NMR-Based Structural Glycomics for High-Throughput Screening of Carbohydrate-Active Enzyme Specificity

ARTICLE *in* ANALYTICAL CHEMISTRY · FEBRUARY 2011

Impact Factor: 5.64 · DOI: 10.1021/ac1032148 · Source: PubMed

9 AUTHORS, INCLUDING:

Romain Irague
University of Nantes
8 PUBLICATIONS **74** CITATIONS
SEE PROFILE

Alain Milon
IPBS - Institut de Pharmacologie et de Biolo…
134 PUBLICATIONS **2,362** CITATIONS
SEE PROFILE

Pierre Monsan
University of Toulouse
249 PUBLICATIONS **5,578** CITATIONS
SEE PROFILE

Gabrielle Veronese
INSA
53 PUBLICATIONS **907** CITATIONS
SEE PROFILE

analytical chemistry

LETTER

pubs.acs.org/ac

NMR-Based Structural Glycomics for High-Throughput Screening of Carbohydrate-Active Enzyme Specificity

Romain Irague,[†,‡,§] Stéphane Massou,[†,‡,§] Claire Moulis,[†,‡,§] Olivier Saurel,[⊥] Alain Milon,[⊥] Pierre Monsan,[†,‡,§] Magali Remaud-Siméon,[†,‡,§] Jean-Charles Portais,[*,†,‡,§] and Gabrielle Potocki-Véronèse[*,†,‡,§]

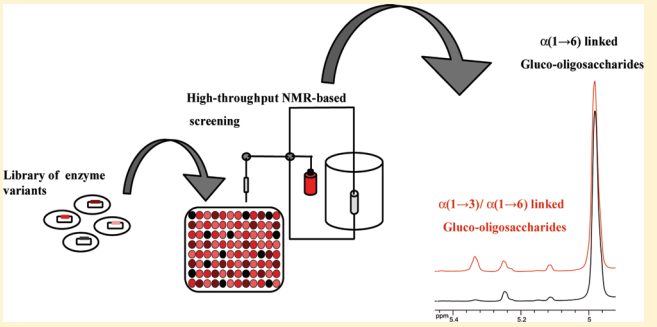
[†]Université de Toulouse, INSA, UPS, INP, LISBP, 135 avenue de Rangueil, F-31077 Toulouse, France

[‡]UMR5504 and [§]UMR792 Ingénierie des Systèmes Biologiques et des Procédés, CNRS, INRA, F-31400 Toulouse, France

[⊥]CNRS, UPS, Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, Toulouse, France

Supporting Information

ABSTRACT: We report here the development of a straightforward, sensitive, and quantitative NMR-based method for high-throughput characterization of carbohydrate structure and screening of carbohydrate active enzyme (CAZyme) specificity. Automated assays starting from gene library expression to carbohydrate structure determination directly from crude reaction media have been established and successfully used to screen a library of 4032 CAZymes obtained by combinatorial engineering, at a rate of 480 enzyme variants per day. This allowed one to accurately discriminate 303 enzyme variants with altered specificity. The results demonstrate the potential of high-throughput NMR technology in glycomics, to mine artificial and natural enzyme diversity for novel biocatalysts.



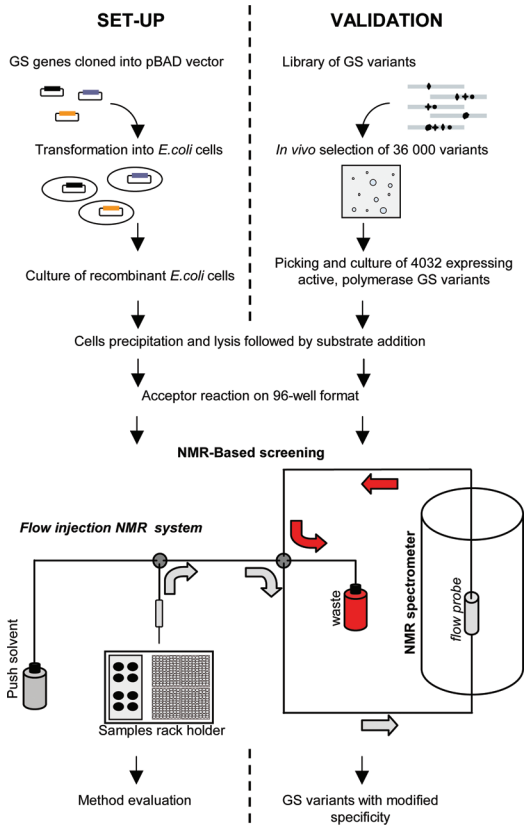
The development of glycoenzyme-based processes for the production of novel glycans, oligosaccharides, or glycoconjugates has gained interest for applications in human health, food and feed industries, fine chemistry, and white biotechnologies. Transglycosidases and glycosyltransferases, in particular, constitute very efficient tools for glycodiversification.[1,2] In addition, protein engineering has shown to be extremely powerful to expand their substrate or product specificities and adapt them to synthetic substrates.[3,4] However, searching for carbohydrate-active enzymes (CAZymes) showing required specificities in libraries of thousand mutants is still a challenge. This is principally due to (i) the high structural complexity and diversity of carbohydrates (three hexoses can yield more than 30 000 different oligosaccharidic structures in theory) and (ii) the difficult acquisition of carbohydrate structural information via a rapid and automated method. These limitations often force the biologists to drastically reduce the size of the library to be screened, especially when searching for glycoenzymes that catalyze novel carbohydrate substitution patterns. The detailed characterization of carbohydrate structures usually requires the combination of several analytical methods including NMR spectroscopy, mass spectrometry, enzymatic and chemical degradation, and methylation and acetylation analysis. Among this panel of techniques, ^{1}H NMR spectroscopy, which is both nondestructive and rapid, can provide key data for the structural characterization of carbohydrates.[5] Indeed, the concept of the "structural reporter group", which encompasses the assignment of well-resolved ^{1}H NMR chemical shifts to specific carbohydrate structural features including the type of sugar units, the linkage composition, or the substitution pattern, was introduced 30 years ago.[6] Recently, this concept was revived and applied to the identification and quantification of the various types of glucosidic linkages found in α-glucans and glucooligosaccharides produced by glucansucrases.[7−9] NMR spectroscopy thus appears very attractive to screen and profile CAZyme specificities. However, to be applicable to large library screening, the technology must allow automated, rapid, sensitive, reproducible, and quantitative monitoring of enzymatic reactions from crude and miniaturized reaction media. In that regard, flow injection technology may answer most of those prerequisites. It was successfully used for combinatorial chemistry,[10] metabolic profiling,[11] racemic mixture analysis,[12] and enzyme inhibitor identification.[13] However, it was never tested to screen large libraries of enzyme variants directly from crude medium.

Here, we describe a flow injection ^{1}H NMR screening method used to screen libraries of glucansucrases for regiospecificity of

Received: December 8, 2010
Accepted: January 13, 2011
Published: January 27, 2011

dx.doi.org/10.1021/ac1032148 | *Anal. Chem.* 2011, 83, 1202–1206

Scheme 1. Outline of NMR-Based Screening for Carbohydrate Active Enzyme Specificity[a]



[a] The method was setup with *E.coli* cells encoding empty plasmid (negative control) and parental glucansucrases (GS) with known specificities. Cells were cultured in a 96-deepwell format for growth and gene expression. After lysis, recombinant enzymes were incubated with 292 mM sucrose and 50 mM iso-maltooligosaccharides to produce oligosaccharides. (Trimethylsilyl)propionic acid-2,2,3,3-d4 acid (37 mM), used as internal standard, was added to each sample before 1D ^{1}H NMR analysis carried out on the crude reaction media. The method was validated by screening a library of glucansucrase variants. Mutants were first selected on their abilities to cleave sucrose. Then, for each of the 4032 selected active glucansucrase variants, the synthesized oligosaccharide structure was screened at a flow rate of one sample per 3 min linkages.

glycidic bond formation. The automated and miniaturized method encompasses (i) the expression of the gene library and production of the protein variants in 96-deepwell microplates, (ii) the enzymatic synthesis of glucooligosaccharides from a nonlabeled substrate, and (iii) the direct 1D ^{1}H NMR analysis of the crude reaction medium using automated flow injection to reveal the key structural features of the produced carbohydrate(s) (Scheme 1). The nondestructive, quantitative, and straightforward method we developed permits one to sort out enzyme variants with modified specificity (in terms of type and/or amount of osidic linkage content) directly from crude reaction medium at a throughput exceeding 400 samples per day. This approach opens new perspectives for screening complex carbohydrate structures and for sorting highly specific CAZyme biocatalysts.

Our model enzymes are glucansucrases. These transglucosidases belong to the Glycoside−Hydrolase[14] family 70 (see Carbohydrate Active Enzymes database at http://www.cazy.org/) and catalyze the synthesis of α-glucans from sucrose. When glycosides or noncarbohydrate molecules are used as exogenous acceptors, these enzymes catalyze the transfer of the glucosyl moiety from sucrose onto the acceptor molecules leading to oligosaccharide or glucoconjugate production. Depending on the enzyme regiospecificity, various types of glucosidic linkages (i.e., α(1→2), α(1→3), α(1→4), α(1→6)) occur in the resulting sugar structures.

To set up the method (Scheme 1), we first evaluated the ability of NMR to provide relevant structural information within short times, i.e., minutes, and for small scale assays. Three glucansucrases with known and different specificities were selected: (i) DSR-S vardel Δ4N,[15] which synthesizes α(1→6) glucosidic linkages, (ii) ASR C-del,[16] which synthesizes both α(1→6) and α(1→3) glucosidic linkages, and (iii) DSR-E,[17] which synthesizes α(1→6), α(1→3) and α(1→2) glucosidic linkages. All of the three glucansucrase encoding genes had previously been cloned into pBAD vector. The enzymes were produced using small-scale cultures of recombinant *E. coli* cells in 96-well microtiter plates. Following gene expression and cell disruption, the lysates were directly used as a source of enzyme to catalyze transglucosylation reaction from sucrose donor to isomaltooligosaccharide acceptors (IMOS). The crude reaction mixtures (500 μL), containing the reaction products, were directly injected into an NMR flow probe. Process was automated using a flow injection NMR system (Tables S-1 and S-2 (Supporting Information) relate sample preparation and NMR measurement parameters in detail) and monitored by IconNMR software. All the experiments, starting from small-scale cultures to NMR analysis, were carried out in triplicate.

The anomeric protons of glucosyl moieties gave clear NMR signals that were assigned as described in previous studies (Figure 1 and Table S-3, Supporting Information).[7−9,18] The chemical shifts at 4.98, 5.09, 5.19, and 5.34 ppm correspond to the anomeric proton of the α(1→6) linked D-Glc*p*, α(1→2) linked D-Glc*p*, α(1→6) linked 2,6-di-O-substituted α-D-Glc*p*, and α(1→3) linked D-Glc*p* units, respectively, and were used as structural-reporter groups. NMR spectra revealed that the glucooligosaccharides produced in the different reaction mixtures contained new types of osidic linkages in addition to the α(1→6) linkages of the IMOS acceptor (Figure 1). Indeed, signals corresponding to anomeric protons of units linked by α(1→2) or α(1→3) arose on the spectra in addition to the signal due to the anomeric proton of the α(1→6) linked moiety. Their integration revealed that DSR-S vardel Δ4N and ASR C-del produced glucooligosaccharides in which the ratios of α(1→3) linkages versus α(1→6) linkages are of 1:99 and 10:90, respectively. The resonance complexity of the α(1→3) linked D-Glc*p* residues is probably due to the neighboring of α(1→6) linked D-Glc*p* residues, as previously described.[8,19] Besides, DSR-E enzyme produced glucooligosaccharides containing α(1→2), α(1→3), and α(1→6) linkages in a ratio of 5:1:94. These results were fully consistent with the known regiospecificity of the three transglucosidases. A signal-to-noise value of 160 was calculated from the weaker signal detected (i.e., anomeric proton of the α(1→3) linked D-Glc*p* units of the glucooligosaccharides synthesized by DSR-S vardel Δ4N; Figure 1). The corresponding concentration of this residue was 5.5 mM, so considering that a 16-fold lower signal-to-noise ratio would still be acceptable for a correct detection, we estimated that the method sensitivity was in the range of a few hundreds of micromolar. This was satisfactory

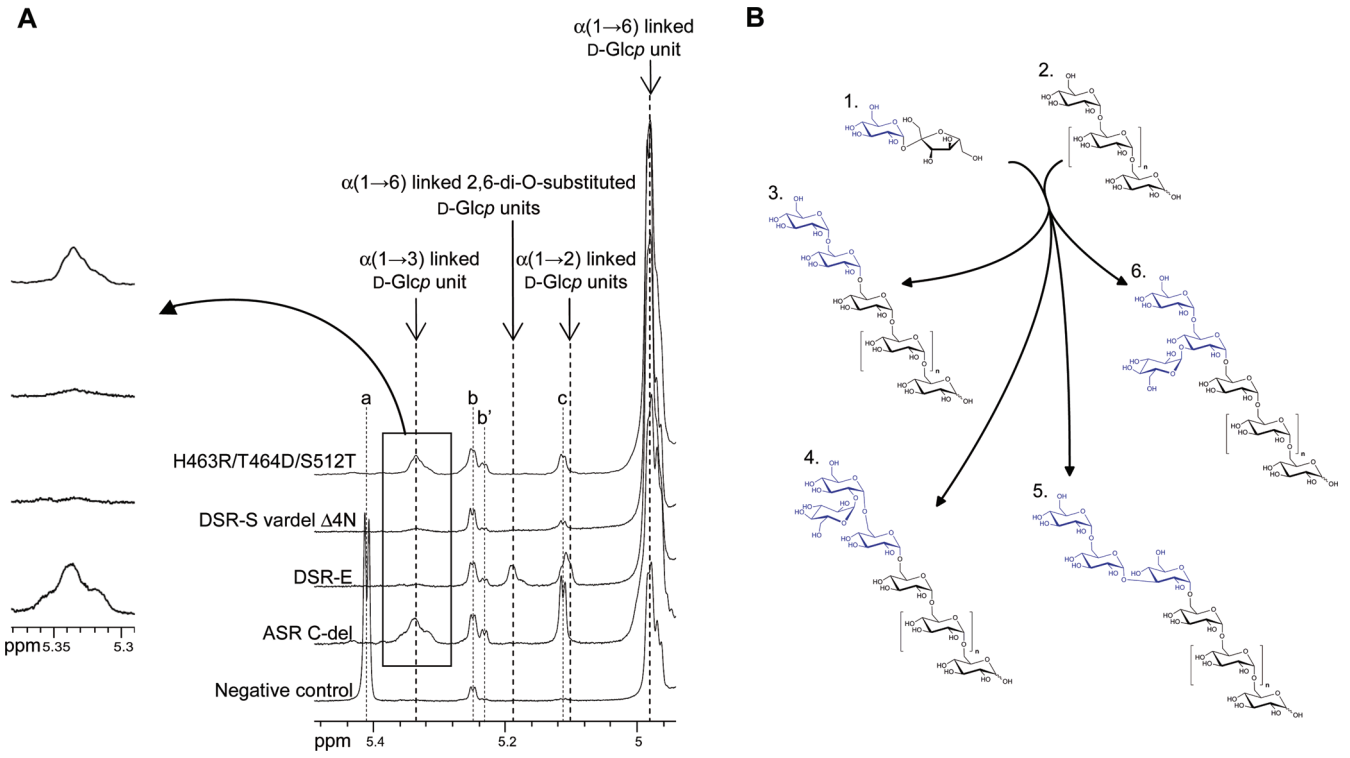


Figure 1. (A) "Structural reporter group" used to screen the carbohydrate structures synthesized by the glucansucrases DSR-S vardel Δ4N, DSR-E, ASR C-del, and by the DSR-S vardel Δ4N H463R/T464D/S512T variant by 1D ^{1}H NMR . Spectra are centered at the anomeric region (4.94−5.48 ppm). Peaks were referenced to internal TSP-d4 (1H = 0 ppm). (a) α-D-Glc*p* unit of the sucrose, (b) α-D-Glc*p* reducing unit, (b′) free D-Glc*p* unit resulting from sucrose hydrolysis reaction, and (c) α(1→5)-D-Glc*p* unit of the leucrose resulting from sucrose isomer synthesis. (B) Examples of glucooligosaccharides synthesized from (1) sucrose and (2) isomaltooligosaccharides by (3) DSR-S vardel Δ4N, (4) DSR-E, (5) ASR C-del, and (6) DSR-S vardel Δ4N H463R/T464D/S512T variant (structural representations show the osidic linkage composition of the products), *n* corresponds to the number of α-D-Glc*p* units ranging from 3 to 5.

considering that these analyses were carried out on crude media without any purification steps. Moreover, the standard deviation of the signal intensities between the triplicates was within 3.3%, which demonstrated the reproducibility of the method. Finally, an entire microtiter plate was analyzed in 288 min (one sample every 3 min). The throughput was thus sufficiently high to analyze 480 samples per day and authorize large library screening in a reasonable duration.

Following the workflow depicted in Scheme 1, the method was thus used to screen for glucosidic bond specificity, a library of 36 000 *E. coli* clones expressing DSR-S vardel Δ4N variants. As described in the Supporting Information, this library was obtained by combinatorial mutagenesis of eight residues (D306, F353, N404, W440, D460, H463, T464, and S512) located in the catalytic domain of the parental enzyme, using the ISOR method.[20] A selection pressure previously described[21] and consisting of growing the recombinant clones onto sucrose substrate was first applied to isolate 5232 clones able to cleave sucrose substrate. Among this selected population, 4032 clones were retained due to the detection of a polysaccharide bubble formed around the colony, indicating a glucansucrase activity. Then, the flow NMR-based protocol was applied to this library of variants which were screened for their glucosidic bond specificity in 8 days. The 1D ^{1}H NMR spectra of the oligosaccharides produced by each mutant were standardized using the TSP-d4 signal intensity and compared to that obtained for the parental DSR-S vardel Δ4N. Of the 4032 screened mutants, 303 (7.5% of the assayed variants) synthesized gluco-oligosaccharides containing a 2−6-fold increase of α(1→3) linked D-Glc*p* units compared to those produced by the parental enzyme (Figure 2a).

Of the 303 positive hits, two single mutants (S512C and F353W) and one triple mutant (H463R/T464D/S512T) showing increasing amount of α(1→3) linked D-Glc*p* units were selected to assess the reliability of the method (Figure 2a). They were produced in larger scale and isolated for further characterization. The 1D ^{1}H NMR structural analysis of the oligosaccharides produced by the three variants confirmed their aptitude to synthesize more α(1→3) linkages than the parental DSR-S vardel Δ4N. Integration of the anomeric proton signals of the α(1→3) and α(1→6) D-Glc*p* units (5.34 and 4.98 ppm, respectively) revealed α(1→3)/α(1→6) ratios of 1.4:98.6, 2.2:97.8, 3.1:96.9, and 6.9:93.1 for parental DSR-S vardel Δ4N, S512C, F353W, and H463R/T464D/S512T, respectively (Figure S-1, Supporting Information). These values are in agreement with the screening results, showing that the miniaturization and automation of oligosaccharide analysis was reliable (Figure 2b−e). Structural differences between the oligosaccharides produced by the variants and by the parental enzyme were also confirmed by HPAEC-PAD (Figure S-2, Supporting Information). Besides, we noticed that the single mutants S512C and F353W also produced higher amounts of leucrose (5-O-α-D-glucosyl fructose) compared to the parental enzyme. Indeed, as shown on the HPAEC-PAD profiles, leucrose yield increased from 3% for DSR-S vardel Δ4N to 6 and 20% for S512C and F353W, respectively. This

dx.doi.org/10.1021/ac1032148 |*Anal. Chem.* 2011, 83, 1202–1206

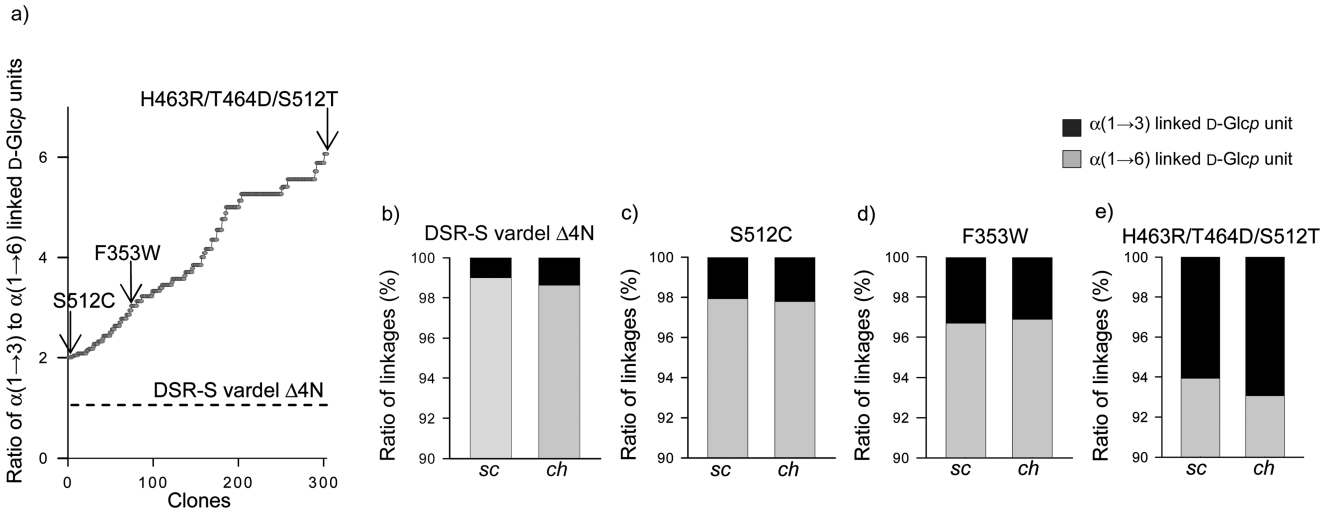


Figure 2. Identification of glucansucrase mutants with modified specificity by NMR-based screening. (a) Ratio of α(1→3) to α(1→6) linked D-Glc*p* units in the glucooligosaccharides produced by the 303 isolated variants compared to DSR-S vardel Δ4N (black dotted line). These results are also presented in details in Supplementary Table 3, Supporting Information. Comparison of synthesized oligosaccharides primary structures determined after screening using flow injection technology (sc) and large-scale characterization (ch) (b) for DSR-S vardel Δ4N, (c) for mutant S512C, (d) for mutant F353W, and (e) for mutant H463R/T464D/S512T.

indicates that the mutations also affected fructose acceptor recognition. This phenomenon was also evidenced from ^{1}H NMR data, as the intensity of the peak corresponding to the D-Glc*p* unit anomeric proton of leucrose (δ 5.11 ppm) clearly increased (Figure S-1, Supporting Information). These results also proved that the method is useful to detect changes in glycoenzyme stereoselectivity, since it provided more information than initially expected. Finally, α-glucans were produced by these three variants using sucrose as sole substrate, purified and subjected to 1D ^{1}H NMR analysis. The NMR data showed that the α(1→3)/α(1→6) ratios in the glucans produced by the parental DSR-S vardel Δ4N and the variants S512C, F353W, and H463R/T464D/S512T were of 4:96, 6:94, 9:91, and 12:88, respectively (Figure S-3, Supporting Information), thus confirming the change of variant specificity and the potential of the methods to screen for novel biopolymer synthesizing enzymes.

In 3 min per sample, this flow NMR-based method allowed the detection and quantification of the type of osidic linkages in glucooligosaccharides with a high sensitivity and reproducibility. The method not only is faster than the LC or LC-MS methods used for carbohydrate analysis but also provides structural information without any compound separation. Finally, the method is nondestructive and can be applied directly on crude cell lysates and natural substrates, as it does not necessitate any labeling strategy. Provided that one is able of identifying an appropriate structural reporter, the methodology can be extended to screen substrate selectivity and linkage specificity of other CAZymes. In a general way, this approach opens promising perspectives in glycomics, to mine artificial and natural enzyme diversity for novel biocatalysts.

■ ASSOCIATED CONTENT

Supporting Information. Description of the materials and methods used and additional figures and tables as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jean-charles.portais@insa-toulouse.fr (J.-C.P); veronese@insa-toulouse.fr (G.P.-V.). Tel: +33 (0)561 559 689 (J.-C.P); +33 (0)561 559 487 (G.P.-V.). Fax: +33 (0)561 559 400 (J.-C.P and G.P.-V.).

■ ACKNOWLEDGMENT

The high throughput screening work was carried out at the Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP, Toulouse, France) with the equipment of the ICEO facility dedicated to the screening and the discovery of new and original enzymes and at the Institut de Pharmacologie et de Biologie Structurale (IPBS, Toulouse, France). ICEO is supported by grants from the Région Midi-Pyrénées, the European Regional Development Fund, and the Institut National de la Recherche Agronomique (INRA). The NMR spectra were recorded on spectrometers financed with the help of European Structural funds, Région Midi-Pyrénées, and Centre National de la Recherche Scientifique (CNRS). They are part of the Integrated Screening Platform of Toulouse (PICT, IBISA).

■ REFERENCES

(1) Monsan, P.; Remaud-Siméon, M.; André, I. *Curr. Opin. Microbiol.* **2010**, *13*, 293.
(2) Weijers, C. A. G. M.; Franssen, M. C. R.; Visser, G. M. *Biotechnol. Adv.* **2008**, *26*, 436.
(3) Homann, A.; Seibel, J. *Nat. Prod. Rep.* **2009**, *26*, 1555–71.
(4) Kittl, R.; Withers, S. G. *Carbohydr. Res.* **2010**, *345*, 1272.
(5) Duus, J. O.; Gotfredsen, C. H.; Bock, K. *Chem. Rev.* **2000**, *100*, 4589.
(6) Vliegenthart, J. F. G.; Dorland, L.; Van Halbeek, H. *Adv. Carbohydr. Chem. Biochem.* **1983**, *41*, 209.
(7) Van Leeuwen, S. S.; Leeflang, B. R.; J., G. G.; Kamerling, J. P. *Carbohydr. Res.* **2008**, *343*, 1114–11.
(8) Van Leeuwen, S. S.; Kralj, S.; Van Geel-Schutten, I. H.; Gerwig, G. J.; Dijkhuizen, L.; Kamerling, J. P. *Carbohydr. Res.* **2008**, *343*, 1237.

(9) Van Leeuwen, S. S.; Kralj, S.; Eeuwema, W.; Gerwig, G. J.; Dijkhuizen, L.; Kamerling, J. P. *Biomacromolecules* **2009**, *10*, 580–8.
(10) Keifer, P. A. *Curr. Opin. Chem. Biol.* **2003**, *7*, 388.
(11) Dumas, M. E.; Barton, R. H.; Toye, A.; Cloarec, O.; Blancher, C.; Rothwell, A.; Fearnside, J.; Tatoud, R.; Blanc, V.; Lindon, J. C.; Mitchell, S. C.; Holmes, E.; McCarthy, M. I.; Scott, J.; Gauguier, D.; Nicholson, J. K. *Proc. Natl. Acad. Sci.* **2006**, *103*, 12511–12516.
(12) Reetz, M. T.; Eipper, A.; Tielmann, P.; Mynott, R. *Adv. Synth. Catal.* **2002**, *344*, 1008–1016.
(13) Dalvit, C.; Ardini, E.; Flocco, M.; Fogliatto, G. P.; Mongelli, N.; Veronesi, M. *J. Am. Chem. Soc.* **2003**, *125*, 14620–5.
(14) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233–238.
(15) Moulis, C.; Joucla, G.; Harrison, D.; Fabre, E.; Potocki-Veronese, G.; Monsan, P.; Remaud-Simeon, M. *J. Biol. Chem.* **2006**, *281*, 31254–67.
(16) Joucla, G.; Pizzut, S.; Monsan, P.; Remaud-Simeon, M. *FEBS letters* **2006**, *580*, 763.
(17) Bozonnet, S.; Dols-Laffargue, M.; Fabre, E.; Pizzut, S.; Remaud-Simeon, M.; Monsan, P.; Willemot, R. M. *J. Bacteriol.* **2002**, *184*, 5753–5761.
(18) Maina, N. H.; Tenkanen, M.; Maaheimo, H.; Juvonen, R.; Virkki, L. *Carbohydr. Res.* **2008**, *343*, 1446.
(19) Cheetham, N. W.; Fiala-Beer, E.; Walker, G. J. *Carbohydr. Polym.* **1991**, *14*, 149–158.
(20) Herman, A.; Tawfik, D. S. *Protein Eng. Des. Sel.* **2007**, *20*, 219–26.
(21) Van Der Veen, B. A.; Potocki-Veronese, G.; Albenne, C.; Joucla, G.; Monsan, P.; Remaud-Simeon, M. *FEBS Lett.* **2004**, *560*, 91–7.