ORGANIC LETTERS

2004
Vol. 6, No. 12
2031−2033

Discovery of the Chemical Function of Glycosidases: Design, Synthesis, and Evaluation of Mass-Differentiated Carbohydrate Libraries

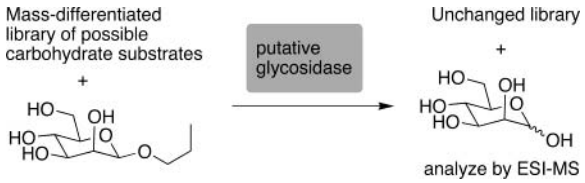
Yang Yu, Kwang-Seuk Ko, Corbin J. Zea, and Nicola L. Pohl*

Department of Chemistry and the Plant Sciences Institute, Gilman Hall, Iowa State University, Ames, Iowa 50011-3111

npohl@iastate.edu

Received April 3, 2004

ABSTRACT



Discovery of the catalytic chemical function of the many putative glycosidases coded in genomes currently relies on individual testing of possible substrates, usually as their *p*-nitrophenol conjugate. Herein, we present an alternative chemical proteomics approach using a synthetic mass-differentiated heat-stable substrate library with mass spectrometry readout. Library components do not serve as reaction inhibitors and both primary and secondary enzyme substrates can be delineated.

Experimental evidence of the chemical function of new proteins with appropriate ligands—chemical proteomics[1]—is of particular importance for the large numbers of proteins that bind and catalyze reactions involving carbohydrates. The human genome has 106 genes tentatively assigned functions as glycosidases and transglycosidases, enzymes that cleave carbohydrate linkages; the first plant genome sequence of *Arabidopsis thaliana* contains 381 such genes.[2] Currently, the only method to delineate the chemical function of one of these gene products is by individually testing the new protein with every possible substrate, usually as its *p*-nitrophenol adduct—a time and material intensive process. Herein, we present an alternative chemical proteomics approach with mass spectrometry readout to delineate the catalytic chemical function of putative glycosidases.

Carbohydrate utilization is believed to be the oldest metabolic pathway of life; moreover, carbohydrates are the building blocks for an incredible variety of structures key to the physical strength, intracellular recognition, and natural defenses of organisms.[3] Our understanding of this complexity is in its infancy. For example, alteration of only four of 354 amino acids entirely changes the carbohydrate recognition properties of the two glycosyltransferases responsible for the A versus B blood group antigen determinants.[4] Hence, analysis of primary sequence information of putative carbohydrate-recognizing enzymes alone does not reveal functional or substrate information. This substrate information

(1) (a) Zea, C. J.; Pohl, N. L. *Anal. Biochem.* **2004**, *327*, 107−113. (b) Jeffery, D. A.; Bogyo, M. *Curr. Opin. Biotechnol.* **2003**, *14*, 87−95. (c) Oda, Y.; Owa, T.; Sato, T.; Boucher, T.; Daniels, S.; Yamanaka, H.; Shinohara, Y.; Yokoi, A.; Kuromitsu, J.; Nagasu, T. *Anal. Chem.* **2003**, *75*, 2159−2165. (d) Edwards, A. M.; Arrowsmith, C. H.; Pallieres, B. *Mod. Drug Discovery* **2000**, *3*, 35−44. (e) Cravatt, B. F.; Sorensen, E. J. *Curr. Opin. Chem. Biol.* **2000**, *4*, 663−668.

(2) (a) Coutinho, P. M.; Deleury, E.; Henrissat, B. *J. Appl. Glycosci.* **2003**, *50*, 241−244. (b) Coutinho, P. M.; Stam, M.; Blanc, E.; Henrissat, B. *Trends Plant Sci.* **2003**, *8*, 563−565. (c) Henrissat, B.; Coutinho, P. M.; Davies, G. J. *Plant Mol. Biol.* **2001**, *47*, 55−72.

(3) (a) Baum, L. G. *Immunity* **2002**, *16*, 5−8. (b) Scanlin, T. F.; Glick, M. C. *Glycoconjugate J.* **2001**, *17*, 617−626. (c) Johnson, D. A.; Liu, H.-W. Deoxysugars: Occurrence, Genetics, and Mechanisms of Biosynthesis. In *Comprehensive Natural Products Chemistry*; Elsevier: New York, 1999; Vol. 3, pp 311−365. (d) Trefzer, A.; Bechthold, A.; Salas, J. A. *Nat. Prod. Rep.* **1999**, *16*, 283−299. (e) Alberts, B.; Bray, D.; Lewis, L.; Martin, R.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 3rd ed.; Garland Publishing: New York, 1994; pp 13−14.

(4) Seto, N. O. L.; Palcic, M. M.; Compston, C. A.; Li, H.; Bundle, D. R.; Narang, S. A. *J. Biol. Chem.* **1997**, *272*, 14133−14138.

10.1021/ol049389b CCC: $27.50
Published on Web 05/07/2004

is crucial to categorize these enzymes, however; glycosidases and glycosyltransferases are named according to the sugar whose glycosidic linkage is cleaved or joined.

The basic strategy of a chemical proteomics approach (Figure 1) relies on the differentiation of all possible

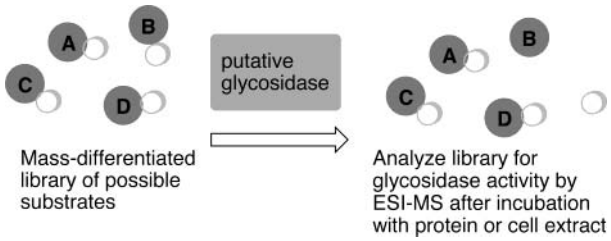


Figure 1. New strategy to determine the chemical function of the protein products of genes putatively assigned functions as glycosidases relies on incubation of the new protein with libraries of possible substrates that are inert to high background hydrolysis rates.

substrates for a given glycosidase by mass rather than by an ultraviolet, radioactive, or fluorescent tag, which would be difficult to distinguish in one reaction vial. In contrast to peptides that have established sequencing protocols, monosaccharides are more difficult to identify as many structures differ only by stereochemistry and not by molecular weight or functional group.[5] To break this mass redundancy, carbohydrates with various mass-differentiated linkages would be needed. Unlike *p*-nitrophenol analogues, these glycosides should not be susceptible to high background hydrolysis rates, even at the elevated temperatures required to analyze putative proteins from hyperthermophilic sources. Finally, the library cannot contain any inhibitors of glycosidase activity for the success of this one-pot approach.

To test the feasibility of this strategy, small libraries of substrates were designed and synthesized (Figure 2). Although glycosidases are commonly assayed with bulky phenols as leaving groups, smaller anomeric leaving groups were desirable to avoid possible steric conflicts in the many potential active sites. Common sugars that would target the bulk of known glycosidases were chosen for an initial library and linkers were chosen for their stability as well as ease of synthesis. To this end, monosaccharides were heated with a variety of alcohols under acidic Fischer−Helferich conditions to produce the α-anomers.[6] In several cases, both the α-linked and *β*-linked anomers were produced and were separated after acetylation. Deacylation with sodium methoxide provided the desired compounds. The remaining *β*-linked substrates were synthesized under Koenigs−Knorr conditions from acylated precursors; neighboring group participation assured high yields of the desired anomer.[6]

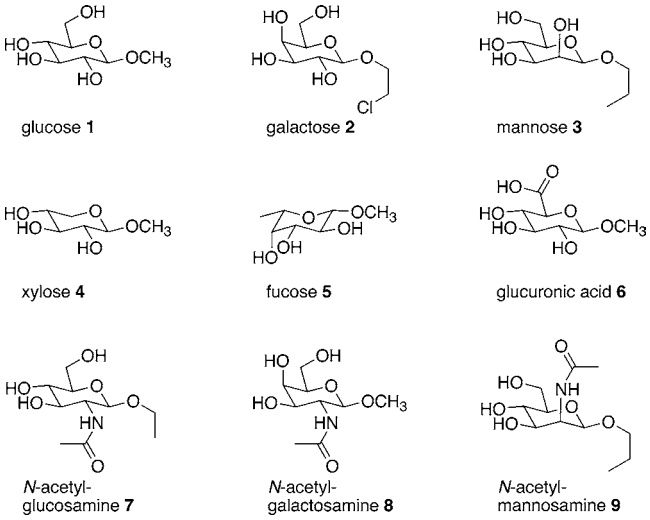


Figure 2. *β*-Linked saccharides that are common known glycosidase substrates were synthesized with various linkers to avoid any mass redundancy. A similar library of α-linked substrates was also made. (See the Supporting Information.)

With small libraries of both α- and *β*-linked carbohydrates in hand, we needed to discover conditions amenable to both buffered enzymatic reactions as well as compound ionization via electrospray for identification. Unfortunately, neutral carbohydrates are difficult to ionize.[7] The analysis of reaction aliquots diluted in aqueous trifluoroacetic acid was feasible in positive-ion mode; however, the *N*-acetylated sugars dominated the mass spectra. Addition of 3-aminophenylboronic acid (3-APBA), which forms complexes with neutral sugars at low pH, to the quench solution prior to analysis improved the spectral readout of the sugar library dissolved in water as previously reported[8] but worked poorly when the substrates were in typical enzyme reaction buffers. Addition of ammonium sulfate to the 3-APBA quench allowed both the amine-containing and neutral sugars to ionize at comparable magnitudes in positive-ion mode. In fact, altering the concentration of ammonium sulfate changed the relative ratios of amine-containing to neutral sugars in the spectra. Peaks for sugars lacking a primary hydroxyl group such as xylose (**4**) and fucose (**5**) and negatively charged sugars such as glucuronic acid (**6**) ionize to a lesser degree as expected, but were still in the range of the other sugars for easy analysis.

Carbohydrate libraries subsequently were tested for their ability to identify glycosidase enzymes by which substrate(s) they cleaved using the above method to consistently ionize the library components. The libraries were incubated with

(5) (a) Veness, R. G.; Evans, C. S. *J. Chromatogr., A* **1996**, *721*, 165−172. (b) Que, A. H.; Novotny, M. V. *Anal. Bioanal. Chem.* **2003**, *375*, 599−608.

(6) (a) Robyt, J. F. *Essentials of Carbohydrate Chemistry*; Springer-Verlag: New York, 1998, Chapter 6. (b) Melvin, F.; McNeill, A.; Henderson, P. J. F.; Herbert, R. B. *Tetrahedron Lett.* **1999**, *40*, 1201−1202. (c) Utille, J. P.; Priem, P. *Carbohydr. Res.* **2000**, *329*, 431−439. (d) McGeary, R. P.; Wright, K.; Toth, I. *J. Org. Chem.* **2001**, *66*, 5102−5105. (e) Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B. *Tetrahedron: Asymmetry* **2000**, *11*, 245−262. (f) Kikuchi, H.; Saito, Y.; Komiya, J.; Takaya, Y.; Honma, S.; Nakahata, N.; Ito, A.; Oshima, Y. *J. Org. Chem.* **2001**, *66*, 6982−6987. See also the Supporting Information.

(7) Sheeley, D. M.; Reinhold: V. N. *Anal. Chem.* **1998**, *70*, 3053−3059.

(8) Willams, D.; Lee, T. D.; Dinh, N.; Young, M. K. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 1530−1537.

known glycosidases to probe four scenarios: (1) glycosidases with one known substrate (β-galactosidase from *Aspergillus oryzae*, α-glucosidase from rice, α-mannosidase from jack beans); (2) glycosidases with multiple known substrates or secondary activities (β-*N*-acetyl glucosaminidase from jack beans and the enzymes used in subsequent scenarios); (3) a glycosidase from crude versus pure sources (β-glucosidase from almond meal and chromatographically pure); and (4) a glycosidase from high-temperature sources (β-glucosidase from *Pyroccocus furious*). In each case, substrates for the enzymes significantly diminished after several hours of incubation with the libraries when compared to control reactions containing either no enzyme or heat-denatured enzyme. At longer reaction times, secondary activities of the enzymes could be distinguished. Although not diagnostic, the cleaved sugar product could also be identified as both the mono- and diadduct of 3-APBA (Figure 3; see the Supporting Information also). More surprisingly, contaminating glycosidase activities easily could be identified from a crude protein mixture of a β-glucosidase, thereby providing a one-shot assessment of protein purity. Also, in contrast to *p*-nitrophenol-linked sugar substrates, the libraries were stable for hours at 80 °C to measure glycosidase activity of proteins from hyperthermophiles.

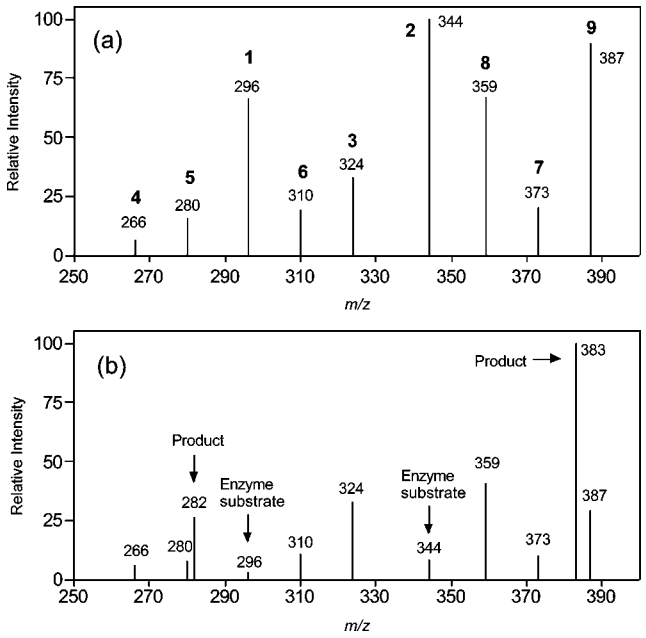


Figure 3. ESI-MS spectra in selected ion-monitoring positive-ion mode of reaction after 3 h at 80 °C of a glucosidase from *P. furiosus* with library from Figure 1 after quenching with 3-APBA: (a) without enzyme; (b) with enzyme. Glucose (**1**) and galactose (**2**) are the primary substrates of this enzyme as seen by the significant diminishment of those peaks. *N*-Acetylmannosamine (**9**) also shows some change over time.

A chemical proteomics approach can clearly help identify small molecule substrates for the large class of enzymes that cleaves glycosidic linkages and thereby identify and name these newly sequenced proteins. In addition, these libraries are relatively stable even at elevated temperatures for several hours to allow chemical function identification without requiring the optimal full-length substrate or buffer conditions of a glycosidase and to allow identification of functions of proteins from hyperthermophilic sources, too. Obviously, once the cleaved glycosidic linkage is discovered for a particular enzyme, libraries containing that sugar linked to a wide variety of compounds can be used to further narrow down the substrate specificity of the protein if desired. Protein mutants could also be screened for altered substrate specificities. Although the current libraries contain many of the sugars commonly found in mammals, the libraries can be expanded to include common plant and bacterial sugars. This competition library approach also can dissect glycosidase activity from crude extracts. Because four bacterial strains previously have been differentiated solely by their ability to cleave three different protease substrates,[9] a larger library of glycosidase substrates promises a mechanism to rapidly diagnose various cell types based on their differential ability to cleave carbohydrates. In addition, a mass-differentiated library approach should also be applicable to the study of a wide range of other enzyme classes.

Acknowledgment. We acknowledge an NSF-CAREER award, the Plant Sciences Institute, the USDA, a Shimadzu University Research Grant, and the Herman Frasch Foundation (ACS-PRF) for support of this research. N.L.P. is a Cottrell Scholar of Research Corp.

Supporting Information Available: Experimental procedures and copies of mass spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

OL049389B

(9) Basile, F.; Ferrer, I.; Furlong, E. T.; Voorhees, K. J. *Anal. Chem.* **2002**, *74*, 4290−4293.