTATGC-3' and 5'-GGTGCTCGAGTCATTAGTGATGATGATGATGATGCGCAA CAACCGGCGGATTAG-3'. Each digested PCR product was inserted between the NcoI and XhoI restriction sites of a linearised pET28b vector. *E. coli* BL21(DE3) competent cells (Novagen) were transformed with the resulting material. Clones were selected using LB-agar medium supplemented with 25 μg/mL kanamycin. Constructions were confirmed by Sanger sequencing (Eurofins Genomics).

1.3 Proteins production and purification

Transformed bacteria were grown overnight with shaking at 37°C in 5 mL LB-medium supplemented with 25 μg/mL kanamycin and 0.5% (w/v) glucose. On the next day, 200 mL of LB autoinducible medium containing 1% (w/v) glucose and the same antibiotic concentration was incubated with 2 mL of overnight culture. Cells were grown with shaking at 25°C. After 24 h, cells were centrifuged (ThermoScientific, Sorvall RC6 plus, rotor SLC 4000, 30 min, 4 150g, 19°C) and pellet was resuspended in NPI-5 buffer (5 mL/g of pellet) [50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole-HCl, pH 8.0] in the presence of 5 μg/mL DNAse I, 250 μg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride (PMSF). Total protein extracts were obtained by sonication. Suspension was centrifuged (ThermoScientific, Sorvall Legend X1R centrifuge, rotor FIS-8x50cy, 20 min, 12 000g, 4°C) to remove cell debris. Proteins were purified from supernatants by immobilised metal ion affinity chromatography (IMAC), using Protino Ni-NTA agarose beads (Macherey-Nagel) equilibrated with NPI-5. After washing with 2x10 column volume (CV) of NPI-10 buffer [50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole-HCl, pH 8.0] and 10 CV of NPI-20 buffer [50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole-HCl, pH 8.0], purified

protein was eluted 5-fold with 1 CV of NPI-250 buffer [50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole-HCl, pH 8.0]. Protein concentration was determined by UV absorbance at 280 nm (NanoDrop 1 000, Thermo Scientific) and purity was confirmed by Coomassie-stained 12% SDS-PAGE (Figure S1). Buffer of the purified protein was exchanged on a Zeba spin desalting column (7 000 Da MWC, Pierce) with 50 mM MOPS-solution at pH 8.0.

1.4 Protein thermal shift assay

Thermal shift assay was conducted as described by the supplier of the real-time PCR machine (Bio-Rad CFX96). Assays were performed in 96-well PCR plate (Bio-Rad, MLL9601) in a final volume of 25 μL. Enzyme (3 μM) and SYPRO Orange (5x concentrated in DMSO) were incubated with concentrated buffers and salts, and volume was completed with water. Plate was sealed with highly transparent optical-clear quality sealing tape. After 5 min of equilibration time at 5°C to allow SYPRO Orange to diffuse, the plate was heated from 10°C to 95°C with increments of 0.5°C for 10 seconds, followed by the fluorescence reading with FRET channel (excitation wavelength: 470/20 nm, emission wavelength: 570/10 nm). Melting curves were obtained with the CFX96 Touch System Software and drawn with Microsoft Excel (Figure S2). Melting temperatures were calculated with the first derivative -(dRFU)/dT (Table S2).

1.5 Verification of hydrolytic activity

Hydrolytic activity was verified on *p*-nitrophenyl-α-D-glucopyranoside (*p*NP-α-glc) and *p*-nitrophenyl-β-D-glucopyranoside (*p*NP-β-glc). Assays were conducted in 50 mM MOPS-NaOH-solution at pH 8.0 in a total volume of 200 μL. Reaction mixture containing 20 mM *p*NP-α-glc or *p*NP-β-glc in DMSO was incubated with a final concentration of 3 μM of purified *Ba*SP WT, Q345F or Q345F/P134D at 37°C for 2 h. Released yellow *p*NP was detected by absorbance measurement at 405 nm using a Labsystems integrated EIA Management System. Activity curves were drawn with Microsoft Excel (Figure S3).

1.6 Screening of (+)-catechin glucosylation by capillary electrophoresis

Assays were conducted in 50 mM MOPS-NaOH solution at pH 8.0 in a total volume of 1 mL. Reaction mixture containing 10 mM (+)-catechin or resveratrol in DMSO (1 equivalent, 100 μL), 10% DMSO (100 μL, (v/v)), 80 mM sucrose in H_2O (8 equivalent, 100 μL) was incubated with a final concentration of 3 μM of purified

*Ba*SP mutant at 37°C under slight agitation for 24 h. After centrifugation (ThermoScientific, Heraeus Pico17 Centrifuge, rotor 75003424, 20 min, 12 000g, 20°C), supernatant of the enzymatic synthesis was analysed by capillary electrophoresis (Beckman Coulter, P/ACE^{TM} MDQ Capillary electrophoresis system) at 214 nm in 20 mM borate 50 mM SDS NaOH-buffer pH 9.0, capillary temperature 25°C, voltage 25 kV. Curves were drawn with Microsoft Excel (Figure S4).

1.7 Kinetics of sucrose hydrolysis

Hydrolysis assays were performed at 37°C in 50 mM MOPS-NaOH buffer at pH 8.0 in a total volume of 1 mL. Final concentrations of 3 μM for Q345F and Q345F/P134D and 0.3 μM for WT, and 0.5 mM to 100 mM of sucrose as substrate were used. Reactions were followed on 90 min for WT and on 2h30 for the two variants. At different times, 150 μL was sampled from each reaction, inactivated at 95°C for 5 min. A volume of 25 μL of samples or standards (0.3 mM to 4.8 mM of glucose) was added to 200 μL ABTS solution [5 mg/mL ABTS, 5 mg/mL glucose oxidase, 5 mg/mL peroxidase, 30 mM citrate-NaOH buffer pH 6.0] in 96-well plates. Plates were covered with an aluminium fold and incubated at ambient temperature for 10 min. Concentrations of released glucose were determined at 405 nm using a Labsystems integrated EIA Management System. Lineweaver-Burk plots and linear plots were drawn with Microsoft Excel (Figure S5). The assay was performed in triplicate.

1.8 Kinetic study by analytical HPLC

Assays were conducted in 50 mM MOPS-NaOH solution at pH 8.0 in a total volume of 1 mL. Reaction mixture containing 10 mM (+)-catechin in DMSO (1 equivalent, 100 μL), 10% DMSO (100 μL, (v/v)), 80 mM sucrose in H_2O (8 equivalent, 100 μL) was incubated with a final concentration of 3 μM of purified Q345F or Q345F/P134D at 37°C under slight agitation for 24 h. After centrifugation (ThermoScientific, Heraeus Pico17 Centrifuge, rotor 75003424, 20 min, 12 000g, 20°C), supernatant of the enzymatic synthesis was analysed by analytical HPLC at 280 nm on a C-18 column (Interchim, 5 μm, 250 x 4.6 mm, US5C18HQ-250/046) with an isocratic flow of 80% H_2O, HCOOH 0.1% (v/v) and 20% MeOH, HCOOH 0.1% (v/v) for 20 min. Curves were drawn with Microsoft Excel (Table S3).

1.9 Purification and analysis of glucosylated (+)-catechin

Assays were conducted in 50 mM MOPS-NaOH solution at pH 8.0 in a total volume of 1 mL. Reaction mixture containing 10 mM (+)-catechin in DMSO (1 equivalent,

100 μL), 10% DMSO (100 μL, (v/v)), 80 mM sucrose in H_2O (8 equivalent, 100 μL) was incubated with a final concentration of 3 μM of purified Q345F or Q345F/P134D at 37°C under slight agitation for 24 h. After centrifugation (ThermoScientific, Heraeus Pico17 Centrifuge, rotor 75003424, 20 min, 12 000g, 20°C), supernatant of the enzymatic synthesis was purified by HPLC at 280 nm on a C-18 column (Interchim, 5 μm, 250 x 21.2 mm, US5C18HQ-250/212) with a gradient system (solvent A: H_2O HCOOH 0.1% (v/v); solvent B: MeOH, HCOOH 0.1% (v/v); t0 min = 70/30, t10 min = 70/30, t70 min = 10/90) (Figure S6). Products were identified by mass spectroscopy (Na^+ mode, Waters, UPLC-MS^2 high resolution) and NMR ^{1}H and ^{13}C in MeOD (Bruker, Pulse 400 MHz RS^{2}D). Chemical shifts are quoted in parts per million (ppm) relative to the residual solvent peak. Coupling constant J are quoted in Hz. Multiplicities are indicated as d (doublet), t (triplet), m (multiplet). NMR peak assignments were confirmed using 2D ^{1}H correlated spectroscopy (COSY), 2D ^{1}H nuclear Overhauser effect spectroscopy (NOESY) and 2D ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC).

1.10 Molecular modelling

Glucosyl-enzyme intermediate 3D-models were built for Q345F and Q345F/P134D using the following procedure and the Rosetta software.[1] Glucosylated-aspartyl 192 residue from chain A of crystal structure of *Ba*SP (PDB: 2GDV-A) was inserted into the crystal structure of Q345F (PDB: 5C8B),[2] that served as initial template for both variants. As this glucosylated aspartyl is a non-standard residue, it was not known in the database of the Rosetta software. Using Pymol,[3] the initial coordinates of this modified residue was retrieved. While this residue (ASP192) and the glucose moiety (BGC) are covalently linked in the crystal structure, the Pymol software considered them as two distinct residues. Thus, they were merged into a single non-standard residue, which was called with a new ID, DGC. Associated charges and rotamers were calculated for this new residue using the Rosetta software. All those data were merged in a single file and added into the Rosetta database (Section 2.8).

With the DGC residue ready to be used, glucosyl-intermediates were built for the two variants. From the crystal structure of Q345F (PDB: 5C8B), the native aspartyl residue in position 192 was mutated by the glucosylated-aspartyl DGC residue using Rosetta. A sample of 50 conformers was generated thanks to the program Backrub from Rosetta suite,[4] with 10 000 tries. A sample of 50 conformers of Q345F/P134D was built from the glucosyl-intermediates of Q345F and by mutating the prolyl residue in position 134 into an aspartyl residue. In parallel, 12 conformers of (+)-catechin were also generated using the Mercury software (CCDC)[5] from the crystal

structure OZIDOR of (+)-catechin.

1.11 Molecular docking

All docking experiments were performed with GOLD software[6] using the glucosyl-intermediates and (+)-catechin conformers built above. Docking perimeter was limited to the residues of the active site of the enzyme. Details of these residues are provided in Table S4. Each of the 12 conformers of (+)-catechin were docked on every conformer of the two variants. This amounts a total of 600 (50x12) docking experiments for each variant of the enzyme. Then, only the productive poses that could lead to a glucosylation of (+)-catechin at position 3' or 5 were selected as they were the positions that were glucosylated experimentally. To do so, docking poses were filtered using the following distance constraints: distances within 3.3 and 4.3 Å between OH-3' or OH-5 and carbon atom CD of Glu232, and within 3.0 and 4.0 Å between OH-3' or OH-5 and anomeric carbon atom C1 of the glucosyl moiety (Table S5). Docking scores were compiled for these productive poses and compared between the two variants. R statistics software package[7] was used to perform the boxplots analysis and Student t-tests. Productive poses were further characterized using LIGPLOT software[8] to obtain details of the molecular interactions within the active site of the variants.

2 Supporting information

2.1 Enzymes parameters

Table S1 Proteins characteristics from the online tool Expansy ProtParam. (https://web.expasy.org/protparam/)

		*Ba*SP WT	Q345F	Q345F/P134D
Number of cysteine moiety		2	2	2
Molecular weight	(g/mol)	55823.5	57088.9	55860.5
Theoretical isoelectric point pI		5.0	5.2	4.9
Molar extinction coefficient (cysteines)	($M^{-1}.cm^{-1}$)	69915	69915	69915
Molar extinction coefficient (reduced cys.)	($M^{-1}.cm^{-1}$)	69790	69790	69790

2.2 Protein production and purification

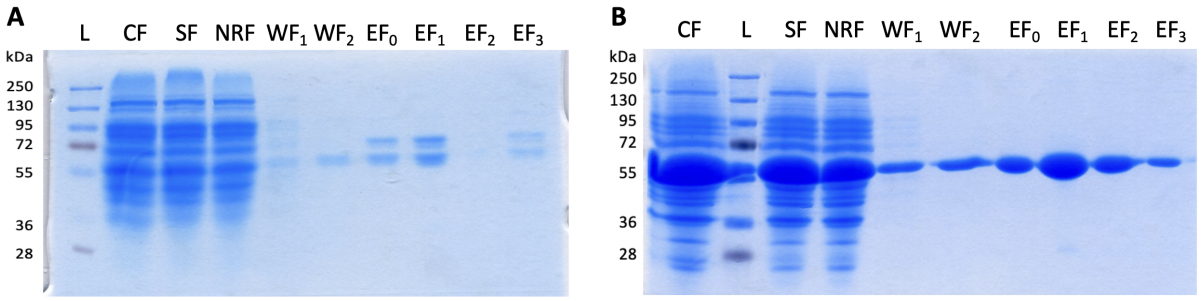


Figure S1 Coomassie-stained 12% SDS-PAGE for Q345F (A) in pCXP34h with an hexahistidine-tag at N-terminal and (B) in pET28b with the tag at C-terminal. Changes of the conditions of production increased the amount and the purity of the enzymes. L: ladder; CF: crude fraction ; SF: soluble fraction; NRF: not retained fraction; WF_1: first washing fraction with 10 mM imidazole-HCl; WF_2: second washing fraction with 20 mM imidazole-HCl; EF_n: elution fraction with 250 mM imidazole-HCl, n = [0, 3]. Elution fractions were deposed with a concentration of 2 μg of enzyme. (Bacteria were grown overnight at 37°C in 10 mL LB-medium supplemented with 10% kanamycin (w/v). On the next day, 400 mL of LB autoinducible media containing 1% glucose (w/v) and the same antiobiotic concentration was incubated with 8 mL of overnight culture at 37°C for 6 h.)[9]

2.3 Protein thermal shift assay

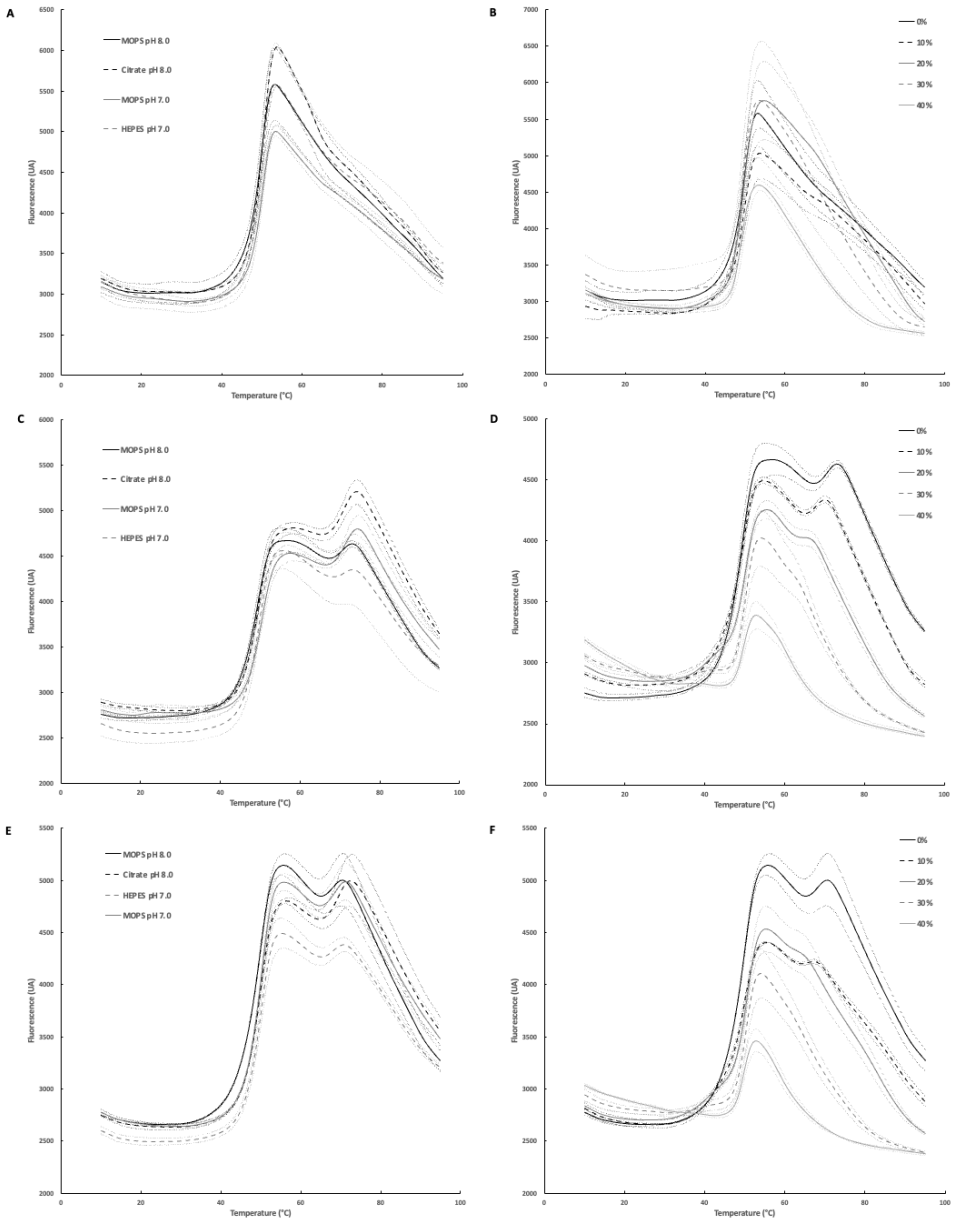


Figure S2 Thermal denaturation assays of *Ba*SP and its variants. Thermostability depends on pH as showed in buffers MOPS-NaOH 50 mM pH 8.0, citrate-NaOH 100 mM pH 8.0, MOPS-NaOH 50 mM pH 7.0 and HEPES-NaOH 25 mM pH 7.0 for (A) *Ba*SP WT, (C) Q345F and (E) Q345F/P134D. Thermostability depends on DMSO concentration as showed in buffer MOPS-NaOH 50 mM pH 8.0 for (B) *Ba*SP WT, (D) Q345F and (F) Q345F/P134D. Dashed lines are standard deviation curves. (3 μM enzyme, 5x Sypro Orange in buffers.)

Table S2 Melting temperatures obtained from the derivative curves -(dRFU)/dT. Pourcentage of DMSO are given in (v/v). Errors correspond to standard deviations.

Buffers:	MOPS 50 mM pH 8.0	Citrate 100 mM pH 8.0	MOPS 50 mM pH 7.0	HEPES 25 mM pH 7.0
WT	50.0 ± 0.1	50.5 ± 0.1	50.2 ± 0.3	50.0 ± 0.1
Q345F	49.5 ± 0.1	49.7 ± 0.3	50.5 ± 0.1	50.2 ± 0.3
Q345F/P134D	50.0 ± 0.1	50.0 ± 0.1	50.3 ± 0.3	50.0 ± 0.1

Buffers: MOPS 50 mM pH 8.0				
	10% DMSO	20 % DMSO	30 % DMSO	40% DMSO
WT	49.7 ± 0.3	49.5 ± 0.1	49.3 ± 0.3	49.0 ± 0.5
Q345F	49.5 ± 0.1	50.2 ± 0.3	50.2 ± 0.3	49.8 ± 0.3
Q345F/P134D	50.0 ± 0.1	50.7 ± 0.3	50.3 ± 0.3	50.5 ± 0.1

2.4 Verification of hydrolytic activity

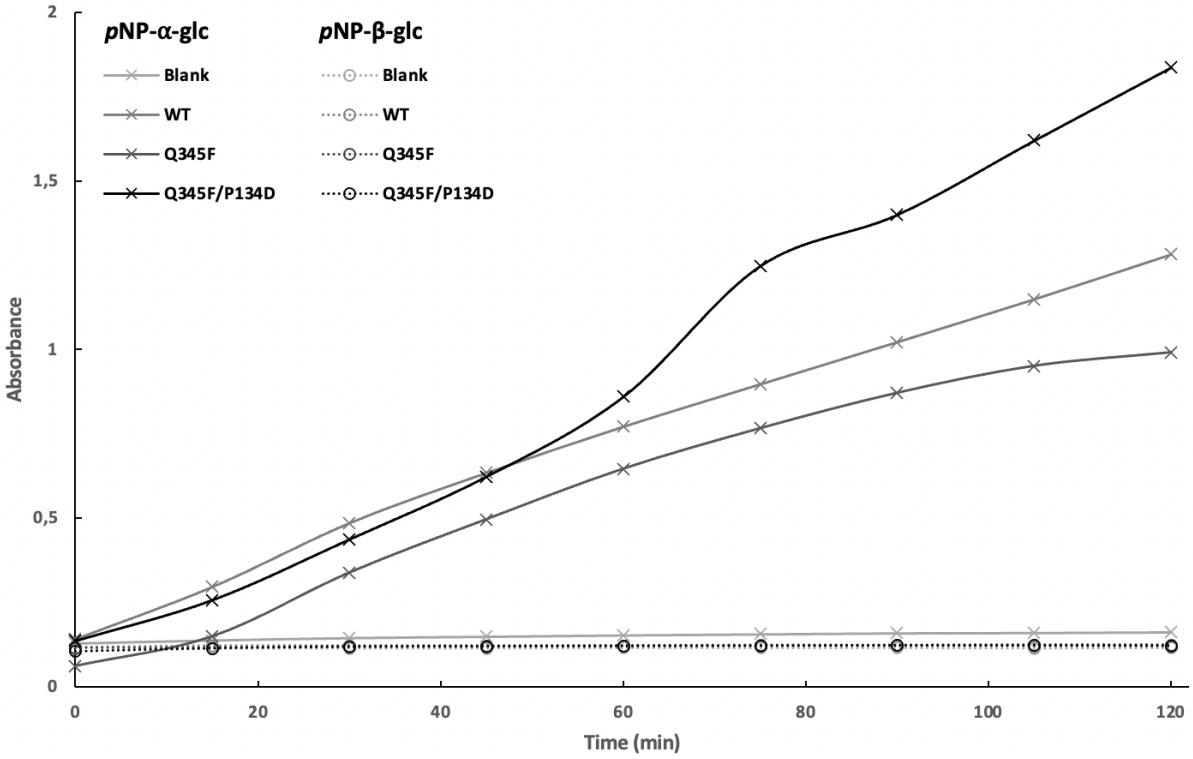


Figure S3 Hydrolytic activity of *p*NP-α-glc (plain lines) and *p*NP-β-glc (dashed lines) by *Ba*SP WT, Q345F and Q345F/P134D. Blanks contain no enzyme. The three enzymes are active on *p*NP-α-glc and inactive on *p*NP-β-glc, validating their stereoselectivy. (20 mM *p*NP-α-glc or *p*NP-β-glc, 3 μM enzyme in MOPS-NaOH 50 mM pH 8.0.)

2.5 Screening of (+)-catechin glucosylation by capillary electrophoresis

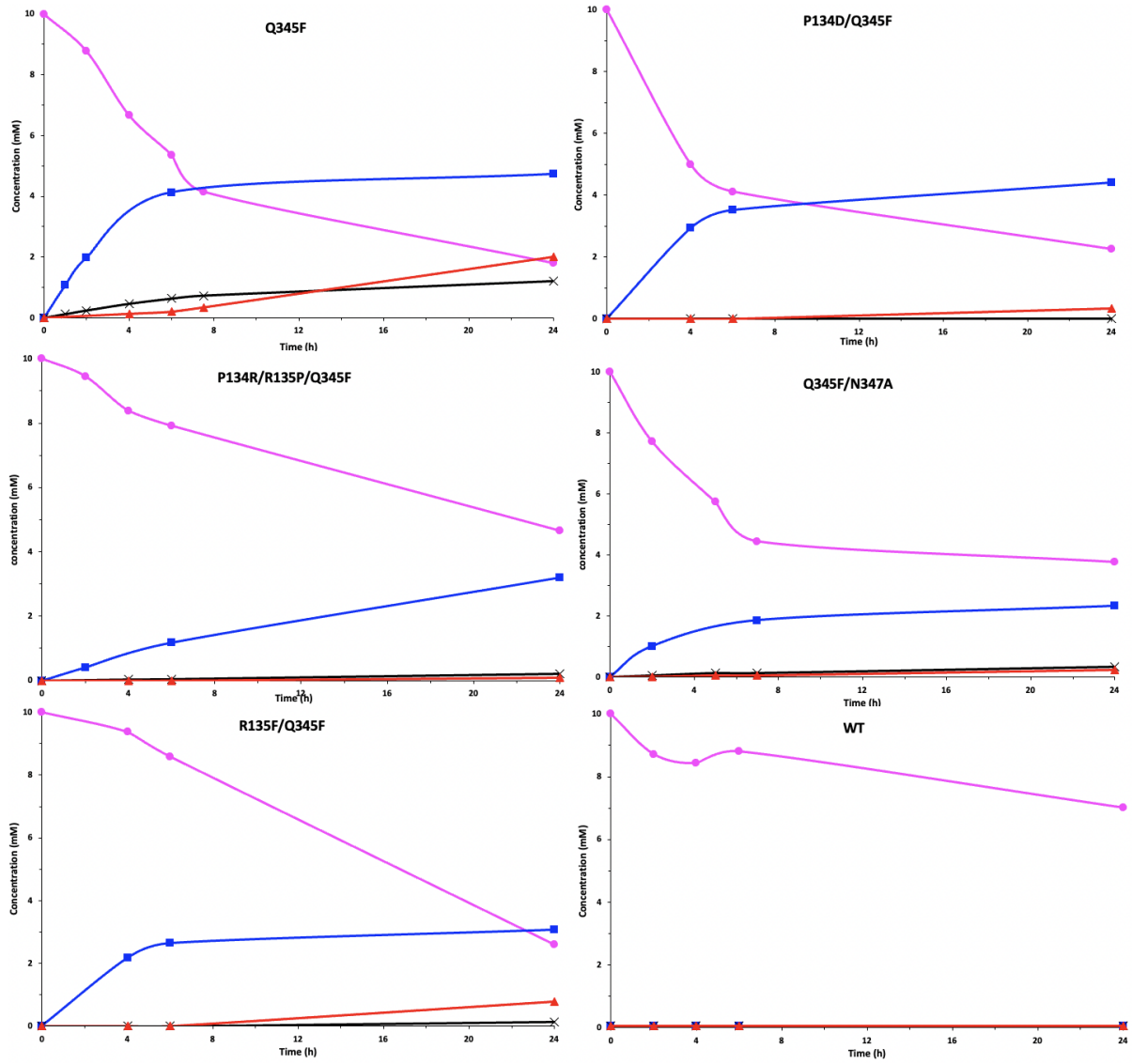


Figure S4 Screening of (+)-catechin glucosylation by capillary electrophoresis. Magenta rounds: (+)-catechin, blue squares: (+)-catechin-3'-O-α-D-glucopyranoside, black cross: (+)-catechin-5-O-α-D-glucopyranoside, red triangles: (+)-catechin-3',5-O-α-D-diglucopyranoside. (Reaction medium: 3 μM enzyme, 80 mM sucrose, 20% DMSO (v/v), 10 mM (+)-catechin in MOPS 50 mM pH 8.0 at 37°C for 24 h; analysis conditions: borate-NaOH buffer 20 mM SDS 50 mM pH 9.0, capillary temperature 25°C, voltage 25 kV, detection at 214 nm).

2.6 Kinetics of sucrose hydrolysis

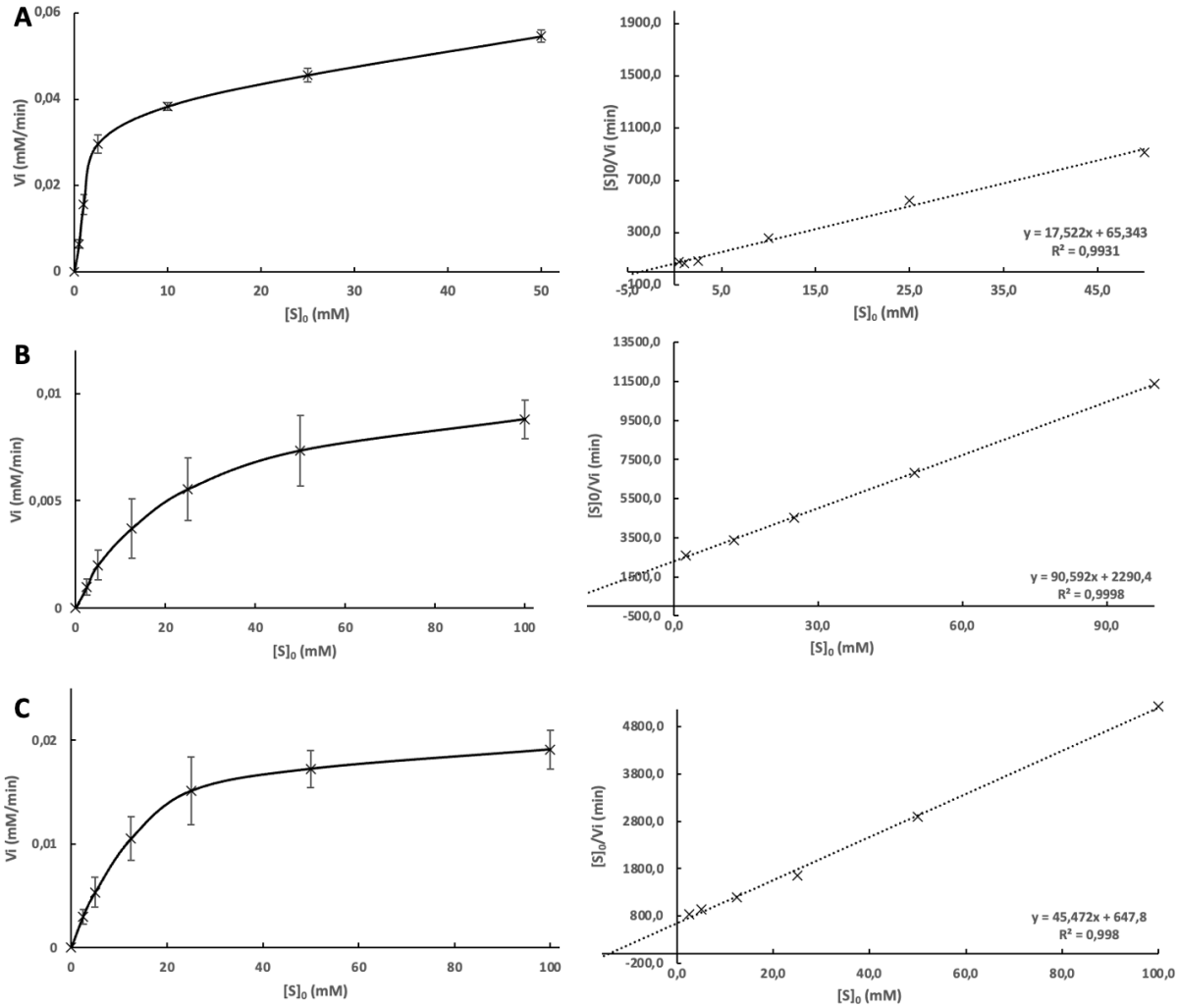


Figure S5 Kinetics and Lineweaver-Burk plots of sucrose hydrolysis by (A) *Ba*SP WT, (B) Q345F and (C) Q345F/P134D. Data were obtained by glucose titration. Errors correspond to standard deviations. (0.3 μM *Ba*SP WT, 3 μM Q345F or Q345F/P134D, 0.5 to 100 mM sucrose in MOPS-NaOH 50 mM pH 8.0.)

2.7 Purification and analysis of glucosylated (+)-catechin

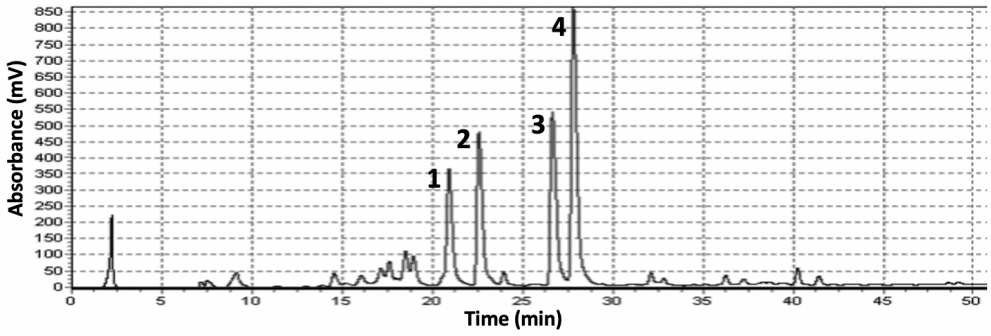
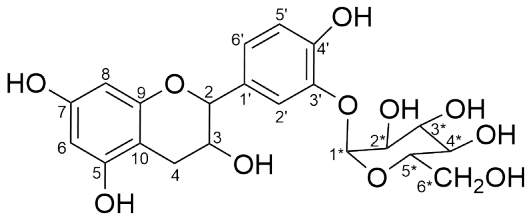


Figure S6 HPLC chromatogram of a 24 h reaction medium of (+)-catechin glucosylation by Q345F. Peak 1: (+)-catechin-5-*O*-α-D-glucoside, peak 2: (+)-catechin-3',5-*O*-α-D-diglucoside, peak 3: (+)-catechin, peak 4: (+)-Catechin-3'-*O*-α-D-glucoside. Unassigned peaks correspond to impurities and degradation of (+)-catechin. (3 μM enzyme, 10 mM (+)-catechin, 10 mM sucrose, 20% DMSO (v/v) in MOPS-NaOH 50 mM pH 8.0 at 37°C for 24 h. Column C-18, detection at 280 nm, gradient of H_2O (formic acid 0.1% (v/v)) and MeOH (formic acid 0.1% (v/v)).)

Table S3 Regioselectivity for (+)-catechin glucosylation by Q345F and Q345F/P134D. Compound concentrations were calculated from the area under the curves obtained by analytical HPLC on a C-18 column for 7 h and at 24 h. Errors correspond to standard deviations. (Reaction medium: 3 μM enzyme, 80 mM sucrose, 20% DMSO (v/v), 10 mM (+)-catechin in 50 mM MOPS-NaOH pH 8.0 at 37°C over 24 h; HPLC conditions: isocratic mode at 80% H_2O, 0.1% HCOOH (v/v) and 20% MeOH, 0.1% HCOOH (v/v).)

Q345F				
Time (h)	CAT-3' (mM)	CAT-5 (mM)	CAT-3',5 (mM)	CAT (mM)
0	0.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01	10.00 ± 0.01
1	0.07 ± 0.02	0.39 ± 0.09	0.00 ± 0.01	9.30 ± 0.18
2	0.15 ± 0.02	0.82 ± 0.08	0.00 ± 0.01	8.25 ± 0.31
3	0.24 ± 0.02	1.23 ± 0.13	0.00 ± 0.01	7.37 ± 0.29
4	0.33 ± 0.01	1.66 ± 0.05	0.02 ± 0.02	6.66 ± 0.23
5	0.43 ± 0.02	2.08 ± 0.13	0.05 ± 0.01	6.04 ± 0.11
6	0.39 ± 0.19	2.36 ± 0.04	0.08 ± 0.01	5.18 ± 0.16
7	0.55 ± 0.04	2.55 ± 0.15	0.12 ± 0.01	4.28 ± 0.16
24	0.94 ± 0.07	3.38 ± 0.16	1.51 ± 0.08	1.74 ± 0.27
Q345F/P134D				
Time (h)	CAT-3' (mM)	CAT-5 (mM)	CAT-3',5 (mM)	CAT (mM)
0	0.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01	10.00 ± 0.01
1	0.07 ± 0.01	0.59 ± 0.13	0.00 ± 0.01	9.06 ± 0.51
2	0.13 ± 0.01	1.31 ± 0.12	0.00 ± 0.01	7.66 ± 0.48
3	0.18 ± 0.01	1.85 ± 0.07	0.01 ± 0.01	6.62 ± 0.50
4	0.23 ± 0.01	2.34 ± 0.01	0.04 ± 0.01	5.66 ± 0.45
5	0.26 ± 0.01	2.67 ± 0.10	0.06 ± 0.01	4.81 ± 0.29
6	0.29 ± 0.01	3.02 ± 0.04	0.08 ± 0.01	4.21 ± 0.36
7	0.30 ± 0.01	3.14 ± 0.05	0.10 ± 0.01	3.70 ± 0.16
24	0.41 ± 0.01	4.05 ± 0.24	0.46 ± 0.01	1.77 ± 0.09

(+)-Catechin-3'-*O*-α-D-glucoside



MS (ESI positive):

Ion Formula: $C_{21}H_{24}O_{11}Na^+$ $[M+Na]^+$
m/z calculated: 475.1216
m/z experimental: 475.1220
error [ppm]: 0.8

^{1}H NMR (MeOD, δ): 7.20 (d, $^3J_{2'-6'}$ = 2.0 Hz, 1H, $H_{2'}$), 6.89 (dd, $^3J_{6'-2'}$ = 2.0 Hz, $^3J_{6'-5'}$ = 8.4 Hz, 1H, $H_{6'}$), 6.75 (d, $^3J_{5'-6'}$ = 8.2 Hz, 1H, $H_{5'}$), 5.83 (d, $^3J_{6-8}$ = 2.3 Hz, 1H, H_6), 5.75 (d, $^3J_{8-6}$ = 2.2 Hz, 1H, H_8), 5.22 (d, $^3J_{1*-2*}$ = 3.7 Hz, 1H, H_{1*}), 4.48 (d, $^3J_{2-3}$ = 7.8 Hz, 1H, H_2), 3.88 (td, $^3J_{3-2}$ = 8.2 Hz, $^3J_{3-4a}$ = 5.5 Hz, $^3J_{3-4b}$ = 8.5 Hz, 1H, H_3), 3.80-3.73 (m, $^3J_{3*-2*}$ = 9.3 Hz, $^3J_{6a*-6b*}$ = 20.1 Hz, 2H, H_{3*}, H_{6a*}), 3.72-3.62 (m, $^3J_{5*-4*}$ = 8.6 Hz, 2H, H_{5*}), 3.48 (dd, $^3J_{2*-1*}$ = 3.7 Hz, $^3J_{2*-3*}$ = 9.7 Hz, 1H, H_{2*}), 3.37-3.32 (m, $^3J_{4*-5*}$ = 8.6 Hz, $^3J_{6b*-6a*}$ = 18.4 Hz, 2H, H_{4*}, H_{6b*}), 2.80 (dd, $^3J_{4a-3}$ = 5.6 Hz, $^2J_{4a-4b}$ = 16.1 Hz, 1H, H_{4a}), 2.40 (dd, $^3J_{4b-3}$ = 8.5 Hz, $^2J_{4b-4a}$ = 16.1 Hz, 1H, H_{4b}).

^{13}C NMR (MeOD, δ): 157.9 (C_7), 157.6 (C_5), 156.9 (C_{10}), 148.8 ($C_{4'}$), 146.3 ($C_{3'}$), 132.3 ($C_{1'}$), 124.0 ($C_{6'}$), 118.7 ($C_{2'}$), 116.9 ($C_{5'}$), 101.7 (C_{1*}), 101.0 (C_9), 96.4 (C_6), 95.6 (C_8), 82.8 (C_2), 74.9 (C_{3*}), 74.6 (C_{5*}), 73.5 (C_{2*}), 71.3 (C_{4*}), 68.8 (C_3), 62.3 (C_{6*}), 29.0 (C_4).

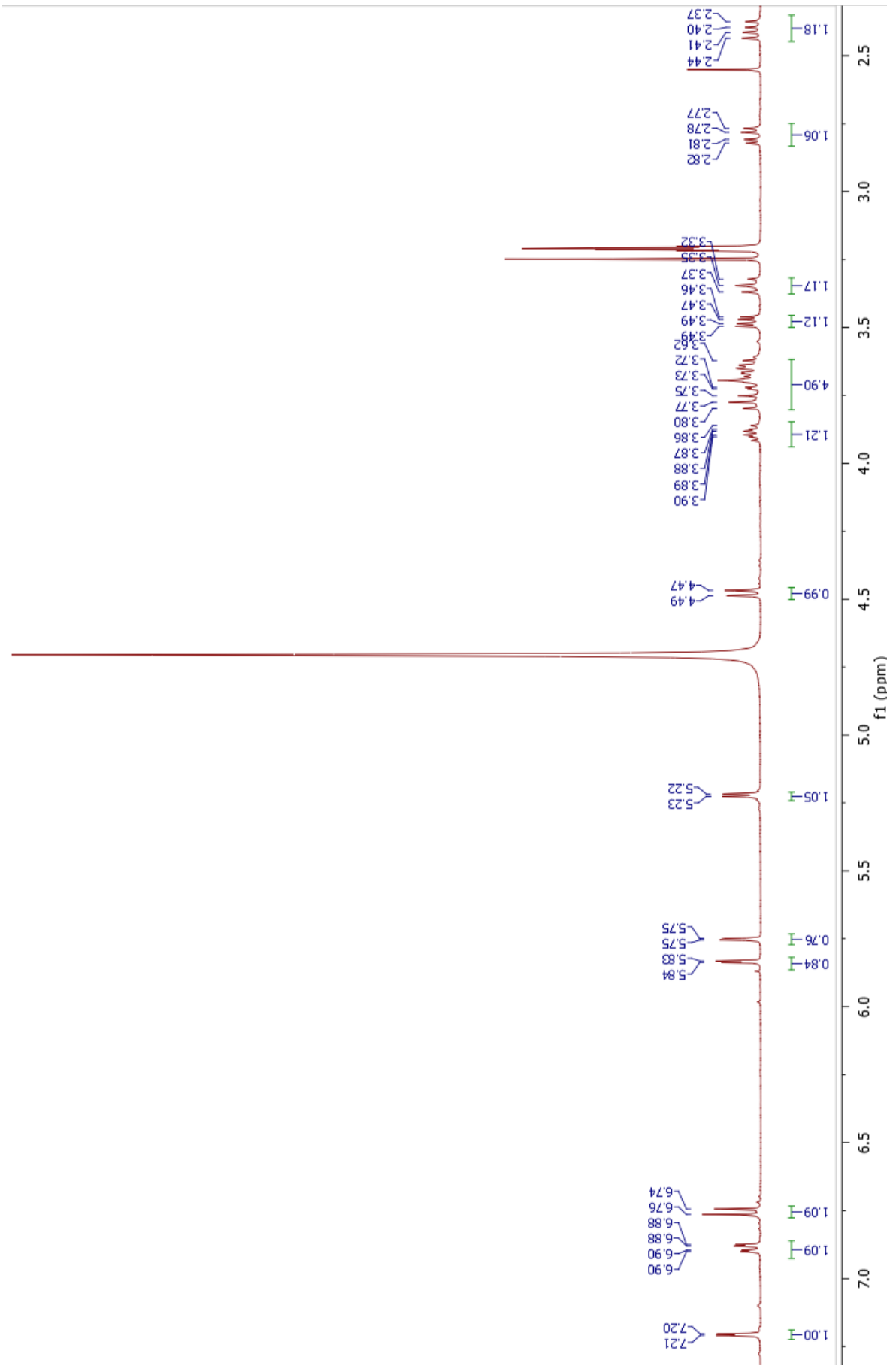


Figure S7 ^{1}H-spectrum of (+)-catechin-3'-*O*-α-D-glucoside.

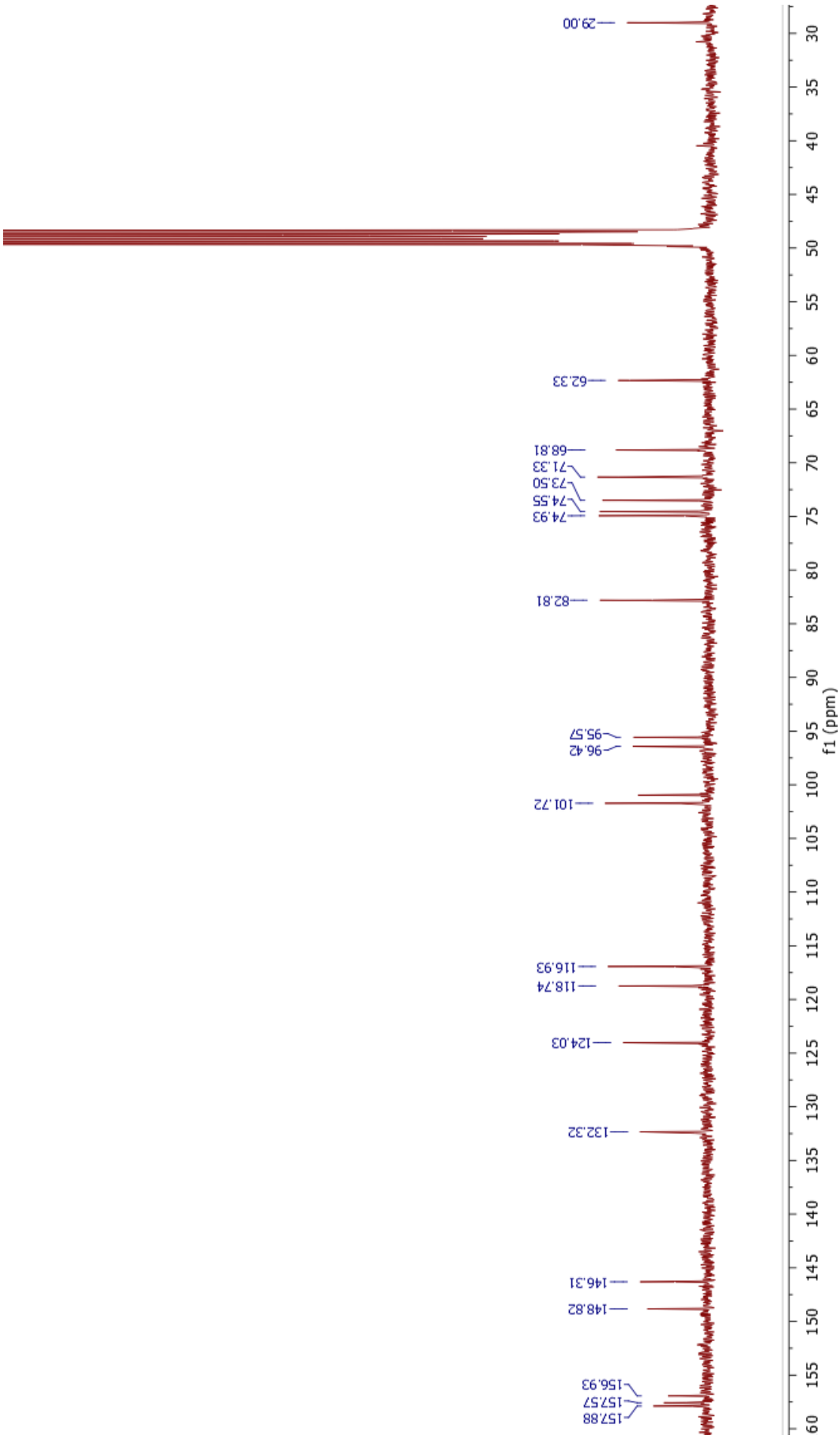
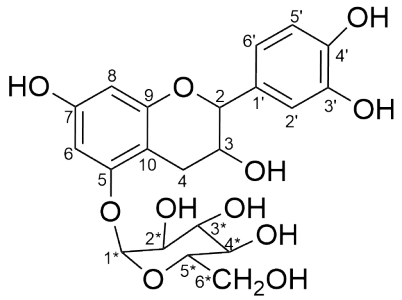


Figure S8 ^{13}C-spectrum of (+)-catechin-3'-*O*-α-D-glucoside.

(+)-Catechin-5-*O*-α-D-glucoside



MS (ESI positive):

Ion Formula: $C_{21}H_{24}O_{11}Na^+$ $[M+Na]^+$
m/z calculated: 475.1216
m/z experimental: 475.1219
error [ppm]: 0.6

^{1}H NMR (MeOD, δ): 6.72 (d, $^3J_{2'-6'}$ = 2.0 Hz, 1H, $H_{2'}$), 6.66 (d, $^3J_{5'-6'}$ = 8.1 Hz, 1H, $H_{5'}$), 6.60 (dd, $^3J_{6'-2'}$ = 2.0 Hz, $^3J_{6'-5'}$ = 8.2 Hz, 1H, $H_{6'}$), 6.23 (d, $^3J_{6-8}$ = 2.3 Hz, 1H, H_6), 5.93 (d, $^3J_{8-6}$ = 2.2 Hz, 1H, H_8), 5.39 (d, $^3J_{1*-2*}$ = 3.6 Hz, 1H, H_{1*}), 4.55 (d, $^3J_{2-3}$ = 6.9 Hz, 1H, H_2), 3.90 (td, $^3J_{3-4a}$ = 5.3 Hz, $^3J_{3-4b}$ = 7.6 Hz, $^3J_{3-2}$ = 7.2 Hz, 1H, H_3), 3.78-3.74 (m, $^3J_{3*-2*}$ = 9.3 Hz, $^3J_{6a*-6b*}$ = 17.8 Hz, 2H, H_{3*}, H_{6a*}), 3.63-3.61 (m, 2H, H_{4*}), 3.51-3.46 (m, $^3J_{5*-4*}$ = 6.6 Hz, $^3J_{2*-1*}$ = 3.6 Hz, $^3J_{2*-3*}$ = 9.3 Hz, 2H, H_{2*}, H_{5*}), 3.37-3.32 (m, $^3J_{6b*-5*}$ = 6.6 Hz, $^3J_{6b*-6a*}$ = 17.4 Hz, 1H, H_{6b*}), 2.76 (dd, $^3J_{4a-3}$ = 5.3 Hz, $^2J_{4a-4b}$ = 16.5 Hz, 1H, H_{4a}), 2.61 (dd, $^3J_{4b-3}$ = 7.6 Hz, $^2J_{4b-4a}$ = 16.5 Hz, 1H, H_{4b}).

^{13}C NMR (MeOD, δ): 158.1 (C_5), 157.2 (C_7), 156.5 (C_{10}), 138.3 ($C_{4'}$), 135.7 ($C_{3'}$), 131.4 ($C_{1'}$), 119.1 ($C_{6'}$), 116.1 ($C_{5'}$), 115.0 ($C_{2'}$), 103.2 (C_9), 98.7 (C_{1*}), 98.0 (C_8), 96.8 (C_6), 82.7 (C_2), 75.1 (C_{3*}), 74.5 (C_{5*}), 73.4 (C_{2*}), 71.4 (C_{4*}), 68.6 (C_3), 62.3 (C_{6*}), 30.0 (C_4).

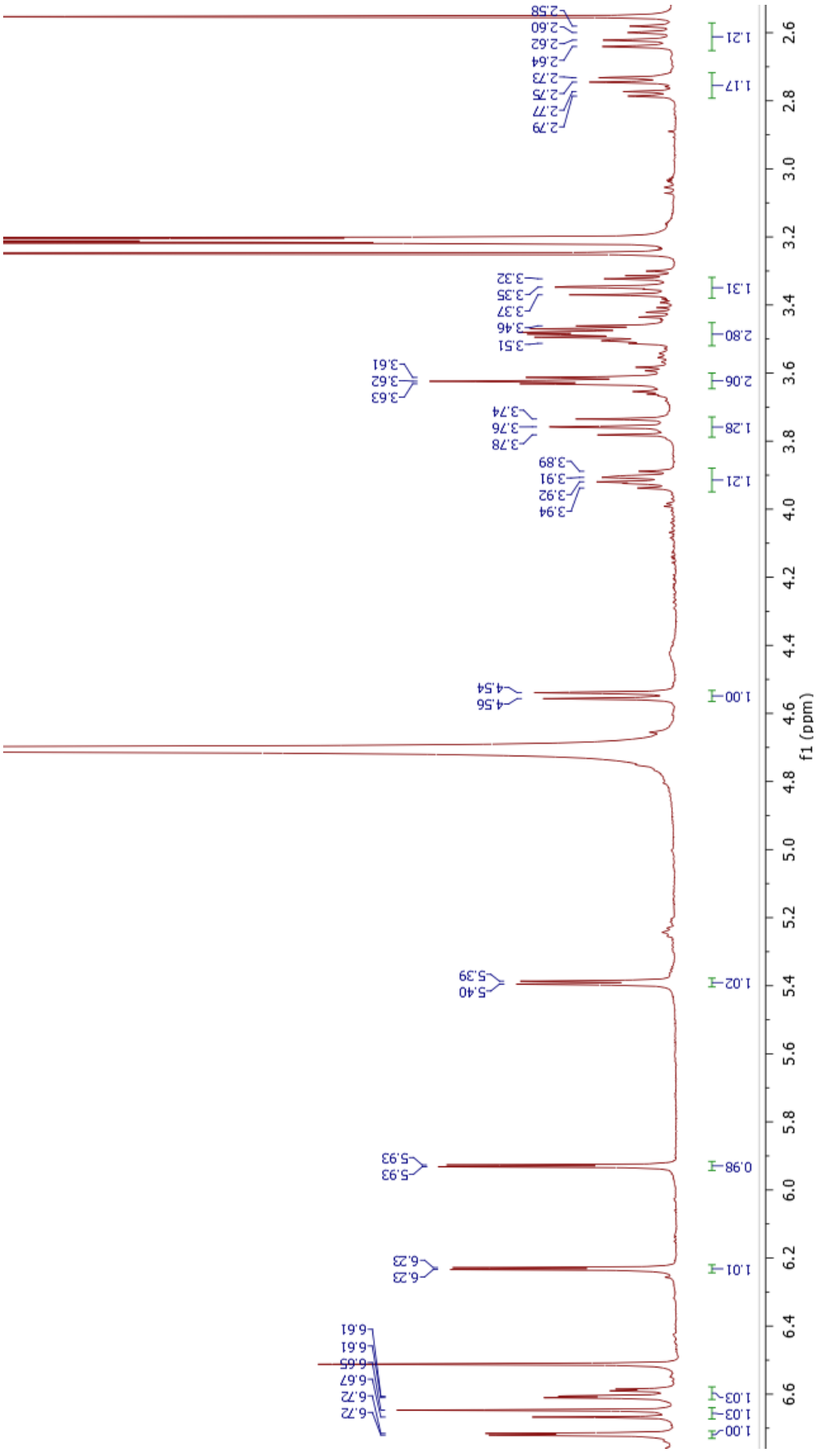


Figure S9 ^{1}H-spectrum of (+)-catechin-5-*O*-α-D-glucoside.

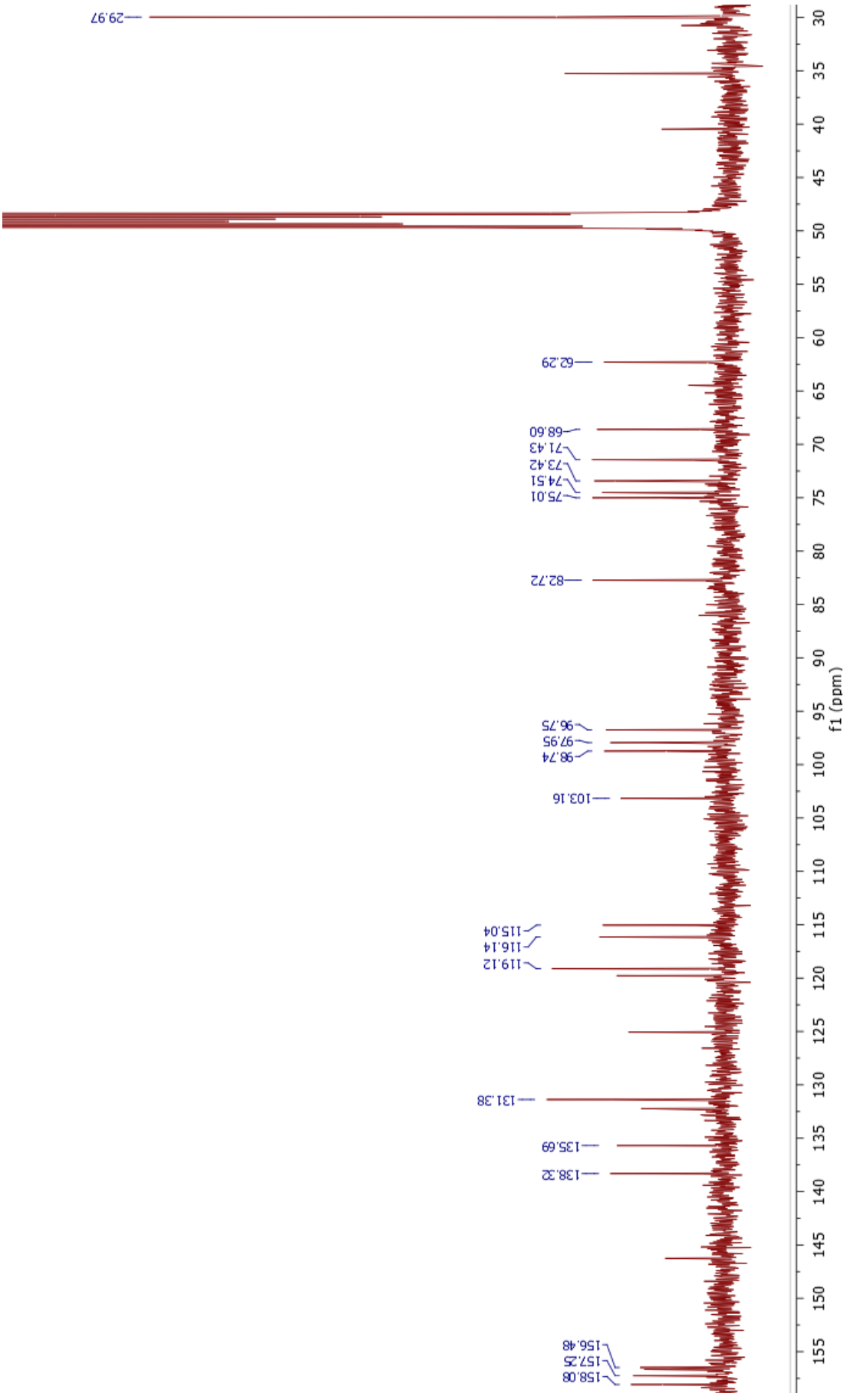
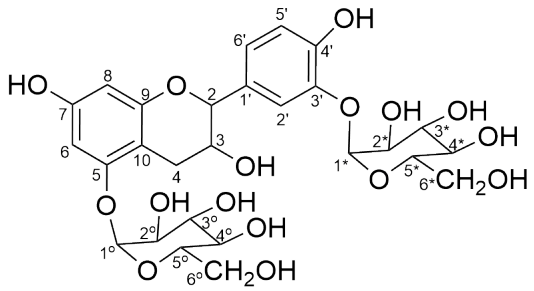


Figure S10 ^{13}C-spectrum of (+)-catechin-5-*O*-α-D-glucoside.

(+)-Catechin-3'.5-*O*-α-D-diglucoside



MS (ESI positive):

Ion Formula: $C_{27}H_{34}O_{16}Na^+$ $[M+Na]^+$
m/z calculated: 637.1745
m/z experimental: 637.1754
error [ppm]: 1.4

^{1}H NMR (MeOD. δ): 7.18 (d, $^3J_{2'-6'}$ = 1.8 Hz, 1H, $H_{2'}$), 6.86 (dd, $^3J_{6'-2'}$ = 1.9 Hz, $^3J_{6'-5'}$ = 8.3 Hz, 1H, $H_{6'}$), 6.74 (d, $^3J_{5'-6'}$ = 8.2 Hz, 1H, $H_{5'}$), 6.23 (d, $^3J_{6-8}$ = 1.9 Hz, 1H, H_6), 5.91 (d, $^3J_{8-6}$ = 2.1 Hz, 1H, H_8), 5.38 (d, $^3J_{1*-2*}$ = 3.6 Hz, 1H, H_{1*}), 5.20 (d, $^3J_{1^o-2^o}$ = 3.4 Hz, 1H, H_{1^o}), 4.54 (d, $^3J_{2-3}$ = 7.5 Hz, 1H, H_2), 3.93 (td, $^3J_{3-2}$ = 7.5 Hz, $^3J_{3-4a}$ = 5.4 Hz, $^3J_{3-4b}$ = 8.0 Hz, 1H, H_3), 3.79-3.73 (m, 2H, H_{3*}, H_{3^o}), 3.69-3.61 (m, 4H, H_{6*}, H_{6^o}), 3.51-3.45 (m, $^3J_{2*-1*}$ = 3.6 Hz, $^3J_{2^o-1^o}$ = 3.4 Hz, 4H, H_{5*}, H_{5^o}, H_{2*}, H_{2^o}), 3.37-3.31 (m, 2H, H_{4*}, H_{4^o}), 2.80 (dd, $^3J_{4a-3}$ = 5.4 Hz, $^2J_{4a-4b}$ = 16.5 Hz, 1H, H_{4a}), 2.60 (dd, $^3J_{4b-3}$ = 8.0 Hz, $^2J_{4b-4a}$ = 16.4 Hz, 1H, H_{4b}).

^{13}C NMR (MeOD, δ): 158.1 (C_5), 157.2 (C_7), 156.7 (C_{10}), 148.8 ($C_{4'}$), 146.3 ($C_{3'}$), 132.3 ($C_{1'}$), 132.3 ($C_{6'}$), 118.5 ($C_{2'}$), 117.0 ($C_{5'}$), 103.3 (C_9), 101.7 (C_{1^o}), 98.8 (C_{1*}), 98.0 (C_8), 96.9 (C_6), 82.7 (C_2), 75.0 (C_{3*}, C_{3^o}), 74.5 (C_{5*}, C_{5^o}), 73.4 (C_{2*}, C_{2^o}), 71.3 (C_{4*}, C_{4^o}), 68.6 (C_3), 62.3 (C_{6*}, C_{6^o}), 28.5 (C_4).

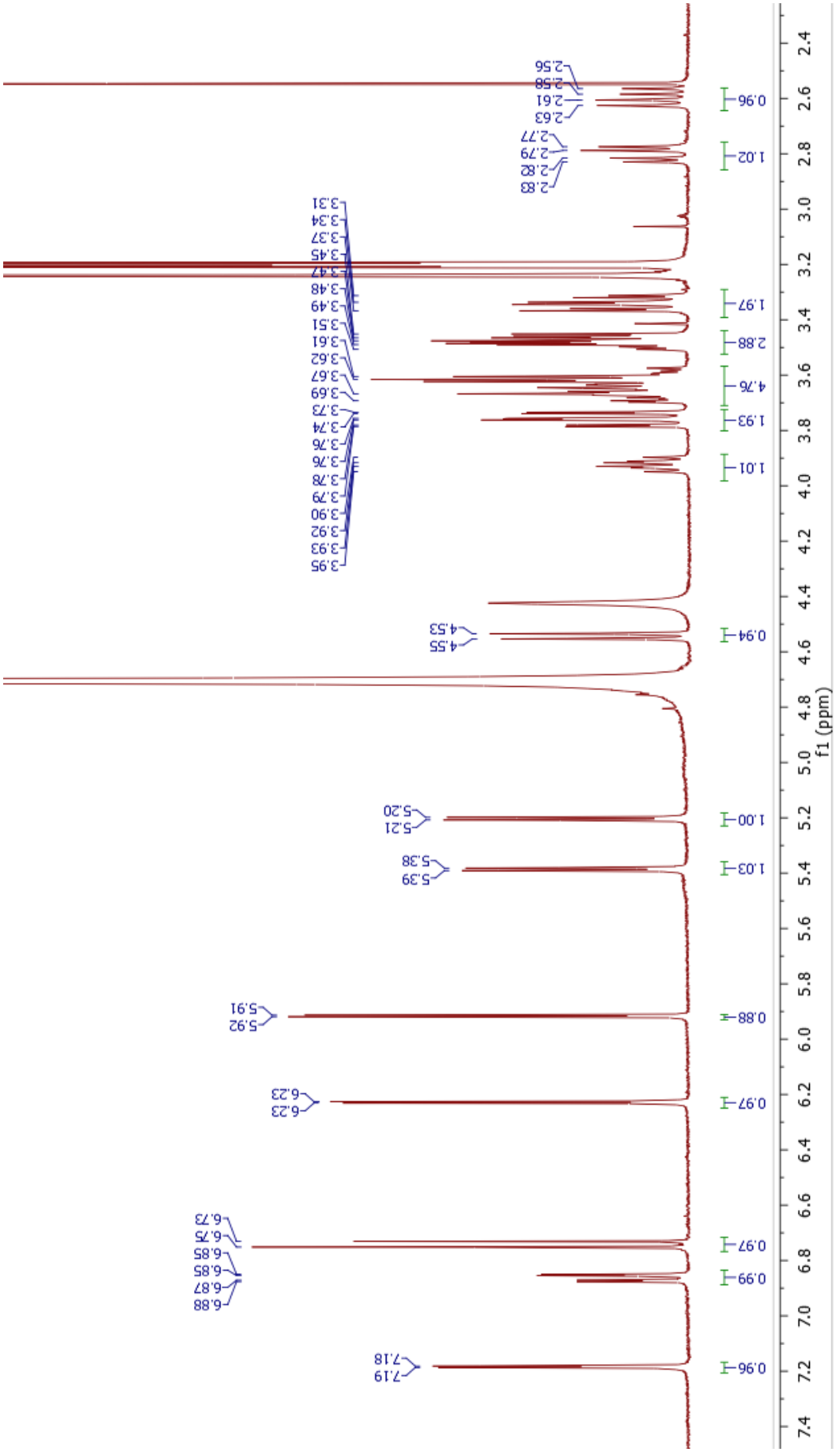


Figure S11 ^{1}H-spectrum of (+)-catechin-3' 5-*O*-α-D-diglucoside.

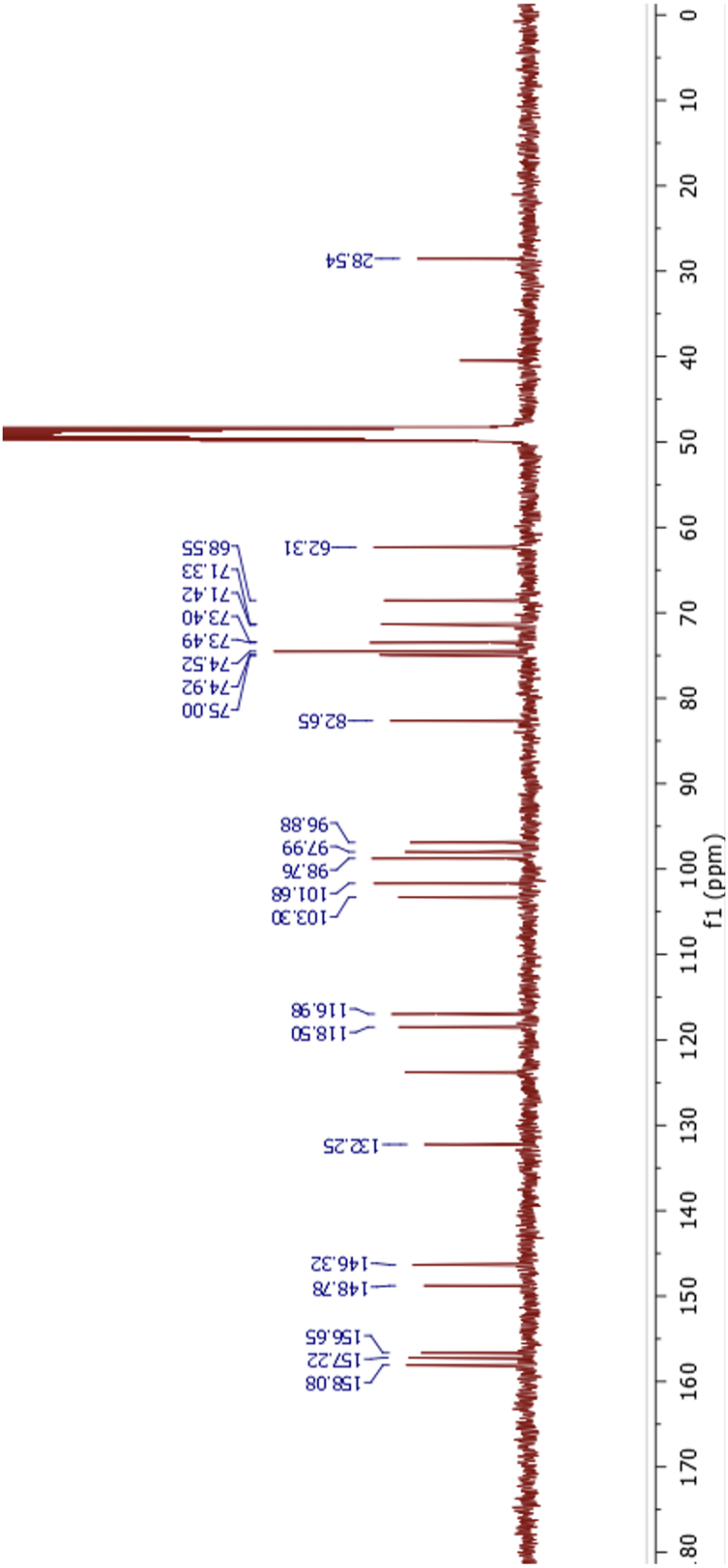


Figure S12 ^{13}C-spectrum of (+)-catechin-3',5-*O*-α-D-diglucoside.

2.8 Molecular modeling

File *DGC.params*:

```
  NAME DGC
IO_STRING DGC X
TYPE POLYMER
AA UNK
ATOM N Nbb NH1 -0.42
ATOM CA CAbb CT1 0.12
ATOM C CObb C 0.56
ATOM O OCbb O -0.46
ATOM HO4 Hapo X 0.15
ATOM CB CH2 X -0.13
ATOM CG COO X 0.67
ATOM OD1 OOC X -0.71
ATOM OD2 OH X -0.61
ATOM C1 CH1 X -0.04
ATOM C2 CH1 X -0.04
ATOM O2 OH X -0.61
ATOM HO9 Hpol X 0.48
ATOM C3 CH1 X -0.04
ATOM O3 OH X -0.61
ATOM H11 Hpol X 0.48
ATOM C4 CH1 X -0.04
ATOM O4 OH X -0.61
ATOM H13 Hpol X 0.48
ATOM C5 CH1 X -0.04
ATOM O5 OH X -0.61
ATOM C6 CH2 X -0.13
ATOM O6 OH X -0.61
ATOM H17 Hpol X 0.48
ATOM H15 Hapo X 0.15
ATOM H16 Hapo X 0.15
ATOM H14 Hapo X 0.15
ATOM H12 Hapo X 0.15
ATOM H10 Hapo X 0.15
ATOM HO8 Hapo X 0.15
ATOM HO7 Hapo X 0.15
ATOM HO5 Hapo X 0.15
ATOM HO6 Hapo X 0.15
ATOM HA Hapo HB 0.15
ATOM HO2 Hpol X 0.48
ATOM H HNbb H 0.36
LOWER_CONNECT N
UPPER_CONNECT C
BOND_TYPE N CA 1
BOND_TYPE N H 1
BOND_TYPE N HO2 1
BOND_TYPE CA C 1
BOND_TYPE CA CB 1
BOND_TYPE CA HA 1
BOND_TYPE C O 2
BOND_TYPE C HO4 1
BOND_TYPE CB CG 1
BOND_TYPE CB HO5 1
BOND_TYPE CB HO6 1
BOND_TYPE CG OD1 2
BOND_TYPE CG OD2 1
BOND_TYPE OD2 C1 1
BOND_TYPE C1 C2 1
BOND_TYPE C1 O5 1
BOND_TYPE C1 HO7 1
BOND_TYPE C2 O2 1
BOND_TYPE C2 C3 1
BOND_TYPE C2 HO8 1
BOND_TYPE O2 HO9 1
BOND_TYPE C3 O3 1
BOND_TYPE C3 C4 1
BOND_TYPE C3 H10 1
BOND_TYPE O3 H11 1
BOND_TYPE C4 O4 1
BOND_TYPE C4 C5 1
BOND_TYPE C4 H12 1
BOND_TYPE O4 H13 1
BOND_TYPE C5 O5 1
BOND_TYPE C5 C6 1
BOND_TYPE C5 H14 1
BOND_TYPE C6 O6 1
BOND_TYPE C6 H15 1
BOND_TYPE C6 H16 1
BOND_TYPE O6 H17 1
CHI 1 C1 C2 O2 HO9
PROTON_CHI 1 SAMPLES 3 60 -60 180 EXTRA 0
CHI 2 C2 C3 O3 H11
```

```
PROTON_CHI 2 SAMPLES 3 60 -60 180 EXTRA 0
CHI 3 C3 C4 O4 H13
PROTON_CHI 3 SAMPLES 3 60 -60 180 EXTRA 0
CHI 4 C5 C6 O6 H17
PROTON_CHI 4 SAMPLES 3 60 -60 180 EXTRA 0
CHI 5 N CA CB CG
CHI 6 CA CB CG OD1
PROPERTIES PROTEIN ALPHA_AA L_AA POLAR SC_ORBITALS
NBR_ATOM CB
NBR_RADIUS 10.918407
FIRST_SIDECHAIN_ATOM CB
NCAA_ROTLIB_PATH ncaa_rotamer_libraries/alpha_amino_acid/DGC.rotlib
NCAA_ROTLIB_NUM_ROTAMER_BINS 2 3 3
ICOOR_INTERNAL N 0.000000 0.000000 0.000000 N CA C
ICOOR_INTERNAL CA 0.000000 180.000000 1.460744 N CA C
ICOOR_INTERNAL C 0.000001 67.981434 1.523479 CA N C
ICOOR_INTERNAL O -44.247331 59.122068 1.238053 C CA N
ICOOR_INTERNAL HO4 -179.974631 60.462024 1.089949 C CA O
ICOOR_INTERNAL CB -123.761918 69.694379 1.543315 CA N C
ICOOR_INTERNAL CG 174.451314 66.711594 1.539869 CB CA N
ICOOR_INTERNAL OD1 -104.244818 61.909481 1.256230 CG CB CA
ICOOR_INTERNAL OD2 -178.600094 60.503989 1.431834 CG CB OD1
ICOOR_INTERNAL C1 -168.188337 47.730932 1.494219 OD2 CG CB
ICOOR_INTERNAL C2 -132.156458 68.705682 1.579054 C1 OD2 CG
ICOOR_INTERNAL O2 -91.422558 69.654790 1.412575 C2 C1 OD2
ICOOR_INTERNAL HO9 92.467830 70.496148 0.959841 O2 C2 C1
ICOOR_INTERNAL C3 -120.767066 65.857353 1.564914 C2 C1 O2
ICOOR_INTERNAL O3 -158.429267 69.872526 1.433261 C3 C2 C1
ICOOR_INTERNAL H11 -173.934081 70.486734 0.959968 O3 C3 C2
ICOOR_INTERNAL C4 120.743338 68.299903 1.526118 C3 C2 O3
ICOOR_INTERNAL O4 179.961806 70.502297 1.416475 C4 C3 C2
ICOOR_INTERNAL H13 -65.703315 70.533868 0.960042 O4 C4 C3
ICOOR_INTERNAL C5 -120.810247 69.077787 1.539931 C4 C3 O4
ICOOR_INTERNAL O5 -28.368901 64.671731 1.495147 C5 C4 C3
ICOOR_INTERNAL C6 -124.578167 67.827997 1.522159 C5 C4 O5
ICOOR_INTERNAL O6 49.352513 66.441896 1.418902 C6 C5 C4
ICOOR_INTERNAL H17 90.677875 70.463658 0.959979 O6 C6 C5
ICOOR_INTERNAL H15 -117.912592 72.937116 1.089833 C6 C5 O6
ICOOR_INTERNAL H16 -121.423558 71.529190 1.089931 C6 C5 H15
ICOOR_INTERNAL H14 -119.019831 77.127980 1.090201 C5 C4 C6
ICOOR_INTERNAL H12 -119.143102 71.571586 1.089861 C4 C3 C5
ICOOR_INTERNAL H10 119.113538 72.771955 1.089604 C3 C2 C4
ICOOR_INTERNAL HO8 -118.251414 75.121390 1.090370 C2 C1 C3
ICOOR_INTERNAL HO7 -131.052278 95.570395 1.090017 C1 OD2 C2
ICOOR_INTERNAL HO5 -118.083294 72.755505 1.090092 CB CA CG
ICOOR_INTERNAL HO6 -121.298012 71.428994 1.090261 CB CA HO5
ICOOR_INTERNAL HA -118.732314 72.431874 1.090272 CA N CB
ICOOR_INTERNAL HO2 -49.006579 60.000318 1.009686 N CA C
ICOOR_INTERNAL H -179.975613 60.010592 1.009564 N CA HO2
ICOOR_INTERNAL UPPER 149.999985 63.800007 1.328685 C CA N
ICOOR_INTERNAL O -180.000000 59.200005 1.231015 C CA UPPER
ICOOR_INTERNAL LOWER -150.000000 58.300003 1.328685 N CA C
ICOOR_INTERNAL H -180.000000 60.849998 1.010000 N CA LOWER
```

2.9 Molecular docking

Table S4 List of residues of the active site of the enzyme that were considered for the docking experiments. Pro134 was replaced by Asp134 for Q345F/P134D.

ILE8	TYR10	HIS38	LEU40	PRO41	PHE43	GLY48	ALA49	ASP50	ALA51
GLY52	PHE53	ASP54	PRO55	HIS58	ASP83	ALA84	ILE85	VAL86	ASN87
HIS88	MET89	MET115	SER116	SER117	VAL118	PHE119	PRO120	ASN121	ALA123
THR124	GLU125	GLU126	ASP127	LEU128	ILE131	TYR132	ARG133	PRO134	ARG135
PRO136	PHE140	LEU151	TRP153	SER155	PHE156	THR157	GLN159	GLN160	ASP162
ARG190	LEU191	DGC192	ALA193	VAL194	GLY195	TYR196	GLY197	LYS199	ALA201
GLY202	PHE206	MET207	THR211	ILE215	LEU230	ILE231	GLU232	VAL233	HIS234
SER235	TYR236	LYS239	GLN240	ILE243	TYR251	PHE253	THR288	HIS289	ASP290
GLY291	GLY293	VAL294	ILE295	ASP296	GLY298	SER299	ASP300	GLN301	LEU302
ARG304	LEU341	ASP342	LEU343	TYR344	PHE345	ASN347	TYR378	ASN397	ARG399
ASN402	ARG403								

Table S5 Constraints of distance for the selection of the productive poses obtained by molecular docking experiments.

(+)-Catechin hydroxyl	Enzyme carbon atom	Distance range
OH-3' or OH-5	carbon atom CD of GLU232	3.8 ± 0.5 Å
OH-3' or OH-5	anomeric carbon atom C1 of DGC192	3.5 ± 0.5 Å

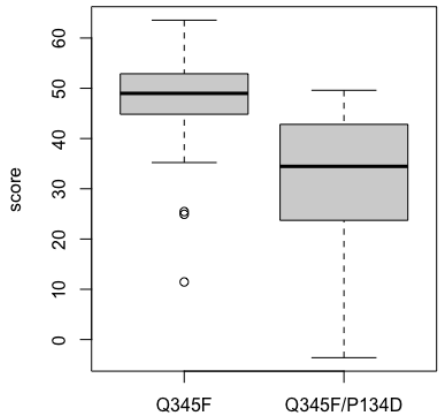


Figure S13 Double variant is less prone for (+)-catechin glucosylation at OH-5 position as showed by docking experiments. Comparison of the productive poses for glucosylation at OH-5 position for Q345F and Q345F/P134D. Difference in scores is significant as evaluated by a Student t-test ($p<0.01$).

Notes and references

[1] C. A. Rohl, C. E. Strauss, K. M. Misura and D. Baker, in *Methods in Enzymology*, Elsevier, 2004, vol. 383, pp. 66–93.

[2] M. Kraus, C. Grimm and J. Seibel, *ChemBioChem*, 2016, **17**, 33–36.

[3] W. L. DeLano *et al.*, *CCP4 Newsletter on Protein Crystallography*, 2002, **40**, 82–92.

[4] F. Lauck, C. A. Smith, G. F. Friedland, E. L. Humphris and T. Kortemme, *Nucleic Acids Research*, 2010, **38**, W569–W575.

[5] F. Allen and O. Johnson, *NATO ASI Series C Mathematical and Physical Sciences-Advanced Study Institute*, 1996, **480**, 55–66.

[6] G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *Journal of Molecular Biology*, 1997, **267**, 727–748.

[7] R. Schumacker and S. Tomek, *Springer*, 2013.

[8] A. C. Wallace, R. A. Laskowski and J. M. Thornton, *Protein engineering, design and selection*, 1995, **8**, 127–134.

[9] D. Aerts, T. Verhaeghe, M. D. Mey, T. Desmet and W. Soetaert, *Engineering in Life Sciences*, 2011, **11**, 10–19.