Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6338–6342, July 1992
Biochemistry

Fast sequencing of oligosaccharides: The reagent-array analysis method

(carbohydrates/glycoproteins/enzyme array)

C. J. EDGE*, T. W. RADEMACHER*, M. R. WORMALD*, R. B. PAREKH*†, T. D. BUTTERS‡, D. R. WING*, AND R. A. DWEK*

*Oxford Glycobiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU United Kingdom; and ‡Searle Research Group, Oxford Glycobiology Unit, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU United Kingdom

Communicated by Baruch S. Blumberg, March 12, 1992

ABSTRACT **A method of oligosaccharide analysis involving controlled fragmentation resulting from enzymatic digestion is presented. The principle involves generating a set of fragments from the original oligosaccharides, characterizing them in terms of their hydrodynamic volumes, determining their molar proportions, and identifying the oligosaccharides by comparison with a computer-generated data base. Experimentally, this technique involves incubation of aliquots of a sample with a set of defined mixtures of exoglycosidases followed by pooling of the products and a single analysis on the product pool. This method has several practical advantages over current techniques, including speed and the ability to use smaller amounts of starting material. The detection of the intensity-versus-hydrodynamic volume profile is limited only by the specific activity of the labeling method. The ability to perform the enzyme digestions is limited by the individual K_m values of the enzymes.**

With the increased awareness of the biological significance of protein glycosylation (1), rapid and sensitive methods for obtaining information on the sequences of oligosaccharides are needed. Several physical, chemical, and enzymatic methods have been developed and refined (2, 3), differing in sensitivity and the type of information provided. Of these, one of the most commonly used is sequential digestion of the oligosaccharide with exoglycosidases of known and well-defined specificities (4). This method provides information about the sequence of the oligosaccharide and is extremely sensitive when applied to radiolabeled substrates, which is important when considering the amounts of material routinely obtained from biological sources. (The method would not, of course, work with metabolically labeled substrates unless the label were present in the reducing-terminus monosaccharide.) The principal disadvantages of sequential digestion are that its application is an *ad hoc* process and requires the repeated isolation and determination of the product hydrodynamic volume (prior to each incubation). In this paper a method that differs fundamentally from the sequential method is described for the application of exoglycosidase digestion to oligosaccharide structural analysis. This technique, called the reagent-array analysis method, involves dividing the purified oligosaccharide sample into aliquots, incubating each aliquot with a precisely defined mixture of exoglycosidases, recombining the products of each incubation, and performing a single analysis on the product pool. A specific set of enzyme mixtures is called an enzyme array.

In the sequential method of oligosaccharide sequencing as originally described (5), the presence of a monosaccharide or oligosaccharide unit linked in a specific manner at a nonreducing terminus is determined by the ability of a given enzyme to cause its cleavage. This procedure thus makes use of positive information (a cleavage has occurred) relating to the nonreducing termini, negative information (a cleavage

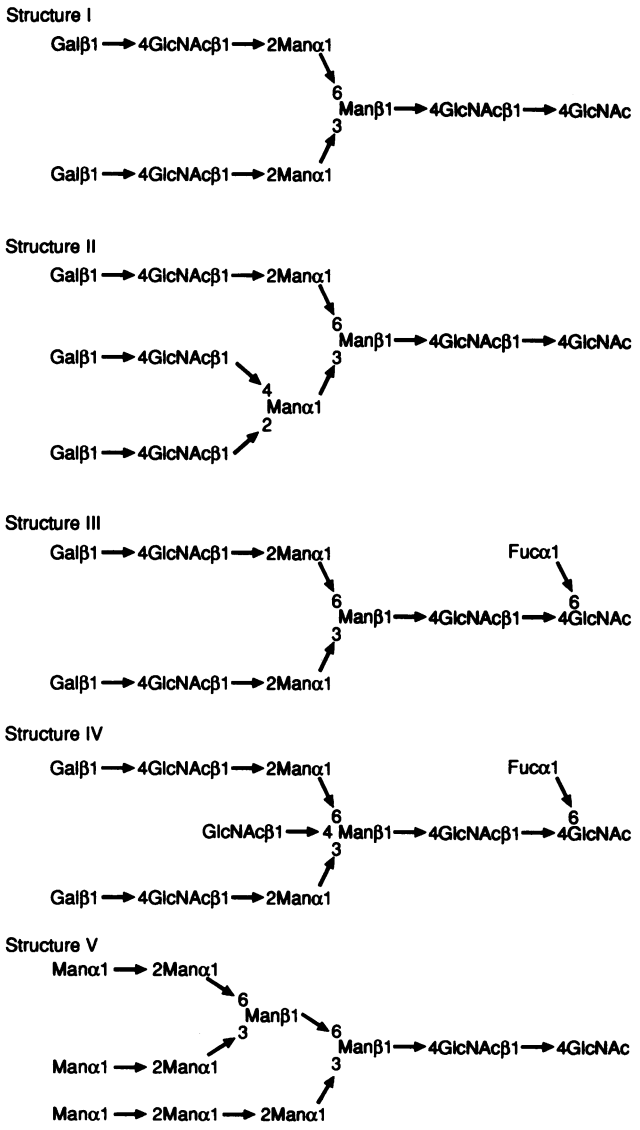

FIG. 1. Structures of the complex and oligomannose N-linked oligosaccharides referred to in the text.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: Oxford Glycosystems Ltd., Unit 4, Blacklands Way, Hitching Court, Abingdon Business Park, Abingdon, Oxon., OX14 1RG United Kingdom.

has not occurred) being included only in a rather generalized manner; e.g., a cleavage not occurring with jack bean β-1,4-galactosidase implies the absence of any terminal galactose linked β1-4 to the rest of the structure. The use of positive digestion information alone is sufficient to sequence a linear structure, but negative information, or information from another source (in the case of oligosaccharides, often mass spectroscopy or prior knowledge about common structures), is necessary to characterize a branched structure.

Conversely, in the enzyme-array method, the presence of a specific linkage anywhere in the oligosaccharide is determined by the inability of an enzyme mix lacking a given enzyme to cleave that linkage (a stop point) and the ability of the other enzymes to cleave the linkages up to that point. Thus, systematic use is made of both positive and negative information, the information obtained relating directly to the whole oligosaccharide.

Enzymes	Enzyme mixes
A	1 2 3 4 5 6 7
Jack bean β-galactosidase	○ ○ ● ● ● ● ●
Jack bean β-N-acetylhexosaminidase	○ ● ○ ● ● ● ●
Diplococcus pneumoniae β-N-acetylhexosaminidase	○ ● ● ○ ● ● ●
Jack bean α-mannosidase (arm specific)	○ ● ● ● ○ ● ●
Achatina fulica β-mannosidase	○ ● ● ● ● ○ ●
B	1 2 3 4 5 6 7 8
Bovine epididymal α-fucosidase	○ ○ ● ● ● ● ● ●
Jack bean β-galactosidase	○ ● ○ ● ● ● ● ●
Jack bean β-N-acetylhexosaminidase	○ ● ● ○ ● ● ● ●
Diplococcus pneumoniae β-N-acetylhexosaminidase	○ ● ● ● ○ ● ● ●
Jack bean α-mannosidase (arm specific)	○ ● ● ● ● ○ ● ●
Achatina fulica β-mannosidase	○ ● ● ● ● ● ○ ●
C	1 2 3 4 5 6 7 8 9
Almond meal α-3/4-fucosidase	○ ○ ● ○ ● ● ● ● ●
Charonia lampas α-fucosidase	○ ● ○ ● ● ● ● ● ●
Jack bean β-galactosidase	○ ● ● ○ ● ● ● ● ●
Jack bean β-N-acetylhexosaminidase	○ ● ● ● ○ ○ ● ● ●
Diplococcus pneumoniae β-N-acetylhexosaminidase	○ ● ● ● ● ○ ● ● ●
Jack bean α-mannosidase (non-arm specific)	○ ● ● ● ● ● ○ ● ●
Achatina fulica β-mannosidase	○ ● ● ● ● ● ● ○ ●

FIG. 2. Enzyme arrays used for analysis of structures I (*A*), II (*B*), and III–V (*C*). Results are shown in Figs. 3 and 4. Numbers refer to the separate enzyme mixes; ○, absence of the enzyme; ●, presence of the enzyme.

The method differs from the previously described use of an enzyme mix applied to a pool of known oligosaccharides to determine the prevalence of a common epitope (6). Here, an array of enzyme mixes is used on a single, pure oligosaccharide to establish the whole sequence of that oligosaccharide.

This paper presents experimental results obtained from application of the enzyme-array method to five N-linked oligosaccharides. Computer simulations of these experiments are also presented. The advantages and disadvantages of the method are discussed.

MATERIALS AND METHODS

^{3}H-radiolabeled alditols I–V (Fig. 1) were purchased from Oxford Glycosystems Ltd., with the ^{3}H label at C-1 of the reducing terminus.

Oligosaccharide I was divided into seven aliquots and digested with the exoglycosidase mixes shown in Fig. 2*A*, oligosaccharide II was divided into eight aliquots and digested with the exoglycosidase mixes shown in Fig. 2*B*, and oligosaccharides III–V were digested with the exoglycosidase mixes shown in Fig. 2*C*. The enzyme mixes in Fig. 2 *A* and *B* were prepared in 0.1 M citric acid/0.2 M disodium phosphate/0.001% sodium azide, pH 5.0; the enzyme mixes in Fig. 2*C* were prepared in 0.1 M citrate buffer/0.001% sodium azide, pH ≈5.3. The enzyme activities are shown in Table 1. Note that the concentration of jack bean α-mannosidase in the arrays shown in Fig. 2 *A* and *B* is arm-specific whereas the concentration in Fig. 2*C* is non-arm-specific. The substrate concentration for structures I and II was 30 μM. The substrate concentration for structures III–V was 5–10 μM. Incubations were carried out for 48 hr at 37°C and were terminated by heating at 100°C for 20 sec.

The resulting product solutions were pooled and passed through beds of Chelex-100 (Na^+ form), Dowex AG-50 W-X12 (H^+ form), AG-3-X4A (OH^- form) (Bio-Rad), and QAE-Sephadex (Pharmacia) resins. The resultant solution

Table 1. Activities of enzymes used in the experimental analyses

Enzyme*	Activity,† units/ml
Analysis of structures I and II	
Bovine epididymal α-L-fucosidase	1.0
Jack bean β-galactosidase	9.0
Jack bean β-*N*-acetylhexosaminidase	11.0
Diplococcus pneumoniae β-*N*-acetylhexosaminidase	0.01
Jack bean α-mannosidase	12.0
Achatina fulica β-mannosidase	0.3
Analysis of structures III–V	
Jack bean β-galactosidase	10.0
Jack bean β-*N*-acetylhexosaminidase	40.0
Diplococcus pneumoniae β-*N*-acetylhexosaminidase	0.005
Achatina fulica β-mannosidase	0.3
Jack bean α-mannosidase	40.0
Almond meal α-3/4-L-fucosidase	0.02
Charonia lampas α-L-fucosidase	0.08

*Details of enzyme sources are given in ref. 6. Details of enzyme specificities except for the fucosidases are given in ref. 7. The specificity of *Charonia lampas* α-L-fucosidase is given in ref. 8, that of bovine epididymal α-L-fucosidase is given in refs. 9 and 10, and that of almond α-L-fucosidase is given in refs. 8 and 11.

†One unit of enzyme activity is defined as the amount of enzyme required to hydrolyze the appropriate 3 mM *p*-nitrophenyl glycoside at a rate of 1 μmol/min at 37°C, except for almond meal α-3/4-L-fucosidase, where one unit of activity is defined as that amount of enzyme required to release 1 μmol of [^{14}C]fucose per min at 37°C from [fucose-^{14}C]asialo-α_1-acid glycoprotein.

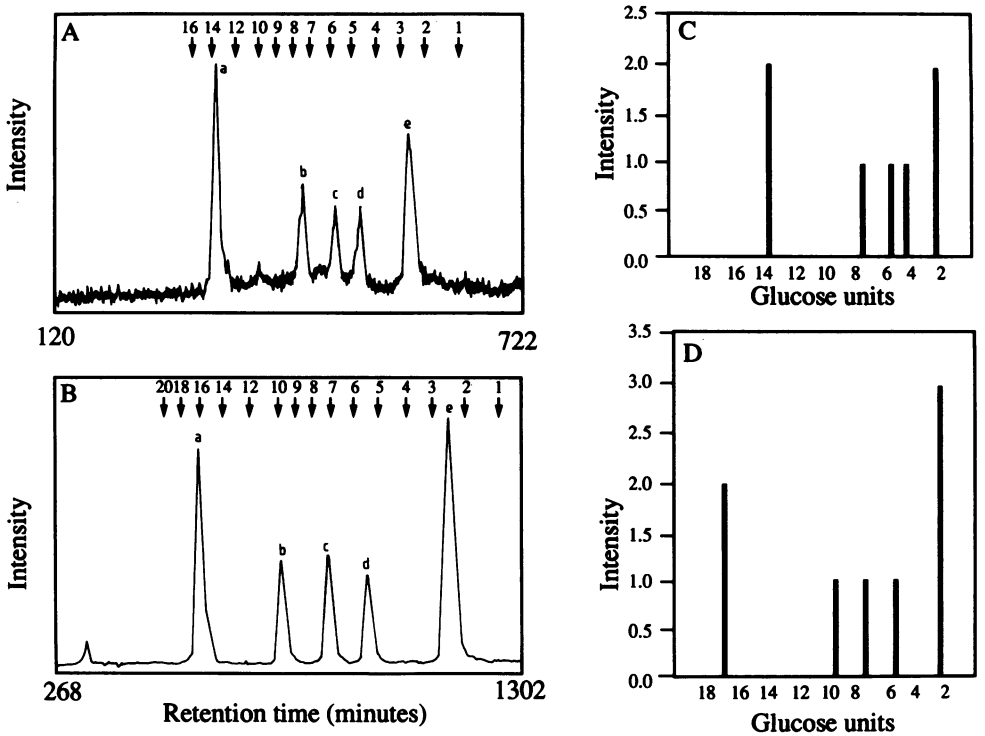

FIG. 3. Experimental (*Left*) and theoretical (*Right*) Bio-Gel P4 elution profiles for digests of structures I (*A* and *C*) and II (*B* and *D*) with the enzyme arrays shown in Fig. 2 *A* and *B*, respectively. Scales at the top of the experimental profiles show glucose units. Data are summarized in Table 2.

was analyzed on a Bio-Gel P4 column (12). The intensities of the resulting peaks were determined by pooling the fractions corresponding to each peak and measuring the radioactivity by scintillation counting.

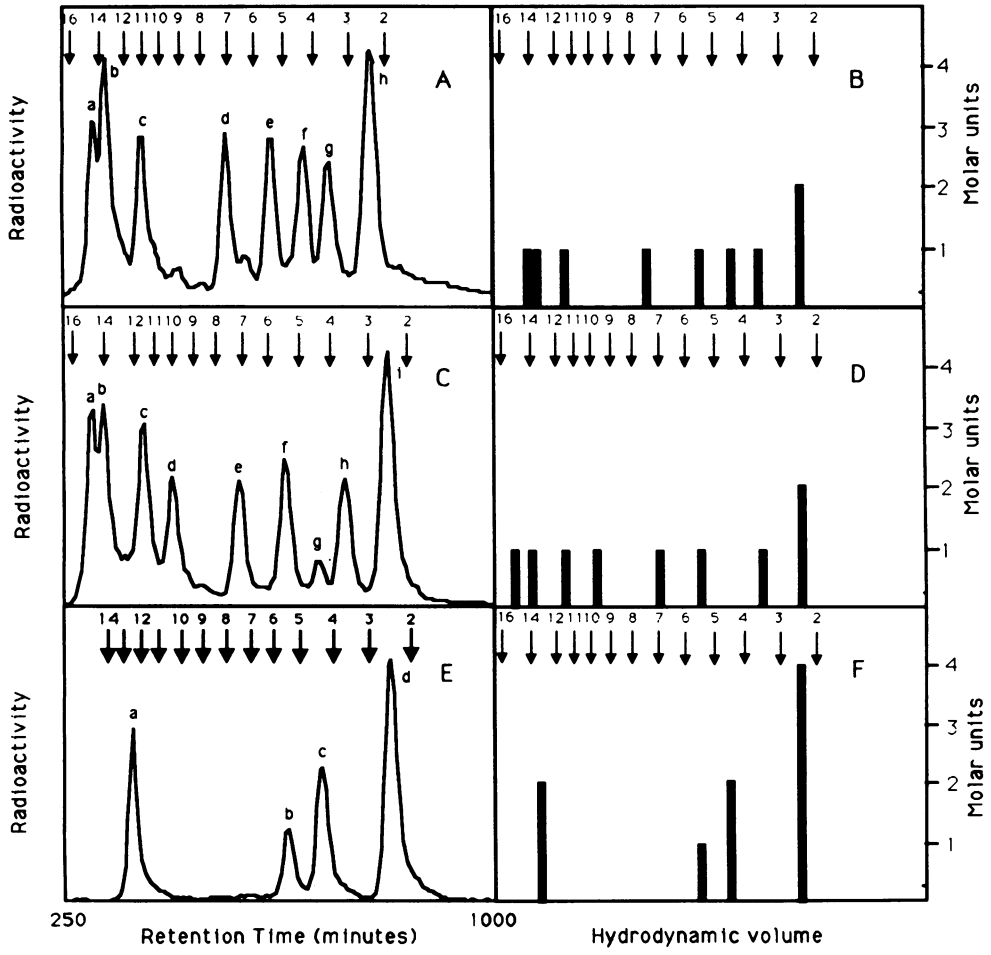

FIG. 4. Experimental (*Left*) and theoretical (*Right*) Bio-Gel P4 elution profiles for digests of structures III (*A* and *B*), IV (*C* and *D*), and V (*E* and *F*). Scales at the top of both sets of profiles show glucose units. Data are summarized in Table 2.

A computer program was written that allows the simulation of the action of any set of exoglycosidases upon a given substrate to give a final product together with its calculated hydrodynamic volume. The change in hydrodynamic volume corresponds to the loss of integral numbers of monosaccharide residues. In general, the loss of a hexose will reduce the hydrodynamic volume by one glucose unit and the loss of an *N*-acetylhexosamine will reduce the hydrodynamic volume by two glucose units. Simulations of the digestions by the enzyme arrays given in Fig. 2 were carried out on an initial data base of possible oligosaccharide structures generated by using the current knowledge of N-linked oligosaccharide biosynthetic glycosyltransferase activities. (Note, however, that the data base is not restricted to known biosynthetic intermediates.) The resultant theoretical fragments from the mixes were added together to give a profile of molar proportions of the fragments versus hydrodynamic volume for each entry in the data base.

The data base used to match the experimental profiles of structures I–V consisted of the calculated profiles of 63,508 complex, 5044 hybrid, and 42 oligomannose structures. The tolerance limit for matching hydrodynamic volumes is proportional to the hydrodynamic volume of each peak and is set to ±20%. The tolerance limit for matching intensities is set to ±0.5 molar unit.

The accuracy of any given match was given by the parameter

$$\chi^2 = \sum_{\text{All peaks}} \frac{(\text{experimental HV} - \text{calculated HV})^2}{\text{calculated HV}},$$

where HV is the hydrodynamic volume of the peak.

Table 2. Experimental and calculated results of Bio-Gel P4 chromatography of digests of structures I (enzyme array A), II (enzyme array B), and III–V (enzyme array C)

Structure I		a	b	c	d	e				
R_t	Exp.	13.5	7.3	5.7	4.5	2.6				
	Calc.	13.5	7.5	5.5	4.5	2.5				
I	Exp.	2.0	1.0	1.0	1.0	2.0				
	Calc.	2.0	1.0	1.0	1.0	2.0				
Structure II		a	b	c	d	e				
R_t	Exp.	16.1	9.7	7.3	5.5	2.5				
	Calc.	16.5	9.5	7.5	5.5	2.5				
I	Exp.	2.0	1.0	1.0	0.9	3.1				
	Calc.	2.0	1.0	1.0	1.0	3.0				
Structure III		a	b	c	d	e	f	g	h	
R_t	Exp.	14.4	13.5	11.2	7.2	5.5	4.5	3.6	2.5	
	Calc.	14.5	13.5	11.5	7.5	5.5	4.5	3.5	2.5	
I	Exp.	0.9	1.4	1.1	0.9	1.0	0.9	1.0	1.8	
	Calc.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	
Structure IV		a	b	c	d	e	f	g	h	i
R_t	Exp.	15.0	14.1	11.6	10.1	7.2	5.5	4.5	3.6	2.5
	Calc.	15.0	14.0	12.0	10.5	7.5	5.5	4.5	3.5	2.5
I	Exp.	0.9	1.1	0.9	1.0	0.9	1.0	0.3	0.9	2.0
	Calc.	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	2.0
Structure V		a	b	c	d					
R_t	Exp.	12.4	5.5	4.5	2.6					
	Calc.	13.5	5.5	4.5	2.5					
I	Exp.	2.0	1.0	2.0	4.0					
	Calc.	2.0	1.0	2.0	4.0					

Elution positions in glucose units and intensities (*I*) are given for peaks a–i (see Figs. 3 and 4).

RESULTS

The experimental results for structures I and II are given in Fig. 3 *Left*, and those for structures III–V are given in Fig. 4 *Left*. The theoretical predictions for the structures are shown in Figs. 3 and 4 *Right*. Table 2 summarizes the results for all the structures. The results of the searches of the data bases for the profiles matching those obtained experimentally from structures I–V are shown in Table 3. The comparison is good in all cases.

All five profiles contain peaks at 2.5 glucose units, corresponding to complete digestion of the native structures by the given enzyme array to the final common product [1-^{3}H]-*N*-acetylglucosaminitol. This set of enzymes is therefore sufficient to completely digest the structures. The integer intensity of the peaks (normalized to the total number of mixes) also indicates that a single reducing-terminus product is obtained from each enzyme mix. This is important, because otherwise the assumption would have to be made that either one or more enzymes had more than one specific activity or that incomplete digestion had taken place in one or more of the mixes.

DISCUSSION

From the above results, it is clear that the enzyme-array method of sequencing of oligosaccharides is a viable experimental technique, capable of providing a great deal of primary structural information in a single chromatographic run. It does this by making systematic use of all positive and negative information, compared with the sequential technique, which makes use of only positive information in an *ad hoc* fashion.

The method can be accurately modeled on a computer, thus allowing the exhaustive data bases of results necessary for the "fingerprint matching" to be generated theoretically

Table 3. Results of a search of data bases to match the experimental data obtained on structures I–V

No. of structures in data base	68,594
Matching tolerance limits	
Hydrodynamic volume	±20%
Intensity	±0.5%
Structure I	
No. of structures found	1
χ^2 for correct structure	0.026
Structure II	
No. of structures found	12
χ^2 for correct structure	0.021 (2 structures)
χ^2 for next best structures	0.080 (4 structures)
Structure III	
No. of structures found	1
χ^2 for correct structure	0.023 (1 structure)
Structure IV	
No. of structures found	103
χ^2 for correct structure	0.070 (3 structures*)
χ^2 for next best structures	0.096 (12 structures)
Structure V	
No. of structures found	4
χ^2 for correct structure	0.007 (1 structure)
χ^2 for next best structures	0.039 (3 structures)

The data bases used contain the theoretical profiles, obtained using the enzyme arrays shown in Fig. 2, of all complex (63,508 structures), hybrid (5044 structures), and oligomannose (42 structures) N-linked oligosaccharides. Some structures are repeated in more than one data base.

*Structure IV is one of the three structures found with the lowest value for χ^2. The three structures arise as a result of enzyme array C being unable to distinguish between a GlcNAc monosaccharide linked at position 2, 4, or 6 on the six-arm mannose residue.

and enabling easy technique development and refinement. Experimental results can be incorporated into these data bases, as they become available, to increase further the accuracy of the technique. The use of theoretical data bases associated with the analysis also removes any chance of user bias in interpreting what may be ambiguous results and will ensure that all structures compatible with the experimental data are considered.

The enzyme-array technique has other practical advantages over the sequential digestion technique in being very much faster, requiring a smaller amount of sample for complete sequencing (there is no sample loss associated with repeated analysis and repurification) and having a very well-defined protocol. This removes the need for specific prior knowledge in designing the experimental procedure, making it easier to use routinely and also amenable to automation.

The main limitations arise from the nature of the enzyme array and the data base used to search the structures. Clearly, more sophisticated arrays can be developed than those illustrated here, and there will be a requirement for further development of a bank of highly specific (bond–arm–aglycon) exoglycosidases, as well as neuraminidases, sulfatases, and phosphatases. An example of such a refinement is seen in the enzyme array in Fig. 2*C*, in which almond meal α-3/4-L-fucosidase was included. This will allow the sequencing of oligosaccharides in which fucose is present on the outer arms. Such fucose-containing structures are not digested by the *Charonia lampas* α-L-fucosidase at the concentrations present in the array, which is there for the purpose of digesting the core α-1,6-fucose present in structures III and IV. However, the use of a fairly simple array can always be complemented by information from other techniques. For techniques amenable to computer simulation (such as partial acetolysis or permethylation analysis, which can also use radiolabeled substrates), a similar computer-based analysis approach can be used to select an exhaustive list of structures compatible with all the available information.

Future developments of the enzyme array method will be driven by the requirement to analyze ever smaller quantities of oligosaccharides. In a recent publication (13) the oligosaccharides from 20 μg of the cytokine interleukin 6 from normal human blood mononuclear cells were sequenced at the picomolar level. Clearly, there will be many cases in the future where such quantities of material will not be available but yet oligosaccharide analysis will be required. The enzyme-array method will allow such analysis to be undertaken in an efficient, rapid manner.

We thank Dr. P. Williams, Dr. D. L. Fernandes, B. L. Matthews, and Miss V. Gamble for help and advice. The Oxford Glycobiology Unit is supported by the Monsanto Company.

1. Rademacher, T. W., Parekh, R. B. & Dwek, R. A. (1988) *Annu. Rev. Biochem.* **57**, 785–838.
2. Dell, A. (1987) *Adv. Carbohydr. Chem. Biochem.* **45**, 19–72.
3. Vliegenthart, J. F. G., Dorland, L. & van Halbeek, H. (1983) *Adv. Carbohydr. Chem. Biochem.* **41**, 209–374.
4. Kobata, A. (1979) *Anal. Biochem.* **100**, 1–14.
5. Mizuochi, T., Yonemasu, K., Yamashita, K. & Kobata, A. (1978) *J. Biol. Chem.* **253**, 7404–7409.
6. Parekh, R. B., Tse, A. G. C., Dwek, R. A., Williams, A. F. & Rademacher, T. W. (1987) *EMBO J.* **6**, 1233–1244.
7. Kobata, A. (1984) in *Biology of Carbohydrates*, eds. Ginsberg, V. & Robbins, P. W. (Wiley, New York), Vol. 2, pp. 87–162.
8. Butters, T. D., Scudder, P., Rotsaert, J., Petursson, S., Fleet, G. W. J., Willenbrock, F. W. & Jacob, G. S. (1991) *Biochem. J.* **279**, 189–195.
9. DiCioccio, R. A., Barlow, J. J. & Matta, K. L. (1982) *J. Biol. Chem.* **257**, 714–718.
10. Bernard, M., Percheron, F. & Foglietti, M. J. (1983) *Int. J. Biochem.* **15**, 1385–1387.
11. Scudder, P., Neville, D. C. A., Butters, T. D., Fleet, G. W. J., Dwek, R. A., Rademacher, T. W. & Jacob, G. S. (1990) *J. Biol. Chem.* **265**, 16472–16477.
12. Yamashita, K., Mizuochi, T. & Kobata, A. (1982) *Methods Enzymol.* **83 (D)**, 105–126.
13. Parekh, R. B., Dwek, R. A., Rademacher, T. W., Opdenakker, G. & van Damme, J. (1992) *Eur. J. Biochem.* **203**, 135–141.