Synthesis of a non-natural glucose-2-phosphate ester able to dupe the acc system of Agrobacterium fabrum

Si-Zhe Li, Armelle Vigouroux, Mohammed Ahmar, Abbas El Sahili, Laurent Soulere, Laila Sago, David Cornu, Solange Moréra, Yves Queneau

Synthesis of a non-natural glucose-2-phosphate ester able to dupe the acc system of *Agrobacterium fabrum*

Si-Zhe Li, [a,#] Armelle Vigouroux, [b,#] Mohammed Ahmar, [a] Abbas El Sahili, [b] Laurent Soulère, [a] Laïla Sago, [b] David Cornu, [b] Solange Moréra, [b,*] Yves Queneau [a,*]

a. Univ Lyon, Institut de Chimie et Biochimie Moléculaires et Supramoléculaires, CNRS, Université Lyon 1, INSA Lyon, CPE Lyon, ICBMS, UMR 5246; Université Claude Bernard, Bâtiment Lederer, 1 Rue Victor Grignard, 69622 Villeurbanne Cedex, France. Yves.queneau@insa-lyon.fr

b. Institute for Integrative Biology of the Cell (I2BC), CNRS CEA Université Paris-Sud, Université Paris-Saclay, Avenue de la Terrasse, Gif-sur-Yvette 91198, France. Solange.morera@i2bc.paris-saclay.fr

# Equal contribution

Abstract. The first non-natural derivative of the rare D-glucose-2-phosphate (G2P), namely glucose-2-(*O*-lactic acid phosphate) (G2LP), has been synthesized. When used as sole carbon source, G2LP enables bacterial growth of the plant pathogenic strain *Agrobacterium fabrum C58* (formerly referred to as *Agrobacterium tumefaciens*). X-ray crystallography and affinity measurements investigations reveal that G2LP binds the periplasmic binding protein (PBP) AccA similarly to the natural compounds and with the same affinity. Moreover, enzymatic assays show that it is able to serve as substrate of the phosphodiesterase AccF. The properties found for G2LP demonstrates that the very unusual glucose-2-phosphoryl residue, present in G2LP, can be used as structural feature for designing non-natural systems fully compatible with the Acc cascade of *A. fabrum.*

Introduction

Carbohydrate phosphates and phosphodiesters are essential metabolites, being substrates, intermediates or products involved in a huge number of biological processes. With respect to glucose phosphates, the most common ones are glucose-6-phosphate (G6P) and glucose-1-phosphate (G1P), involved in the conversion of glucose to glycogen or starch, or their metabolism in most multicellular animals and plants. Comparing bibliographic databases using glucose-x-phosphate keywords is a sign of how much phosphorylation at other sites than O-1 or O-6 are far much less encountered (see for example hits in database searches for references mentioning

glucose-X-phosphate).‡ Whereas G1P and G6P are accepted abbreviations for glucose 1 and 6 phosphates, G2P or G3P more often refer to glycerol or glyceraldehyde derivatives. As a building block present in more complex systems, the glucose-2-phosphoryl residue was identified or suggested in some phosphorylated starches and glycogens,[1-3] and with respect to smaller systems, it was found only in a few biological molecules, namely the phosphodiesters agrocinopines C and D and the phosphoramidate agrocin 84 (Figure 1), produced by various agrobacteria stains.[4-9] Agrocin 84 is a natural antibiotic produced by the non-pathogenic bacterial strain *Agrobacterium radiobacter* K84 used as a biocontrol agent, being also imported by *Agrobacterium fabrum* C58 by the periplasmic binding protein (PBP) AccA associated to its ABC transporter. As *Agrobacterium tumefaciens* is now recognized as a complex of several species including *A. fabrum* to which belongs *A. fabrum* C58, here we use its new name although the most abundant literature is associated with the *A. tumefaciens* species complex, and especially with the strain C58.

Agrocin 84 structure is a phosphoramidate connecting a glucose moiety with the toxic moiety TM84, a di-substituted adenine nucleotide analog which inhibits leucine synthase once imported and delivered inside the cell, thus killing the bacteria. Initially proposed with a glucofuranosyl-1-phosphoryl moiety,[10] [11] its structure was then suggested to be possibly a pyranose system though with imprecise connection position,[11] and recently fully ascertained with the help of the X-ray structure of agrocin 84 bound to AccA.[12] Assessing this structure was essential for explaining how the same PBP AccA is able to bind several types of ligands. A few other G2P derivatives, namely the phosphodiester connecting two glucoses at *O*-2 and *O*-6, and a phenyl glucoside derivative were reported. [13-15]

The quite rare D-glucose-2-phosphate, which we will refer to as G2P in this paper, is known since the late 1940's from the work by Fleury et al and by Farrar, [16, 17] and has been observed as a product of the opening of the cyclic glucose-1,2-phosphate (GCP).[18] G2P was included in several studies on phosphate hydrolysis rates and phosphatases characterizations,[19] and used for analytical purposes for assessing starch or glycogen phosphorylation positions.[3, 20] As an isolated molecule, apart from the studies mentioned above, only one specific biological property has been established for G2P by El Sahili *et al*,[12] being able to activate quorum sensing (QS) and Ti plasmid transfer responsible for the virulence of the pathogenic bacteria *Agrobacterium fabrum* causing the crown gall. This property is associated with its ability to bind the AccR transcriptional repressor just like the natural ligand for this process, L-arabinose-2-phosphate (A2P)(Figure 1), does.[12] This common property of G2P and A2P was one of several clues unveiling the complete role of agrocinopine A, the 2-*O*-L-arabinose-4'-*O*-sucrose phosphodiester produced in the tumors which develop in plants

infected by *Agrobacterium fabrum* and re-imported in the bacteria by the PBP AccA to serve as nutrient and precursor of signals of QS regulated virulence.[12]

Overall, the ability of the rarely encountered glucosyl-2-*O*-phosphoryl moiety to bind AccA suggested that the two first steps in the *acc* cascade (AccA-AccF, import then degradation) could be used as a strategy for delivering molecules inside *Agrobacterium fabrum*. The work presented herein aims at probing this strategy, by preparing the first, non-natural, glucose-2-*O*-phosphodiester, namely D-glucose-2-(*O*-L-lactic acid phosphate) (G2LP), and by studying its ability to be used as carbon source by *A. fabrum*, its binding with AccA and its ability to be hydrolyzed by the phosphodiesterase AccF.

Results and discussion

G2LP synthesis relied on the preparation of an intermediate phosphite **5** bearing one benzylated glucosyl residue, one lactic acid benzyl ester residue and one benzyl residue (Scheme 1). Thus, benzyl (2S)-2-(((benzyloxy)(diisopropylamino)-phosphanyl)oxy)propanoate (**3**) was prepared from bis(diisopropylaminochlorophosphine) (**1**) and benzyl L-lactate, leading to the intermediate bisaminated phosphoramidite **2** which gave the monoaminated phosphoramidite **3** after substitution of only one diisopropylamino group by benzyl alcohol, thanks to the use of diisopropylammonium tetrazolide as mild and hindered base.[21]

Coupling of phosphoramidite **3** with the known benzyl 2,3,4,6-tetra-*O*-benzyl-α-D-glucopyranoside **4** (obtained by TIBAL-mediated selective 2-*O*-debenzylation[22] of perbenzylated glucose) proceeded in 84% yield in the presence of tetrazole as activating agent, giving the desired phosphite **5** in a 1:1 mixture of the two diastereoisomers at the chiral phosphorus atom. The phosphite **5** showed limited thermal stability and should be used rapidly for the subsequent oxidation step performed at 0°C using *tert*-BuOOH, affording the fully protected phosphate **6** in 80% yield. The alternative order for constructing the phosphite, namely coupling the glucose with the bisaminated phosphoramidite **2** followed by substitution with benzyl alcohol was found less efficient overall.

Palladium catalyzed hydrogenolysis of all 5 benzyl groups towards the target compound G2LP was found problematic due to the formation of the free lactic carboxylic acid which could be either hydrolyzed or transformed to an undesired methyl ester when the reaction was performed in methanol. Several conditions were tested (Table 1), showing that salinization of the medium from

the beginning using sodium or potassium carbonate was advantageous, as previously reported in the case of unstable substrates,[23] however in our case the reaction became incomplete. Finally the most favorable conditions were to use neutral aqueous ethanol as solvent then to salinize the medium with potassium carbonate after filtration of the catalyst and before any evaporation of the solvent.

Further ion exchange chromatography on a DEAE-Sephadex column led to D-glucose-2-(O-L-lactic acid phosphate) ammonium salt in 63% yield. In addition to mass spectroscopy characterization of G2LP, detailed NMR investigations ascertained the position of the phosphodiester on the 2-O-glucose and 2-O-lactic acid moieties (See 2D NMR in SI).

Having G2LP in hands, several biological investigations were conducted with respect to its uptake as nutrient by *Agrobacterium fabrum*, its ability to bind the PBP AccA and its ability to be hydrolyzed by the phosphodiesterase AccF, which are reported below. When present as the sole carbon source for *A. fabrum* C58 strain, G2LP was used as nutrient by the bacteria as indicated by bacterial growth (Figure 2). A similar behaviour was observed with L-arabinose-2-phosphate used as a sole carbon source meaning that G2LP was degraded in the bacterial cytoplasm once imported by AccA and its ABC transporter.

For assessing the ability of G2LP to bind AccA, an X-ray diffraction investigation was then performed and the structure of the PBP AccA in complex with G2LP at 1.8 Å was solved by molecular replacement using the structure of AccA in complex with G2P (PDB code 4RA1).[12] Both structures are very similar displaying an average root mean square deviation of 0.38 Å over all C□ atoms. As expected, G2LP is bound between the two closed lobes of AccA. The G2LP is well defined in its electron density maps and can present an □ or □-conformation for the O1 atom of the $^4C_1$ "chair" glucose moiety (Figure 3A). This was also observed for the bound G2P (PDB code 4RA1). Notably, the □-conformation O1 atom makes protein interactions while the □-conformation none. The G2P part of the bound G2LP overlaps the bound G2P when the two structures are superimposed making very numerous similar protein interactions (Figures 3B-D). In contrast, the lactate part of the G2LP interacts with AccA through water molecules. G2LP binding to AccA was investigated using tryptophan fluorescence spectroscopy. Titration experiments yielded an apparent K_D value of 2 ± 0.2 μM (Figure 4) which is very similar to the reported K_D value of 2.5 ± 0.5 μM for G2P.[12]

Finally, the ability of G2LP to serve as a substrate for the phosphodiesterase AccF was investigated. A mass spectrometry-based enzyme assay was performed by measuring substrate and product after

mixing AccF and G2LP. A close comparison of mass spectra showed that G2LP peak at 331.04 m/z was not detected in the presence of AccF and revealed a hydrolysis product at 281.24 m/z corresponding to a sodium adduct of G2P (Figure 5).

Conclusions

The newly synthesized D-glucose-2-(O-L-lactic acid phosphate) (G2LP) was found to enable *Agrobacterium fabrum* C58 growth, bind AccA and serve as substrate of AccF. This supports the idea that, mimicking *Agrobacterium radiobacter* which uses Agrocin 84 for competing with *Agrobacterium fabrum*, other derivatives possessing the glucose-2-phosphate moiety can mystify the PBP AccA, based on the unusual glucose-2-phosphoryl key for entering *Agrobacterium fabrum*.

Experimental

Chemistry – General information and synthesis of G2LP and intermediates 2, 3, 5 and-6

Reagents and solvents were purchased from Aldrich, Acros, Lancaster, Alfa Aesar, Fluka or TCI to be used without further purification. NMR spectra were recorded on a Bruker 300 MHz, Bruker 400 MHz or Bruker 500 MHz spectrometers. The chemical shifts (ppm) are referenced to the solvent residual peak and coupling constants are reported in Hz. The following abbreviations are used to explain the multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. Electrospray ionization (ESI) mass spectrometry (MS) experiments were performed on a Thermo Finnigan LCQ Advantage mass. High-resolution mass spectra (HRMS) were recorded on a Finnigan Mat 151xL mass spectrometer using electrospray. Analytical thin-layer chromatography was carried out on silica gel Merck 60 D254 (0.25 mm). Flash chromatography was performed on Merck Si 60 silica gel (40–63 μm). Optical rotations were measured on a Perkin Elmer 241 or Jasco P1010 polarimeter with a 10 cm cell (concentration c expressed as g/100 mL).

Benzyl (S)-2-((bis(diisopropylamino)phosphanyl)oxy)propanoate (2)

To a suspension of bis(diisopropylaminochlorophosphine) **(1)** (1.0 g, 3.76 mmol) in dry Et_2O (40 mL), a solution of benzyl L-lactate (812 mg, 4.512 mmol) and Et_3N (1.05 mL, 7.52 mmol) in dry Et_2O (10 mL) was added dropwise at 0 ^oC and then the white solid were removed by filtration, the solvent of the filtrate was evaporated, and the residue was further purified by chromatography

column (pentane/Et_3N=30:1) to yield benzyl (S)-2-((bis(diisopropylamino)phosphanyl)oxy)propano-ate **2** (1.21 g, yield 79%) as a colorless oil.

^{1}H NMR (300 MHz, Chloroform-*d*) δ 7.39 – 7.03 (m, 5H, ArH), 5.07 (s, 2H, $PhCH_2$), 4.40 – 4.04 (m, 1H, CH-O), 3.63 – 3.16 (m, 4H, 4CH), 1.36 (d, *J* = 6.8 Hz, 3H, CH_3), 1.19 – 0.80 (m, 24H, $8CH_3$). **^{31}P NMR (122 MHz, Chloroform-*d*)** δ 118.79. **^{13}C NMR (75 MHz, Chloroform-*d*)** δ: 173.6 (d, *J* = 3.9 Hz, C=O), 136.0 (ArC), 128.4 (2ArC), 128.1 (2ArC), 69.0 (PO-CH-CH_3), 66.2 ($PhCH_2$), 44.7 (CH-N-CH), 44.5 (CH-N-CH) 24.4 (CH_3), 24.3 (CH_3), 24.24 (CH_3), 24.19 (CH_3), 24.1 ($2CH_3$), 24.0 ($2CH_3$), 20.23 (d, *J* = 4.5 Hz, CH_3). **HRMS m/z (ESI):** Calcd. for $C_{22}H_{39}NaN_2O_3P$ $(M+Na)^+$: 433.2698. Found: 433.2616.

Benzyl (2S)-2-(((benzyloxy)(diisopropylamino)phosphanyl)oxy)pro-panoate (3)

To a solution of 1*H*-tetrazole (3.88 mL 0.45 M CH_3CN solution, 1.746 mmol) in 5 mL of anhydrous CH_2Cl_2 was added diisopropylamine (0.286 mL, 2.037 mmol) under argon at room temperature, the mixture was allowed to stir at the same temperature for 30 min, and the solvent was removed under reduced procedure to afford diisopropylammonium tetrazolide as a white solid. Then a solution of benzyl (S)-2-((bis(diisopropylamino)phosphanyl)oxy)propanoate (**2)** (1.21 g, 2.91 mmol) in dry CH_2Cl_2 (10.0 mL) was added dropwise to a stirred solution of diisopropylammonium tetrazolide (0.380 g, 2.22 mmol), which was dried by repeated co-evaporation with dry MeCN prior to use, and benzyl alcohol (0.318 mL, 3.054 mmol) in dry CH_2Cl_2 (20.0 mL) at 0 °C under argon. After being stirred for 4 h at 0 °C, the mixture was diluted with CH_2Cl_2 (30 mL) and washed with a saturated $NaHCO_3$ aqueous solution (30 mL). The aqueous layer was separated and extracted with CH_2Cl_2 (10 mL). The organic layers were combined, dried over Na_2SO_4, filtered and concentrated under reduced pressure. The residue was purified by silica gel column chromatography (hexane/AcOEt=30/1, v/v with 2% TEA) to afford benzyl (2S)-2-(((benzyloxy)(diisopropylamino)phosphanyl)oxy)propanoate as a colorless oil. The spectrum showed that the product is a pair of enantiomers (ratio 1/1).

^{1}H NMR (300 MHz, Chloroform-*d*) δ: 7.40 – 7.25 (m, 10H, ArH), 5.24 – 5.11 (m, 2H, $ArCH_2$), 4.79 – 4.64 (m, 2H, $PhCH_2OOC$-), 4.54 – 4.37 (m, 1H, -CH), 3.68 (m, 2H, N-CH), 1.49 (d, *J* = 6.9 Hz, 3H, CH_3), 1.25 – 1.10 (m, 12H, $4CH_3$). **^{13}C NMR** (75 MHz, Chloroform-*d*) δ: 173.1 (d, *J* = 2.8 Hz, C=O), 172.7 (d, *J* = 2.8 Hz, C=O), 139.5 (d, *J* = 7.8 Hz, ArC), 139.4 (d, *J* = 7.7 Hz, ArC), 135.7 (ArC), 128.5 (ArC), 128.2 (ArC), 128.1 (ArC), 127.3 (ArC), 127.1 (ArC), 127.0 (ArC), 126.9 (ArC), 68.9 (d, CH-O), 68.2 (d, CH-O), 66.6 (d, *J* = 7.0 Hz, $COOCH_2Ph$), 65.7 (d, $PhCH_2$), 65.4 (d, $PhCH_2$), 43.4 (CH), 43.2 (2CH), 43.0 (CH), 24.7 (CH_3), 24.6 (CH_3), 24.5 (CH_3), 24.4 (CH_3),

20.1 (d, *J* = 3.6 Hz, CH_3), 20.0 (d, *J* = 4.8 Hz, CH_3). **^{31}P NMR** (122 MHz, Chloroform-*d*) δ 148.63, 148.47. **HRMS** m/z (ESI): Calcd. for $C_{23}H_{32}NaNO_4P$ $(M+Na)^+$: 440.2069. Found: 433.2026.

Benzyl (2S)-2-(((benzyloxy)(((2S,3R,4S,5R,6R)-2,4,5-tris(benzyloxy)-6-((benzyloxy)methyl)tetrahydro-2H-pyran-3-yl)oxy)phosphanyl)-oxy)propanoate (5)

Benzyl (2S)-2-(((benzyloxy)(diisopropylamino)phosphanyl)oxy)-propanoate **(3)** (520 mg, 1.247 mmol) was stirred with 1*H*-tetrazole (0.45 mol/L in acetonitrile, 2.8 mL, 1.25 mmol) in dry CH_2Cl_2 (20 mL) under argon for 30 min at room temperature. Compound **4** (225 mg, 0.416 mmol) dissolved in dry CH_2Cl_2 (5 mL) was added and the resulting mixture was stirred for 2h. TLC (petroleum ether/ethyl acetate 10:1) showed that the reaction was completed. The reaction was diluted in 50 mL DCM and washed with 50 mL of saturated $NaHCO_3$ solution, the aqueous layer was washed with DCM (2x 20 mL), the combined organic layer was washed with brine (50 mL) and dried over Na_2SO_4, filtered, and concentrated under reduced pressure. Purification by silica gel column chromatography (petroleum ether/ethyl acetate 13:1) yield **5** (300 mg, 84%) as a colorless oil.
^{1}H NMR (400 MHz, Chloroform-*d*) δ: 7.45 – 7.28 (m, 27H, ArH), 7.24 (dd, *J* = 2.0, 1.6 Hz, 1H, ArH), 7.18 (m, 2H, ArH), 5.19 (d, *J* = 4.0 Hz, 0.5H, H-1), 5.19 – 5.16 (d, 1H, $ArCH_2$), 5.12 – 5.07 (m, 1H, $ArCH_2$), 5.09 (d, *J* = 4 Hz, 0.5H, H-1), 4.98 (d, *J* = 8.0 Hz, 0.5H, $ArCH_2$), 4.92 – 4.82 (m, 3.5H, $ArCH_2$), 4.81 – 4.67 (m, 3H, $ArCH_2$), 4.78 (dd, 0.5H, *J* = 2 Hz, CH_3CH-O), 4.75 (dd, 0.5H, *J* = 2 Hz, CH_3CH-O), 4.61 – 4.51 (m, 3H, $ArCH_2$), 4.35 – 4.28 (m, 1H, H-2), 4.08 (m, 1H, H-3), 3.92 (m, 1H, H-5), 3.79 (m, 1H, H-6), 3.76 – 3.71 (m, 1H, H-4), 3.70 – 3.64 (m, 1H, H-6), 1.39 (m, 3H, CH_3). **^{13}C NMR** (101 MHz, Chloroform-*d*) δ: 172.0 (d, *J* = 4.0 Hz, C=O), 171.9 (d, *J* = 2.0 Hz, C=O), 138.8 (2Glu-OCH_2-CPh), 138.7 (2Glu-OCH_2-CPh), 138.2 (P-OCH_2-CPh), 138.0 (d, *J* = 6.7 Hz, P-OCH_2-CPh), 137.5 (2Glu-OCH_2-CPh), 137.4 (2Glu-OCH_2-CPh), 135.5 (d, *J* = 2.0 Hz, $COOCH_2$CPh), 128.7 (ArC), 128.6 (2ArC), 128.5 (ArC), 128.4 (6ArC), 128.3 (2ArC), 128.2 (2ArC), 128.1 (ArC), 128.0 (3ArC), 127.9 (2ArC), 127.8 (4ArC), 127.7 (ArC), 127.6 (2ArC), 127.5 (2ArC), 127.16 (ArC), 97.8 (d, *J* = 2.0 Hz, C-1), 97.6 (d, *J* = 2.0 Hz, C-1), 81.4 (t, *J* = 4.0 Hz, 2C-3) , 78.0 (C-4), 77.9 (C-4), 75.5 (2OBn), 75.1 (2OBn), 73.9 (d, *J* = 10.0 Hz, C-2) , 73.6 (d, *J* = 9.0 Hz, C-2), 73.5 (2OBn), 70.6 (d, *J* = 6.5 Hz, 2C-5), 69.6 (OBn), 69.5 (OBn), 68.5 (C-6), 68.4 (C-6), 67.9 (d, *J* = 11.6 Hz, CH-O-P), 67.3 (d, *J* = 11.6 Hz, CH-O-P), 66.8 (d, *J* = 4.0 Hz, 2OBn), 64.1 (d, *J* = 10.0 Hz, P-OBn), 64.0 (d, *J* = 13.0 Hz, P-OBn), 20.0 (t, *J* = 3.8 Hz, $2CH_3$). **^{31}P NMR** (122 MHz, Chloroform-d) δ: 139.12, 138.88. **$[\alpha]_D$** = +22.5 (*c* 3.0, $CHCl_3$). **HRMS** m/z (ESI): Calcd. for $C_{51}H_{54}O_{10}P$ $(M+H)^+$: 857.3376. Found: 857.3449.

Benzyl (2S)-2-(((benzyloxy)(((2S,3R,4S,5R,6R)-2,4,5-tris(benzyloxy)-6-((benzyloxy)methyl)tetrahydro-2H-pyran-3-yl)oxy)phosphoryl)-oxy)propanoate (6)

To a solution of compound **5** (220 mg, 0.257 mmol) in 10 mL anhydrous DCM was added t-BuOOH (47 uL, 6 mol/L in decane, 0.283 mmol) at 0 ^{o}C, the reaction was allowed to stir at 0 ^{o}C for 2h, then room for 1h until the TLC (pentane/AcOEt=3:1) showed the starting material was consumed completely, then the reaction solvent was evaporated and the residue was purified by chromatography column (pentane/ether=1:1) to afford phosphate **6** (180 mg, yield=80%) as colorless syrup. The spectrum showed that the product is a pair of diastereomers with ratio 3/2.

^{1}H NMR (400 MHz, Chloroform-*d*) δ: 7.34 – 7.14 (m, 28H, ArH), 7.11 – 7.04 (m, 2H, ArH), 5.36 (d, *J* = 3.7 Hz, 0.6H, H-1), 5.22 (d, *J* = 3.7 Hz, 0.4H, H-1), 5.11 (m, 1H, $ArCH_2$), 5.06 – 4.88 (m, 3H, $ArCH_2$), 4.83 (d, *J* = 11.3 Hz, 1H, CHO-P), 4.87-4.81 (m, 0.5H, $ArCH_2$), 4.77 – 4.70 (m, 2H, $ArCH_2$), 4.70 – 4.59 (m, 2H, $ArCH_2$), 4.58 – 4.44 (m, 3.5H, $ArCH_2$), 4.43 (m, 0.6H, H-2), 4.35 (m, 0.4H, H-2), 4.03 – 3.93 (m, 1H, H-3), 3.84 - 3.79 (m, 1H, H-5), 3.68 (dd, *J* = 4.0, 4.0 Hz, 1H, H-6), 3.65 - 3.60 (m, 1H, H-4), 3.59 - 3.55 (m, 1H, H-6), 1.38 (d, *J* = 6.9 Hz, 2H, CH_3), 1.34 (d, *J* = 6.8 Hz, 1H, CH_3). **^{13}C NMR** (101 MHz, Chloroform-*d*) δ: 170.3 (d, *J* = 4.6 Hz, C=O), 170.1 (d, *J* = 5.8 Hz, C=O), 138.5 (d, *J* = 2.0 Hz, 2Glu-OCH_2-CPh), 138.1 (2Glu-OCH_2-CPh), 138.0 (d, *J* = 5.0 Hz, 2Glu-OCH_2-CPh), 137.4 (Glu-OCH_2-CPh), 137.2 (Glu-OCH_2-CPh), 135.8 (d, *J* = 8.0 Hz, P-OCH_2-CPh), 135.7 (d, *J* = 7.6 Hz, P-OCH_2-CPh), 135.3 ($COOCH_2$-CPh), 135.2 ($COOCH_2$-CPh), 128.7 (2ArC), 128.50 (2ArC), 128.4 (d, *J* = 2.3 Hz, 6ArC), 128.3 (d, *J* = 2.1 Hz, 6ArC), 128.2 (ArC), 128.1 (ArC), 128.0 (d, *J* = 2.1 Hz, 4ArC), 127.9 (ArC), 127.8 (d, *J* = 2.4 Hz, 4ArC), 127.7 (ArC), 127.6 (d, *J* = 5.1 Hz, 2ArC), 96.6 (C-1), 96.5 (C-1), 80.7 (C-3), 80.6 (C-3), 77.73 (C-4), 77.67 (C-4), 77.5 (d, *J* = 6.2 Hz, C-2), 77.30 (d, *J* = 6.2 Hz, C-2), 75.5 (OCH_2Ph), 75.4 (OCH_2Ph), 75.21 (d, *J* = 2.0 Hz, $2OCH_2Ph$), 73.52 (d, *J* = 2.9 Hz, $2OCH_2Ph$), 72.26 (d, *J* = 5.1 Hz, P-OCH), 72.04 (d, *J* = 5.2 Hz, P-OCH), 70.6 (C-5), 70.5 (C-5), 70.1 (OCH_2Ph), 70.0 (OCH_2Ph), 69.6 (d, *J* = 5.8 Hz, P-OCH_2Ph), 69.4 (d, *J* = 5.5 Hz, P-OCH_2Ph), 68.4 (C-6), 68.3 (C-6), 67.2 (d, *J* = 5.3 Hz, $2OCH_2Ph$), 19.1 (d, *J* = 5.2 Hz, CH_3). **^{31}P NMR** (122 MHz, Chloroform-*d*) δ: -2.38, -2.61. **$[\alpha]_D$** = +40.4 (*c* 5.5, $CHCl_3$). **HRMS** m/z (ESI): Calcd. for $C_{51}H_{53}ClO_{11}P$ $(M+Cl)^-$: 907.3020. Found: 907.3017.

D-Glucose-2-(*O*-L-lactic acid phosphate) ammonium salt (G2LP)

Pd/C (70 mg) was added to a solution of compound **6** (70 mg, 0.082 mmol) in ethanol and water (10 mL, v/v) and the mixture is stirred under H_2 atmosphere for 24 h at room temperature. The solution was then filtered through celite and K_2CO_3 was added to the filtrate, the mixture was stirred for 30 min, then the solvent was evaporated to give a residue as white solid, which was further

purified by Chromabond C18 column to afford 21mg desired compound as a potassium salt. The solid was dissolved in water and further purified by ion exchange chromatography on a DEAE-Sephadex column eluted with NH_4CO_3, to give **G2LP (**19 mg) in 63% yield (α/β=5/3).

^{1}H NMR (400 MHz, Deuterium oxide) δ: 5.36 (d, *J* = 3.6 Hz, 1H, H-1α), 4.68 (d, *J* = 7.9 Hz, 0.6H, H-1β), 4.60 – 4.46 (m, 1.6H, P-OCHαβ), 3.95 (td, *J* = 4.0, 8.0, 4.0Hz, 1H, H-2α), 3.89 – 3.82 (m, 1.6H, H-6αβ), 3.81 – 3.75 (m, 2.6H, H-5α, H-3α, H-2β), 3.73 – 3.65 (m, 1.6H, H-6αβ), 3.63 – 3.56 (m, 0.6H, H-3β), 3.47 – 3.42 (m, 2.2H, H-4αβ, H-5β), 1.40 (dd, *J* = 6.9, 2.8 Hz, 5H, CH_3αβ). **^{13}C NMR** (101 MHz, Deuterium oxide) δ: 180.4 (d, *J* = 4.7 Hz, C=O), 180.1 (d, *J* = 5.8 Hz, C=O), 95.0 (d, *J* = 5.1 Hz, C-1β), 90.9 (d, *J* = 2.0 Hz, C-1α), 78.9 (d, *J* = 6.6 Hz, C-2β), 75.9 (C-5β), 75.7 (d, *J* = 6.8 Hz, C-2α), 75.0 (d, *J* = 2.7 Hz, C-3β), 73.1 (t, *J* = 5.7 Hz, P-OCH), 71.5 (d, *J* = 5.6 Hz, C-3α), 71.2 (C-5α), 69.2 (C-4α), 69.0 (C-4β), 60.6 (d, *J* = 5.0 Hz, C-6β), 60.5 (C-6α), 20.2 (d, *J* = 2.3 Hz, CH_3β), 19.8 (d, *J* = 3.2 Hz, CH_3α). **^{31}P NMR** (122 MHz, Deuterium Oxide) δ: -0.86(β), -1.16 (α). **HRMS** m/z (ESI): Calcd. for $C_9H_{16}O_{11}P$ $(M-H)^-$: 331.0436. Found: 331.0441. **$[\alpha]_D$** = +45.0 (*c* 1.0, H_2O).

Biology – Materials and Methods

G2LP uptake assay

A single colony of the plant pathogen *Agrobacterium fabrum* C58 derivative pTi::Gm[24] was grown overnight at 28°C in AB media supplemented with mannitol (2 g/L) and gentamicin 25 µg.ml^{-1}. The bacteria pellet was washed twice with 0.8% NaCl. 500 □l of AB media in presence of a carbon source (G2LP or L-arabinose-2-phosphate at a final concentration of 1 mM) or in absence of any carbon source and supplemented with gentamicin (25 µg.ml^{-1}) were inoculated at an initial OD_{600} of 0.1 and the OD was monitored for 56 hours.

Structure of AccA in complex with G2LP

The AccA protein at 14 mg/ml, purified as described,[12] with 5 mM of G2LP cocrystallized in 25% PEG 4000, 0.2 M sodium acetate and 0.1 M sodium citrate pH 5.6. Crystals were transfered to a cryo-protectant solution containing the mother solution supplemented with 22 % (w/v) PEG 400 and flash frozen in liquid nitrogen. A 1.8 Å resolution diffraction data was collected at 100 K on the PROXIMA I beamline at SOLEIL synchrotron (Saint-Aubin, France). Data collection and processing statistics are given in Table 1 of ESI. Data processing was performed using the XDS package.[25] The structure determination of AccA-G2LP complex was performed by molecular

replacement with PHASER[26] using the structure of AccA in complex with G2P (PDB code 4RA1)[12] as a search model. The structure was refined with BUSTER-2.10[27] and inspection of the density maps and manual rebuilding were performed using COOT.[28] The three dimensional model of G2LP was generated using the ProDRG webserver.[29] Molecular graphics images were generated using PYMOL software (http://www.pymol.org).

Measurement of G2LP affinity with AccA

G2LP bound to AccA was monitored by autofluorescence by excitating the protein at a wavelength of 295 nm and monitoring the quenching of fluorescence emission of tryptophans at 335 nm. The experiment was performed at 22°C in microplates using Cary Eclypse spectrofluorometer (Varian) in 25 mM Tris-HCl pH 8.0 and 150 mM NaCl with a fixed amount of proteins (2 µM) and increasing concentrations of G2LP. G2LP has no emission signal at 335 nm. The data were analysed using Origin 7 software and fitted to the equation $f=\Delta Fluorescence_{max}*abs(x)/(KD+abs(x))$.

Investigation of G2LP activity with AccF by mass spectrometry

AccF purification. AccF expression plasmid was chemically synthesized using codon optimization for the expression in E. coli and inserted into pET-21b plasmid using NdeI and BamHI restriction enzyme (Genscript, Piscataway, NJ). E. coli BL21 competent cells transformed with pET21b-AccF were grown in LB media at 37°C until OD600 of 0.6. 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the culture for overnight expression at 25°C. The cells were pelleted by centrifugation at 8000 g for 15 min at 4°C, resuspended in 50 mM Tris-HCl pH 8, 300 mM NaCl and 20 mM imidazole and disrupted by sonication. After centrifugation at 25000 g for 30 minutes, the filtered supernatant is injected on a nickel affinity column (HiTrap 5 ml, GE Healthcare). After a washing step of 6% of 50 mM Tris-HCl pH 8, 150 mM NaCl and 300 mM imidazole (Buffer B), the protein is eluted with Buffer B and injected on a gel filtration Superdex 200 26/60 (GE Healthcare) using 50 mM Tris-HCl pH 8 and 150 mM NaCl. The protein fractions are pooled, concentrated at 5 mg/ml and stored at -80°C.

Mass spectrometry experiments. Mass spectrometry measurements were performed with an electrospray Triple-TOF 4600 mass spectrometer (ABSciex) used in negative polarity mode. For ESI–MS measurements, the Triple-TOF instrument was operated in RF quadrupole mode with the TOF data being collected between m/z 50–1000. Collision energy was set to 10 eV, and nitrogen was used as collision gas. External calibration was performed with PPG 300 µM in the m/z 100–

1000 mass range. The Analyst 1.6 and PeakView 1.2 softwares were used for acquisition and data processing, respectively. The monoisotopic masses are annotated in the spectrum, and the estimated mass accuracy is ±0.01 Da. For mass spectra acquisitions, direct analyzes were performed after sample dilution in 50% acetonitrile and 1% formic acid.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge SOLEIL for provision of synchrotron radiation facilities (proposals ID 20160782) in using Proxima 1 beamline. This work was supported by ANR-Blanc SENSOR (ANR-12-BSV8-0003-01/02/03), and benefited from the I2BC crystallization platform supported by FRISBI ANR-10-INSB-05-01 and the I2BC mass spectrometry SICaPS platform. A PhD grant to SZL from the China Scholarship Council is gratefully acknowledged, as well as financial support from CNRS and MESR.

Accession numbers

Coordinates and structure factors have been deposited at the Protein Data Bank (PDB) under accession code 6I7W.

Notes and references

‡ Hits in database searches for glucose-X-phosphate, (December numbers) in SciFinder: X=1: 7066; X=2: 19; X=3: 93; X=4: 13; X=5 : 14; X=6: 59332 ; and in Web of Knowledge : X=1: 3509; X=2: 9; X=3: 38; X=4: 7; X=5: 14; X= 6: 34806.

1 V. S. Tagliabracci, C. Heiss, C. Karthik, C. J. Contreras, J. Glushka, M. Ishihara, P. Azadi, T. D. Hurley, A. A. DePaoli-Roach and P. J. Roach, *Cell Metab.*, 2011, **13**, 274-282.

2 V. M. Chikwana, M. Khanna, S. Baskaran, V. S. Tagliabracci, C. J. Contreras, A. DePaoli-Roach, P. J. Roach and T. D. Hurley, *Proc. Natl. Acad. Sci. U S A*, 2013, **110**, 20976-20981.

3 S. Lim and P. A. Seib, *Cereal Chem.*, 1993, **70**, 145-152.

4 J. G. Ellis and P. J. Murphy, *Mol. Gen. Genet.*, 1981, **181**, 36-43.

5 J. G. Ellis, A. Kerr, A. Petit and J. Tempe, *Mol. Gen. Genet.*, 1982, **186**, 269-274.

6 G. T. Hayman and S. K. Farrand, *Mol. Gen. Genet.*, 1990, **223**, 465-473.

7 P. Oger and S. K. Farrand, *Mol. Microbiol.*, 2001, **41**, 1173-1185.
8 M. Lindberg, T. Norberg and S. Oscarson, *J. Carbohydr. Chem.*, 1992, **11**, 243-253.
9 M. Lindberg and S. Oscarson, *J. Carbohydr. Chem.*, 1993, **12**, 1139-1147.
10 W. P. Roberts, M. E. Tate and A. Kerr, *Nature*, 1977, **265**, 379-381.
11 R. J. Thompson, R. H. Hamilton and C. F. Pootjes, *Antimicrob. Agents Chemother.*, 1979, **16**, 293-296.
12 A. El Sahili, S. Z. Li, J. Lang, C. Virus, S. Planamente, M. Ahmar, B. G. Guimaraes, M. Aumont-Nicaise, A. Vigouroux, L. Soulere, J. Reader, Y. Queneau, D. Faure and S. Moréra, *PLoS Pathog.*, 2015, **11**, e1005071.
13 O. Holst, H. Moll, H. Brade and J. Thiem, *Carbohydr. Res.*, 1989, **188**, 219-221.
14 M. Franzkowiak, J. Thiem and C. Demoulin, *Carbohydr. Res.*, 1986, **158**, 13-35.
15 P. Rivaill and L. Szabo, *Bull. Soc. chim. Fr.*, 1963, 716-721.
16 P. Fleury, J. Courtois and A. Desjobert, *Bull. Soc. chim. Fr.*, 1948, 694-700.
17 K. R. Farrar, *J. Chem. Soc.*, 1949, 3131-3135.
18 F. C. Kokesh, D. A. Cameron, Y. Kakuda and P. V. Kuras, *Carbohydr. Res.*, 1978, **62**, 289-300.
19 F. C. Kokesh, R. K. Stephenson and Y. Kakuda, *Biochim. Biophys. Acta*, 1977, **483**, 258-262.
20 T. Whiteside, L. Carreira and S. Hilal, *QSAR Comb. Sci.*, 2007, **26**, 587-595.
21 N. Oka, Y. Morita, Y. Itakura and K. Ando, *Chem. Commun.*, 2013, **49**, 11503-11505.
22 T. Lecourt, A. Herault, A. J. Pearce, M. Sollogoub and P. Sinay, *Chemistry*, 2004, **10**, 2960-2971.
23 L. Y. Wu, M. O. Anderson, Y. Toriyabe, J. Maung, T. Y. Campbell, C. Tajon, M. Kazak, J. Moser and C. E. Berkman, *Bioorg. Med. Chem.*, 2007, **15**, 7434-7443.
24 E. Haudecoeur, S. Planamente, A. Cirou, M. Tannieres, B. J. Shelp, S. Moréra and D. Faure, *Proc. Natl. Acad. Sci. U S A*, 2009, **106**, 14587-14592.
25 Kabsch, Acta Crystallogr. D Biol. Crystallogr., 2010, **66**, 125-132.
26 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658-674.
27 E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea and G. Bricogne, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 2210-2221.
28 P. Emsley and K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 2126-2132.
29 A. W. Schuttelkopf and D. M. van Aalten, *Acta Crystallogr. D Biol. Crystallogr.*, 2004, **60**, 1355-1363.

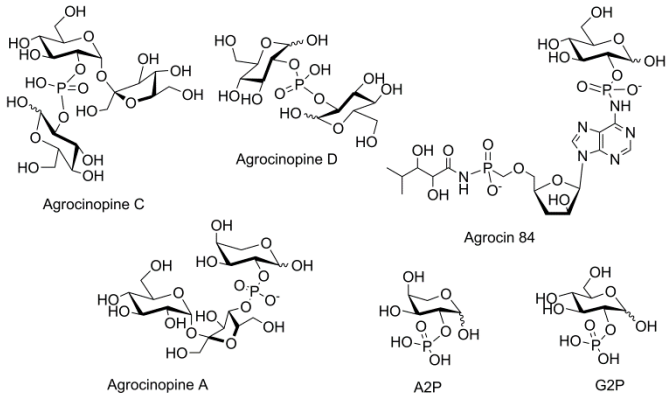
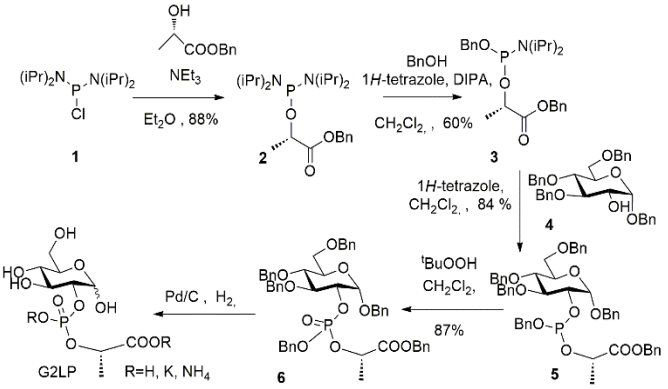


Figure 1. Structures of D-glucose-2-phosphate (G2P), its esters agrocinopines C and D, its amide agrocin 84, and D-arabinose-2-phosphate (A2P) and its ester agrocinopine A.



Scheme 1. Preparation of 2-O-D-glucose-2-O-L-lactic acid phosphate.

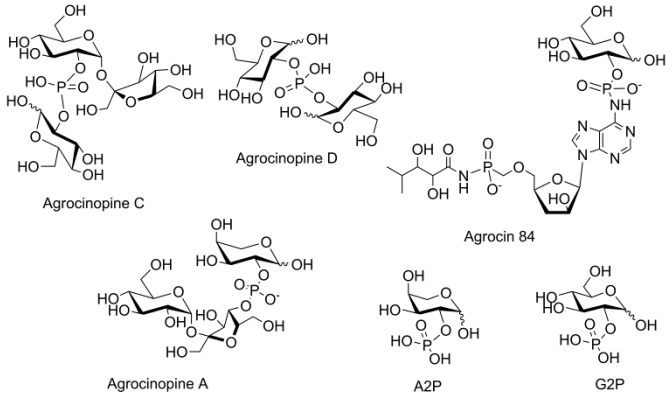


Figure 1. Structures of D-glucose-2-phosphate (G2P), its esters agrocinopines C and D, its amide agrocin 84, and D-arabinose-2-phosphate (A2P) and its ester agrocinopine A.

Table 1. Optimization of hydrogenolysis conditions leading to G2LP

Entry	Conditions	Results
1	CH_3OH	$R=CH_3$ [a]
2	THF/H_2O (1/1, v/v)	M=390 [b]; and M=332 [b]
3	Na_2CO_3/methanol/H_2O (1 eq, 1/1, v/v)	incomplete debenzylation
4	$K_2CO_3/THF/H_2O$ (1 eq, 1/1, v/v)	incomplete debenzylation
5	ethanol/H_2O (1/1, v/v)	M=332 [c]
6	Ethanol/H_2O (1/1, v/v), K_2CO_3 (1 eq)	R=K, and $R=NH_4^+$ [d]

a. transesterification product. b. the mass was measured in both positive and negative ionization mode. c. decomposition after purification. d. purification by ion exchange chromatography on a DEAE-Sephadex column eluted with NH_4CO_3.

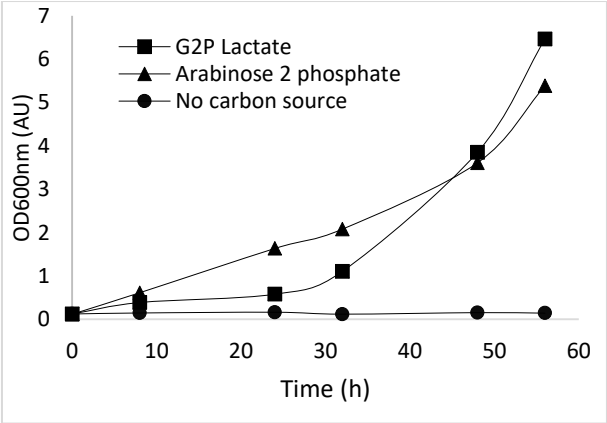


Figure 2: G2LP is used as a carbon source. OD monitoring (600 nm) of cultures in presence of G2LP (black squares) L-arabinose-2-phosphate (black triangles) as a carbon source and in absence of any carbon source (black circles) in AB minimum media.

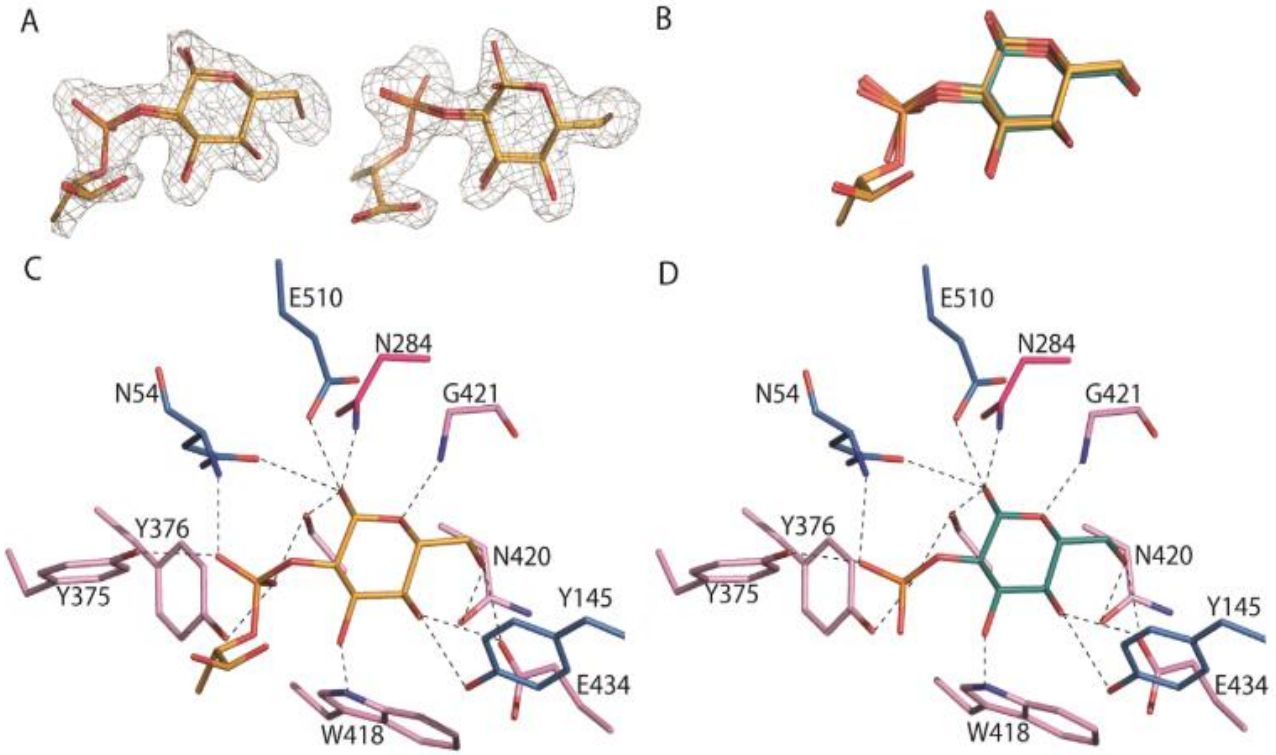


Figure 3: G2PL bound to the ligand-binding site of AccA (A) Two views of G2PL shown in sticks in its annealing Fo-Fc omit map contoured at 3σ. (B) Superposition of the bound G2PL and G2P shown in yellow and green sticks, respectively, in the binding site of AccA (C) G2PL, (D) G2P bound to the binding site of AccA are shown in the same code colour as in B. Hydrogen bonds between AccA (labelled amino acids shown in pink and slate) and each ligand are shown as dashed lines in black (distances are up to 3.2 Å).

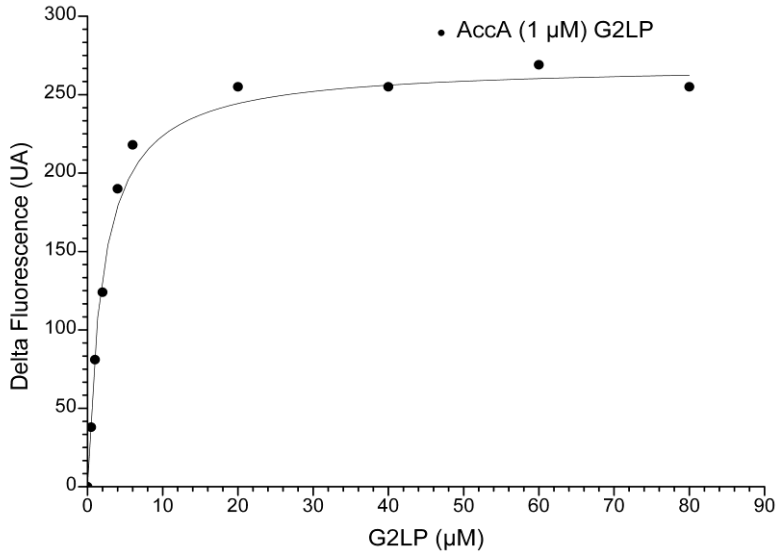


Figure 4. AccA fluorescence monitoring upon titration with G2PL and fit (solid line) to a single binding model using Origin. Measures were done in triplicates.

Overall, a clear proof of the hydrolysis of the phosphodiesters linkage of G2LP by the phosphodiesterase AccF is provided. This, combined with the ability to bind AccA, is in adequation with the ability of *A. fabrum* to grow when G2LP is the unique source of carbon.

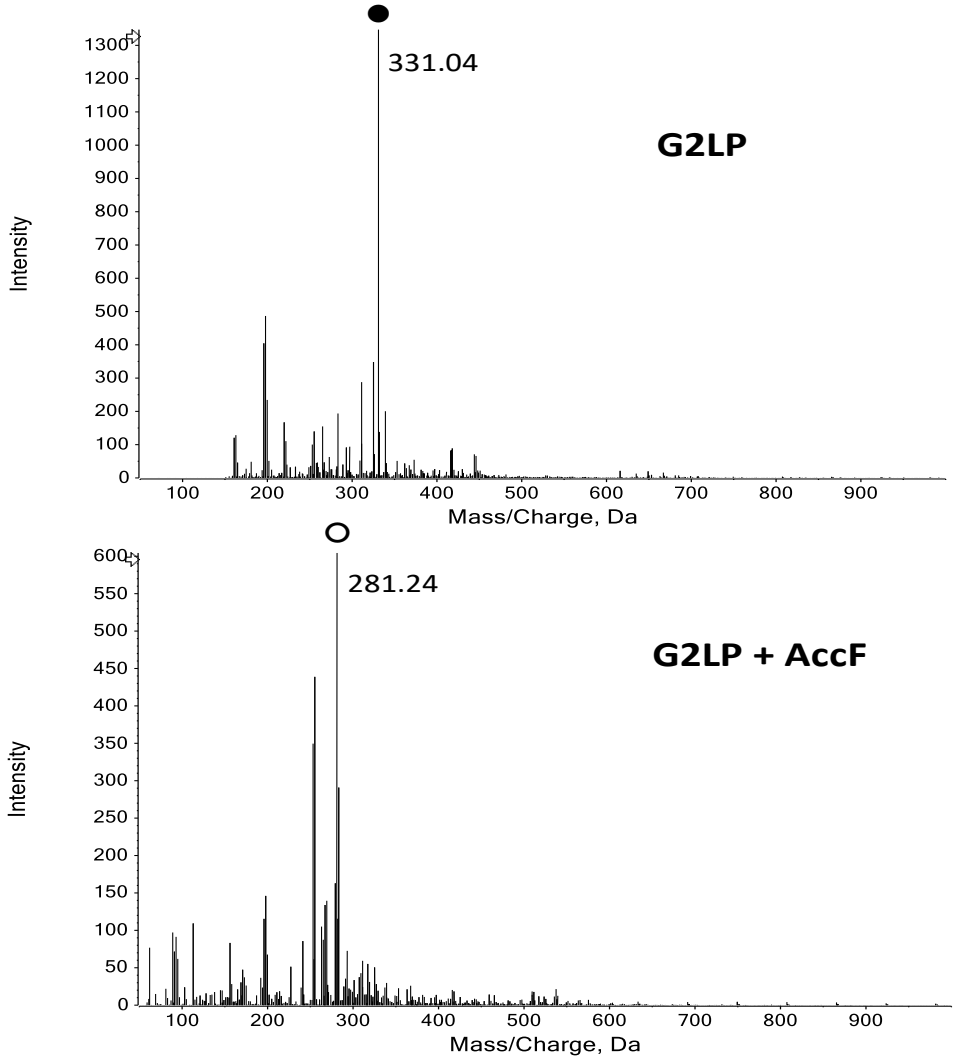


Figure 5. Enzymatic activity of AccF monitored by mass spectrometry: Mass spectra of G2LP (upper panel) and G2LP mixed with AccF (lower panel) are shown. Peaks marked with filled and empty circles correspond to G2LP and a G2LP hydrolysis product, respectively. In lower spectrum, G2LP peak at 331.04 m/z is not detected while a peak at 281.24 is exclusively observed. This latter peak derived from the hydrolysis of G2LP corresponds to a sodium adduct of G2P (G2PNa).