Characterization of the Disulfide Bonds and the N-Glycosylation Sites in the Glycoprotein from Rathke's Gland Secretions of Kemp's Ridley Sea Turtle (Lepidochelys kempi)

ARTICLE *in* ANALYTICAL BIOCHEMISTRY · FEBRUARY 1996

Impact Factor: 2.22 · DOI: 10.1006/abio.1996.0026 · Source: PubMed

4 AUTHORS, INCLUDING:

Paul Weldon
Smithsonian Institution
105 PUBLICATIONS **1,213** CITATIONS

ANALYTICAL BIOCHEMISTRY **233,** 181–187 (1996)
Article No. 0026

Characterization of the Disulfide Bonds and the N-Glycosylation Sites in the Glycoprotein from Rathke's Gland Secretions of Kemp's Ridley Sea Turtle (*Lepidochelys kempi*)[1]

Christopher C. Q. Chin, Radha G. Krishna, Paul J. Weldon,*[,2] and Finn Wold
*Department of Biochemistry and Molecular Biology, University of Texas Medical School, P.O. Box 20708, Houston, Texas 77225; and *Department of Biology, Texas A&M University, College Station, Texas 77843*

Received July 5, 1995

The disulfide bonds and N-glycosylation sites in a glycoprotein from the Rathke's gland secretion of the Kemp's ridley turtle (*Lepidochelys kempi*) have been characterized with respect to peptide sequences and glycan structures. The glycoprotein constitutes about 70% of the total protein in the secretion, and based on partial sequence information, it shows more than 20% identity with both the catalytic (esterases) and the noncatalytic (thyroglobulin) members of the esterase/lipase family of proteins. For the determination of the disulfide locations, the glycoprotein was digested with chymotrypsin, and the three HPLC peptide peaks yielding fluorescent products after treatment with tributylphosphine (Bu_3P) and 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD-F) were collected. The three fractions were treated with the same reagents in separate experiments, the resulting pairs of ABD-Cys-containing peptides were separated by HPLC, and the sequence of each individual peptide was determined. The peptide identity established that three disulfide bonds existed in the glycoprotein: Cys 65-Cys 91, Cys 254-Cys 265, and Cys 130-Cys 404; the first two of these are conserved in all the members of the esterase family. For the study of the glycosylation sites, the glycoprotein was reduced with Bu_3P and the SH groups covalently blocked with ABD-F, and the resulting product was digested with chymotrypsin. The glycopeptides were isolated by affinity chromatography, separated by reverse-phase HPLC, and subjected to sequence analysis and fast atom bombardment mass spectrometry before and after separation of the glycans and the peptides through the action of glycoamidase. Three separate glycosylation sites were identified, each containing multiple glycans. The sugar analyses of the hydrolysates of the glycoprotein indicated that only GlcNAc and Man were present as building blocks, and the mass spectrometric data showed that $Man_3GlcNAc_2$-, $GlcNAc_{2-4}Man_3GlcNAc_2$-, and possibly $GlcNAc_2Man_2GlcNAc_2$- were the major glycan structures, distributed differently at the three sites. The three glycosylation sites match three of the nine sites glycosylated in human serum choline esterase, and one of them, Asn 106, is also found as one of two glycosylation sites in the homologous segment of thyroglobulin. © 1996 Academic Press, Inc.

A glycoprotein of about 55-kDa mass makes up 60–70% of the proteins in the secretion produced by the Rathke's glands of the Kemp's ridley sea turtle (*Lepidochelys kempi*) (1). Although no specific function can be assigned to either the secretion or the glycoprotein, preliminary studies show that the protein belongs to the esterase family of proteins (2) made up of 29 esterases and 3 noncatalytic proteins, including thyroglobulin. Testing the glycoprotein for esterase and choline esterase activities has not revealed any biological activity. A comparison of partial sequences from the glycoprotein with human serum butyrylcholine esterase (3) and segment 2183–2750 of bovine thyroglobulin (4) reveal 22% identity among the three proteins. When the Cys-peptides from the glycoprotein were determined, it was found that five of six Cys residues in the glycoprotein are conserved in the other two structures, corresponding to two conserved disulfide bonds in the folded structures. An examination of possible glycosyl-

[1] This work was supported in part by grants from the Robert A. Welch Foundation (AU-916 and AU-0009).

[2] Present address: Conservation and Research Center, Smithsonian Institution, 1500 Remount Rd., Front Royal, VA 22630.

0003-2697/96 $12.00

ation sites in the three proteins, however, suggested that the N-glycosylation sites are not conserved to the same extent. Of the 10 -Asn-X-Ser(Thr)- in the esterase sequence, 9 of which are glycosylated (3), only one is found in the corresponding thyroglobulin segment, and three in the partial glycoprotein sequence. These three sites are all glycosylated in the glycoprotein.

In this paper we report the elucidation of the three disulfide bonds and the three glycosylation sites in the turtle glycoprotein. Since the complete sequence of the glycoprotein has not been determined, the numbering system established for human serum choline esterase (3) is used to designate the glycoprotein sequences involved.

MATERIALS AND METHODS

General Preparative and Analytical Procedures

The secretion was collected and the glycoprotein was purified from secretion samples pooled from more than 2000 animals according to methods described previously (1, 5). For purification, the secretion is simply gel filtered on Sephadex G-100 to yield a high M_r, excluded, slightly turbid peak 1; a symmetrical major peak 2 with a minor trailing peak; and a major included peak 3 containing inorganic and organic salts, amino acids, and lipids. Peak 2 is collected, concentrated, and subjected to gel filtration on Sephadex G-75; using appropriate standards, an M_r of 55K–56K can be established for the glycoprotein. SDS–polyacrylamide gel electrophoresis of the product from this second gel filtration step showed that it is at least 95% pure and confirmed the M_r of approximately 55K. This product from the second gel filtration step was used for all the studies reported in this paper.

Fast atom bombardment mass spectrometry (FAB-MS)[3] was performed on a Kratos MS50RF high-resolution mass spectrometer. A concentrated aqueous solution of the lyophilized samples from HPLC or gel filtration columns was used for the MS analyses. A 1-μl aliquot, containing 1–5 nmol of glycan, was mixed well with 1 μl of either a 90:10 mixture of glycerol and a saturated aqueous oxalic acid solution or a 50:50 mixture of glycerol:thioglycerol. The entire sample was applied to the FAB probe of the mass spectrometer. The spectra were obtained using Xe as the bombarding gas and recorded using the wide-range multichannel analyzer program (raw data) of the Kratos Mach 3 data system. Spectra were calibrated with cesium iodide. Amino acid sequencing was done with the Applied Biosystems 477A gas-phase sequencer. Amino acid analyses, including glucosamine analyses, were carried out with the LKB Alpha Plus amino acid analyzer, using standard hydrolysis conditions, 21 h at 110°C in 6 N HCl, for amino acid analysis, and special conditions, 4 h at 110°C in 2 N HCl, for the analysis of glucosamine. A routine test for the presence of glycans in a given fraction was conveniently done in this way; for complete analysis of the sugar content of each fraction, the samples were hydrolyzed with 2 N trifluoroacetic acid for 3 h at 100°C (6) and analyzed by ion-exchange chromatography with pulsed amperometric detection on the Dionex BioLC carbohydrate analyzer system using a CarboPak PA 1 column eluted isocratically with dilute (4–10 mM) NaOH.

[3] Abbreviations used: FAB-MS, fast atom bombardment mass spectrometry; Bu_3P, tributylphosphine; ABD-F, 4(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole; GP, glycopeptide; G, glycan; P, peptide.

Preparation and Analysis of Individual Cys-Peptides and Glycopeptides

For disulfide characterization, the glycoprotein in Tris buffer, pH 8.0, was digested overnight with chymotrypsin (1:20, w/w) at 37°C. The resulting peptides were subjected to HPLC on a C18 column eluted with a 1-h 0–60% acetonitrile gradient in 0.1% trifluoroacetic acid with a 1 ml/min flow rate. Individual fractions were collected and a sample of each was treated with tributylphosphine (Bu_3P) and 4(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD-F) (7). Three fractions, with retention times of 18, 30, and 35 min, gave a fluorescent (activation at 265 nm; fluorescence above 418 nm) product. These three fractions were treated with Bu_3P and ABD-F and were rechromatographed under the same conditions on HPLC. Each peak yielded two new fluorescent peaks; these were collected and subjected to sequencing.

For glycopeptide characterization, the glycoprotein, in 0.5 M Tris buffer, pH 8.4, containing 6 M guanidinium Cl and 2 mM EDTA, was treated with a twofold excess, relative to Cys, of Bu_3P and ABD-F to reduce all disulfide bonds and block all thiols (7). The reaction mixture was dialyzed against water and finally against Tris buffer, pH 8.0. Chymotrypsin was added to the resulting turbid protein solution at the level of about 1:50 (w/w), and the reaction was left at 37°C overnight. The progress of the digestion was conveniently monitored by reverse-phase HPLC on C18 columns (the standard procedure was a 1 h 15–60% acetonitrile gradient in 0.1% trifluoroacetic acid with a 1 ml/min flow rate). The resulting peptide maps recorded both absorbance at 230 nm and fluorescence above 418 nm (activation at 265 nm) of the ABD-Cys-containing peptides. To simplify the characterization of glycopeptides, the peptide mixture was subjected to affinity chromatography on a Con A–Sepharose (Pharmacia) column; after thorough washing, the bound glycopeptides were eluted with αMe-mannoside, and the αMe-mannoside was finally removed on Biogel P-2. Both the nonretarded and the retarded fractions from Con A–Sepha-

rose were analyzed for the presence of carbohydrate. All the carbohydrate was found in the retarded fraction, and the subsequent analyses were thus carried out with this fraction only.

The standard HPLC analysis of the fraction bound to Con A–Sepharose gave four major peaks, two of which showed fluorescence. Multiple HPLC runs permitted accumulation of sufficient amounts of the individual glycopeptides to subject them to further analysis. A fraction was kept intact; another fraction was treated with the glycoamidase, *N*-glycosidase F (Boehringer Mannheim), in 0.1 M (Na) phosphate buffer, pH 7.2, overnight at 37°C. The reaction mixture was again subjected to standard HPLC; the liberated glycans were collected in the breakthrough volume along with buffer salts and other hydrophilic substances and the free peptide in the acetonitrile gradient very close to the original glycopeptide position. The various fractions were analyzed as follows: Individual glycopeptides were subjected to amino acid sequencing and to FAB-MS. After glycoamidase treatment, the individual peptides (containing Asp instead of the glycosyl-Asn in the glycopeptide) were subjected to amino acid sequencing and to FAB-MS. Finally, the glycans, after desalting on Biogel P-2, were subjected to FAB-MS.

RESULTS AND DISCUSSION

Disulfide Bond Determination

The results of the identification of the Cys-containing peptides in the glycoprotein are summarized in Fig. 1. The separate analysis of crosslinked pairs of peptides permitted the identification of the following disulfide bonds: Cys 65-Cys 92 and Cys 254-Cys 265, both of which are present in the choline esterase and thyroglobulin, and also Cys 130-Cys 404, which is not. Cys 404 is conserved in the other two proteins, but Cys 130 is unique to the glycoprotein; it is replaced with Gly in the other two proteins, and Cys 404 is crosslinked with Cys 533 instead. These six conserved cysteine residues, 65, 92, 254, 265, 404, and 533, are the only ones present in thyroglobulin; choline esterase, in addition, has Cys 66 and Cys 585. It has been proposed that the latter one is involved in intersubunit disulfide bond formation. If ABD-Cys is close to aromatic residues in a given peptide, the fluorescence may be quenched (7), and that peptide is not detected by the above method. With only incomplete sequence data for the glycoprotein, we have assumed that there are no quenched, "invisible," Cys-peptides, and that the present data show the presence of only six Cys residues in the glycoprotein. Three of the identified Cys-peptides are derived from very short sequences (Fig. 1), and the conclusions thus need to be carefully evaluated with regard to the validity of the assignments. Cys 65 (CHQW) is the most problematic one, and the assignment is made simply on the conservation of CXQX in the other two proteins. Cys 265 (VACL, with CL conserved), on the other hand, receives additional documentation as a component of a larger tryptic peptide which in addition to Cys 265 contains both Cys 254 and the glycosylation site N 258. Similarly, Cys 130 (SRCEF with C and F conserved) is also part of a large, 47-residue peptide which shows extensive identity among the three proteins. Thus, the unique Cys in the glycoprotein Cys 130 appears to be well documented as a replacement for Cys 533 in the disulfide bond involving Cys 404. The conservation of cysteines and indeed of disulfide bonds in the turtle glycoprotein is consistent with the conclusion that it is a member of the esterase family of proteins.

Characterization of the Glycosylation Sites

The carbohydrate analyses obtained from the hydrolysates of intact glycoprotein showed the presence of only two sugars, GlcNAc and Man; quantitatively the ratio of the two sugars varied for the different glycoprotein preparations, and, as expected, was sensitive to the hydrolysis conditions. All samples showed a GlcNAc:Man ratio in the range from 2:1 to 3:1. The chromatograms were remarkably free of other sugars; only two minor peaks corresponding to less than 5% of the glucosamine peak were discerned in elution positions corresponding to galactosamine and xylose. This observation, along with the consistent cochromatography of the two sugar peaks from the glycoprotein with authentic glucosamine and Man, clearly established that these two sugars are the primary building blocks of the glycans; this is the basis for the assignment of these two sugars to all the observed mass peaks.

The four glycopeptides obtained by affinity chromatography and preparative HPLC were sufficiently pure to permit detailed analysis by sequencing and FAB-MS. In all cases it was possible to determine the entire sequence of the peptides; even with the expected preponderance of hydrophobic residues at the C-terminus of the chymotryptic peptides, the background was low enough to permit unequivocal identification of all the residues. Two of the peptides, GP 1 and GP 3, had fluorescence and contained ABD-Cys. Sequence analysis showed that the two peptides display a similar sequence; they differ only in the presence of an N-terminal Leu in GP 3, which is absent in GP 1. The data for the four glycopeptides are summarized in Table 1. The analysis includes several steps and depends on the separate information from the different components involved. The mass of the glycopeptide (GP) is the sum of the glycan (G) and the peptide (P) (G + P = GP − 18); thus, if one determines the mass of GP and P and subtracts the latter from the former, the mass of the glycan radical (M − 18) is obtained. The results in Table 1 are reported in this way, and deduced glycan

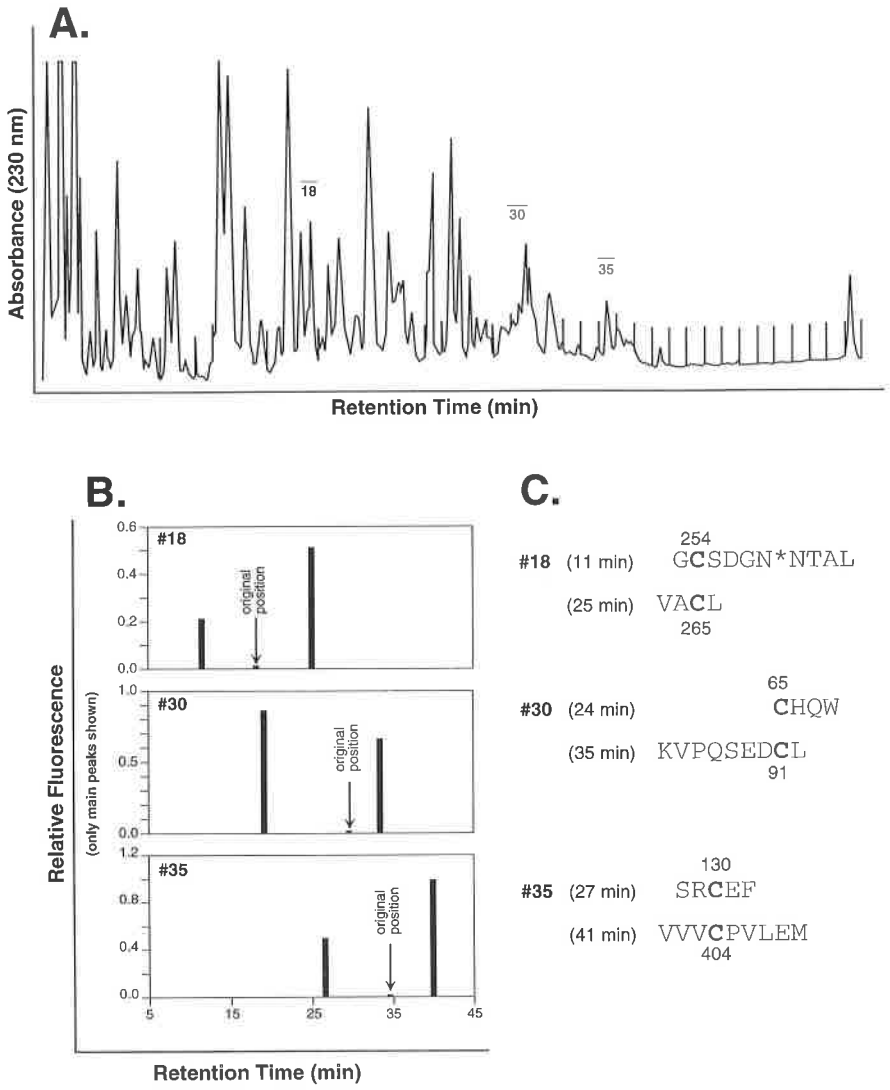

FIG. 1. Identification of Cys-containing peptides. (A) The HPLC pattern of a chymotryptic digest of 7 nmol of turtle glycoprotein. The elution was carried out with a C18 column and a 60-min linear gradient from 0 to 60% acetonitrile in 0.1% trifluoroacetic acid. The elution was monitored for absorbance at 230 nm; the flow rate was 1 ml/min, and 1-ml fractions were collected. (B) The fractions from A were lyophilized, reacted with Bu_3P and ABD-F, and then each fraction was rerun with the same HPLC system as in (A), using an automatic sample injector, and by monitoring the fluorescence above 418 nm (activation at 265 nm). Fractions 18, 30, and 36 each gave two fluorescent peaks with retention times distinct from those of the original peptide. (C) Each of the new peaks (retention times indicated in parentheses) was sequenced; C in the sequence is short for the S-ABD-Cys actually observed in the sequences.

masses are all radical masses. Another correction needs to be noted. GP contains Asn-glycan, while the free peptide after the glycoamidase treatment contains Asp. In the manipulation of the data, when G is subtracted from GP, the resulting mass of P contains Asn (radical mass 114.1), while the mass determined for the peptide is for P containing Asp (radical mass 115.1). Thus, before the observed value of P is subtracted from that of GP, it must be reduced by 1. (For P 1 the value subtracted is 1148.2 rather than 1149.2; for P 2 1483.6 rather than 1484.6, etc.) With these considerations, the elucidation of the glycan structures worked well. The calculated mass for the established peptide sequences agreed extremely well with the observed values, and when these values were subtracted from the observed multiple peak glycopeptides, the resulting glycan radical mass values (bold in Table 1) were internally consistent with those expected from the different glycosylation sites and mostly corresponded to reasonable structures differing only in the GlcNAc content. The one peak in GP 4, 1135.4, given in parentheses, is considered an unusual, but plausible structure containing only two Man residues ($GlcNAc_2Man_2GlcNAc_2$-; $(M + H^+) = 1136.5$). A number of observed mass peaks listed

TABLE 1

Amino Acid Sequence and Molecular Mass of Glycopeptides (GP), Peptides (P), and Glycans from the N-Glycosylation Sites in Turtle Glycoprotein

Peptide/glycan	Obsd sequence	Calcd (M + H^+)	Obsd (M + H^+), relative amount of each peak			
GP 1[a]	G(ABD-)CSDGXNTAL		2042.3, 28%	2448.6, 18%	2650.4, 20%	
P 1[b]	G(ABD-)CSDGNTAL	1149.4	1149.2			
Glycan (deduced radical mass)[c]			**894.1**	**1300.4**	**1502.2**	
GP 2[a]	ASVVGSXHTEAEVAL		2376.1, 44%	2783.0, 11%	2985.5, 32%	
P 2[b]	ASVVGSDHTEAEVAL	1483.9	1484.6			
Glycan (deduced radical mass)[c]			**892.5**	**1299.4**	**1501.9**	
GP 3[a]	LG(ABD-)CSDGXNTAL			2559.6, 15%	2763.5, 21%	
P 3[b]	LG(ABD-)CSDGDNTAL	1262.5		1262.4		
Glycan (deduced radical mass)[c]				**1298.2**	**1502.1**	
GP 4[a]	NLEXASSIGW		(2225.9), 18%	2389.4, 22%	2592.6, 22%	2796.3, 10%
P 4[b]	NLEDASSIGW	1091.5	1091.5			
Glycan (deduced radical mass)[c]			(**1135.4**)	**1298.9**	**1502.1**	**1705.8**

[a] The sequence of the glycopeptide shows a "hole" (X) for the glycosylated Asn residue. FAB-MS data are reported in terms of the identifiable mass peaks indicating the relative amount of each peak as a fraction (%) of the intensity (height) of that peak relative to the sum of the heights of all the recorded peaks. The identified peaks correspond to multiple glycoforms, differing primarily in GlcNAc (radical mass 203) content at each glycosylation site; the 1135.4 peak in GP 4 differs by the mass of one hexose with radical mass 162. Additional peaks were observed, but could not be identified; GP 1: 2165.8, 15%; 2322.6, 9%; 2602.1, 10%. GP 2: 2271.7, 7%; 2648.7, 6%; 2851.7, 6%. GP3: 2186.6, 26%; 2289.9, 11%; 2349.1, 9% 2473.6, 9%; 2633, 9%. GP 4: 2156.0, 22%; 2699.1, 6%.

[b] After treatment with glycoamidase, the glycosyl-Asn in the glycopeptide is converted to and identified as Asp.

[c] The glycan radical mass is deduced by subtracting the mass of the peptide corrected (by subtracting 1) for the presence of Asn instead of Asp in the glycopeptide: Glycan radical mass = observed glycopeptide mass − (observed peptide mass − 1).

in footnote *a* in Table 1 cannot be identified in terms of glycopeptides containing the identified peptides and glycans made up of only GlcNAc and Man. Most of these peaks are minor ones, each representing 10% or less of the total, but some, notably the 2165.8 (15%) peak in GP 1, the 2186.6 (26%) peak in GP 3, and the 2156.0 (22%) peak in GP 4, represent significant components of the reaction mixture. We have no explanation for these peaks. The most likely explanation of the rather small discrepancies in mass is that they reflect peptide microheterogeneity. However, most amino acid variances would cause a significant shift in the retention time of both the glycopeptide and the free peptide and presumably preclude the appearance of the heterogeneous peptides in the same HPLC fraction.

After treating the glycopeptides with glycoamidase, the nonretarded fractions from HPLC were pooled and desalted on BioGel P-2, and the concentrated, desalted fraction was subjected to FAB-MS analysis. As shown in Table 2, two major and two minor fractions were observed. The two major ones (1299.5 and 1502.6) correspond to the two observed in all the individual glycopeptides (Table 1); the minor peak with a deduced radical mass of 1094.5, corresponding to the proposed precursor GlcNAc*$Man_3GlcNAc_2$-, was not observed in any of the glycopeptides, while that with a deduced radical mass of 1705.5 ($GlcNAc_4Man_3GlcNAc_2$-) was observed only in GP 4.

Figure 2 compares the glycosylated peptides from turtle glycoprotein to the corresponding ones in choline esterase and thyroglobulin. In spite of the limited residue identity we feel confident that the comparative assignments are correct. This comparison shows that only 3 of the 9 glycosylation sites in the choline esterase are conserved in the turtle glycoprotein, and that only one of them is common to all three proteins. The 9 sites (of a total of 10 possible ones), Asn 17, 57, 106, 243, 258, 343, 461, 487, and 491, are all glycosylated in choline esterase (3), while the 10th, Asn 492, is not. All three

TABLE 2

Mass of the Pooled Glycans Liberated from the Glycopeptides by Digestion with Glycoamidase[a]

Observed mass (M + Na^+)[b]	1134.5	1339.5	1542.6	1745.7
Relative amount (%)[c]	13	27	51	8
Deduced glycan radical mass	**1094.5**	**1299.5**	**1502.6**	**1705.7**

[a] The glycans were recovered in the early "salt fraction" of the HPLC run. After lyophilization, the fraction was dissolved in water and gel filtered on a column of Biogel P-2 eluted with water. The salt-free excluded fraction was again lyophilized and dissolved in a small volume of water for FAB-MS analysis. The gel filtration step was repeated as required to give salt-free samples for analysis.

[b] The above procedure appeared to yield exclusively the Na^+ adducts of the glycans; only traces of or no masses corresponding to H^+ adducts were observed. To convert the observed mass of (free glycan + Na^+) to the radical mass derived in Table 1, the mass of Na^+ (22) and water (18) must be subtracted.

[c] The relative amount of each of the four major peaks was calculated from the intensity (height) of each peak as a fraction (%) of the sum of the intensities of all peaks.

Turtle	(L)GCSDGNNTAL
	258
Esterase	TGCSRENETEI
Thyroglob.	VGCPSSSVQEM
Turtle	ASVVGSNHTEAEVAL
	461
Esterase	L-ERRDNYTKAEEIL
Thyroglob.	YPAYEGQFTLEEKSL
Turtle	NLENASSIGW
	106
Esterase	KPKNATVLIW
Thyroglob.	MAPNASVLVF

FIG. 2. The sequence alignments of the three glycosylation sites in the turtle glycoprotein with the corresponding sequences in human serum choline esterase and bovine thyroglobulin (2). Identical residues in all three proteins are indicated by underlined bold letters; residues common to any two of the three proteins are indicated by bold letters—indicates a deletion.

sites in the turtle glycoprotein, Asn 106, 258, and 461 are conserved from the esterase and are glycosylated. Thyroglobulin has two potential glycosylation sites, Asn 51 and 106, but it is not known if they are actually glycosylated (4); if the one universally conserved site, Asn 106, is found to be glycosylated, a maximum of a single site is conserved and glycosylated in all three proteins.

The primary conclusion from the results is that the glycans in the turtle glycoprotein, like those in other secreted glycoproteins, are heterogeneous, and that the *in vivo* processing of different glycosylation sites differs. Thus, while the $GlcNAc_2Man_3GlcNAc_2$- and $GlcNAc_3Man_3GlcNAc_2$- structures were seen at all three sites, $Man_3GlcNAc_2$- was found only in GP 1 and GP 2, and $GlcNAc_4Man_3GlcNAc_2$- and the putative $GlcNAc_2Man_2GlcNAc_2$- only in GP 4. Figure 3 presents the most likely structures for the observed glycan masses, assuming that the turtle glycoproteins are produced according to the common pathways of glycoprotein biosynthesis and processing in plants and animals. The structure GlcNAc*$Man_3GlcNAc_2$-, the main product of mannosidase II, appears to be a common intermediate in all eukaryotic organisms other than yeasts and fungi. In many animal cells this intermediate is acted on by GlcNAc transferases to add additional GlcNAc in the positions indicated in Fig. 3 (8, 9), and by Gal transferase to produce Gal-containing bi-, tri-, and tetraantennary complex structures. In some tissues, the Gal transferases appear to be low and the end products of glycan processing are similar to those proposed in Fig. 3. Examples of such end products are the heterogeneous glycans in avian egg white proteins such as ovomucoid (10), ovoinhibitor, and ovotransferrin (11) and minor components in ovalbumin in which high levels of hybrid structures are also observed (10). The structure $Man_3GlcNAc_2$- has only been observed as a major product of glycan processing in mosquito cells (12) and as a minor product in plant glycoproteins, where the major components contain additional fucose and xylose (11). Information available for neither the glycan structure in turtle glycoproteins nor the processing enzymes involved. It is interesting to speculate at this time that the glycan processing in the Rathke's gland may be similar to that in avian oviducts.

It is not possible to draw definitive conclusions regarding evolutionary conservation based on the limited data available. It may be of interest to note, however, that three of five hormonogenic sites in turtle thyroglobulin are homologous to those in mammalian species, suggesting significant conservation of these iodination sites over the considerable evolutionary distance involved (13). Similarly, the significant sequence identity among the three proteins studied here, and especially the extensive Cys and disulfide conservation, together with the apparent conservation of the active site Ser in all the catalytically active members of the family (2) suggest that both functional and structural features of this family of proteins have been extensively con-

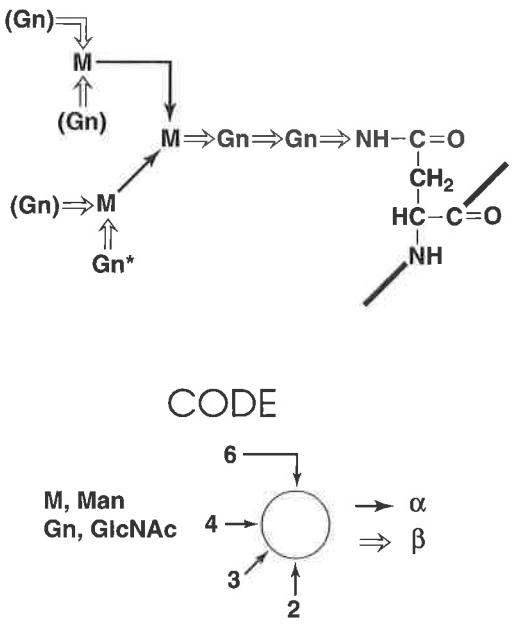

FIG. 3. Hypothetical structures assigned to the N-linked glycans in turtle glycoprotein. Based on the monosaccharide analysis, it is assumed that GlcNAc (Gn) and Man (M) are the only building blocks. It is reasonable to also assume that the common precursor of all the observed structures is the product of mannosidase II, $Gn^*M_3Gn_2$ (radical mass 1096.1), before any additional GlcNAc (Gn) has been introduced. Removal of Gn* from this precursor by a hexosaminidase would yield M_3Gn_2 (radical mass 892.9), while addition of Gn to the common precursor with GlcNAc transferases would give rise to $Gn_2M_3Gn_2$ (radical mass 1299.3), $Gn_3M_3Gn_2$ (radical mass 1502.5), and $Gn_4M_3Gn_2$ (radical mass 1705.7). All these species except the proposed common precursor are observed in the glycopeptides. If the proposed GP 4 structure, $Gn_2M_2Gn_2$, exists, it must have both Gn on the same 1–3-linked M with the 1–6-linked M missing.

served. The much greater variation in the glycosylation pattern reemphasizes the enigma associated with protein glycosylation. If this family of proteins has indeed conserved certain functional features reflecting certain strucural properties, including folding, the glycosylation cannot be an important determinant of these critical properties. Perhaps only one glycosylation site (Asn 106) is required to establish the universal properties of the family, and the others are required for the specific modulation of the basic structure/function.

REFERENCES

1. Radhakrishna, G., Chin, C. C. Q., Wold, F., and Weldon, P. J. (1989) *Comp. Biochem. Physiol. B* **94,** 375–378.
2. Cygler, M., Schrag, J. D., Sussman, J. L., Harel, M., Silman, I., Gentry, M. K., and Doctor, B. P. (1993) *Protein Sci.* **2,** 366–382.
3. Lockridge, O., Bartels, C. F., Vaughan, T. A., Wong, C. K., Norton, S. E., and Johnson, L. L. (1987) *J. Biol. Chem.* **262,** 549–557.
4. Mercken, L., Simons, M.-J., Swillens, S., Massaer, M., and Vassart, G. (1985) *Nature* **316,** 647–651.
5. Krishna, R. G., Chin, C. C. Q., Weldon, P. J., and Wold, F. (1995) *Comp. Biochem. Physiol. B* **111,** 257–264.
6. Hardy, M. R., Townsend, R. R., and Lee, Y. C. (1988) *Anal. Biochem.* **170,** 54–62.
7. Chin, C. C. Q., and Wold, F. (1993) *Anal. Biochem.* **214,** 128–134.
8. Schachter, H. (1984) *Biol. Cell* **51,** 133–146.
9. Kornfeld, R., and Kornfeld, S. (1985) *Annu. Rev. Biochem.* **54,** 631–664.
10. Yet, M.-G., Chin, C. C. Q., and Wold, F. (1988) *J. Biol. Chem.* **263,** 111–117.
11. Yet, M.-G., and Wold, F. (1990) *Arch. Biochem. Biophys.* **278,** 356–364.
12. Hsieh, P., and Robbins, P. W. (1984) *J. Biol. Chem.* **259,** 2375–2382.
13. Roe, M. T., Anderson, P. C., Dunn, A. D., and Dunn, J. T. (1989) *Endocrinology* **124,** 1327–1332.