Towards a classification of glycosyltransferases based on amino acid sequence similarities: prokaryotic α -mannosyltransferases

Roberto A. GEREMIA*, E Alejandro PETRONI†, Luis IELPI† and Bernard HENRISSAT*‡

*Centre de Recherches sur les Macromolécules Végétales (affiliated with the Université Joseph Fourier), C.N.R.S., BP 53, 38041 Grenoble cedex 9, France, and

†Instituto de Investigaciones Bioquímicas 'Fundación Campomar', Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, y CONICET, Patricias Argentinas 435 (1405) Buenos Aires, Argentina

A number of genes encoding bacterial glycosyltransferases have been sequenced during the last few years, but their low sequence similarity has prevented a straightforward grouping of these enzymes into families. The sequences of several bacterial α -mannosyltransferases have been compared using current alignment algorithms as well as hydrophobic cluster analysis (HCA). These sequences show a similarity which is significant but too low to be reliably aligned using automatic alignment methods. However, a region spanning approx. 270 residues in these proteins could be aligned by HCA, and several invariant amino acid

residues were identified. These features were also found in several other glycosyltransferases, as well as in proteins of unknown function present in sequence databases. This similarity most probably reflects the existence of a family of proteins with conserved structural and mechanistic features. It is argued that the present IUBMB classification of glycosyltransferases could be complemented by a classification of these enzymes based on sequence similarities analogous to that which we proposed for glycosyl hydrolases [Henrissat, B. (1991) *Biochem. J.* **280**, 309–316].

INTRODUCTION

An increasing number of sequences of proteins involved in polysaccharide biosynthesis have become available during the last few years, mostly due to the cloning and sequencing of the corresponding genes. A number of these proteins are glycosyltransferases, which catalyse the formation of glycosidic bonds using a nucleotide–diphospho-sugar as the sugar donor. These enzymes are responsible for the sugar sequence and linkage type in oligo- and poly-saccharides. Based on the stereochemistry of the glycosidic bond formed and that of the sugar donor, two types of glycosyltransferases can be defined: inverting ones and retaining ones. Retaining glycosyltransferases result in glycosidic bonds with stereochemistry identical to that of the glycosyl donor, while inverting enzymes produce glycosidic bonds of the opposite stereochemistry [1]. This situation is identical to that described for glycosyl hydrolases, which catalyse the hydrolysis of sugars with either retention or inversion of the anomeric configuration [2]. Glycosyl hydrolases have been classified into several families based on sequence similarity [3,4]. Members of a given glycosyl hydrolase family have a similar three-dimensional structure [5] and use the same molecular mechanism [6]. In the case of glycosyltransferases, the sequence similarity is often very low, and precludes their straightforward grouping by standard sequence alignment algorithms [7,8]. The sensitive hydrophobic cluster analysis (HCA) [9,10] method has been successfully used in several cases for the grouping of proteins of very low sequence similarity [11]. Using this method we have recently succeeded in assigning several inverting β -glycosyltransferases to a single family [1].

Although several bacterial polysaccharides contain mannose (for a review, see [12]), only eight α -mannosyltransferase genes have been identified and characterized so far ([13–15a]; E. A.

Petroni and L. Ielpi, unpublished work). The derived proteins do not display global similarity, but a short consensus peptide has been reported to be conserved in three α -mannosyltransferases [13]. In the present paper we show that α -mannosyltransferases that are involved in the biosynthesis of bacterial lipopolysaccharides (LPSs) and exopolysaccharides, as well as some eukaryotic proteins, in fact share this motif, albeit in a more degenerated form. We also demonstrate that these proteins share a global similarity, and discuss the implications for a classification of glycosyltransferases that would better incorporate the mechanistic and structural features of these enzymes.

MATERIALS AND METHODS

Sequences were retrieved from either the SwissProt or GenBank sequence databanks. Sequence similarity searches were performed using BLAST at the Ecole Polytechnique Fédérale de Lausanne, Switzerland.

HCA plots were obtained using the program HCA-Plot V3.0 (Doriane S.A., Le Chesnay, France). A HCA plot is a bi-dimensional representation in which the amino acid sequence of a protein is displayed as an unrolled and duplicated longitudinal cut of a cylinder where the amino acids follow an α -helical pattern. It has been shown that the clusters of contiguous hydrophobic residues (Val, Ile, Leu, Met, Phe, Trp, Tyr) correspond to secondary structure elements in globular proteins [16]. In the plots, the following amino acids are represented by the symbols: ★ for proline, ◆ for glycine, □ for serine and □ for threonine. Graphical manipulations of the plots were carried out using IslandDraw V3.0 (Island Graphics Corp., Hoofddorp, The Netherlands). The analysis then involved the comparison of

cluster shape and of cluster distribution between several plots in order to find correspondences [9,10].

The significance of the pairwise comparisons was assessed using the PCOMPARE program of PC/Gene (Intelligenetics) using a Dayhoff MDM-78 matrix with a bias value of 60, a gap penalty of 60, and 180 random runs.

RESULTS AND DISCUSSION

A signature for α -mannosyltransferases

The sequences of several bacterial mannosyltransferases have recently become available, mostly through studies on the biosynthesis of O-antigen and exopolysaccharides. These proteins have not been assigned to a single family for lack of obvious sequence similarity. However, Kido et al. [14] have recently reported that the consensus sequence SXXEGFGLPXXE is present in three α -mannosyltransferases: MtfA and MtfB from *Escherichia coli* and RfbW from *Salmonella enterica*. We have searched for this motif in other bacterial α -mannosyltransferases (listed in Table 1). All the enzymes share the peptide

EXFGXXXXE in the C-terminal half of the protein. Interestingly, MtfA shows this motif two times in its sequence (see below). Eukaryotic glycosyltransferases were tested for this peptide, but no similarity was found. RfbO, an *S. enterica* D2 mannosyltransferase, did not display the consensus sequence. However, the only mannose residue in the antigen O of this species is β -linked, supporting the idea that the conserved motif is specific for α -mannosyltransferases. It is not surprising to find that α - and β -mannosyltransferases belong to different families, since the difference in the stereochemistry of the reaction catalysed probably reflects profound differences in the active-site architecture of the enzymes.

The conserved sequence was also found in other proteins or open reading frames. First, Genbank entry M84642 (*orf6.17*) is a gene located in the *rfb* cluster of *S. enterica* group C1. The function of the hypothetical protein was not studied, but it presumably encodes a glycosyltransferase [17]. Since *S. enterica* C1 antigen O contains two α -linked mannoses, it is conceivable that *orf6.17* codes for an α -mannosyltransferase. Surprisingly, the consensus peptide was also found in two eukaryotic proteins, PigA and Gpi3, which are involved in glycosylphosphatidyl-

Table 1 Source, origin and function of the proteins analysed

Under accession no., gb and sp indicate GenBank and SwissProt entries respectively. Size is given in amino acid residues. For definitions of MtfAN and MtfAC see text.

Protein name	Accession no.	Function	Linkage	Organism	Size	Ref.	Identity with RfbW (%)
RfbW	gbX61917	O-antigen C2 biosynthesis	α -(1,2) mannose	<i>S. enterica</i> C2	336	[13]	100
RfbZ	gbX61917	O-antigen C2 biosynthesis	α -(1,3) mannose	<i>S. enterica</i> C2	385	[13]	12.80
RfbU	gbX56793	O-Antigen biosynthesis	α -(1,4) mannose	<i>S. enterica</i> B	353	[13]	14.29
GumH	gbU22511	Xanthan gum biosynthesis	α -(1,3) mannose	<i>Xanthomonas campestris</i>	380	[15]	20.74
AceA	gbU37258	Acetan biosynthesis	α -(1,3) mannose	<i>Acetobacter xylinum</i>	394	[15a]	18.75
MtfA	gbD43637	Antigen O9 biosynthesis	3 $\times$ α -(1,2) mannose	<i>E. coli</i> O9	814	[14]	
MtfAN					380		19.94
MtfAC					361		11.01
MtfB	gbD43637	Antigen O9 biosynthesis	2 $\times$ α -(1,3) mannose	<i>E. coli</i> O9	381	[14]	16.67
MtfC	gbD43637	Antigen O9 biosynthesis	α -(1,3) mannose	<i>E. coli</i> O9	274	[14]	17.88
Orf 16.7	M84642	O-antigen biosynthesis	?	<i>S. enterica</i> C1	336	[17]	
PigA	spP37287	Biosynthesis of N-acetylglucosaminyl-phosphatidylinositol	?	<i>Homo sapiens</i>	485	[18]	18.45
Gpi3	spP32363	Biosynthesis of N-acetylglucosaminyl-phosphatidylinositol	?	<i>Saccharomyces cerevisiae</i>	452	[19]	14.58

Table 2 Scores for pairwise comparisons of the sequences in Table 1

Scores are given in S.D.s (σ). For PigAt protein, the region comprising amino acids 40–420 of the original protein was used in this study.

	RfbW	RfbU	MtfB	GumH	AceA	RfbZ	MtfAN	MtfAC	MtfC
RfbU	9.66								
MtfB	14.56	8.49							
GumH	8.08	6.37	6.27						
AceA	6.58	6.73	4.92	33.13					
RfbZ	4.43	4.26	3.65	4.61	3.67				
MtfAN	12.52	8.24	17.16	4.02	5.15	2.67			
MtfAC	6.45	7.84	7.57	2.31	3.04	2.41	6.90		
ORF6.17	4.35	4.72	5.95	2.46	0.55	1.13	6.21	4.32	0.55
PigAt	4.81	5.90	4.91	5.32	8.31	3.16	3.73	3.73	—0.63
MtfC	3.31	2.01	1.49	2.40	1.36	2.95	2.30	0.16

Figure 1 HCA plots of the conserved region in α -mannosyltransferases and PigA

The protein codes are defined in Table 1, and symbols representing amino acids are defined in the text. Vertical lines indicate the proposed correspondence between proteins. The 16 conserved segments are labelled 1–16. The α -helix is labelled Helix. F, W, Y shows a mosaic of aromatic residues.

inositol anchor biosynthesis [18,19]. PigA and Gpi3 are probably involved in the transfer of *N*-acetylglucosamine to form *N*-acetylglucosaminylinositol. These putative enzymes use an alternative sugar, suggesting that the signature we have found may also occur in glycosyltransferases using other sugar donors.

In order to find other regions of similarity, multiple alignments of the proteins listed in Table 1 were performed. While in all the cases they successfully aligned the region encompassing the signature, they failed to detect overall similarity between these proteins. To assess the significance of the similarity between the proteins, pairwise comparisons of the complete sequences were performed. All the scores exceeded the mean score of alignments with randomly shuffled sequences by at least three S.D.s (σ) (Table 2), except for the case of MtfC that will be discussed later. Among the proteins with a score higher than 3σ , only RfbZ failed to reach a more stringent score of 5σ . These results indicate that these proteins are significantly related, allowing them to be grouped in a single family. However, the similarity

was too low to obtain an automatic multiple alignment which would uncover other regions of similarity.

HCA plots

Because of its established effectiveness in the comparison of widely divergent proteins, the HCA method was used for the sequence comparisons. The sensitivity of this method allows useful global alignments to be obtained at high divergence levels, mostly because the method can reliably identify elements of secondary structure in proteins [16]. The HCA plots of representative bacterial α -mannosyltransferases and of PigA protein are shown in Figures 1 and 2. Analysis of the plots allowed most of the proteins to be divided into 16 segments (labelled 1–16 in Figures 1 and 2). The features of segments 2, 4, 7, 9, 11, 12, 13, 15 and 16 are conserved among all the proteins studied. These segments are followed by a region whose HCA pattern is indicative of an α -helical structure, and by a mosaic of aromatic

Figure 2 HCA plots of the conserved region in α -mannosyltransferase involved in the biosynthesis of lipopolysaccharide O9

The protein codes are defined in Table 1, and symbols representing amino acids are defined in the text. Broken lines indicate weaker similarity. Other details are as in Figure 1.

residues (Figures 1 and 2). Among all the proteins, conserved segments 11–16 share a striking similarity at the HCA level, which correlates with significant scores in pairwise comparisons (results not shown). Segments 7–10 are also very well conserved, but they differ in size. On the other hand, segments 1–6 appear less well conserved, and in the case of MtfC the similarity is hard to detect (Figure 2). This may explain the lower scores obtained for the pairwise comparisons of MtfC with the other proteins. The signature motif EXFGXXXXE, present in segments 13 and 14, contains the two Glu residues which are invariant in this family of proteins. Because they could also be important for structure or catalysis, we searched for other conserved amino acids. For instance, a conserved Pro is found to be almost invariant at the beginning of segment 1. A His and a Ser are conserved in segments 2 and 3 respectively (Figures 1 and 2). An invariant Lys is also found at the beginning of segment 9, while a conserved Leu is found in segment 11.

The sequences of eukaryotic α -mannosyltransferases and of a bacterial β -mannosyltransferase (RfbO from *S. enterica*) were also analysed by HCA, but no similarity was detected, indicating that they probably belong to other families.

MtfA has an internal duplication

MtfA is an α -mannosyltransferase that is involved in the biosynthesis of *E. coli* O9 LPS. It catalyses the addition of three successive α -(1,2)-linked mannose residues [14]. MtfA is twice as large as the other transferases. As this may indicate the presence of two catalytic domains that act sequentially, we searched for

the sequence signature in the additional residues of MtfA. The signature is indeed present twice in MtfA, separated by approx. 400 residues. The protein was arbitrarily divided into two halves, MtfAN (residues 1–380) and MtfAC (residues 440–801). At the DNA level the similarity between these two moieties is 48%, while at the protein level the identity is 19.8%. However, the HCA plots of the two halves are very similar and show all the conserved segments found in the other proteins under study (Figure 2). This indicates that MtfA is a two-domain protein that has probably evolved by gene duplication.

The biosynthetic pathway of O9 LPS has recently been described [14] (Figure 3). The first step is the formation of undecaprenol-*P*-*P*-Glc- α -(1,3)-Man and is catalysed by MtfC. The second step consists of the consecutive addition of two α -(1,3)-linked mannoses by MtfB. Finally, MtfA adds three α -(1,2)-linked mannoses. This last reaction can indeed be divided in two steps: the first Man residue is added to an acceptor that consists of α -(1,3)-linked mannose, while the acceptor for the other two residues is an α -(1,2)-linked mannose (see underlined sugars in Figure 3). The likely presence of two independent catalytic units in MtfA may account for these different acceptors: one unit would recognize and transfer to α -(1,3)-linked mannose, and the second would act twice on the terminal α -(1,2)-linked mannose.

Databank searches

In order to test the reliability of our analysis and to possibly detect other sequences sharing similarity with the proteins listed

Figure 3 Proposed biosynthetic pathway of the O9 polysaccharide from *E. coli* O9 as proposed previously [14]

Man is mannose, Glc is glucose, and Und is undecaprenol. Sugars added by the corresponding enzyme are in bold. The two last mannoses added by MtfA are underlined.

Table 3 Proteins retrieved from databases

The two invariant Glu residues are in bold

PROTEIN NAME	ACCESSION NUMBER ^a	SIGNATURE PEPTIDE	ORGANISM	SIZE ^b	FUNCTION	L ^c	N- ^d
YPJH	spP42982	KLLSEK ES FGLVLL EA MACGVPCI	<i>Bacillus subtilis</i>	357	Unknown	+	+
CAPH	spP39857	FLFPSEL EG LPGAL IE AHHNCIVI	<i>Staphylococcus aureus</i>	355	V Capsular polysaccharide biosynthesis	+	+
CAPJ	spP39859	YIQPSLT EG HCRAT IE AIGNGVPTL	<i>Staphylococcus aureus</i>	391	V Capsular polysaccharide biosynthesis	+	+
CAPM	spP39862	FVFPSTR EG FGNV IE AQALEVPVI	<i>Staphylococcus aureus</i>	380	V Capsular polysaccharide biosynthesis	-	+
AMSK	gbX77921	TAADGDM EG IPVAL ME AMAVGLPVV	<i>Erwinia carotovora</i>	407	Polysaccharide biosynthesis	+	+
RFBC	gbL31762	ALPSVY SE GVPRILL EA SSVGRACI	<i>Klebsiella pneumoniae</i>	377	Galactosyltransferase	+	+
RFBF	gbL41518	ALPSVY SE GVPRILL EA SSVGRPCI	<i>Klebsiella pneumoniae</i>	377	Galactosyltransferase	+	+
VIPC	spQ04975	FILFSR YEG LPNV IE AQMVGVPVI	<i>Salmonella typhimurium</i>	578	VI polysaccharide biosynthesis	+	-
CPSF	gbL36873	VLPSFY RE GVPKTL EA GAMGKPII	<i>Proteus mirabilis</i>	370	Unknown	+	+/-
TAGE	spP13484	TISTSH FEG FLSN ME ALSNGCPVV	<i>Bacillus subtilis</i>	673	α-Glucosyltransferase	-	+
YEFL	spP26388	TGTDGDM EG IPVAL ME AMAVGIPVV	<i>Salmonella enterica</i>	406	Next to <i>rfb</i> region	+	+
RFBF-1	gbL34167	ALPSTY SE GVPRILL EA SSVGRACI	<i>Serratia marcescens</i>	380	Galactosyl transferase	+	+
RFPB	gbS73325	VLPSY RE GVPRST QEA AMAGRPII	<i>Shigella dysenteriae</i>	377	Galactosyltransferase	-	+
TRSD	gbZ47767	YAMTSR FEG FPML EA KASGLPII	<i>Yersinia enterocolitica</i>	358	Probable glycosyltransferase	+	-
HI1698	gbU32842	YCLPSQ TE GLPLVV IE AMAFGLPIV	<i>Haemophilus influenzae</i>	353	Hypothetical protein	-	+/-
RFAK	spP26470	IVPSQV EA FCMVAV EA MAAGKAVI	<i>Salmonella typhimurium</i>	381	α-(1,2)-N-acetylglucosaminyltransferase	+	+
RFBAT_19	gbX59554	FPSHLR SE AFGISL IE AQMYCKAII	<i>Vibrio cholerae</i>	410	Possible outer membrane protein	+	+
TRSE	gbZ47767	FVLSS EWEG FPVLV IE AMACKKIIIV	<i>Yersinia enterocolitica</i>	344	Putative glycosyltransferase	+	+
CPS-10 ORF8	gbD21242	VLPIKW QEP FRVV VE AIFAGKVVL	<i>Klebsiella pneumoniae</i>	373	K2 Capsular polysaccharide biosynthesis	+	+
CPS-9 ORF7	gbD21242	LLMPSR WES FGLVAV EA QLYGVPI	<i>Klebsiella pneumoniae</i>	358	K2 Capsular polysaccharide biosynthesis	+	+
U2168F	gbU00018	VAVPSY SE SFGLVAV EA QACGTPVV	<i>Mycobacterium leprae</i>	409	Similar to RfaG and SusY	+	+
RFAK	spP25740	LLHPAY QEA AGIVL EA ITAGLPVL	<i>Escherichia coli</i>	374	Probable glycosyltransferase	+	+
ALG2	spP43636	LLYTPAY EH FGIVPL EA MKLGKPV	<i>Saccharomyces cerevisiae</i>	529	Glycosyltransferase	+	+
ECRFBG_1	gbX59852	YVCPSKI EG TNPVL ES MACGVPI	<i>Escherichia coli</i>	697	Sugar- or O-antigen transferase	-	-
YEFL_ECOLI	spP37751	VCFPSRL ET WGLPL SE AKERGKWVL	<i>Escherichia coli</i>	372	Probable glycosyltransferase (LPS biosynthesis)	-	-
RFAB_ECOLI	spP27127	LLLTSS FEG FPMTL EA LSWGIPCI	<i>Escherichia coli</i>	369	α-Glucosyltransferase	-	+
RFAK_ECOLI	spP27242	LLTIPIL ET ETLPL DE FIYTVALT	<i>Escherichia coli</i>	357	Probable glycosyltransferase (LPS biosynthesis)	-	+

^a gb and sp indicate GenBank and SwissProt entries respectively.

^b In amino acids residue.

^c Presence (+) or absence (—) of the conserved leucine in segment 11.

^d Similarity in the N-terminal region (segments 1–6) compared with consensus: +, good; +/-, medium; —, poor.

in Table 1, databank searches were undertaken. Because HCA is not adapted to extensive databank searches, a consensus peptide PSLHEGFGLPXVEAMSGLP was defined after inspection of the HCA plots in Figures 1 and 2. Databanks were searched for this peptide using BLAST [20]. In order to maximize the likelihood of finding similarities, BLAST searches were also performed with the RfbU protein. In this case, only those proteins sharing similarity with the consensus peptide were retained. The results are summarized in Table 3. An alignment of the region of the consensus peptide is also shown, where the

XXXXXXXXXE segment is strictly conserved (Table 3). Analysis of HCA plots indicates that, in all of the listed proteins, the HCA pattern around the signature peptide is similar to that of segments 11–16 defined earlier. However, some proteins lack the semi-conserved Leu residue in fragment 11 (Table 3). In other proteins the N-terminal region (segments 1–6) is less conserved (see Table 3). Among the sequences retrieved using BLAST, we found several glycosyltransferases with different donor specificity, which indicates that the features we found for α-mannosyltransferases are common to other glycosyltransferases. Some

of these proteins are known to act with retention of the stereochemistry at the anomeric carbon of the sugar; for the others, the mechanism is not known. It is of note that all the proteins which were retrieved are either putative glycosyltransferases or proteins involved in polysaccharide biosynthesis whose exact function is not yet known.

We have noticed that the consensus peptide is also present in sucrose and sucrose-phosphate synthases, which are also retaining α -glycosyltransferases. However, the similarity could not be extended to other regions of these proteins, making further studies necessary to confirm the resemblance (if any) between the family described here and sucrose synthases.

The peptide EXXXXXXXE was also found in other proteins, i.e. the SwissProt entries P31612 (a viral protein) and P21302 (an ornithine carbamoyltransferase). However, visual examination of the corresponding HCA plots did not reveal any similarity. Furthermore, analysis of the region that would correspond to segments 11–16 was performed by calculating the HCA scores against the same segment taken from RfbU. The calculated score was 21 % for P31612 and 38 % for P21302, confirming that there is no similarity in this domain, since HCA similarity scores become significant only above 60 % [10].

Implications for the mechanism of action

Irrespective of the direction of the reaction (i.e. synthesis or hydrolysis), there are two main catalytic mechanisms for glycosyl transfer, leading to either retention or inversion of the anomeric configuration of the sugar involved in the reaction. In the glycosyltransferases studied here, catalysis takes place by a retaining mechanism, since the sugar donor has the same stereochemistry as the glycosidic bond that is formed. Such a mechanism, which is likely to proceed via a glycosyl-enzyme intermediate [1], requires two catalytic residues (an acid and a base). Since the catalytic residues of enzymes are usually subjected to intense conservation pressure, they are probably located in the best conserved region of the present family of proteins, i.e. the region that spans segments 7–16. Only three polar amino acids are invariant in this region, a Lys and two Glu residues. Only Glu residues have the appropriate chemical characteristics to account for this type of catalysis. The two conserved Glu residues may, therefore, be of particular importance for catalysis in these enzymes.

Towards a classification of glycosyltransferases

Here we demonstrate that several enzymes that share the same molecular mechanism can be grouped in one family, and that the best conserved region of these proteins can be used to search for other glycosyltransferases in sequence databanks. Several of the retrieved sequences represent enzymes that use different sugar donors. It is, therefore, possible that the hydrophobic cluster organization and conserved glutamate residues are indicative of a broader family that includes several other retaining α -D-glycosyltransferases. Biochemical analysis of proteins that belong to this family and whose function is unknown will serve as a test for this hypothesis.

The number of sequences of glycosyltransferases present in the databases has dramatically increased during the last 5 years, and it is expected that this number will continue to increase steadily. Little is known, however, about the structure–function relationships of glycosyltransferases, mainly because of the scarcity of three-dimensional structures. Since sequence similarity is a good indication of structural similarity, a systematic comparison of the primary structures of glycosyltransferases should enable the definition of families of structurally related proteins. The IUBMB classification of glycosyltransferases is presently based on the sugar transferred, the sugar donor and the acceptor. A useful and complementary classification of glycosyltransferases should incorporate the following criteria: (i) sequence (hence structural) similarity, and (ii) stereochemistry of the reaction. The classification system that we propose will also help to handle sequences that become available in the future. Based on the analysis of other sequences, we have previously identified a family of β -glycosyltransferases which correlated with the stereochemistry of the reaction and the processivity of the enzymes [1]. In this paper we present another family of proteins containing α -glycosyltransferases and related proteins. There are still many glycosyltransferase sequences that do not belong to these two established families, and their categorization will be the objective of our future work.

We thank Dr. P. Reeves for useful suggestions.

REFERENCES

- Saxena, I. M., Brown, R. M. J., Fèvre, M., Geremia, R. A. and Henrissat, B. (1995) *J. Bacteriol.* **177**, 1419–1424
- Sinnott, M. L. (1991) *Chem. Rev.* **90**, 1171–1202
- Henrissat, B. (1991) *Biochem. J.* **280**, 309–316
- Henrissat, B. and Bairoch, A. (1993) *Biochem. J.* **293**, 781–788
- Davies, G. and Henrissat, B. (1995) *Structure* **3**, 853–859
- Gebler, J., Gilkes, N. R., Claeysens, M., Wilson, D. B., Béguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Warren, R. A. J. and Withers, S. G. (1992) *J. Biol. Chem.* **267**, 12559–12561
- Paulson, J. C. and Colley, K. J. (1989) *J. Biol. Chem.* **264**, 17615–17618
- Joziasse, D. H. (1992) *Glycobiology* **2**, 271–277
- Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J. P. (1987) *FEBS Lett.* **224**, 149–155
- Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J. P. (1990) *Biochimie* **72**, 555–574
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. and Davies, G. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7090–7094
- Whitfield, C. and Valvano, M. A. (1993) *Adv. Microb. Physiol.* **35**, 135–246
- Liu, D., Haase, A. M., Lindqvist, L., Lindenberg, A. A. and Reeves, P. R. (1993) *J. Bacteriol.* **175**, 3408–3413
- Kido, N., Torgov, V. I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N. and Jann, K. (1995) *J. Bacteriol.* **177**, 2178–2187
- Capage, M. A., Doherty, D. H., Betlach, M. R. and Vanderslice, R. W. (1987) Patent, International Publication Number WO 87/05938
- Petroni, E. A. and Ielpi, L. (1996) *J. Bacteriol.*, in the press
- Woodcock, S., Mornon, J.-P. and Henrissat, B. (1992) *Protein Eng.* **5**, 629–635
- Lee, S. J., Romana, L. K. and Reeves, P. R. (1992) *J. Gen. Microbiol.* **138**, 1843–1855
- Miyata, T., Takeda, J., Iida, Y., Yamada, N., Inoue, N., Takahashi, M., Maeda, K., Kitani, T. and Kinoshita, T. (1993) *Science* **259**, 1318–1320
- Schonbachler, M., Horvath, A., Fassler, J. and Riezma, H. (1995) *EMBO J.* **14**, 1637–1645
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410