

Review

Structure and biosynthesis of human salivary mucins[★]

Anna Zalewska^{1☒}, Krzysztof Zwierz¹, Krzysztof Żółkowski² and Andrzej Gindzieński³

¹Department of Pharmaceutical Biochemistry Institute of Chemistry, ²Department of Maxillofacial Surgery, ³Department of General Chemistry, Medical Academy, Białystok, Poland

Received: 9 March, 2000; revised: 1 September, 2000; accepted: 4 December, 2000

Key words: mucins, saliva, glycoproteins, human

Human salivary glands secrete two types of mucins: oligomeric mucin (MG1) with molecular mass above 1 MDa and monomeric mucin (MG2) with molecular mass of 200–250 kDa. Monomers of MG1 and MG2 contain heavily O-glycosylated tandem repeats located at the central domain of the molecules. MG1 monomers are linked by disulfide bonds located at sparsely glycosylated N- and C-end. MG1 are synthesized by mucous cells and MG2 by the serous cells of human salivary glands.

Human salivary glands secrete 1000–1500 ml per day of saliva composed of water, proteins and low molecular mass substances (mainly electrolytes). Up to 26% of the salivary proteins are mucins [1]. The mucins of human saliva are extremely effective lubricants, which provide an effective barrier against desiccation and environmental insult. They control permeability of mucosal surface, limit penetration of potential irritants and toxins to mucous cells, protect mucosal cell membranes against proteases generated by bacteria in the bacterial plaques around the

teeth and regulate colonization of the oral cavity by bacteria and viruses [2]. The mucins are high molecular mass glycoproteins [3, 4], in which proline and serine/threonine constitute up to 20–55% of total amino acids and are concentrated in one or several regions of the polypeptide. These serine/threonine residues are heavily glycosylated, and 40–80% of the mass of such mucins consists of O-linked oligosaccharides [5]. The cysteines at the N- and C-ends may link mucin monomers by disulfide bridges forming linear mucin oligomers [6]. Two types of mucins are present in human sa-

[★]The paper was presented at the 35th Annual Meeting of the Polish Biochemical Society, Olsztyn, 1999.

☒ Anna Zalewska, Department of Pharmaceutical Biochemistry, Medical Academy, A. Mickiewicza 2a, 15-222 Białystok, Poland; tel: (48 85) 742 2021 ext. 2236 or 2252; e-mail: kzwie@amb.ac.bialystok.pl

Abbreviations: Le^a, Le^b, Le^x, Le^y, Lewis blood groups; MG1, mucin glycoprotein 1; MG2, mucin glycoprotein 2; NeuNAc, N-acetyl-neuraminic acid (sialic acid).

liva: oligomeric mucin glycoprotein (MG1) with molecular mass above 1 MDa, and monomeric mucin glycoprotein (MG2) with molecular mass of 200–250 kDa [1]. The submandibular glands containing mucous cells (producing MG1) and serous cells (producing MG2) secrete 30% of the salivary mucins, while sublingual, labial and palatal glands (which contain mainly mucous cells) secrete 70% [7–11]. Concentration of mucin secreted by sublingual glands is higher than that secreted by submandibular glands, while the secretion of parotid glands is devoid of mucins [7, 12].

POLYPEPTIDES OF SALIVARY MUCINS

Three genes encoding mucins in human salivary glands have been reported [5]: *MUC1* localized to chromosome 1q21-24, which encodes a transmembrane mucin [13], and two genes encoding secretory mucins, *MUC5B* localized to chromosome 11p15.5 encoding oligomeric mucin (MG1) [7] and *MUC7* localized to chromosome 4q13-21 encoding monomeric MG2 [13]. The limited expression of *MUC5B* in sublingual and submandibular glands secreting huge amounts of oligomeric mucins suggests that MG1 is encoded by two or more mucin genes [10]. The peptide sequence of MG1, deduced from the nucleotide sequence of the *MUC5B* gene, consists of about 5000 amino acids grouped in three domains: N-terminal, central and C-terminal [10] (Fig. 1A). It is suggested that the N-terminal domain of *MUC5B* gene product comprising about 450 amino acids is cysteine-rich and contains potential N-glycosylation sites [10].

A single large exon encoding the central domain of MG1 comprising 3570 amino acids (rich in threonine (27%), serine (12.9%) and proline (10.6%)) has been described. The domain consists of 19 subdomains: seven cysteine-rich, 3 subdomains with no repeats,

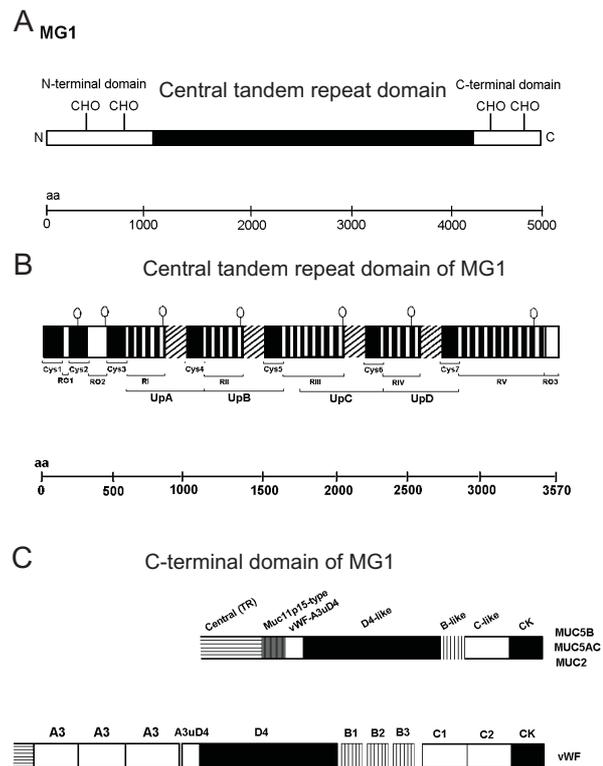


Figure 1. Structure of human sublingual oligomeric mucin (MG1)-the *MUC5B* gene product.

A. General view (CHO, potential N-glycosylation site). B. Structure of the central domain of MG1: Cys (1–7), cysteine-rich subdomains; R01, R02, R03, subdomains with no repeats; RI–V, imperfect repeat subdomains; Re = R-end subdomains. UpA, UpB, UpC and UpD, super repeats of 528 amino acids (*lollipop symbol* potential N-glycosylation site). C. Comparison of the structure of the C-terminal domain of MG1 and C-terminal domains of other mucins (*MUC5AC* and *MUC2* gene products), with the structure of the von Willebrand factor. Central (TR), end of central domain of MG1. Muc 11p15 type, the subdomain of the *MUC5B* gene product similar to the products encoded by appropriate segments of the *MUC2* and *MUC5AC* genes clustered to chromosome 11p15.5, vWF-A3u4D, D-like, B-like, C-like and CK, subdomains of the C-terminal domain of the *MUC5B* gene product similar to the appropriate subdomains found in von Willebrand factor. (Based on the data presented by Desseyn *et al.* [13, 17] by permission of the American Society for Biochemistry and Molecular Biology, and Troxler *et al.* [10] by permission of Oxford University Press).

5 irregular tandem repeats and 4 unique R-end sequences conserved over a long evolutionary time scale, with no typical repeats.

(Fig. 1B). There is a high similarity between cysteine-rich subdomains of human MG1 and homologous subdomains in other mucins (e.g. those coded for by the human *MUC5AC* and *MUC2* genes, mouse *MUC5AC*, rat *MUC2* and pig gastric mucin). All but two of the cysteine residues of the central domain of MG1 are concentrated at highly conserved sites in seven cysteine-rich subdomains. Cysteine-rich subdomains contain numerous amino acids such as tryptophan, proline, arginine and glycine at conserved sites and only one potential O-glycosylation site [13]. The first three cysteine-rich subdomains of the central domain of MG1 are separated by two other subdomains (R01 and R02) enriched in threonine, serine, proline and alanine, without any repeats [13]. Five cysteine-rich subdomains (Cys3 to Cys7) are separated by five subdomains (RI–RV) composed of Thr/Ser-rich tandem repeats and four subdomains called R-end [13]. Three Thr/Ser-rich tandem repeats (RI, RII, RIV) are composed of 11 irregular repeats of 29 amino acids, RIII is composed of 17 and RV of 22 such repeats. The Thr/Ser-rich tandem repeats (RI–RV) show an extremely high content of threonine (37.2%), serine (15.3%), proline (11.5%) and alanine (10.3%) and numerous potential O-glycosylation sites. The sequence TXXP (X, any amino acid with the exception of amino acids with large side chains, probably due to steric hindrance, or high hydrophobicity [14]) has been suggested as a major site for O-linked GalNAc attachment to polypeptide, and is found in the majority of the 72 Thr/Ser-rich tandem repeats. In each of the RI–RV subdomains, one potential N-glycosylation site is found. The central domain of MG1 has four R-end subdomains. Each R-end subdomain is composed of 111 amino acids and has a high content of threonine (30.4%) and serine (18.2%) with numerous potential O-glycosylation sites. Each R-end subdomain has five TXXP sequences, which are major sites for O-linkage of GalNAc to polypeptide. The central domain ends with a peptide called

R03, which is enriched in threonine (20%) and serine (20%), proline, phenylalanine and valine. In the central domain of MG1 one may distinguish four super-repeats called UpA–UpD. Each super-repeat (528 amino acid long) consists of the R, R-end and cysteine-rich subdomains [13]. In the central domain of MG1 there are 7 potential N-glycosylation sites [13].

MG1 is an oligomeric mucin composed of subunits linked by disulfide bridges [15], similarly to gastric mucin [6]. MG1 is heavily O-glycosylated and contains numerous hydrophobic sites possessing hydrophobic as well as negatively charged amino acids in subdomains with no oligosaccharide chains [16]. The deduced amino-acid sequence of the central domain of MG1 based on the structure of the *MUC5B* gene is consistent with the biochemical data. Potential sites for O-glycosylation of the central domain of MG1 are located in five Ser/Thr-rich 20–29 amino acid RI–RV tandem repeats (Ser/Thr 52.5%), four Ser/Thr-rich R-end subdomains without repeats and three subdomains R01–R03 [13]. Disulfide bridges and hydrophobic sites possessing negatively charged amino acids of the central domain of MG1 occupy seven cysteine-rich subdomains (Cys1–Cys7) rich in cysteine (70 residues), aspartic acid (37 residues) and glutamic acid (58 residues). Cysteine-rich subdomains contain also a large number of aromatic amino acids, i.e., tyrosine (31 residues), phenylalanine (31 residues) and tryptophan (17 residues) [13].

Two laboratories reported that the C-terminal domain of MG1 comprises 804 [10] or 808 [17] amino acids, respectively. It is relatively rich in cysteine and proline, but relatively poor in threonine and serine [10, 17] and has a similar structure to von Willebrand factor (vWF) [19]. Based on the similarity to this factor was encoded six subdomains in the C-terminal domain of MG1 (Fig. 1C): Muc 11p 15 type subdomain (analogical to a subdomain found in products of genes *MUC2* and *MUC5AC* clustered on chromosome 11p

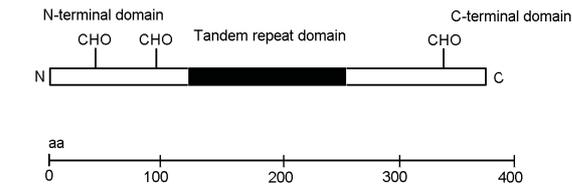
15.5), vWF-A3uD₄-like, vWF4D-like, vWFB-like, vWFC-like and CK [17]. The Muc 11p 15 type subdomain consisting of 40 amino acids (similar to the polypeptide encoded by appropriate sequences of genes *MUC2* and *MUC5AC*, but absent in vW factor) follows the central domain of MG1. The second, 69 amino-acid subdomain called vWF-A3uD₄, similar to the subdomain found in vW factor, was also found in the products of the genes *MUC2* and *MUC5AC*. The third 378 amino-acid subdomain called vWFD4-like is similar not only to subdomains found in vWF, subdomains of the products of genes *MUC2* and *MUC5AC*, but also to subdomains in other proteins, such as zonadhesin [20] and vitellogenin. The vWFD-like subdomain contains vicinal cysteine motifs (CGXC) similar to the sequences found at the active sites of the enzyme that catalyses thiol protein disulfide interchange and may be involved in dimer formation between the C-terminal domains of MG1 [17, 20]. The C-terminal domain of MG1 contains one vWF B-like subdomain 40 amino acid long, instead of three B subdomains found in vWF. It also contains one vWFC-like subdomain, related to the C2 subdomain in vWF, instead of two corresponding subdomains in vWF. A hydrophobic segment (YXX₆CCX₇C) located between aa 658–678 in the vWFC-like subdomain (similar to that present in serum albumin) may be responsible for binding hydrophobic compounds to MG1 [17]. The last subdomain of the C-terminal domain of MG1 similar to vWF and other mucins, is named CK or “cystine knot”. This “knot” is probably constructed of three internal disulfide bridges linking cysteines 741–788, 764–813 and 768–820. Cysteine 787 may be involved in the formation of intermolecular bridges linking MG1 dimers [10, 17]. The C-terminal domain of MG1 shows alternating hydrophobic and hydrophilic sequences. The most hydrophilic segment occurs in the CK domain [10]. Six subdomains of the C-terminal region of the *MUC5B* gene product have conserved se-

quences similar to those present in vWF, the *MUC2* and *MUC5AC* gene products. Nearly all cysteine residues are conserved along with several other amino acids, suggesting a very similar tertiary structure for the C-terminal domain of MG1 compared to the above proteins. Based on the deduced amino-acid sequence of the C-terminal domain of MG1 computer simulation predicts 62% of β -turn and 13% of α -helix located between a 219–250 and 407–421 in the vWFD-like subdomain and 776–801 in the CK subdomain [17]. The remaining 25% of the C-terminal domain consist of extended and coil structures. The rigid, rod-shaped conformation of the C-terminal domain is essential for oligomerization of MG1. It is noteworthy that 88% of cysteines in the C-terminal domain are localised in or at the vicinity of a β -turn in 11 out of 15 potential N-glycosylation sites [17]. The structure of MG1 deduced from the nucleotide sequence of the *MUC5B* gene is consistent with the immuno-electron-microscopy picture in which filamentous structures of 0.5–10 μ m were observed. The subunits of MG1 are linked by disulfide bridges and possess an exposed (not covered by carbohydrate chains) area 100–150 nm long at the ends and exposed intervals 100–180 nm long in the central part of the molecule. In the native molecule of MG1 some of these intervals are not exposed on the surface of the molecule [21].

The amino-acid sequence (deduced from the *MUC7* gene) of monomeric apomucin from human saliva (MG2) is composed of three domains: the N-terminal domain consisting of 144 amino acids, the central domain comprising amino acids 145–283, including six tandem repeats of 23 amino acids each, and the C-terminal domain consisting of 74 amino acids [22] (Fig. 2). The first 20 amino acids of the N-terminal domain of MG2 make up a leader peptide [23]. The remaining portion of the N-terminal domain has two cysteine residues, three potential N-glycosylation sites [23] and 9 potential O-glycosylation sites (4 serine and 5 threonine residues). It was suggested

that the N-terminal domain of secreted MG2 was involved in the formation of disulfide bridges causing self-association [24]. It is worthy of note that the amino-acid sequence of the N-terminal domain of MG2 encoded by the *MUC7* gene (Fig. 2B) is consistent with the

A MG 2



B

E-G-R-E-R-D-H-E-L-R-H-R-R-H-H-H-Q-S-P-K-S-H-F-E-L-P-H-Y-P-G- (1-30)
 L-L-A-H-Q-K-P-F-I-R-K-S-Y-K-C-L-H-K-R-C-R-P-K-L-P-S-P-N-K- (31-60)
 P-P-K-F-P-N-P-H-Q-P-P-K-H-P-D-K-N-S-S-V-V-N-P-T-L-V-A-T-T-Q- (61-90)
 I-P-S-V-T-F-P-S-A-S-T-K-I-T-T-L-P-N-V-T-F-L-P-Q-N-A-T-T-I-S- (91-12)
 S-R-E-N-V-N-T-S-S-S-V-A-T-L-A-P-V-N-S-P-A-P-Q-D-T-T-A-A-P-P- (121-1)
 T-F-S-A-T-T-P-A-P-P-S-S-S-A-P-P-E-T-T-A-A-P-P-T-P-S-A-T-T-Q- (151-1)
 A-P-P-S-S-S-A-P-P-E-T-T-A-A-P-P-T-P-A-T-T-P-A-P-P-S-S-S-A- (181-2)
 P-P-E-T-T-A-A-P-P-T-P-S-A-T-T-P-A-P-L-S-S-S-A-P-P-E-T-T-A-V- (211-2)
 P-P-T-P-S-A-T-T-L-D-P-S-S-A-S-A-P-P-E-T-T-A-A-P-P-T-P-S-A-T- (241-2)
 T-P-A-P-P-S-S-P-A-P-Q-E-T-T-A-A-P-I-T-T-P-N-S-S-P-T-T-L-A-P- (271-2)
 D-T-S-E-T-S-A-A-P-T-H-Q-T-T-T-S-V-T-Q-T-T-T-K-Q-P-T-S-A- (301-2)
 P-G-Q-N-K-I-S-R-F-L-L-Y-M-K-N-L-L-N-R-I-I-D-D-M-V-E-Q (331-2)

Figure 2. Structure of human salivary monomeric mucin (MG2) the *MUC7* gene product.

A. General view of MG2. B. Amino-acid sequence of human monomeric mucin MG2 (the six tandem repeat sequences are underlined). (Based on the data reported by Antonyraj *et al.* [22] and Troxler *et al.* [10] by permission of Oxford University Press).

data obtained by Edman degradation of the purified protein [25]. The central domain of the MG2 polypeptide is rich in potential O-glycosylation sites, including a tandem repeat region of 138 residues. Evidence suggests that the tandem repeat sequences serve as constrained structural elements to stabilize poly-L-proline II conformation of the central domain of MG2. In the extended poly-L-proline conformation, the side chains are segregated farther apart from each other [22] than in other conformations. This segregation can promote O-glycosylation. If this central domain was extensively O-glycosylated and extended, for example as a β -structure, it would have an estimated length of approximately 65 nm. Results of electron microscopy and light scattering suggest that the mole-

cules are random-coil structures with a radius of gyration below 15 nm [24]. Trypsin digestion of native MG2 releases a 90 kDa fragment of central tandem repeats, which carries primarily O-linked oligosaccharides [25]. The C-terminal domain of MG2 is devoid of cysteine residues but is rich in proline (5 residues), and has one potential N-glycosylation and 26 potential O-glycosylation sites (7 serine and 19 threonine residues) [22].

OLIGOSACCHARIDES OF SALIVARY MUCINS

Human salivary mucins have O-linked (MG1 and MG2) and N-linked (MG2) oligosaccharide chains [26]. In MG1 from human saliva of blood group B individuals 55% of neutral and 45% of acidic oligosaccharide chains were found [27]. Human oligomeric mucin (MG1) contains 292 O-linked oligosaccharide chains per monomeric unit (that may form a “bottle brush”-like structure) and 118 oligosaccharide chains have sialic acids on non-reducing ends [26]. Oligosaccharide chains of MG1 ranging from 1–20 residues (Fig. 3) are located mostly in clusters on the super-repeat subdomains of the central domain [1, 18, 26]. The complex oligosaccharide chains of MG1 may be divided into three regions: the core, backbone and peripheral domains [28]. Of the eight core classes described to date, four have been found in oligosaccharides of human MG1 [18]. The backbone structures of glycoproteins are formed by disaccharides which differ in the type of glycosidic bond, i.e., type 1 (Gal β 1-3 GlcNAc) or type 2 (lactosamine, Gal β 1-4 GlcNAc). Some peripheral domains of mucin glycoproteins create such antigenic determinants as T, sialylT, A, B, H, Le^a, Le^b, Le^x, Le^y, i, I (Figs. 3 and 4).

Human monomeric mucin (MG2) was found to contain 170 oligosaccharide chains of which 67 were sialylated [26]. Approximately 80% of mucin oligosaccharides comprise di- and trisaccharide units of the core 1 type

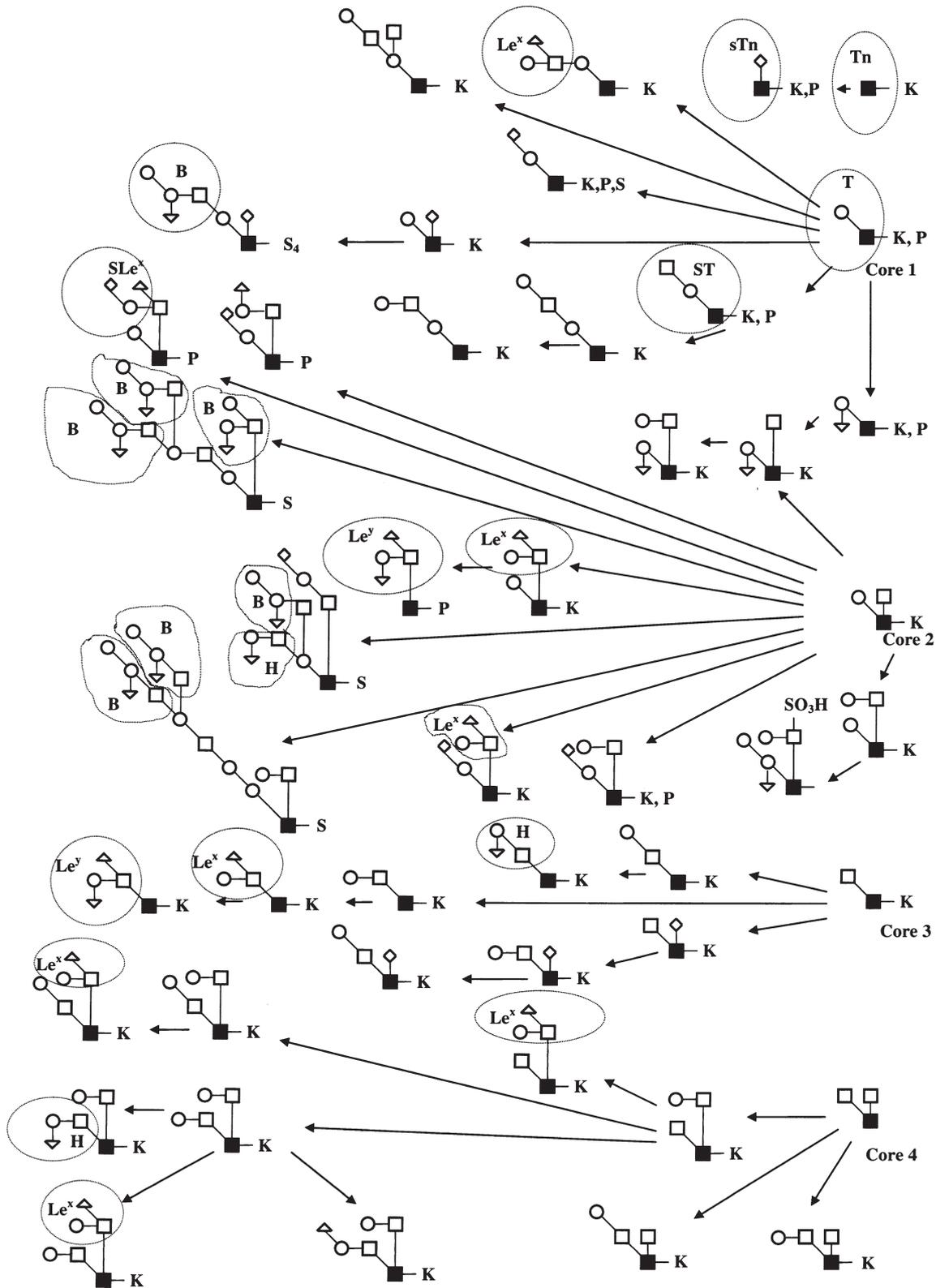


Figure 3. Oligosaccharide structures of human salivary oligomeric mucin (MG1).

(K, according to Klein *et al.* [18], S, Slomiany *et al.* [27] and P, Prakobphol *et al.* [25]. Antigenic determinants are dotted). ϕ , GalNAc, *N*-acetyl-galactosamine; TM , Gal, galactose; f , GlcNAc, *N*-acetyl-glucosamine; N , NeuAc, *N*-acetyl-neuraminic acid (sialic acid); Δ , Fuc, fucose;



(Fig. 5). The remaining 20% of oligosaccharides are dominated by 5–7 saccharide units of core 2 type [25, 26]. Of 41 proposed O-linked oligosaccharide structures 23 were neutral, 15 sialylated and 3 sulfated [25]. Some peripheral regions from O-linked oligosaccharides of MG2 contain antigenic determinants T, Le^x, Le^y, sialyl T and sialyl Le^y [25]. It has been noted that core 2 antigens Le^x and sialyl Le^x are potential ligands for selectins (adhesive molecules controlling leucocyte movement) [30] and some strains of pathogenic bacteria are associated with MG2 through antigen T, sialyl T or lactosamine sequences [29].

There is some controversy concerning oligosaccharide chains of MG1 and MG2. Slomiany *et al.* [27] in a study of five blood group B individuals showed that oligosaccharide chains of MG1 and MG2 have similar proportions of neutral and acidic oligosaccharides. The predominant neutral oligosaccharides in both glycoproteins are composed of 15 and 16 sugar units (Fig. 3). The majority of these oligosaccharides bore the blood group B determinants. Basing on these data Slomiany *et al.* in 1993 [27] suggested that MG1 and MG2 arise through the action of endogenous proteases on one high molecular mass mucus glycoprotein [28]. However according to other authors [5, 8, 13, 25, 26] MG1 and MG2 have different polypeptide chains encoded by different genes [5, 8, 13] and carry very different carbohydrate structures [25]. This opinion is consistent with the proposed polypeptide structures of MG1 [10, 13, 17] and MG2 [22], the known specificity of GalNAc polypeptide transferases [14] and the data suggesting that MG1 and MG2 are synthesized by different populations of cells in human salivary glands [9].

BIOSYNTHESIS OF SALIVARY MUCINS OLIGOSACCHARIDES

The biosynthesis and processing of the N-linked glycoproteins is now almost com-

pletely understood [14]. It has been shown that although low molecular mass rat submandibular mucin is N-glycosylated, this N-glycosylation is not required for its secretion [31]. Recently, however, it has been reported that some mucins, e.g. N-glycosylated human colonic mucin (MUC2), become associated with the lectin-like chaperones (calreticulin and calnexin) by disulfide bridges, before folding to their correct structure in the ER [32]. There is no data available on such chaperone action on MG1 or MG2. Recently new information concerning O-glycosylation appeared [14]. Most studies indicate that O-glycosylation is initiated in the *cis* Golgi compartment. Glycosyltransferases can be divided into two types: those transferring the sugar residue with retention of its anomeric configuration, and those transferring with inversion. Both types of glycosyltransferases have similar nucleotide recognition domains [33]. The biosynthesis of O-linked glycans begins with addition of GalNAc to serine or threonine residues in a polypeptide chain [34], catalysed by a family (6 members reported to the end of 1999 [35]) of GalNAc polypeptide transferases (UDPGalNAc:polypeptide *N*-acetyl-galactosaminyltransferases, GalNAcT (EC 2.4.1.41)), which are conserved in evolution from nematodes to humans [36]. These GalNAc polypeptide transferases consist of short N-terminal cytoplasmic tails, transmembrane domains, extended stem regions and a large carboxyl terminal where the catalytic domains are located and which face the interior of the Golgi compartment [37–39]. Genes for GalNAc1T, -T2 and -T3 are localized to chromosomes 18q12-q21, 1q412-q42, and 2q24-q31, respectively [40]. This family of GalNAc polypeptide transferases has different expression patterns in different organs. The specificity for folded polypeptide chains of proteins and thus potential sites of O-glycosylation are different for each transferase [41]. Glycosylation of certain acceptor sites by one of the GalNAc polypeptide transferases is required before other sites can

be glycosylated by another GalNAc polypeptide transferase [31, 35]. In addition, individual GalNAc polypeptide transferases show different localization within the Golgi compartments [39], moreover, the sizes of their stem regions are different [36]. The repertoire of GalNAc polypeptide transferases varies with cell localization and differentiation [42]. GalNAcT3 is expressed in all layers of normal human mouth epithelia and in most squamous carcinoma cells, GalNAcT2 was found in undifferentiated cell layers in the normal epithelial and in most carcinoma cells except in well differentiated foci. GalNAcT1 shows a low level of expression, markedly enhanced in tumors [42].

Positions C3 and C6 of the core GalNAc may be substituted by Gal, GalNAc, GlcNAc or NeuAc (Fig. 3) and thus create cores 1–8 [14]. Addition of each sugar to the O-linked GalNAc is catalysed by a specific transferase [43]. Much attention is devoted to a family of GlcNAc transferases (UDP-GlcNAc:Gal β 1-3-GalNAc α 1-R β 1-6GlcNAc transferase) that take part in the synthesis of cores 2, 3 and 6.

Introduction of the C2GnT gene (encoding core 2 β 1-6 acetylglucosaminyltransferase) to human pancreatic cancer cells shifts O-glycosylation toward more complex glycans [44]. It was reported that sialyl- and Gal-transferases compete with C2Gn transferase for the OH group at C6 of O-linked GalNAc. Activation of T cells results in induction of C2Gn transferase, which causes a shift of O-linked oligosaccharide chains from tetrasaccharide (NeuAc α 2-3Gal β 1-3 (NeuAc α 2-6) GalNAc α 1Ser/Thr) to hexasaccharide (NeuAc α 2-3Gal β 1-3 (NeuNAc α 2-3 Gal β 1-4GlcNAc β 1-6) GalNAc α 1Ser/Thr) [45]. This C2GnT was expressed in 7 day embryos. In older embryos expression of C2GnT was restricted to the tissues producing mucins and cartilage [46]. P-Selectin glycoprotein ligand 1, a mucin type cell surface glycoprotein has been reported to be a substrate for C2Gn transferase [47]. Another member of C1-6 GlcNAc transferase family is I branching enzyme (IGnT) which transfers GlcNAc to C6 of galactose moiety in backbone structures (Fig. 4). The activity of IGnT is associated with embryo-

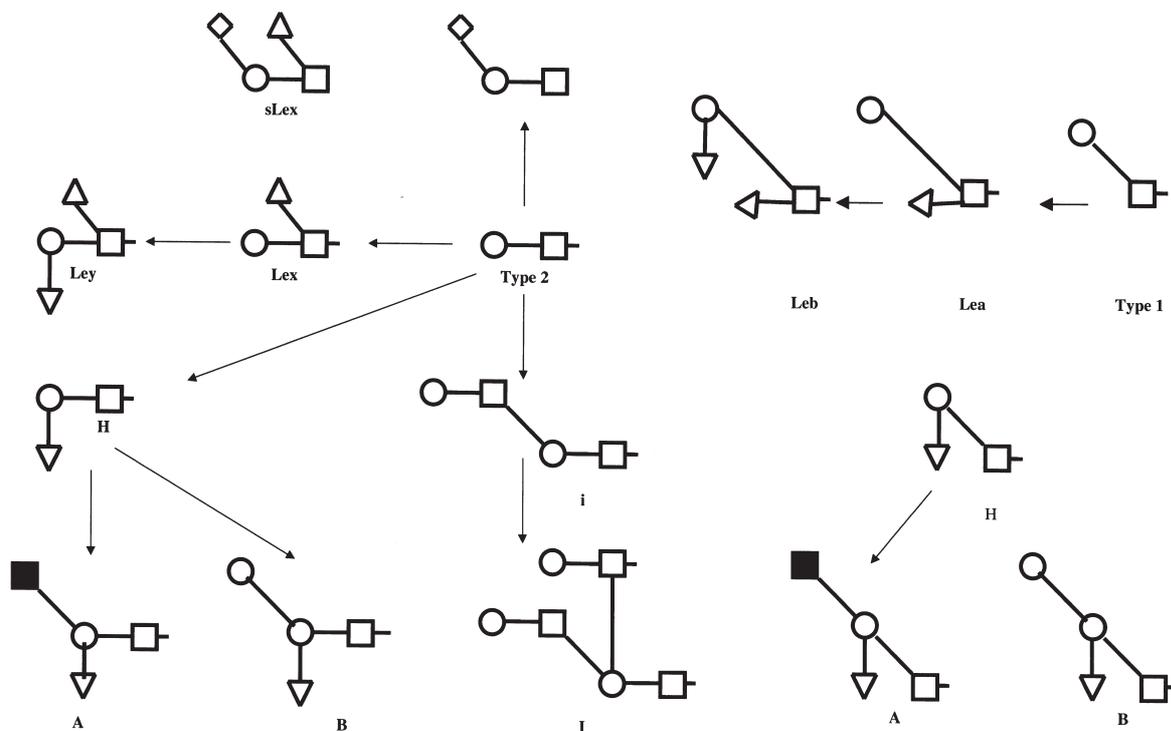


Figure 4. Antigenic determinants of oligosaccharide chains.

Based on the data of Abe *et al.* [52], Grooi *et al.* [50], Lloyd *et al.* [51] and Van den Scheen *et al.* [14].

genesis [48]. The formation of these backbones and peripheral domains of glycans has been described previously [49]. It is noteworthy that determinants Le^x , Le^y , A, B, H type 2 and sialyl Le^y are synthesized from structure 2, determinants Le^a , Le^b , A, B and H type 1 are synthesized from structure 1 [49, 50] (Fig. 5) and that sialylation terminates the synthesis of O-linked oligosaccharides [14].

CONCLUSIONS AND FUTURE PROSPECTS

There is a consensus concerning the structure of the central and C-terminal regions of

sus on the quantity and quality of oligosaccharide chains in human salivary mucins. Initiation of O-glycosylation takes place in the *cis* Golgi compartment, however, the subregions of Golgi where particular glycosylation steps take place are not yet known. The requirements of the tertiary structure of the substrates for particular glycosyltransferase action and eventual role of chaperones in this process are also not known. Rules have to be established for the competition of glycosyltransferases for a particular substrate. The number of mucin genes taking part in biosynthesis of the MG1 polypeptide chains remains to be established. Data indicating which potential O- and N-glycosylation sites

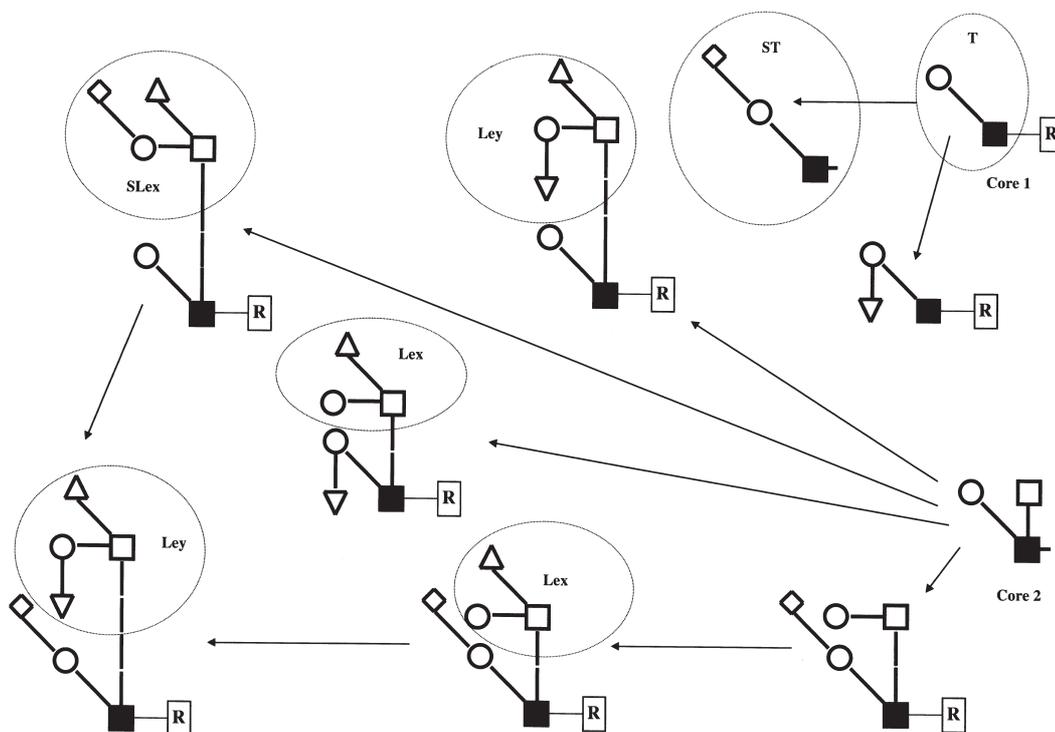


Figure 5. Prominent oligosaccharide chains of human salivary monomeric mucin (MG2).

According to Levine *et al.* [26] and Prakobphol *et al.* [25].

MG1. Soon we may expect a report on the N-terminal sequence of MG1. Also there is an agreement concerning the deduced sequence for MG2. Thirty seven O-linked oligosaccharide chains in MG1 and 41 in MG2 have been described, however, not all data on this subject are unanimous. There is a lack of consen-

are glycosylated and what is the nature of regulatory factors are urgently awaited. The source of the microheterogeneity of the oligosaccharide chains of human salivary mucins is unknown. Regulation of the biosynthesis of the polypeptide and oligosaccharide chains and the changes taking place in pathological

conditions, have yet to be established. Also there is no data available on the relationship between secondary structure and biological function of salivary mucins. An evaluation of which cysteine residues take part in the formation of disulfide bridges and the conditions necessary for this linkage have not yet been published.

We are indebted to Dr. Tony Merry of Oxford Glycobiology Institute for help in improving the style of the manuscript, Dr. Pauline Rudd of the same Institute and Dr. Bronislaw Slomiany of UMDNJ-New Jersey Dental School for sending reprints of their papers.

REFERENCES

- Nieuw Amerongen, A.V., Bolscher, J.G.M. & Veerman, E.C.I. (1995) Salivary mucins: Protective functions in relation to their diversity. *Glycobiology* **5**, 733–740.
- Mandel, I.D. (1987) The functions of saliva. *J. Dent. Res.* (Spec. Iss.) **66**, 623–627.
- Gindzieński, A. & Zwierz, K. (1987) Isolation and fractionation of human gastric mucus gel. *Biomed. Biochim. Acta* **46**, 165–176.
- Paszkiwicz-Gadek, A., Gindzieński, A. & Porowska, H. (1995) The use of preparative polyacrylamide gel electrophoresis and electroelution for purification of mucus glycoproteins. *Anal. Biochem.* **226**, 263–267.
- Van Klinken, B.J.-W., Dekker, J., Büller, H.A. & Einerhand, A.W.C. (1995) Mucin gene structure and expression: Protection *vs.* adhesion. *Am. J. Physiol. (Gastroint. Liver Physiol. 32)* **269**, G613–G627.
- Minkiewicz-Radziejewska, I., Gindzieński, A. & Zwierz, K. (2000) Disulfide-bound fragments of gastric mucin are 100- and 140-kDa proteins. *Biochem. Biophys. Res. Commun.* **270**, 722–727.
- Wu, A.M., Csako, G. & Herp, A. (1994) Structure, biosynthesis, and function of salivary mucins. *Mol. Cell. Biochem.* **137**, 39–55.
- Thornton, D.J., Khan, N., Mehrotra, R., Howard, M., Veerman, E., Packer, N.H. & Sheehan, J.K. (1999) Salivary mucin MG1 is comprised almost entirely of different glycosylated forms of MUC5B gene product. *Glycobiology* **9**, 293–302.
- Nielsen, P.A., Mandel, U., Therkildsen, M.H. & Clausen, H. (1996) Differential expression of human high-molecular-weight salivary mucin (MG1) and low-molecular-weight salivary mucin (MG2). *J. Dent. Res.* **75**, 1820–1826.
- Troxler, R.F., Iontcheva, I., Oppenheim, F.G., Nunes, D.P. & Offner, G.D. (1997) Molecular characterization of a major molecular weight mucin from human sublingual gland. *Glycobiology* **7**, 965–973.
- Milne, R.W. & Dawes, C. (1973) The relative contributions of different salivary glands to the blood group activity of whole saliva in humans. *Vox Sang.* **25**, 298–307.
- Edgerton, M. & Levine, M.J. (1992) Characterization of acquired denture pellicle from healthy and stomatitis patients. *J. Prosthet. Dent.* **68**, 683–691.
- Desseyne, J.-L., Guyonnet-Duperat, V., Porchet, N., Aubert, J.-P. & Laine, A. (1997) Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternate subdomains resulting in a super-repeat. Structural evidence for a 11p.15.5 gene family. *J. Biol. Chem.* **272**, 3168–3178.
- Van den Steen, Rudd, P.M., Dwek, R. & Opdenakker, G. (1998) Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33**, 151–208.
- Wickström, C., Davies, J.R., Eriksen, G.V., Veermans, E.C. & Carlsted, I. (1998) MUC5B is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and

- endocervix: Identification of glycoforms and C-terminal cleavage. *Biochem. J.* **334**, 685–693.
- 16.** Loomis, R.E., Prakobphol, A., Levine, M.J., Reddy, M.S. & Jones, P.C. (1987) Biochemical and biophysical comparison of two mucins from human submandibular-sublingual saliva. *Arch. Biochem. Biophys.* **258**, 452–464.
- 17.** Desseyne, J.-L., Aubert, J.-P., VanSeuninghen, I., Porchet, N. & Laine, A. (1997) Genomic organization of the 3' region of the human mucin gene MUC5B. *J. Biol. Chem.* **272**, 16873–16883.
- 18.** Klein, A., Carnoy, Ch., Wieruszewski, J.-M., Strecker, G., Strang, A.-M., van Halbeck, H., Roussel, Ph. & Lamblin, G. (1992) The broad diversity of neutral and sialylated oligosaccharides derived from human salivary mucin. *Biochemistry* **31**, 6152–6165.
- 19.** Mancuso, D.J., Tuley, E.A., Westfield, L.A., Worrall, N.K., Shelton-Inloes, B., Sorace, J.M., Alevy, Y.G. & Sadler, J.E. (1989) Structure of the gene for human von Willebrand factor. *J. Biol. Chem.* **264**, 19514–19527.
- 20.** Hardy, D.M. & Garbers, D.L. (1995) A sperm membrane protein that binds in a species-specific manner to the egg extracellular matrix is homologous to von Willebrand factor. *J. Biol. Chem.* **270**, 26025–26028.
- 21.** Sheehan, J.K., Boot-Handford, R.P., Chantler, E., Carlstedt, I. & Thornton, D.J. (1991) Evidence for shared epitopes within the “naked” domains of human mucus glycoproteins. A study performed using polyclonal antibodies and electron microscopy. *Biochem. J.* **274**, 293–296.
- 22.** Antonyraj, K.J., Karunakaran, T. & Raj, P.A. (1998) Bactericidal activity and poly-L-proline II conformation of the tandem repeat sequence of human salivary glycoprotein (MG2). *Arch. Biochem. Biophys.* **356**, 197–206.
- 23.** Bobek, L.A., Tsai, H., Biesbrock, A.R. & Levine, M.J. (1993) Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (MUC 7). *J. Biol. Chem.* **268**, 20563–20569.
- 24.** Mehrotra, R., Thornton, D.J. & Sheehan, J.K. (1998) Isolation and physical characterization of MUC 7 (MG2) mucin from saliva: Evidence for self association. *Biochem. J.* **334**, 415–422.
- 25.** Prakobphol, A., Thomsson, K.A., Hansson, G., Rosen, S.D., Singer, M.S., Phillips, N.J., Meddihradzky, K.F., Burlingame, A.L., Leffler, H. & Fisher, S. (1998) Human low-molecular-weight salivary mucin expresses the sialyl Lewis^x determinant and has L-selectin ligand activity. *Biochemistry* **37**, 4916–4927.
- 26.** Levine, M.J., Reddy, M.S., Tabak, L.A., Loomis, R.E., Bergey, E.J., Jones, P.C., Cohen, R.E., Stinson, M.W. & Al-Hashimi, I. (1987) Structural aspects of saliva glycoproteins. *J. Dent. Res.* **66**, 436–441.
- 27.** Slomiany, B.L., Murthy, V.L.N. & Slomiany, A. (1993) Structural features of carbohydrate chains in human salivary mucins. *Int. J. Biochem.* **25**, 259–265.
- 28.** Slomiany, B.L., Piotrowski, J., Czajkowski, A. & Slomiany, A. (1993) Control of mucin molecular forms expression by salivary protease: Difference with caries. *Int. J. Biochem.* **25**, 681–687.
- 29.** Prakobphol, A., Tangemann, K., Rosen, S.D., Hoover, Ch.I., Leffler, H. & Fisher, S.J. (1999) Separate oligosaccharide determinants mediate interactions of the low-molecular-weight salivary mucins with neutrophils and bacteria. *Biochemistry* **38**, 6817–6825.
- 30.** Li, F., Wilkins, P.P., Crawley, S., Weinstein, J., Cummings, R.D. & McEver, R.P. (1996) Post-translational modifications of recombinant P-selectin glycoprotein ligand-1 required for binding to P- and E-selectin. *J. Biol. Chem.* **271**, 3255–3264.

31. Nehrke, K. & Tabak, L.A. (1997) Biosynthesis of a low-molecular-mass rat submandibular gland mucin glycoprotein in COS 7 cells. *Biochem. J.* **323**, 497–502.
32. McCool, D.J., Okada, Y., Forstner, J.F. & Forstner, G.G. (1999) Roles of calreticulin and calnexin during mucin synthesis in LS180 and HT29/A1 human colonic adenocarcinoma cells. *Biochem. J.* **341**, 593–600.
33. Kapitonov, D. & Yu, R.K. (1999) Conserved domains of glycosyltransferases. *Glycobiology* **9**, 961–978.
34. Porowska, H., Paszkiewicz-Gadek, A. & Gindziński, A. (1999) Activity of partially purified UDP-*N*-acetyl- α -D-galactosamine: Polypeptide *N*-acetylgalactosaminyltransferase with different peptide acceptors. *Acta Biochim. Polon.* **46**, 365–360.
35. Bennet, E.P., Hassan, H., Mandel, U., Hollingsworth, M.A., Akisawa, N., Ikematsu, Y., Merckx, G., van Kelle, A.G., Olofsson, S. & Clausen, H. (1999) Cloning and characterization of a close homologue of human UDP-*N*-acetyl- β -D-galactosamine: Polypeptide *N*-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. *J. Biol. Chem.* **274**, 25362–25370.
36. Hagen, K.G.T., Hagen, F.K., Balys, M.M., Beres, T.M., van Wuyckhuysse, B. & Tabak, L.A. (1998) Cloning and expression of a novel, tissue specifically expressed member of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase family. *J. Biol. Chem.* **273**, 27749–27754.
37. Clausen, H. & Bennet, E.P. (1996) A family of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology* **6**, 635–646.
38. Paulson, J.C. & Colley, K.J. (1989) Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J. Biol. Chem.* **264**, 17615–17618.
39. Rottger, S., White, R., Wandall, H.H., Bennet, E.A., Stark, A., Olivio, J., Whitehouse, C., Berger, E.G., Clausen, H. & Nilson, T. (1998) Differential localization of three human GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation occurs through the Golgi apparatus. *J. Cell. Sci.* **111**, 45–60.
40. Bennet, E.P., Weghuis, D.D., Merckx, G., van Kessel, A.G., Eiberg, H. & Clausen, H. (1998) Genomic organization and chromosomal localization of three members of the UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosamine transferase family. *Glycobiology* **8**, 547–555.
41. Wandall, H.H., Hassan, H., Mirgorodskaya, E., Kristensen, A.K., Roepstorff, P., Bennet, E.P., Nielsen, P.A., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J. & Clausen, H. (1997) Substrate specificities of three members of the human UDP-*N*-acetyl- α -D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase family. *J. Biol. Chem.* **272**, 23503–23514.
42. Mandel, U., Hassan, H., Therkildsen, M.H., Rygaard, J., Jakobsen, M.H., Juhl, B.R., Dabelsteen, E. & Clausen, H. (1999) Expression of polypeptide GalNAc-transferases in stratified epithelia and squamous cell carcinomas: Immunohistological evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. *Glycobiology* **9**, 43–52.
43. Field, M.C. & Wainwright, L.J. (1995) Molecular cloning of eucaryotic glycoprotein and glycolipid glycosyltransferases: A survey. *Glycobiology* **5**, 463–472.
44. Beum, P.V., Singh, J., Burdick, M., Hollingsworth, M.A. & Cheng, P.W. (1999) Expression of core 2 β 1,6-*N*-acetylglucosaminyltransferase in a human pancreatic cancer cell line results in altered expression of MUC1 tumor-associated epitopes. *J. Biol. Chem.* **274**, 24641–24648.

45. Barran, P., Fellingner, W., Warren, Che, E., Dennis, J.W. & Ziltener, H.J. (1997) Modification of CD 43 and other lymphocyte O-glycoproteins by core 2 *N*-acetylglucosaminyl transferase. *Glycobiology* **7**, 129–136.
46. Granovsky, M., Fode, C., Warren, C.E., Campbell, R.M., Marth, J.D., Pierce, M., Fregien, N. & Dennis, J.W. (1995) GlcNAc-transferase V and core GlcNAc-transferase expression in the developing mouse embryo. *Glycobiology* **5**, 797–806.
47. Li, F., Wilkins, P.P., Crawley, S., Weinstein, J., Cummings, R.D. & McEver, R.P. (1996) Post-translational modifications of recombinant P-selectin glycoprotein ligand-1 required for binding to P- and E-selectins. *J. Biol. Chem.* **271**, 3255–3264.
48. Kapadia, A., Feizi, T. & Evans, M.J. (1981) Changes in the expression and polarization of blood group I and i antigens in post-implantation embryos and teratocarcinomas of mouse associated with cell differentiation. *Exp. Cell Res.* **131**, 185–195.
49. Schachter, H. & Brockhausen, I. (1989) The biosynthesis of branched O-glycans. *Symp. Soc. Exp. Biol.* **43**, 1–26.
50. Gooi, H.C., Feizi, T., Kapadia, A., Knowles, B.B., Solter, D. & Evans, M.J. (1981) Stage-specific embryonic antigen involves α 1-3 fucosylated type 2 blood group chains. *Nature* **292**, 156–158.
51. Lloyd, K.O., Kabat, E.A. & Licerio, E. (1968) Immunochemical studied on blood groups XXXVIII. Structures and activities of oligosaccharides produced by alkaline degradation of blood-group Lewis^a substance. Proposed structure of the carbohydrate chains of human blood-group A, B, H, Le^a, and Le^b substance. *Biochemistry* **7**, 2976–2990.
52. Abe, K., McKibbin, J.M. & Hakomori, S.I. (1983) The monoclonal antibody directed to difucosylated type 2 chain (Fuc α 1-2 Gal β 1-4-[Fuc α 1-3] GlcNAc), Y determinant. *J. Biol. Chem.* **258**, 11793–11797.