Location of the Carbohydrate Groups of Ovomucoid

By JOHN G. BEELEY

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

(Received 12 April 1976)

Tryptic glycopeptides were purified from the sialic acid-free variant of ovomucoid, O₁, and its CNBr fragments. The amino acid sequences adjacent to the four major sites of carbohydrate (Carb.) attachment were: (1), Phe-Pro-Asn(Carb.)-Ala-Thr-Asp-Lys-Glu-Gly-Lys; (2), Ala-Tyr-Ser-Ile-Glu-Phe-Gly-Thr-Asn(Carb.)-Ile-Ser-Lys; (3), Glu-Thr-Val-Pro-Met-Asn(Carb.)-Cys-Ser; (4), Ser-Ser-Tyr-Ala-Asn(Carb.)-Thr-Thr-Ser-Glu-Asp-Gly-Lys. Glycosylated Asn residues were located at position 10, between residues 49 and 60, and at positions 69 and 75, in the primary sequence. All of these carbohydrate groups contained GlcNAc, Man and Gal in the approximate molar proportions 5:3:0.5. A further glycopeptide containing His was isolated in low yield, suggesting that some carbohydrate is attached at a fifth site. Two of the carbohydrate-attachment sites (Asn-10 and Asn-75) occur in sequences that show internal homologies. These are presumed to have evolved as a consequence of partial gene duplication. Three of the carbohydrate-attachment sites occur in similar positions to the carbohydrate groups in quail ovomucoid [Laskowski (1976) *Protides Biol. Fluids Proc. Colloq.* **23**, in the press]. Prediction of peptide conformation from the sequence data by the method of Chou & Fasman [(1974) *Biochemistry* **13**, 222–225] indicated that four glycosylated Asn residues in hen ovomucoid are very close to groups of amino acids that occur with high frequency in β -turns. The possible significance of peptide-chain conformation in the attachment of carbohydrate to glycoproteins is briefly discussed.

The ovomucoids are a family of glycoproteins with anti-tryptic activity which have been isolated from the egg white of several avian species (Feeney & Allison, 1969). Because of its ready availability and high carbohydrate content, the ovomucoid from domestic hen's eggs has been extensively studied (for reviews see Melamed, 1966; Lin & Feeney, 1972). Hen ovomucoid contains a single polypeptide chain to which several carbohydrate groups are attached. However, there is extensive heterogeneity of sialic acid and neutral sugars within ovomucoid preparations (Beeley, 1971). Analysis of glycopeptides obtained by digestion with proteolytic enzymes of broad specificity (Montgomery & Wu, 1963; Neuberger & Papkoff, 1963; Monsigny *et al.*, 1968) led to agreement over the conclusion that the protein-carbohydrate linkages were of the glycosylamine type involving *N*-acetylglucosamine and asparagine, but whereas Montgomery & Wu (1963) proposed that there were three carbohydrate groups all of similar composition, Monsigny *et al.* (1968) found two major carbohydrate-attachment sites.

The work reported here was undertaken to ascertain the number of carbohydrate moieties in ovomucoid, their distribution on the peptide chain and the sequence of amino acid residues adjacent to the sites of carbohydrate attachment. To do this, tryptic pep-

tides obtained from reduced, alkylated ovomucoid or CNBr fragments of ovomucoid (Beeley, 1976) were fractionated and sequenced. The ovomucoid used was the major variant ovomucoid O₁ (Beeley, 1971), which is free of sialic acid and was obtained from the eggs of an individual hen.

Methods

Ovomucoid preparation

Eggs were obtained from individual hens and sialic acid-free ovomucoid variant O₁ was isolated as described previously (Beeley, 1976). All preparations used gave a single band on gel isoelectric focusing (Beeley, 1971), indicating that they were homogeneous in charge.

Reduction and aminoethylation

The procedure used was based on that of Cole (1967). Ovomucoid O₁ (500 mg) was dissolved in 50 ml of buffer containing 0.1 M-Tris, 8 M-urea and 25 mM-EDTA adjusted to pH 8.6 with HCl. O₂-free N₂ was bubbled through the solution before addition of 2-mercaptoethanol (0.5 ml) through a N₂ barrier. After incubation for 18 h at 20°C, freshly redistilled ethyleneimine was added in 3 × 1 ml portions at intervals of 10 min and after a further 30 min the

solution was acidified with acetic acid (6ml). The reaction mixture was applied to a column (5cm × 91cm) of Sephadex G-25 equilibrated and eluted with 0.2M-acetic acid at a flow rate of 40ml/h; fractions (10ml) were collected. The broad peak of aminoethylated ovomucoid, detected by measuring E_{280} , was pooled and freeze-dried (yield 490mg).

Digestion with trypsin

Aminoethylated ovomucoid (250mg) was dissolved in 25ml of 0.1M-NH₃ adjusted to pH8.5 with CO₂ gas. Trypsin (5mg), treated with 1-chloro-4-phenyl-3-tosylamidobutan-2-one (TPCK), prepared from twice-crystallized bovine pancreatic trypsin (Koch-Light Laboratories, Colnbrook, Bucks., U.K.) by the method of Kostka & Carpenter (1964), dissolved in 10mM-HCl (0.5ml) was added and the mixture incubated for 4h at 25°C under toluene. More modified trypsin (5mg) dissolved in 10mM-HCl (0.5ml) was added and digestion was continued for 20h. The digest was heated for 20min at 100°C, then centrifuged and the supernatant was freeze-dried three times (yield 245mg).

Glycopeptide isolation

A portion of the tryptic digest (175mg) was dissolved in 5ml of 0.0075M-NH₃ adjusted to pH9.3 with formic acid and the pH was re-adjusted to 9.3. This sample was applied to a column (1.5cm × 30cm) of QAE [diethyl(2-hydroxypropyl)aminoethyl]-Sephadex A-50 (Pharmacia, Uppsala, Sweden) equilibrated and initially eluted with the pH9.3 ammonium formate buffer. After collection of 15 fractions, a linear gradient was applied, produced by mixing 200ml of the 0.0075M-ammonium formate buffer and 200ml of the 0.5M-NH₃ adjusted to pH9.3 with formic acid. When 75 fractions had been collected, elution was continued with 0.1M-formic acid. The flow rate was 11ml/h and the fractions were all of 4ml. The E_{230} and E_{280} of samples was determined and portions (50μl) were taken for hexose assay. Pooled glycopeptides were freeze-dried repeatedly to remove ammonium formate and were dissolved in 2ml of 0.1M-acetic acid.

Glycopeptides were further purified by gel filtration on columns (1.5cm × 90cm) of Sephadex G-25 equilibrated and eluted with 0.1M-acetic acid at a flow rate of 20ml/h; 4ml fractions were collected. Samples (100μl) were taken for hexose determination and the E_{230} was determined. The glycopeptide peaks were pooled and freeze-dried.

Each of the glycopeptide fractions was dissolved in water (2.0ml); 0.1ml of 0.2M-pyridine adjusted to pH2.8 with formic acid and 20μl of formic acid were added. These samples were applied to columns (1.5cm × 30cm) of SP (sulphopropyl)-Sephadex C-25 equilibrated with 0.01M-pyridine/formate, pH2.8. A linear three-chamber gradient was used. The

three buffers used were 0.01M-pyridine adjusted to pH2.8 with formic acid, 0.2M-pyridine adjusted to pH2.8 with formic acid and 0.4M-pyridine adjusted to pH5.0 with formic acid. The volume of buffer in each reservoir was 150ml, except for the re-chromatography of glycopeptide T-2c, where the volume of buffer in each chamber was 300ml. Columns were eluted at a flow rate of 20ml/h and fractions (5ml) were collected. The E_{230} of samples was read and portions were assayed for hexose content.

Analytical methods

Amino acid analyses were performed after hydrolysis at 110°C for 24h in 6M-HCl in evacuated sealed tubes containing 100μg of phenol/ml. Hexosamine was determined by using an amino acid analyser after hydrolysis with 4M-HCl for 3h at 100°C or with 3M-methanesulphonic acid at 100°C for 24h (Beeley, 1976). Amino acid and hexosamine determinations were related by the inclusion of norleucine as internal standard in both determinations.

Hexose was measured by the method of Dubois *et al.* (1956) and individual monosaccharides were assayed as described by Chambers & Clamp (1971).

End groups were determined as described by Hartley (1970).

Sequence determination

Subtractive Edman degradation was done by a method based on that of Konigsberg (1972). Glycopeptides (10nmol of each residue to be sequenced) were dried in a vacuum desiccator maintained at 45°C by a heating mantle. To the peptide dissolved in 0.25ml of aq. 70% (v/v) pyridine, 0.025ml of phenyl isothiocyanate (sequenator grade, obtained from Eastman, Rochester, NY, U.S.A.) was added; the tubes were flushed with O₂-free N₂, stoppered and incubated at 37°C for 2h. The solvent was removed by evaporation under vacuum and the samples were dried at 45°C overnight in a vacuum desiccator. Dried samples were dissolved in anhydrous trifluoroacetic acid (0.2ml) and were incubated under N₂ at 37°C for 1h. Trifluoroacetic acid was removed in a stream of nitrogen and 0.2M-acetic acid (1.0ml) was added. After incubation at 60°C for 10min, tubes were cooled and extracted with benzene (2.0ml) and centrifuged at 1000g for 5min to break the emulsion that formed. Three further extractions were done, each with 1ml of benzene. A sample of the aqueous phase was hydrolysed for amino acid analysis and the remaining material was dried at 45°C *in vacuo* in preparation for the next degradation cycle.

Hydrazinolysis

Glycopeptides (20–30nmol) were dried over P₂O₅ at 80°C and treated with 0.2ml of hydrazine (95%; Eastman) for 24h at 80°C, as described by

Fraenkel-Conrat & Tsung (1967). Hydrazine was removed *in vacuo* over conc. H_2SO_4 and P_2O_5 . Norleucine (10nmol) was added and the sample was analysed directly on the long column of the amino acid analyser.

Digestion of glycopeptides

Digestion of glycopeptides with carboxypeptidase B or carboxypeptidase A treated with di-isopropyl phosphorofluoridate (Worthington Biochemical Corp., Freehold, NJ, U.S.A.) was carried out at 37°C in 0.2M-*N*-ethylmorpholine adjusted to pH 8.5 with acetic acid (Ambler, 1972).

Pronase (Calbiochem, San Diego, CA, U.S.A.) or chymotrypsin (Worthington) digests were carried out

in 0.05M-*N*-ethylmorpholine/0.01M- CaCl_2 adjusted to pH 7.8 with acetic acid. Glycopeptides (240nmol) were incubated with enzyme (0.1 mg) in a total volume of 0.3 ml for 24 h at 37°C under toluene.

Fig. 1. Chromatography of tryptic digest of aminoethylated ovomucoid on QAE-Sephadex A-50

The tryptic digest (175mg) was applied to a column (1.5cm $\times$ 30cm) equilibrated with 0.0075M- NH_3 adjusted to pH 9.3 with formic acid. A linear gradient from 0.075M-ammonium formate to 0.5M-ammonium formate was applied at fraction 15 and at fraction 75 elution was continued with 0.1M-formic acid. The flow rate was 11 ml/h and fractions were 4ml. The E_{230} of samples was determined ($\bullet$), and portions were assayed for hexose as described in the text and the resulting E_{485} ($\circ$) was recorded. Glycopeptide fractions (T1-T4) were pooled as indicated by the bars.

Fig. 2. Chromatography of tryptic glycopeptides on SP-Sephadex C-25

The glycopeptides (T1-T4) were applied to columns (1.5cm $\times$ 30cm) equilibrated with 0.01M-pyridine adjusted to pH 2.8 with formic acid and were eluted with a gradient produced from three chambers containing 0.01M-pyridine adjusted to pH 2.8 with formic acid (150ml), 0.2M-pyridine adjusted to pH 2.8 with formic acid and 0.4M-pyridine adjusted to pH 5.0 with formic acid. The E_{230} of samples was measured ($\bullet$), and portions were assayed for hexose by the method of Dubois *et al.* (1956) and their E_{485} was recorded ($\circ$). Purified glycopeptides (T-1a etc.) were pooled as indicated.

Table 1. *Composition of tryptic glycopeptides isolated from aminoethylated ovomucoid*

Results are expressed as mol of residues/mol of peptide and were calculated relative to integral values of Asp, Lys, or Ala. Values in parentheses give the probable integral number of each residue. The charge at pH 6.5 was determined by the method of Offord (1966). Abbreviation: Cys(CH₂)₂NH₂, 2-aminoethylcysteine. The ovomucoid (O₁) used was obtained from the eggs of a single hen (no. 13).

Amino acid		Composition (mol of residue/mol of peptide)							
Glycopeptide	...	T-1a	T-2c	T-2cR	T-2d	T-3a	T-3b	T-3c	T-4a
Lys			1.0	1.0 (1)	1.0 (1)	0.1	0.7 (1)	2.0 (2)	0.9 (1)
His					1.0 (1)				
Cys(CH ₂) ₂ NH ₂	0.8 (1)					0.8 (1)			
Asp	1.0 (1)	1.7	1.8 (2)	1.9 (2)	1.0 (1)	1.1 (1)	1.9 (2)	2.0 (2)	
Thr		1.0	1.0 (1)	1.8 (2)	0.9 (1)	1.0 (1)	1.0 (1)	1.9 (2)	
Ser		0.8	0.4	2.0 (2)	0.1	1.7 (2)	0.2	2.7 (3)	
Glu		0.5	0.2	1.0 (1)	1.0 (1)	0.9 (1)	1.0 (1)	1.0 (1)	1.0 (1)
Pro		0.8	0.9 (1)	0.2	1.0 (1)				0.2
Gly		0.5	0.2	2.0 (2)	0.1	1.0 (1)	1.1 (1)	1.1 (1)	
Ala		0.9	0.8 (1)	1.1 (1)	0.1	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)
Val				1.2 (1)	1.1 (1)				0.2
Met					0.7 (1)				
Ile		0.8	0.4				1.6 (2)		
Leu					1.8 (2)				
Tyr							0.8 (1)		0.9 (1)
Phe			1.1	0.9 (1)	0.9 (1)		1.0 (1)	0.9 (1)	
GlcNAc	4.6				4.7	4.5	5.0	4.5	5.2
Hexose					5.0			3.0	
Man			2.9			2.4	2.7		2.7
Gal			0.5			0.5	0.5		0.5
Charge	+1		0	0	0	0	0	0	-1

Electrophoretic mobility

Mobilities of glycopeptides were determined by high-voltage paper electrophoresis in pyridine/acetic acid/water (25:225:1, by vol.), pH 6.5, at 40 V/cm (Offord, 1966).

Results

Tryptic glycopeptides from ovomucoid

After treatment of aminoethylated ovomucoid with trypsin, four glycopeptide fractions (T1-T4) were separated by chromatography on QAE-Sephadex A-50 (Fig. 1). Each of these fractions was chromatographed on Sephadex G-25 and then on SP-Sephadex C-25 (Fig. 2) and the seven glycopeptides, whose compositions are given in Table 1, were isolated. With one exception (T-2c), the composition of these glycopeptides suggested that they were homogeneous in their peptide moieties.

By the dansyl-Edman (Hartley, 1970) and subtractive Edman methods peptide T-1a gave the sequence Asx-Cys(CH₂)₂NH₂. The electrophoretic mobility of the peptides at pH 6.5 showed that it was positively charged. It was concluded that the structure of this peptide is Asn(Carb.)*-Cys(CH₂)₂NH₂.

Fraction T-2c contained a mixture of glycopeptides.

* Abbreviation: Carb. (in sequences), carbohydrate.

Table 2. *Subtractive Edman degradation of peptide T-2cR*

Portions of glycopeptide T-2cR were subjected to hydrolysis and amino acid analysis after successive cleavages by Edman degradation. The compositions of peptides are related to integral values for Ala or Asp. Underlined values indicate the *N*-terminal residue lost at each cleavage.

	No. of cleavages	Composition (mol of residue/mol of peptide)					
		0	1	2	3	4	5
Lys	1.0	—	—	—	—	—	—
Asp	2.1	2.2	2.3	1.2	1.0	1.0	—
Thr	1.1	1.0	1.1	1.1	0.9	0.2	—
Pro	0.9	0.9	0.2	—	—	—	—
Ala	1.0	1.0	1.0	1.0	0.2	—	—
Phe	1.1	0.2	—	—	—	—	—
Sequence		Phe	Pro	Asx	Ala	Thr(Asx,Lys)	

Rechromatography of this fraction on SP-Sephadex C-25 by using an extended gradient led to the isolation of glycopeptide T-2cR, which was sufficiently homogeneous for determination of the sequence Phe-Pro-Asx-Ala-Thr(Asx,Lys) by the subtractive Edman procedure (Table 2). Digestion with carboxypeptidases A and B brought about the release of

Ala (0.2), Thr (0.6), Asp (0.8), and Lys (1.0mol of residues/mol of peptide) and the peptide carried no net charge at pH6.5. The structure of the glycopeptide is therefore Phe-Pro-Asn(Carb.)-Ala-Thr-Asp-Lys (Fig. 3).

Glycopeptide T-2d was the only histidine-containing peptide isolated. Subtractive Edman degradation gave Asx as the N-terminal residue. Because this peptide contained 16 residues and was obtained in low yield it was not sequenced further.

The composition of glycopeptide T-3a is shown in Table 1 and its complete sequence determined by Edman degradation is given in Fig. 4. Because the sequence Asx-Cys(CH₂)₂NH₂ occurred in both this glycopeptide and peptide T-1a, the possibility that the latter might arise from enzymic cleavage of peptide T-3a was examined. Glycopeptide T-3a (70 nmol) was incubated at 25°C for 24h with treated trypsin (0.5mg) in 250 µl of 0.1M-NH₃ adjusted to pH8.5 with CO₂ gas. The freeze-dried digest was chromatographed on SP-Sephadex C-25. A peptide having the composition Glx (1), Thr (1), Val (1), Pro (1), Met (1) and a glycopeptide with the same composition as peptide T-1a [Asx (1), Cys(CH₂)₂NH₂ (1)] were isolated from this digest. It was concluded that glycopeptide T-1a arose from hydrolysis of the amide bond involving the carbonyl group of a methionine residue. This cleavage is outside the normal range of specificity for trypsin and may result from pseudotrypsin (Smith & Shaw, 1969) in the enzyme preparation.

The first six residues from the N-terminal end of glycopeptide T-3b (Table 1 and Fig. 5) were obtained by sequential Edman degradation. Digestion of peptide T-3b (50 nmol) with carboxypeptidases A

(20 µg) and B (1 µg) in 0.2ml of N-ethylmorpholine adjusted to pH8.5 with acetic acid at 37°C resulted in the release, after 2.5h, of Ile (0.2), Ser (0.3) and Lys (1.0mol of residues/mol of peptide). After 18h digestion the residues released were Ile (0.6), Ser (0.8) and Lys (1.0). After digestion of peptide T-3b with chymotrypsin and Pronase respectively, the glycopeptides T-3b-C and T-3b-P with the composition given in Table 3 were isolated after gel filtration. After one round of Edman degradation on glycopeptide T-3b-P, a glycopeptide T-3b-P-E was isolated by gel filtration and was found to contain only carbohydrate and aspartic acid (Table 3). The proposed sequence of glycopeptide T-3b is shown in Fig. 5. Glycopeptide T-3c contained two residues of lysine (Table 1) and the composition suggested that it arose, by incomplete tryptic cleavage, from the same carbohydrate attachment site as peptide T-2c. The N-terminal residue was phenylalanine. Digestion of peptide T-3c with carboxypeptidase B resulted in the release of 1 mol of lysine/mol of peptide, and gel filtration on Sephadex G-25 yielded a glycopeptide, T-3c-CPB, that differed from T-3c in that it contained only a single residue of lysine. This peptide was resistant to carboxypeptidase A, but hydrazinolysis of 22 nmol yielded 12.7 nmol of glycine. The sequence of this peptide is given in Fig. 3.

Subtractive Edman degradation yielded the sequence of nine residues of peptide T-4a (Fig. 4) which had the composition shown in Table 1. Treatment with carboxypeptidase B and gel filtration

Fig. 3. Amino acid sequence of glycopeptides T-2cR, T-3c and T-3c-CPB

The sequence of underlined residues was determined by subtractive Edman degradation. ----, Hydrazinolysis was used to establish the C-terminal residue.

Table 3. Composition of glycopeptides isolated from peptide T-3b

The glycopeptides were isolated by gel filtration after treatment of peptide T-3b with either chymotrypsin (T-3b-C), Pronase (T-3b-P) or Pronase followed by one Edman degradation (T-3b-P-E).

	Composition (mol of residue/mol of peptide)			
	T-3b	T-3b-C	T-3b-P	T-3b-P-E
Lys	0.7	—	—	—
Asp	1.0	1.0	1.0	1.0
Thr	1.0	1.0	1.0	0.2
Ser	1.7	1.0	—	—
Glu	0.9	—	—	—
Gly	1.0	1.2	0.2	0.1
Ala	1.0	—	—	—
Ile	1.6	1.2	—	—
Tyr	0.8	—	—	—
Phe	1.0	—	—	—
GlcN	5.0	2.8*	3.1*	2.4*

* These glucosamine values were determined after hydrolysis at 110°C in 6M-HCl and are not corrected for destruction.

Fig. 4. Amino acid sequence of tryptic glycopeptides T-1a, T-3a, T-4a, T-4a-CPB and T-4a-P and CNBr fragment S
The sequence of underlined residues was determined by subtractive Edman degradation.

Fig. 5. Amino acid sequence of glycopeptides T-3b, T-3b-C and T-3b-P

The sequence of underlined residues was determined by subtractive Edman degradation. ----, Sequences established by digestion with carboxypeptidases B and A.

Table 4. Glycopeptides isolated from tryptic digests of ovomucoid CNBr fragments M-AE and S-AE

To determine location of the tryptic glycopeptides obtained from aminoethylated ovomucoid (Table 1), glycopeptides were isolated from the *N*-terminal (M-AE) and central (S-AE) CNBr fragments of ovomucoid (Beeley, 1976).

Amino acid	Composition			
	(mol of residue/mol of peptide)			
Component ...	M-AE-T ₁	M-AE-T ₂	S-AE-T ₁	S-AE-T ₂
Lys	1.0	1.0	—	0.9
Cys(CH ₂) ₂ NH ₂	—	—	1.0	—
Asp	1.7	1.2	1.0	1.9
Thr	1.0	1.1	—	2.0
Ser	0.7	1.7	—	2.9
Glu	0.3	1.0	—	0.9
Pro	0.6	—	—	—
Gly	0.4	1.0	—	1.0
Ala	0.8	1.0	—	1.0
Ile	0.8	1.8	—	—
Tyr	—	0.5	—	0.8
Phe	1.0	1.0	—	—
GlcN	3.8*	3.6*	4.4*	3.5

* See footnote to Table 3.

yielded a glycopeptide (T-4a-CPB) which contained less than 0.1 mol of lysine/mol, but was otherwise identical in composition with peptide T-4a. Hydrazinolysis of 27 nmol of peptide T-4a released 12.8 nmol of glycine. Pronase digestion of peptide T-4a and gel filtration resulted in the isolation of a glycopeptide containing equimolar amounts of aspartic acid and threonine and traces of alanine, serine and glutamic acid. Edman degradation resulted

in loss of aspartic acid, and analysis of the unhydrolysed products indicated that threonine was the only detectable free amino acid. The sequence of peptide T-4a is shown in Fig. 4.

Tryptic glycopeptides from CNBr fragments of ovomucoid

The isolation of three reduced aminoethylated CNBr fragments of ovomucoid (L1-AE, M-AE and S-AE) and the evidence establishing their order of occurrence in the peptide sequence of ovomucoid have been described previously (Beeley, 1976).

After digestion of the CNBr fragments with trypsin, glycopeptides were isolated by similar chromatographic methods to those used for isolating glycopeptides from the intact glycoprotein. The compositions of the peptides isolated are shown in Table 4.

Two glycopeptide fractions were isolated from CNBr fragment M-AE (Table 4). The behaviour on ion-exchange chromatography and the composition of fraction M-AE-T₂ were identical with those of peptide T-3b. Fraction M-AE-T₁ was closely similar in composition and chromatographic properties to fraction T-2c and on further fractionation yielded a glycopeptide indistinguishable from peptide T-2c-R. It was concluded that the carbohydrate-attachment sites shown in Fig. 3 (site 1) and Fig. 5 (site 2) both originated from CNBr fragment M-AE.

Two tryptic glycopeptides (S-AE-T₁ and S-AE-T₂) were isolated from fragment S-AE (Table 4). These had the same compositions as glycopeptides T-1a and T-4a respectively. Subtractive Edman degradation of fragment S-AE allowed the determination of the sequence of nine residues from the *N*-terminal

Table 5. Recovery of glycopeptides from tryptic digest of aminoethylated ovomucoid

Recovery is expressed as (mol of glycopeptide recovered/mol of aminoethylated ovomucoid) $\times$ 100 and correction was made for samples removed for analysis.

Glycosylated sequence ... Glycosylation site ...	Recovery (%)				
	Pro-Asn-Ala 1	Thr-Asn-Ile 2	Met-Asn-Cys 3	Ala-Asn-Thr 4	? 5
Glycopeptide	T-3c (5.3) T-2c (10.7)	T-3b (21.9)	T-1a (5.6) T-3a (15.5)	T-4a (29.2)	T-2d (4.2)
Total recovery (%)	16.0	21.9	21.1	29.2	4.2

end (Fig. 4). This information, together with the sequences of peptides T-3a and T-4a made it possible to locate precisely the two carbohydrate-attachment sites (sites 3 and 4) in the central CNBr fragment of ovomucoid.

No homogeneous glycopeptides were recovered from the tryptic digest of the C-terminal CNBr fragment L1-AE.

Recovery of glycopeptides from ovomucoid

The peptide sequences determined in the tryptic peptides isolated from ovomucoid or CNBr fragments of ovomucoid appear to arise from four major sites of glycosylation (Figs. 3–5). Recoveries of glycopeptides originating from these sites were within the range 16–29% (Table 5). Without quantitative recovery of glycopeptides it is not possible to determine whether all these sites are fully glycosylated in all molecules of ovomucoid. However, if it is assumed that the average carbohydrate compositions of the glycopeptides at these four sites [GlcNAc, 5: Man, 3: Gal, 0.5 (mol of residues/mol of glycopeptide)] are representative of the sites in ovomucoid from which they are derived, approx. 20 *N*-acetyl-D-glucosamine and 14 hexose residues per molecule of ovomucoid can be accounted for of the 23 *N*-acetyl-D-glucosamine and 15 hexose residues per molecule (mol.wt. 27000) of glycoprotein. The fractional values obtained for the carbohydrate compositions of the glycopeptides indicate that microheterogeneity occurs at all of the sites.

Glycopeptide T-2d originates from a peptide sequence distinct from the four major glycosylation sites. The low recovery of this glycopeptide (Table 5) might indicate that it originates from an impurity, that large losses occur during chromatographic procedures or that it arises from a fifth site which is particularly heterogeneous or which is only partly glycosylated. Marked heterogeneity of the carbohydrate attached to CNBr fragment L-AE has been observed (Beeley, 1976), suggesting that a very heterogeneous or partially glycosylated region occurs in ovomucoid.

Discussion

Although proteolytic enzymes of broad specificity have been widely used to produce glycopeptides from glycoproteins, more specific methods such as cleavage with trypsin or CNBr have advantages in structural studies when multiple carbohydrate groups are present. Thus it was possible to isolate, from a tryptic digest of aminoethylated ovomucoid, glycopeptides that were sufficiently homogeneous in their peptide moiety and large enough to yield distinct sequences for four different carbohydrate-attachment sites, and the occurrence of an additional site, or sites, of glycosylation was suggested. Previous investigations of Pronase or papain digests of ovomucoid gave glycopeptides containing few amino acids, and only two (Monsigny *et al.*, 1968) or three (Montgomery & Wu, 1963) major sites of carbohydrate attachment were recognized.

The sequences of tryptic glycopeptides reported here can be correlated with some of the glycopeptides characterized in previous investigations of ovomucoid. Glycopeptides with the sequence Asn(Carb.)-Thr and Thr-Asn(Carb.) were identified by Montgomery & Wu (1963) and by Monsigny *et al.* (1968). These sequences occur at attachment sites 4 and 2 respectively (Figs. 4 and 5). A glycopeptide Ala-Asn(Sac)-Thr, corresponding to the sequence at site 4, was also reported by Tanaka (1961). The sequence at site 1 resembles that of a glycopeptide Phe-Pro-Ala-Asn(Carb.) reported by Montgomery & Wu (1963) except that the position of Ala and Asn(Carb.) are transposed.

The amino acid sequences adjacent to four carbohydrate-acceptor sites all conform to the structural pattern Asn(Carb.)-X-Ser/Thr which has been proposed as a necessary requirement for glycosylation of asparagine (Neuberger & Marshall, 1968). However, the occurrence of cystine in the site-3 glycopeptide sequence Asn(Carb.)-Cys-Ser has not been reported and further extends the diversity of side chains found in this position.

The location of four of the carbohydrate-attach-

Site 1	Hen	<div>Carb.</div> <div>Phe-Pro-Asn-Ala-Thr-Asp-Lys-Glu-Gly-Lys</div> <div>10</div> <div>15</div>
	Quail	<div>Phe-Pro-Asn-Thr-Thr-Asn-Glu-Glu-Gly-Lys</div> <div><div>Carb.</div><div>Carb.</div></div> <div>75</div> <div>80</div>
Site 4	Hen	<div>Tyr-Ala-Asn-Thr-Thr-Ser-Glu-Asp-Gly-Lys</div> <div>75</div> <div>80</div>
	Quail	<div>Tyr-Pro-Asn-Thr-Thr-Ser-Glu-Asp-Gly-Lys</div> <div>Carb.</div>
Site 2	Hen	<div>Ala-Tyr-Ser-Ile-Glu-Phe-Gly-Thr-Asn-Ile-Ser-Lys</div> <div>45</div> <div>50</div> <div>55</div> <div>Carb.</div>
	Quail	<div>Phe-Tyr-Asn-Lys-Glu-Tyr-Gly-Thr-Asn-Ile-Ser-Lys</div> <div>Carb.</div>
Site 3	Hen	<div>Carb.</div> <div>Glu-Thr-Val-Pro-Met-Asn-Cys*-Ser-Ser</div> <div>65</div> <div>70</div>
	Quail	<div>Glu-Pro-Val-Thr-Met-Asp-Cys-Ser-Arg</div>

Fig. 6. Homologies in sequence between carbohydrate-acceptor sequences in quail (Laskowski, 1976) and hen ovomucoid. The numbering of residues is based on the sequence of quail ovomucoid (Laskowski, 1976). Abbreviation: Cys*, 2-aminoethylcysteine.

ment sites in ovomucoid has been determined by isolation of tryptic glycopeptides from CNBr fragments. Two of the glycosylated asparagine residues can be identified precisely as positions 69 and 75 in the intact protein in a region between the two methionine residues (68 and 84) at which cleavage occurs (Beeley, 1976). Two further glycosylated sites are located in the 68-residue *N*-terminal CNBr fragment. The sequence of the first 40 residues of ovomucoid has been determined by Murthy *et al.* (1973) using a sequenator. If a single amendment (Thr for Pro at position 12) is made to their sequenator results, the nine-residue glycopeptide T-4a can be placed in the sequence with carbohydrate-attachment site 1 at position 10 (Fig. 6). No sequence resembling that of the site-2 glycopeptide appears in the *N*-terminal 40 residues, indicating that the site-2 glycopeptide (T-3b), containing 12 amino acid residues, occurs in a portion of the molecule between residues 49 and 60. The site of origin of glycopeptide T-2d has not been directly established. However, it appears probable that it occurs in the glycosylated C-terminal portion of the molecule (Beeley, 1976), as the composition of the glycopeptide is not consistent with the sequence of either the *N*-terminal or central CNBr fragments.

The location of carbohydrate-attachment sites in hen's-egg ovomucoid can be compared with those reported to occur in quail ovomucoid (Laskowski, 1976). At least three carbohydrate groups (1, 2 and 4 of hen ovomucoid; see Fig. 6) are in closely similar or identical positions in both quail and hen glycoproteins. However, in quail ovomucoid, glycosylation does not occur in the sequence homologous to attachment-site 3 in hen ovomucoid. The glycosylated Asn-69 in hen ovomucoid is apparently replaced by Asp-69 in quail ovomucoid with consequent loss of carbohydrate attachment.

The occurrence of internal homologies in the sequence of hen ovomucoid was proposed on the basis of similarities between sequences adjacent to carbohydrate-attachment sites 1 and 4 (Beeley, 1974). Extensive sequence analysis of quail ovomucoid supports this conclusion and indicates the presence of three homologous domains (Laskowski, 1976). In both species carbohydrate is attached at internally homologous peptide sequences in two of the domains (Fig. 6). This repetition of glycosylation sites has presumably evolved as a consequence of partial gene duplication. This type of process may be of considerable significance in the evolution of proteins with multiple carbohydrate-attachment sites.

Further studies of the glycosylation sites of ovomucoids of different avian species will be of value in tracing the evolution of the ovomucoid gene (or genes) and in assessing the selective pressures that may be exerted in the retention or elimination of glycosylated sequences.

Fig. 7. Predicted β -turns and the sites of glycosylation in ovomucoid

Values of p_i for tetrapeptide sequences were calculated as described by Chou & Fasman (1974). $p_i = f_i \cdot f_{i+1} \cdot f_{i+2} \cdot f_{i+3}$ when $f_i, f_{i+1}, f_{i+2}, f_{i+3}$ represent the frequency of occurrence of particular amino acid residues at the first, second, third and fourth position of the β -turn. The beginning of a β -turn is predicted by the condition $p_i \geq 0.5 \times 10^{-4}$. Possible β -turns containing the four major sites of carbohydrate attachment are indicated.

A selective factor that has not previously been evaluated is that of the secondary structure of the peptide chain in the region of carbohydrate attachment. The location of α -helical, β -sheet and β -turn structures have been predicted with considerable success in several proteins from sequence data by the application of empirical methods (Chou & Fasman, 1974). Fig. 7 shows the result obtained when a test for the presence of β -turns was applied to the four major glycosylated sites of ovomucoid. Each of the carbohydrate-attachment sites lie in the same relation to a residue at which a four-residue β -turn would be predicted to begin with Asn(Carb.) at the third position in the turn. Thus if the prediction were correct the Ser/Thr residue in the sequence Asn(Carb.)-X-Ser/Thr would occur as the residue after the β -turn. However, some overprediction has been observed when this method is applied to proteins of known structure (Chou & Fasman, 1974). Thus it cannot be assumed that glycosylation sites necessarily occur close to β -turns. However, it is clear that in ovomucoid, a group of amino acids that occur with high frequency in β -turns are very close

to four of the sites of glycosylation. These residues may be significant in determination of the conformation of the peptide chain adjacent to the glycosylation site (whether or not this is a β -turn), which may be significant in the process of carbohydrate attachment. Peptide-chain conformation may also have a role in the evolution of carbohydrate-attachment sites. It is noteworthy that Bunting *et al.* (1972) have suggested from studies on immunoglobulins that preservation of predicted β -turn regions may be one mechanism by which selective pressure is exerted on chain conformation.

Skilful technical assistance was provided by Mrs. A. Lindsay and Mr. R. Blackie. I am grateful to Dr. G. Leaf for provision of facilities for amino acid analysis and to Dr. M. Laskowski for providing information on the sequence of quail ovomucoid before publication. This project was supported in part by a grant from the Medical Research Council.

References

- Ambler, R. P. (1972) *Methods Enzymol.* **25B**, 143–154
 Beeley, J. G. (1971) *Biochem. J.* **123**, 399–405
 Beeley, J. G. (1974) *Biochem. Soc. Symp.* **40**, 27–36
 Beeley, J. G. (1976) *Biochem. J.* **155**, 345–351
 Bunting, J. R., Athey, T. W. & Cathou, R. E. (1972) *Biochim. Biophys. Acta* **285**, 60–71
 Chambers, R. E. & Clamp, J. R. (1971) *Biochem. J.* **125**, 1009–1018
 Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
 Cole, R. D. (1967) *Methods Enzymol.* **11**, 315–317
 Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956) *Anal. Chem.* **28**, 350–356
 Feeney, R. G. & Allison, R. G. (1969) *Evolutionary Biochemistry of Proteins*, Wiley-Interscience, New York
 Fraenkel-Conrat, H. & Tsung, C. M. (1967) *Methods Enzymol.* **11**, 151–154
 Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822
 Konigsberg, W. (1972) *Methods Enzymol.* **25B**, 326–332
 Kostka, V. & Carpenter, F. H. (1964) *J. Biol. Chem.* **239**, 1799–1803
 Laskowski, M. (1976) *Protides Biol. Fluids Proc. Colloq.* **23**, in the press
 Lin, Y. & Feeney, R. E. (1972) in *Glycoproteins, their Composition, Structure and Function* (Gottschalk, A., ed.), 2nd edn., part B, pp. 762–781, Elsevier, Amsterdam
 Melamed, M. D. (1966) in *Glycoproteins, their Composition, Structure and Function* (Gottschalk, A., ed.), 1st edn., pp. 317–334, Elsevier, Amsterdam
 Monsigny, M., Adam-Chosson, A. & Montreuil, J. (1968) *Bull. Soc. Chim. Biol.* **50**, 857–866
 Montgomery, R. & Wu, Y. (1963) *J. Biol. Chem.* **238**, 3547–3554
 Murthy, G. S., Gilardeau, C., Chrétien, M. (1973) *Can. J. Biochem.* **51**, 1548–1551
 Neuberger, A. & Marshall, R. D. (1968) in *Carbohydrates and their Roles* (Schultze, H. W., Cain, R. F., Molstad, R. W., eds.), p. 115, Avi, Westport, CO, U.S.A.
 Neuberger, A. & Papkoff, H. (1963) *Biochem. J.* **87**, 581–587
 Offord, R. E. (1966) *Nature (London)* **211**, 591–593
 Smith, R. L. & Shaw, E. (1969) *J. Biol. Chem.* **244**, 4704–4712
 Tanaka, M. (1961) *J. Pharm. Soc. Jpn.* **81**, 1464–1467