Green Chemistry

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/greenchem

Journal Name

RSCPublishing

ARTICLE

Cite this: DOI: 10.1039/x0xx00000x
Received 00th January 2012,
Accepted 00th January 2012
DOI: 10.1039/x0xx00000x
www.rsc.org/

Efficient chemoenzymatic synthesis of lipo-chitinoligosaccharides plant growth promotors

R. Chambon,[a] G. Despras,[b] A Brossay,[b] B. Vauzeilles,[b,c] D. Urban,[b] J.-M. Beau,[b,c] S. Armand,[a] S. Cottaz,*[a] and S. Fort[a]

Lipo-chitinoligosaccharides (Nod and Myc LCOs) are molecules involved in symbiotic phenomena in the plant kingdom. They play a major role in the process of atmospheric nitrogen fixation and mineral soil nutrients uptake both in legumes and in non-legumes, and are active at extremely low concentrations down to the nano- and even picomolar range. These compounds contain various substitutions along the oligosaccharidic backbone of the molecule including an essential fatty acid on the non-reducing unit and are considered as environmentally-friendly fertilizers. Currently, chemical synthesis cannot produce precursors of Nod and Myc LCOs at a large scale and an *in vitro* chemo-enzymatic pathway is presented here as a new and efficient method for preparing quantities of these high-value oligosaccharides. VC1280 (*Vibrio cholerae*) is a chitin deacetylase (CD) able to regioselectively cleave an acetate from the non-reducing penultimate *N*-acetyl D-glucosaminyl (GlcNAc) unit of chitinoligosaccharides (COs). This provides a free amino group which can be further *N*-acylated with a fatty-acid chain to give analogues of LCOs. Alternatively the non-reducing GlcNAc unit can be removed by β-*N*-acetylglucosaminidase treatment, followed by *N*-acylation to give natural LCOs. VC1280 CD was produced in the periplasm of *E. coli*. In the conditions used, 120 mg of pure enzyme was recovered from 1L of culture medium. For the first time, *in vitro* production of a library of natural LCOs as well as analogues has been carried out at preparative scale from biosourced chitinoligosaccharides constituting an approach of major interest for sustainable agriculture.

Introduction

Nodulation factors (Nod factors) are key signal molecules for nodule development and bacterial invasion during initial stages in the *Rhizobium*-legume symbiosis.[1-3] They consist of lipo-chitinoligosaccharides (LCOs) and are produced by rhizobia, including the genera *Allorhizobium, Azorhizobium, Bradyrhizobium, Mesorhizobium, Rhizobium* and *Sinorhizobium.* Such symbioses result in the formation of root nodules, new organs colonized by differentiated bacteria, which fix atmospheric nitrogen and provide it to their respective host plant, thereby promoting plant growth independently of the available soil nitrogen.

While Nod LCOs are involved in the *Rhizobium*-legume symbiosis, another kind of symbiosis, between mycorrhizal fungi and plants implicates LCOs.[4] Arbuscular mycorrhiza is a root endosymbiosis between fungi of the ancient phylum Glomeromycota and terrestrial plants. The arbuscular mycorrhiza fungus *G. intraradices* secretes a mixture of sulfated and non-sulfated LCOs (called Myc-LCOs because they stimulate mycorrhiza formation in legumes and non-legumes) that have structural similarities with

rhizobial Nod LCOs. Nod and Myc LCOs are extremely active molecules, down to the nano- even picomolar range, and quantities required for fertilization are less than 1 mg per hectare of planted seed!

All these properties make Nod and Myc LCOs natural and biodegradable molecules with a great promise to improve the yield of crops and reduce the use of fertilizers.

Nod and Myc LCOs consist of an oligomeric backbone of β-1,4-linked *N*-acetyl-D-glucosaminyl residues, *N*-acylated with a fatty acid chain at the non-reducing terminal residue.[4-6] Their structures differ in the number of GlcNAc residues present in the chitinoligosaccharidic backbone, with a degree of polymerization (DP) of generally four and five glucosaminyl units. There are also variations in the nature of the fatty acyl group, and in the substitution of the reducing and non-reducing terminal residues (e.g. *O*-acetyl, *O*-fucosyl, *O*-carbamoyl or *O*-sulfate group).

The availability of LCOs and their chitinoligosaccharidic precursors is a critical issue. Rizhobia culture at industrial scale is a process dedicated more for their use as inoculants rather than for extraction and purification of LCOs.[7] Chemical synthesis has only enabled access to minute quantities of such compounds so far, despite its potential for large scale production thanks to considerable progress in synthetic methodologies.[8] Precursors of Nod factors have been produced by high-cell-density culture of metabolically engineered *E.coli*.[9] This latter approach is hampered by the difficulty to purify mixtures of partially de-*N*-acetylated chitinoligosaccharides from the bacterial cells. As an alternative, we selected chitin as starting renewable biomolecule. Chitin is one of the most abundant naturally occurring polymers since it is mainly found in the cuticles of all arthropods, the endoskeletons of all cephalopods, but also in the cell walls of fungi and yeasts. Chitin is a linear copolymer of β(1→4)-linked D-glucosamine (GlcN) and GlcNAc units in varying proportions.[10] Chemical or enzymatic depolymerization of chitin is hampered by its highly crystalline structure and water insolubility, leading mainly to GlcNAc and chitinbiose (CO-II)[11] while higher oligomers are obtained only in low yield. Partially deacetylated chitin, namely chitosan, is water soluble and its enzymatic hydrolysis efficiently provides oligosaccharides in a wide range of molecular weights. These oligomers are produced industrially and marketed for cosmetic application. *N*-acetylation of chitosan oligomers would provide an efficient way to produce chitinoligosaccharides (COs) which could serve as precursors of LCOs by *in vitro* selective de-*N*-acetylation catalyzed by chitin deacetylase (CD). CD (EC 3.5.1.-) are produced by marine bacteria, several fungi and insects. They hydrolyze the acetamido group of GlcNAc units of chitin or COs, generating GlcN units and acetic acid. They are members of the carbohydrate esterase family 4 (CE4), as defined in the CAZy database (http://www.cazy.org).

Vibrio cholerae CD is acting on the penultimate *N*-acetylglucosaminyl (GlcNAc) unit of COs. It has been suggested that this enzyme acts in the same way as rhizobium NodB CD in the biosynthesis of Nod LCOs, albeit with a significant difference in the location of the hydrolyzed acetamido group from the COs backbone.[12] The objective of this work was to express and purify *Vibrio cholerae* CD in *E. coli* to examine a preparative scale production of Nod and Myc LCOs precursors, thus providing new route to these important molecules and analogues as well by subsequent fatty acid *N*-acylation.

Experimental

General

Reagents for bacterial media were from Euromedex (Mundolsheim, France) and Invitrogen (Cergy-Pontoise, France). Reagents for molecular biology were obtained from Invitrogen, Euromedex, Macherey-Nagel (Hoerd, France) and Thermo Scientific (Villebon sur Yvette, France). FACOS™ (a low molecular weight chitosan inferior to 2000 $g.mol^{-1}$ with an acetylation degree of 10 %) was purchased from Kitto Life Co (Kyongki-Do, South Korea). Chemicals and *Aspergillus oryzae* β-galactosidase containing β-*N*-acetylglucosaminidase were purchased from Sigma-Aldrich Chimie (Saint Quentin-Fallavier, France).

Thin Layer Chromatography

Reactions were monitored by thin layer chromatography (TLC) using Silica Gel 60 F254 precoated plates (E. Merck, Darmstadt) and propanol/H_2O/ammonia (70:30:1) as a solvent. Detection of carbohydrates was achieved by charring with the diphenylamine/aniline/orthophosphoric acid reagent.[13]

Analytical size-exclusion chromatography (HPSEC)

COs and CO(N)s[11] were analysed by HPSEC. The system consisted of a Shodex OH pack SB-G precolumn in front of a series of a Shodex OH pack SB-802 and Shodex OH pack SB-802.5 columns (8*300 mm), connected to a pump (Waters 510, Millipore). Samples were solubilized in water. Elution was performed with a 100 mM

sodium nitrate solution at a flow rate of 0.5 mL min^{-1} and monitored by a refractometric detector (Waters 410, Millipore).

Mass spectrometry analysis

The mass measurements of chitinoligosaccharides were performed using a MALDI-ToF/ToF AutoFlex I (Bruker) for low resolution analysis, and using a ESI-QToF (Xevo GS2-S QToF-Waters) for high resolution (HRESIMS) analysis (PSM facility, PCN-ICMG, Grenoble).

NMR procedures

NMR spectra were recorded on Bruker AV300, AV360 AV400 and cryo-AV600 spectrometers, at 350 ºK (300 °K for cryo-AV600). Chemical shifts (in ppm) were determined relative to residual non-deuterated solvent as an internal reference. Coupling constant(s) in hertz (Hz) were measured from one-dimensional spectra and multiplicities were abbreviated as following: 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.

Bacterial strains, plasmids and growth conditions

Plasmid pET26b-VC1280 was kindly provided by late Professor Saul Roseman (Department of Biology, The John Hopkins University, Baltimore, USA). *Escherichia coli* strain TOP10 (Invitrogen) was used as host for the plasmid propagation. *Escherichia coli* strain BL21 (DE3) (Invitrogen) was used for the large-scale production of *V. cholerae* CD.

Overexpression and purification of recombinant *V. cholerae* CD

E. coli strain BL21 (DE3) harboring the plasmid pET26b-VC1280 was grown at 37 °C in 1 L of Luria-Bertani broth medium supplemented with 40 μg.mL^{-1} kanamycin to a cell density of 0.5 at 600 nm. The temperature was lowered to 20° C and protein production was induced during 22 h by the addition of 50 μM isopropyl β-D-thiogalactopyranoside. Cells were then harvested by centrifugation at 8,450 x g for 10 min at 4° C and lysed by cell disruption. After centrifugation at 50,000 x g for 30 min at 4° C, CD was purified by applying the supernatant on a Ni^{2+}-chelating resin (Qiagen). The column was then washed with five volumes of 20 mM Tris-HCl pH 8, 500 mM NaCl, 20 mM imidazole (washing buffer) and *V. cholerae* CD was finally eluted with a 20-300 mM imidazole gradient in washing buffer. Fractions containing CD were identified by 10% polyacrylamide gel electrophoresis analysis. They were pooled, dialyzed against 20 mM potassium phosphate buffer pH 8 and stored at – 80 °C. The protein concentration was determined by the Bradford method with BSA as a standard.[14]

SDS-PAGE

Protein samples were heated at 100 °C for 5 min in Laemmli buffer (65 mM Tris/HCl, pH 6.8, 0.3 % SDS, 10 % glycerol, 5 % β-mercaptoethanol and 0.1 mg.mL^{-1} bromophenol blue),[15] followed by electrophoresis in a 10 % polyacrylamide gel. Gels were stained with Instant Blue (Expedeon).

Enzyme assays

Chitin deacetylase (CD) activity was assayed by incubating chitinbiose (CO-II) with 1.5 μM CD in 20 mM HEPES buffer pH 7 at 37°C. At regular time intervals, the amount of amino group was quantified by using 3-methyl-2-benzothiazolinone hydrazone hydrochloride reagent (MBTH).[16] For the determination of kinetic parameters, measurements of initial rates were made at several substrate concentrations ranging from 0.2 to 3 mM. K_M and k_{cat} were determined by non-linear regression using the Grafit program (Erithacus Software)

Preparation of COs

Commercial chitosan oligosaccharides (FACOS^{TM}) were *N*-acetylated by treatment with 5% acetic anhydride in (50:50) water-methanol for 12h at 20°C followed by concentration *in vacuo*. The mixture was dissolved in water and passed through a cationic exchange resin DOWEX 50WX4 (H^+ form) column to eliminate the partially acetylated oligosaccharides. The non-retained fraction containing COs was neutralized with an anionic exchange resin Amberlite IRA400 (OH^- form) and freeze-dried. COs were individually separated by size exclusion chromatography (SEC) consisting of a Biogel P4 colomn (1.5 x 1000 mm) connected to a pump (Smartline semi-preparative, Knauer) and with an inline degasser (Knauer). Elution was performed with water at flow rate of 2.25 mL.min^{-1}, and monitored by a refractometric detector (IOTA-2, Precision Instrument (DIONEX)). COs purity was confirmed by MALDI-ToF MS and HPSEC analyzes.

CD catalyzed de-*N*-acetylation of COs

Standard procedure for 1 mg of pure CO: CO-IV, CO-V or CO-VI was incubated with 50 μg of CD in 0.5 mL of 10 mM ammonium carbonate buffer pH 7.9 during 20 hours at 37 °C. The enzyme was

heat-inactivated (10 min, 100 °C) and the reaction mixture was centrifuged for 10 min at 21,100 g at 4 °C. The resulting supernatant was freeze-dried without further purification to give the corresponding CO(N) de-*N*-acetylated at the penultimate glucosaminyl residue from the reducing end, and characterized by low and high resolution MS, ^{1}H NMR and HPSEC (ESI S1-S3†).

CO-IV(N^{III}). ^{1}H NMR (600 MHz, D_2O) δ 5.16 (d, 0.6H, $J_{1,2}$ = 2.8 Hz, H-1α^I), 4.67 (d, 0.4H, $J_{1,2}$ = 7.7 Hz, H-1β^I), 4.59-4.50 (m, 3H, H-1^{II,III,IV}), 3.95-3.40 (m, 23H, H-2^{I,II,IV}, H-3^{I,II,III,IV}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, CH_2-6^{I,II,III,IV}), 2.74 (m, 1H, H-2^{III}), 2.03 (s, 6H, 2N(C=O)CH_3), 2.01 (s, 3H, N(C=O)CH_3). HRESIMS: *m/z* calcd for $C_{30}H_{52}N_4O_{20}$: 789.3253 $[M+H^+]^+$, found 789.3265.

CO-V(N^{IV}). ^{1}H NMR (360 MHz, D_2O) δ 5.19 (s, 0.5H, H-1α^I), 4.72-4.53 (m, 4.5H, H-1β^I, H-1^{II,III,IV,V}), 3.98-3.42 (m, 29H, H-2^{I,II,III,V}, H-3^{I,II,III,IV,V}, H-4^{I,II,III,IV,V}, H-5^{I,II,III,IV,V}, CH_2-6^{I,II,III,IV,V}), 3.01 (dd, 1H, $J_{2,1}$ = $J_{2,3}$ = 9.3 Hz, H-2^{IV}), 2.07 (s, 9H, 3N(C=O)CH_3), 2.04 (s, 3H, N(C=O)CH_3). HRESIMS: *m/z* calcd for $C_{38}H_{64}N_5NaO_{25}$: 1014.3866 $[M+Na]^+$, found 1014.3875.

CO-VI(N^{V}). MALDI-MS: *m/z* 1217 [M+Na].

Removal of the non-reducing GlcNAc of CO-V(N^{IV}) and CO-VI(N^{V})

Standard procedure for 1 mg of CO(N): *Aspergillus oryzae* β-galactosidase (0.1 mg, 0.8 IU, contaminated with β-*N*-acetylglucosaminidase activity)[17] was solubilized in water and dialyzed (Nanosep, Pall, 10 KDa MWCO) against water to remove salts. The resulting enzyme solution was added to 1 mg of CO-V(N^{IV}) or CO-VI(N^{V}) in a final volume of 0.1 mL of water. The reaction was carried out at 37°C during 20 hours. The reaction mixture was dialyzed with a 10 KDa membrane (Nanosep, Pall) to remove protein. The resulting CO(N) solution was precipitated in water/acetone mixture (1:20) to remove liberated GlcNAc remaining in the supernatant. Pellet constituted of CO(N) was re-dissolved in water then freeze-dried, the resulting white solid was characterized by low and high resolution MS, and HPSEC (ESI S4, S5†).

CO-IV(N^{IV}). ^{1}H NMR (360 MHz, D_2O) δ 5.21 (s, 0.6H, H-1α^I), 4.85-4.69 (m, 0.4H, H-1β^I), 4.65-4.56 (m, 2H, H-1^{II,III}), 4.49 (d, 1H, $J_{1,2}$ = 8.3 Hz, H-1^{IV}), 4.02-3.34 (m, 23H, H-2^{I,II,III}, H-3^{I,II,III,IV}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, CH_2-6^{I,II,III,IV}), 2.69 (dd, 1H, $J_{2,1}$ = $J_{2,3}$ = 8.3 Hz, H-2^{IV}), 2.09 (s, 6H, 2N(C=O)CH_3), 2.06 (s, 3H, N(C=O)CH_3). HRESIMS: *m/z* calcd for $C_{30}H_{52}N_4O_{20}$: 789.3253 $[M+H]^+$, found 789.3250.

CO-V(N^{V}). ^{1}H NMR (360 MHz, D_2O) δ 5.20 (s, 0.5H, H-1α^I), 4.88 (d, 1H, $J_{1,2}$ = 8.2 Hz, H-1),4.84-4.69 (m, 0.5H, H-1β^I), 4.65-4.54 (m, 3H, H-1), 3.99-3.46 (m, 29H, H-2^{I,II,III,IV}, H-3^{I,II,III,IV,V}, H-4^{I,II,III,IV,V}, H-5^{I,II,III,IV,V}, CH_2-6^{I,II,III,IV,V}), 3.15 (dd, 1H, $J_{2,1}$ = $J_{2,3}$ = 9.3 Hz, H-2^{V}), 2.07 (s, 9H, 3N(C=O)CH_3), 2.05 (s, 3H, N(C=O)CH_3). HRESIMS: *m/z* calcd for $C_{38}H_{65}N_5O_{25}$: 992.4047 $[M+H]^+$, found 992.4030.

Acylation of CO(N)s

Acylation reactions involved *N*-hydroxysuccinimide activated esters, which were obtained according to known literature procedures.[18] Triethylamine (1.6 equiv) was added to a solution of CO(N)s (5.4 mM, 1 equiv) and activated ester of oleic acid[19] or palmitic acid[20] (1.2 equiv) in dimethylsulfoxide in a polypropylene sealed tube. The resulting mixture was stirred at 60 °C for 6 days then concentrated under vacuum. The residue was suspended in ethyl acetate/methanol (9/1, 0.5 mL), sonicated for 10 min, centrifuged (7000 rpm, 10 min) and the supernatant was removed. This operation was repeated four times and the residue was re-suspended in water (0.5 mL), freeze-dried to afford the desired LCO without further purification. The resulting lipo-chitinoligosaccharides were characterized by ^{1}H NMR, and HRESIMS for new compounds (ESI S6-S13†).

LCO-IV (C16:0^{III}). ^{1}H NMR (600 MHz, DMSO-*d*6/D_2O, 20:1) δ 4.87 (bs, 0.6H, H-1α^I), 4.43 (d, $J_{1\beta^I,2^I}$ = 5.9 Hz, 0.4H, H-1β^I), 4.38-4.29 (m, 3H, H-1^{II,III,IV}), 3.70 (app d, 1H, $J_{6a,\ 6b}$ = 11.1 Hz, H-6), 3.66-3.11 (m, 22H, H-2^{I,II,III,IV}, H-3^{I,II,III,IV}, 3H-4, H-5^{I,II,III,IV}, 7H-6), 3.02 (dd, $J_{3,4}$ = $J_{4,5}$ = 9.2 Hz, H-4), 2.04 (t, 2H, *J* = 7.6 Hz, NH(CO)$\underline{CH_2}CH_2$), 1.86-1.76 (m, 9H, 3N(C=O)CH_3), 1.50-1.41 (m, 2H, NH(CO)$CH_2C\underline{H_2}$), 1.29-1.15 (m, 24H, 12CH_2), 0.83 (t, 3H, *J* = 6.6 Hz, CH_3). HRESIMS: *m/z* calcd for $C_{48}H_{82}N_4NaO_{21}$: 1049.5364 $[M+Na]^+$, found 1049.5359.

LCO-IV (C18:1Δ9Z^{III}). ^{1}H NMR (600 MHz, DMSO-*d*6/D_2O, 20:1) δ 5.38-5.27 (m, 2H, $CH_2C\underline{H}$=$C\underline{H}CH_2$), 4.88 (bs, 0.6H, H-1α^I), 4.45 (m, 0.4H, H-1β^I), 4.51-4.30 (m, 3H, H-1^{II,III,IV}), 3.71 (app d, 1H, $J_{6a,\ 6b}$ = 10.6 Hz, H-6), 3.68-3.09 (m, 22H, H-2^{I,II,III,IV}, H-3^{I,II,III,IV}, 3H-4, H-5^{I,II,III,IV}, 7H-6), 3.02 (dd, $J_{3,4}$ = $J_{4,5}$ = 9.0 Hz, H-4), 2.05 (t, 2H, *J* = 6.3 Hz, NH(CO)$\underline{CH_2}CH_2$), 2.01-1.95 (m, 4H, $C\underline{H_2}CH$=$CHC\underline{H_2}$), 1.89-1.77 (m, 9H, 3N(C=O)CH_3), 1.53-1.42 (m, 2H, NH(CO)$CH_2C\underline{H_2}$), 1.38-1.19 (m, 20H, 10CH_2), 0.85 (t, 3H, *J* = 6.4 Hz, CH_3). HRESIMS: *m/z* calcd for $C_{48}H_{84}N_4NaO_{21}$: 1075.5520 $[M+Na]^+$, found 1075.5500.

LCO-V (C16:0^{IV}). ^{1}H NMR (400 MHz, DMSO-*d*6/D_2O, 20:1, 350K) δ 4.93 (s, 0.6H, H-1α^I), 4.51-4.39 (m, 4.4H, H-1β^I, H-1^{II,III,IV,V}), 3.77-3.05 (m, 30H, H-2^{I,II,III,IV,V}, H-3^{I,II,III,IV,V}, H-4^{I,II,III,IV,V},

H-5^{I,II,III,IV,V}, CH_2-6^{I,II,III,IV,V}), 2.09 (t, 2H, J = 7.4 Hz, NH(CO)C$\underline{H}_2$CH$_2$), 1.89-1.79 (m, 12H, 4N(C=O)CH$_3$), 1.57-1.47 (m, 2H, NH(CO)CH$_2$C$\underline{H}_2$), 1.34-1.20 (m, 24H, 12CH$_2$), 0.86 (t, J = 6.5 Hz, 3H, CH$_3$). HRESIMS: *m/z* calcd for $C_{54}H_{95}N_5NaO_{26}$: 1252.6163 $[M+Na]^+$, found 1252.6118.

LCO-V (C18:1Δ9Z^{IV}). ^{1}H NMR (400 MHz, DMSO-*d*6/D$_2$O, 20:1, 350K) δ 5.35-5.29 (m, 2H, CH$_2$C$\underline{H}$=C$\underline{H}$CH$_2$), 4.91 (d, 0.7H, $J_{1\alpha^I,2^I}$ = 1.7 Hz, H-1α^I), 4.48-4.35 (m, 4.3H, H-1β^I, H-1^{II,III,IV,V}), 3.76-3.03 (m, 30H, H-2^{I,II,III,IV,V}, H-3^{I,II,III,IV,V}, H-4^{I,II,III,IV,V}, H-5^{I,II,III,IV,V}, CH_2-6^{I,II,III,IV,V}), 2.06 (t, 2H, J = 7.6 Hz, NH(CO)C$\underline{H}_2$CH$_2$), 2.01-1.93 (m, 4H, C$\underline{H}_2$CH=CHC$\underline{H}_2$), 1.87-1.78 (m, 12H, 4N(C=O)CH$_3$), 1.53-1.42 (m, 2H, NH(CO)CH$_2$C$\underline{H}_2$), 1.34-1.18 (m, 20H, 10CH$_2$), 0.84 (t, 3H, J = 6.8 Hz, CH$_3$). HRESIMS: *m/z* calcd for $C_{56}H_{97}N_5NaO_{26}$: 1278.6314 $[M+Na]^+$, found 1278.6343.

NodSm-IV (C18:1Δ9Z).[8] ^{1}H NMR (400 MHz, DMSO-*d*6/D$_2$O, 20:1, 355K) δ 5.35-5.32 (m, 2H, CH$_2$C$\underline{H}$=C$\underline{H}$CH$_2$), 4.94 (d, 0.6H, $J_{1\alpha^I,2^I}$ = 2.7 Hz, H-1α^I), 4.52-4.41 (m, 3.4H, H-1β^I, H-1^{II,III,IV}), 3.74 (dd, 1H, $J_{6a,\,6b}$ = 11.6 Hz, $J_{6a,\,5}$ = 2.1 Hz, H-6), 3.69-3.07 (m, 23H, H-2^{I,II,III,IV}, H-3^{I,II,III,IV}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, 7H-6), 2.09 (t, 2H, J = 7.6 Hz, NH(CO)C$\underline{H}_2$CH$_2$), 2.01-1.97 (m, 4H, C$\underline{H}_2$CH=CHC$\underline{H}_2$), 1.87-1.81 (m, 9H, 3N(C=O)CH$_3$), 1.54-1.48 (m, 2H, NH(CO)CH$_2$C$\underline{H}_2$), 1.34-1.24 (m, 20H, 10CH$_2$), 0.86 (t, 3H, J = 7.0 Hz, CH$_3$); ^{13}C NMR (100 MHz, DMSO + εD$_2$O, 355K) δ 172.3 (N(C=O)CH$_2$), 169.3 (N(C=O)CH$_3$), 169.1 (N(C=O)CH$_3$), 129.6 (CH$_2$C$\underline{H}$=C$\underline{H}$CH$_2$), 101.7 (C-1), 101.6 (C-1), 101.5 (C-1), 90.2 (C-1α^I), 81.5, 81.1, 81.0, 76.7, 74.8, 73.9, 72.3, 72.2, 70.7, 69.9, 68.8, 60.9 (C-6), 60.1 (C-6), 59.9 (C-6), 55.4 (C-2), 54.8 (C-2), 53.6 (C-2), 35.7 (NH(CO)C$\underline{H}_2$CH$_2$), 31.1 (CH$_2$), 29.1 (CH$_2$), 29.0 (CH$_2$), 28.7 (CH$_2$), 28.5 (CH$_2$), 26.6 (C$\underline{H}_2$CH=CH), 26.5 (C$\underline{H}_2$CH=CH), 24.8 (NH(CO)CH$_2$C$\underline{H}_2$), 22.8 (N(C=O)C$\underline{H}_3$), 22.7 (N(C=O)C$\underline{H}_3$), 22.4 (N(C=O)C$\underline{H}_3$), 21.9 (CH$_2$), 13.7 (CH$_3$)

NodSm-V (C18:1Δ9Z).[8] ^{1}H NMR (400 MHz, DMSO-*d*6/D$_2$O, 20:1, 355K) δ 5.35-5.33 (m, 2H, CH$_2$C$\underline{H}$=C$\underline{H}$CH$_2$), 4.93 (s, 0.6H, H-1α^I), 4.51-4.41 (m, 4.4H, H-1β^I, H-1^{II,III,IV,V}), 3.76-3.08 (m, 30H, H-2^{I,II,III,IV,V}, H-3^{I,II,III,IV,V}, H-4^{I,II,III,IV,V}, H-5^{I,II,III,IV,V}, CH_2-6^{I,II,III,IV,V}), 2.10 (t, 2H, J = 7.6 Hz, NH(CO)C$\underline{H}_2$CH$_2$), 2.02-1.98 (m, 4H, C$\underline{H}_2$CH=CHC$\underline{H}_2$), 1.86-1.82 (m, 12H, 4N(C=O)CH$_3$), 1.55-1.48 (m, 2H, NH(CO)CH$_2$C$\underline{H}_2$), 1.33-1.25 (m, 20H, 10CH$_2$), 0.87 (t, 3H, J = 7.0 Hz, CH$_3$); ^{13}C NMR (100 MHz, DMSO + εD$_2$O, 355K) δ 172.3 (N(C=O)CH$_2$), 169.7 (N(C=O)CH$_3$), 169.6 (N(C=O)CH$_3$), 169.3 (N(C=O)CH$_3$), 169.1 (N(C=O)CH$_3$), 129.5 (CH$_2$C$\underline{H}$=C$\underline{H}$CH$_2$), 101.6 (C-1), 101.5 (C-1), 101.4 (C-1), 90.2 (C-1α^I), 81.3, 81.1, 80.9, 80.8, 76.7, 74.8, 74.6, 73.9, 72.4, 72.2, 70.7, 69.8, 68.8, 60.9 (C-6), 60.3 (C-6), 60.1 (C-6), 55.4 (C-2), 55.1 (C-2), 54.9 (C-2), 53.6 (C-2), 31.0 (CH$_2$), 28.9 (CH$_2$), 28.7 (CH$_2$), 28.5 (CH$_2$), 28.3 (CH$_2$), 26.5 (C$\underline{H}_2$CH=CH), 26.4 (C$\underline{H}_2$CH=CH), 24.7 (NH(CO)CH$_2$C$\underline{H}_2$), 22.6 (N(C=O)C$\underline{H}_3$), 21.8 (N(C=O)C$\underline{H}_3$), 13.6 (CH$_3$).

Myc-LCO-IV (C16:0).[8] ^{1}H NMR (400 MHz, DMSO-*d*6/D$_2$O, 20:1, 355K) δ 4.93 (d, 0.6H, $J_{1\alpha^I,2^I}$ = 2.7 Hz, H-1α^I), 4.50-4.42 (m, 3.4H, H-1β^I, H-1^{II,III,IV}), 3.73 (dd, 1H, $J_{6a,\,6b}$ = 11.6 Hz, $J_{6a,\,5}$ = 2.2 Hz, H-6), 3.70-3.04 (m, 23H, H-2^{I,II,III,IV}, H-3^{I,II,III,IV,}, H-4^{I,II,III,IV}, H-5^{I,II,III,IV}, 7H-6), 2.09 (t, 2H, J = 7.6 Hz, NH(CO)C$\underline{H}_2$CH$_2$), 1.88-1.89 (s, 9H, 3N(C=O)CH$_3$), 1.55-1.46 (m, 2H, NH(CO)CH$_2$C$\underline{H}_2$), 1.33-1.19 (m, 24H, 12CH$_2$), 0.85 (t, 3H, J = 6.7 Hz, CH$_3$); ^{13}C NMR (100 MHz, DMSO + εD$_2$O, 355K) δ 172.3 (N(C=O)CH$_2$), 169.3, 169.2, 169.1 (3N(C=O)CH$_3$), 101.7 (C-1), 101.6 (C-1), 101.5 (C-1), 90.2 (C-1α^I), 81.5, 81.1, 81.0, 76.7, 74.8, 73.9, 72.3, 72.2, 70.6, 69.9, 68.8, 60.9 (C-6), 60.1 (C-6), 59.9 (C-6), 55.4 (C-2), 54.8 (C-2), 53.6 (C-2), 35.7 (NH(CO)C$\underline{H}_2$CH$_2$), 31.1 (CH$_2$), 28.9 (CH$_2$), 28.8 (CH$_2$), 28.7 (CH$_2$), 28.5 (CH$_2$), 24.8 (NH(CO)CH$_2$C$\underline{H}_2$), 22.8, 22.7, 22.4 (3N(C=O)C$\underline{H}_3$), 21.9 (CH$_2$), 13.7 (CH$_3$)

Myc-LCO-V (C16:0).[8] ^{1}H NMR (400 MHz, DMSO-*d*6/D$_2$O, 20:1, 355K) δ 4.92 (s, 0.6H, H-1α^I), 4.50-4.38 (m, 4.4H, H-1β^I, H-1^{II,III,IV,V}), 3.75-3.06 (m, 30H, H-2^{I,II,III,IV,V}, H-3^{I,II,III,IV,V}, H-4^{I,II,III,IV,V}, H-5^{I,II,III,IV,V}, CH_2-6^{I,II,III,IV,V}), 2.09 (t, 2H, J = 7.8 Hz, NH(CO)C$\underline{H}_2$CH$_2$), 1.86-1.80 (m, 12H, 4N(C=O)CH$_3$), 1.55-1.47 (m, 2H, NH(CO)CH$_2$C$\underline{H}_2$), 1.26 (s, 24H, 12CH$_2$), 0.87 (t, 3H, J = 6.5 Hz, CH$_3$); ^{13}C NMR (100 MHz, DMSO + εD$_2$O, 355K) δ 172.2 (N(C=O)CH$_2$), 169.0 (N(C=O)CH$_3$), 168.9 (N(C=O)CH$_3$), 101.5 (C-1), 101.4 (C-1), 101.3 (C-1), 90.1 (C-1α^I), 81.3, 80.8, 76.6, 74.8, 74.7, 73.9, 72.1, 72.0, 70.6, 69.8, 68.6, 60.8 (C-6), 60.1 (C-6), 60.0 (C-6), 59.9 (C-6), 55.4 (C-2), 55.0 (C-2), 54.9 (C-2), 54.8 (C-2), 35.5 (NH(CO)C$\underline{H}_2$CH$_2$), 31.0 (CH$_2$), 28.7 (CH$_2$), 28.6 (CH$_2$), 28.3 (CH$_2$), 24.7 (NH(CO)CH$_2$C$\underline{H}_2$), 22.6 (N(C=O)C$\underline{H}_3$), 22.3 (N(C=O)C$\underline{H}_3$), 21.7 (CH$_2$), 13.5 (CH$_3$).

Results and discussion

Expression, purification and characterization of *V. cholerae* CD

V. cholerae CD was obtained from *E. coli* BL21 harbouring the overexpression vector pET26b-VC1280. In this plasmid, the *cd* gene is fused to an N-terminal pelB sequence for a periplasmic localization and to a C-terminal 6-His Tag sequence to allow the enzyme purification by Immobilized Metal ion Affinity Chromatography (IMAC) methodology. The purification

procedure gave 120 mg of pure protein per litre of culture, as judged by SDS-PAGE (data not shown). Determination of the kinetic parameters for de-*N*-acetylation of CO-II, the best substrate, gave the following values: K_M = 3.75 ± 0.75 mM; k_{cat} = 0.45 ± 0.06 s^{-1}. VC1280 is present in different oligomeric forms in solution, the monomeric form being the most active.[21] For our synthesis purpose, we chose to work with the mixture of monomeric and oligomeric forms, and this can explain the high enzyme/substrate ratio used in the de-*N*-acetylation reactions.

Enzymatic conversion of COs to LCOs precursors

The synthetic strategy adopted for the synthesis of LCOs, its precursors and analogues is summarized in figure 1.

Chitin oligomers can be prepared through colloidal chitin chemical or enzymatic depolymerization. However this approach is tedious and cannot afford chitintetraose to chitinhexaose (CO-IV to CO-VI) in large quantities. We found more appropriate to start from a commercial mixture of chitosan oligomers which were re-*N*-acetylated and purified by ion exchange chromatography followed by gel filtration to give individual oligomers (CO-IV, CO-V and CO-VI) in pure form (Table 1).

Table 1 Quantities of isolated COs starting from 1 g of COs mixture:

Chitin oligomer	Quantity (mg)	Isolated yield (%)
CO-II	86.4	8.6
CO-III	157.2	15.7
CO-IV	196.4	19.6
CO-V	148	14.8
CO-VI	78.4	7.8

Each oligomer was treated with *V. cholerae* CD in ammonium carbonate buffer to give the corresponding CO-IV(N^{III}), CO-V(N^{IV}) and CO-VI(N^{V}) regioselectively de-*N*-acetylated at the penultimate GlcNAc unit from the reducing end in quantitative yield after protein removal as judged by MALDI-MS, TLC and HPSEC analysis (ESI S1-S3†). Ammonium carbonate buffer was used because of its ease of elimination by concentration followed by freeze drying without remaining inorganic salts contaminating the reaction product. Isolated yields were 97 %, 91 % and 79 % for CO-IV(N^{III}), CO-V(N^{IV}) and CO-VI(N^{V}) respectively, precursors of LCO analogues bearing a fatty acid chain at the penultimate residue from the reducing unit (Table 2).

Table 2 Enzymatic de-*N*-acetylation of COs

Starting compound	Quantity (mg)	Product[11]	Quantity (mg)	Isolated yield (%)
CO-IV	18.7	CO-IV(N^{III})	17.3	97
CO-V	19.1	CO-V(N^{IV})	16.6	91
CO-VI	5.1	CO-VI(N^{V})	3.9	79

Aspergillus oryzae β-*N*-acetylglucosaminidase (GlcNAcase) is an exo-acting glycoside hydrolase liberating GlcNAc units from the non-reducing end of chitinoligomers. It is found as a contaminant in commercial preparations of β-galactosidase.[17] This enzyme is inactive on glucosaminyl residue and is appropriate to release only the non-reducing GlcNAc residue from chitinoligomers de-*N*-acetylated at the penultimate GlcNAc unit from the reducing end (Figure 1). CO-V(N^{IV}) and CO-VI(N^{V}) where incubated with *A. oryzae* β-*N*-acetylglucosaminidase in water, followed by dialysis and precipitation in water/acetone mixture to remove protein and liberated GlcNAc remaining in the supernatant. Isolated yields were 66 % and 43 % for CO-IV(N^{IV}) and CO-V(N^{V}) respectively, precursors of LCO-IV and LCO-V (Figure 1, ESI S4-S5†).

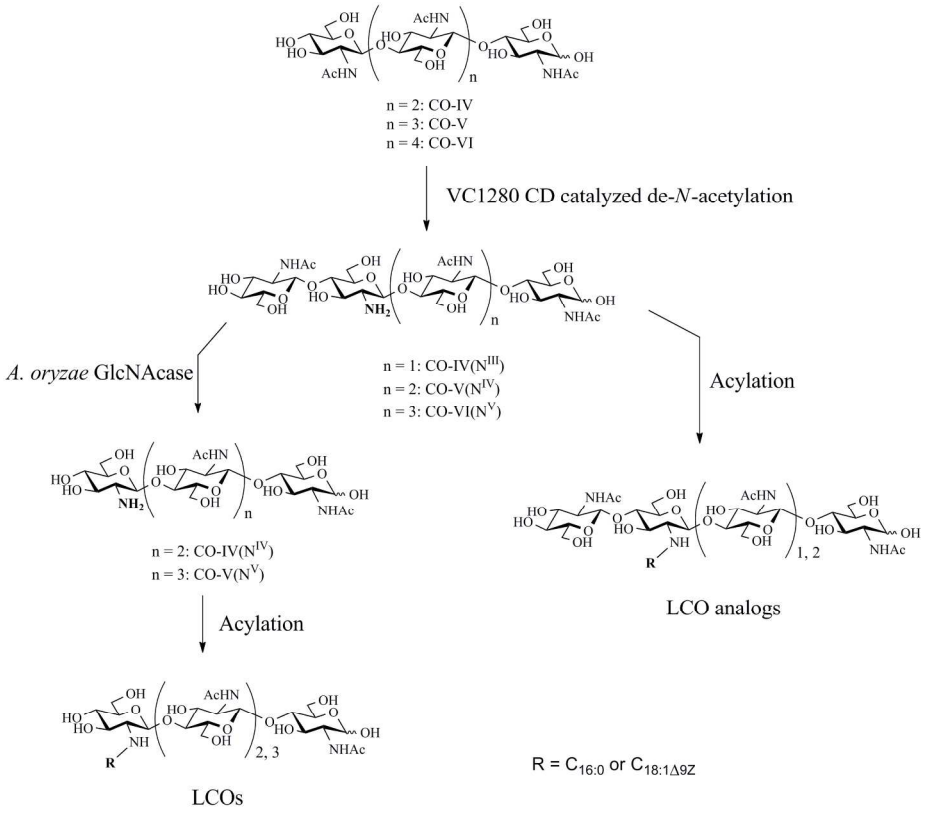

Fig. 1 Access to LCOs and analogues from COs.

Acylation of CO(N)s

The combined *Vibrio cholerae* CD and *Aspergillus oryzae* GlcNAcase approach led selectively to chitinoligosaccharides with the free amine at the non-reducing end, which are advanced precursors of lipo-chitinoligosaccharides such as Nod and Myc-LCOs. The synthesis of these plant growth promotors requires a *N*-acylation reaction, which was performed with the *N*-hydroxysuccinimide ester[18] of oleic acid[19] (C18:1Δ9Z) or of palmitic acid[20] (C16:0) in DMSO at 60 °C in the presence of an excess of triethylamine. The expected Nod and Myc LCOs[1,4,8] were isolated in 68 % to quantitative yields (Figure 2).

Applying the same procedure to CO-IV(N^{III}) and CO-V(N^{IV}) provided LCOs analogues, which were isolated in 85% to quantitative yields, attesting the efficiency of the acylation reaction (Table 3, ESI S6-S13†).

Table 3 Acylation of CO(N)s

Starting compound	Quantity (mg)	Product[22]	Quantity (mg)	Isolated yield (%)
CO-IV(N^{III})	3.1	LCO-IV(C16:0^{III})	4.0	100
CO-IV(N^{III})	3.1	LCO-IV(C18:1Δ9Z^{III})	4.2	100
CO-IV(N^{IV})	2.1	Myc-LCO-IV(C16:0)	2.8	100
CO-IV(N^{IV})	2.7	NodSm-IV(C18:1Δ9Z)	3.6	100
CO-V(N^{IV})	11.0	LCO-V(C16:0^{IV})	11.6	85
CO-V(N^{IV})	6.0	LCO-V(C18:1Δ9Z^{IV})	7.3	96
CO-V(N^{V})	3.3	Myc-LCO-V(C16:0)	2.8	68
CO-V(N^{V})	13.0	NodSm-V(C18:1Δ9Z)	14.6	89

Journal Name RSCPublishing

ARTICLE

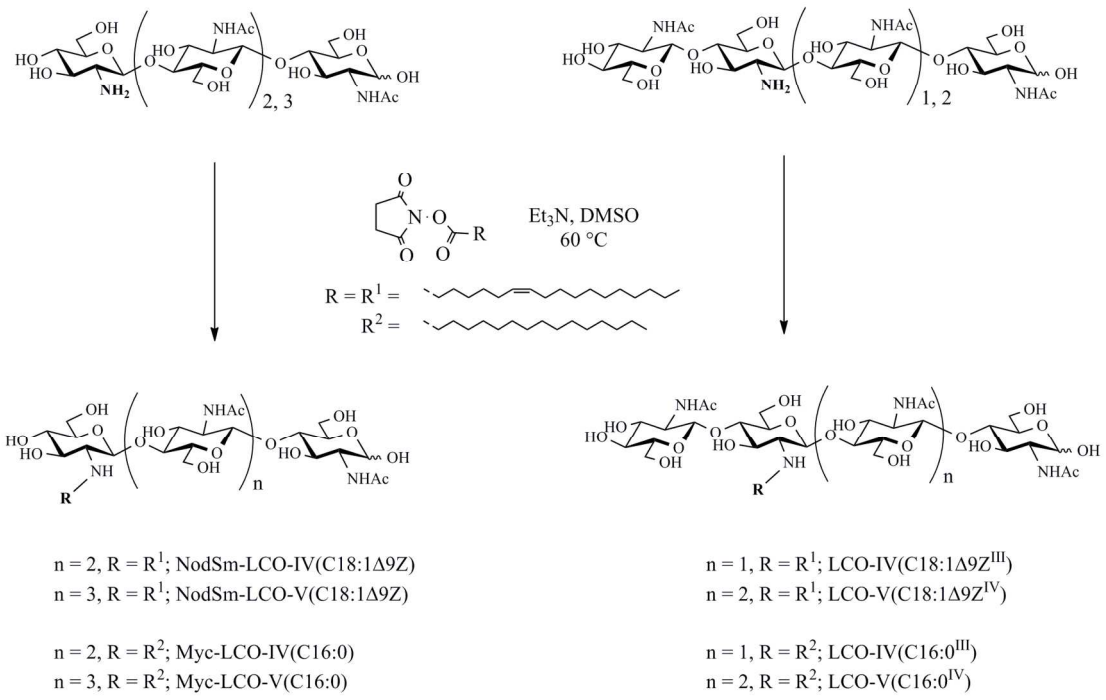

Fig 2. Acylation of COs

Conclusion

The present work reports an efficient and straightforward *in vitro* chemo-enzymatic access to natural LCO fertilizers and analogues from biomass. Commercial source of chitosan oligomers was found to be convenient for the preparation of pure COs by *N*-acetylation and subsequent purification by size-exclusion chromatography.

Selective *N*-deacetylation of theses COs was achieved by VC1280 CD which was produced in high yield (120 $mg.L^{-1}$) by periplasmic expression in *E. coli*. The enzyme was active in carbonate buffer, allowing a straightforward route to COs specifically de-*N*-acetylated at the penultimate residue from the reducing end. Very high yield could be achieved after a simple purification procedure consisting of protein thermal denaturation and centrifugation followed by freeze drying, leading to pure CO(N). These CO(N) could be transformed in Nod and Myc-LCO precursors by controlled enzymatic degradation using *A. oryzae* β-*N*-acetylglucosaminidase. Efficient acylation procedure using *N*-hydroxysuccinimide fatty acid ester led to natural Nod and Myc-LCOs. Alternatively, omitting the β-*N*-acetylglucosaminidase treatment gave access to LCO analogues bearing the fatty acid chain at the penultimate residue from the reducing end. They may play also a role in the plant signaling events.

This methodology opened the way for the large-scale production of lipo-chitinoligosaccharides, high value molecules of major agrochemical interest, starting from chitin an abundant renewable resource.

Acknowledgements

Authors are indebted to late Prof. S. Roseman for providing VC1280 CD gene and are grateful to S. Pradeau, M. Touvrey and P. Chaud for skillful assistance. Authors acknowledge support from ICMG FR 2607, LabEx ARCANE (ANR-11-LABX-0003-01), PolyNat Carnot Institute, French Agence

Nationale de la Recherche (ANR-10-CD2I-0008, BioCOS 2010-2014), and Bill & Melinda Gates Foundation (ENSA 2012-2017).

Abbreviations

Amp, ampicillin; CO, chitinoligosaccharide; CO(N), de-*N*-acetylated chitinoligosaccharide; CD, chitin deacetylase; CO-IV, chitintetraose; CO-V, chitinpentaose; CO-VI, chitinhexaose; GlcN, D-glucosamine; GlcNAc, *N*-acetyl-D-glucosamine; IPTG, isopropyl *β*-D-thiogalactopyranoside; Kan, kanamycin; LCO, lipo-chitinoligosaccharide; LB, Luria Broth; NodSm, Nod factors produced by Sinorhizobium meliloti; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TLC, thin layer chromatography

Notes and references

[a]*Univ. Grenoble Alpes, CERMAV, F-38000 Grenoble, France. CNRS, CERMAV, F-38000 Grenoble, France.*
[b]*CNRS and Université Paris-Sud, Glycochimie Moléculaire et Macromoléculaire, ICMMO, UMR 8182, 91405 Orsay, France.*
[c]*CNRS, ICSN, Centre de Recherche de Gif, 1 avenue de la Terrasse, F-91198 Gif-sur-Yvette, France*

† Electronic Supplementary Information (ESI) available: [Low resolution MALDI-ToF MS, HRESIMS, ^{1}H NMR and HPSEC analyses for purified compounds]. See DOI: 10.1039/b000000x/

1. P. Lerouge, P. Roche, C. Faucher, F. Maillet, G. Truchet, J. C. Promé and J. Denarie, *Nature*, 1990, **344**, 781-784.
2. W. J. Broughton, S. Jabbouri and X. Perret, *J. Bacteriol.*, 2000, **182**, 5641-5652.
3. X. Perret, C. Staehelin and W. J. Broughton, *Microbiol. Mol. Biol. Rev.*, 2000, **64**, 180-201.
4. F. Maillet, V. Poinsot, O. André, V. Puech-Pagès, A. Haouy, M. Gueunier, L. Cromer, D. Giraudet, D. Formey, A. Niebel, E. A. Martinez, H. Driguez, G. Bécard and J. Dénarié, *Nature*, 2011, **469**, 58-64.
5. J. Dénarié, F. Debellé and J.-C. Promé, *Ann. Rev. Biochem.*, 1996, **65**, 503-535.
6. E. Kamst, H. P. Spaink and D. Kafetzopoulos, *Subcell. Biochem.*, 1998, **29**, 29-71.
7. S. Dobbelaere, J. Vanderleyden and Y. Okon, *Crit. Rev. Plant Sci.*, 2003, **22**, 107–149.
8. G. Despras, A. Alix, D. Urban, B. Vauzeilles and J.-M. Beau, *Angew. Chem. Int. Ed.*, 2014, **53**, 11912–11916.
9. E. Samain, V. Chazalet and R. A. Geremia, *J. Biotechnol.*, 1999, **72**, 33-47.
10. S. K. Kim and N. Rajapakse, *Carbohydr. Polym.*, 2005, **62**, 357-368.
11. We followed the nomenclature already adopted for LCOs and derivatives (W. D'Haeze and M. Holsters, *Glycobiology*, 2002, **12**, 79R-105R). Concerning COs de-*N*-acetylated at a specific position (CO(N^x)), the N letter indicates a free amino group at C-2 of the glycosyl unit, and the roman upper case number specifies the unit position starting from the reducing end. i.e. CO-II (N^I) for chitinbiose de-*N*-acetylated at the reducing unit).
12. X. B. Li, L. X. Wang, X. S. Wang and S. Roseman, *Glycobiology*, 2007, **17**, 1377-1387.
13. J. W. Walkley and J. Tillman, *J. Chromatogr.*, 1977, **132**, 172-174.
14. M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248-254.
15. U. K. Laemmli and M. Favre, *J. Mol. Biol.*, 1973, **80**, 575-599.
16. A. Tsuji, T. Kinoshita and M. Hishino, *Chem. Pharm. Bull.*, 1969, **17**, 1505-1510.
17. M. Scigelova, S. Singh, D. H. G. Crout, *J. Mol. Cat. B: Enzym.,*1999, **6**, 483–494.
18. a) Y. Lapidot, S. Rappoport and Y. Wolman, *J. Lipid. Res.*, 1967, **8**, 142-145; b) S. Ikeshita, Y. Nakahara and T. Ogawa, *Glycoconjugate J.*, 1994, **11**, 257-261.
19. K. Vávrová, A. Hrabálek, P. Doležal, L. Šámalová, K. Palát, J. Zbytovská, T. Holas and J. Klimentová, *Bioorg. Med. Chem.*, 2003, **11**, 5381-5390.
20. M. J. Clemente, J. Fitremann, M. Mauzac, J. L. Serrano and L. Oriol, *Langmuir*, 2011, **27**, 15236-15247.
21. E. Andrés, D. Albesa-Jové, X. Biarnés, B. M. Moerschbacher, M. E. Guerin and A. Planas, *Angew. Chem. Int. Ed.*, 2014, **53**, 6882–6887.
22. We propose a specific nomenclature for LCOs analogues where the roman upper case number specifies the position starting from the reducing end (i.e. LCO-IV(C16:0^{III}) for chitintetraose *N*-palmitoylated at the penultimate residue from the reducing unit).

lipo-chitinoligosaccharides (LCOs), keys molecules in plant-symbiotic microorganisms communication, are readily obtained via chemoenzymatic synthesis from chitin, a renewable abundant biopolymer.