

The Complete Amino-Acid Sequence of Hen Ovalbumin

Andrew D. NISBET, Richard H. SAUNDRY, Arthur J. G. MOIR, Linda A. FOTHERGILL, and John E. FOTHERGILL

Department of Biochemistry, University of Aberdeen

(Received September 12/November 27, 1980)

The complete amino acid sequence of hen ovalbumin, comprising 385 residues, has been determined. The sequence was deduced from the 17 cyanogen bromide fragments and from peptides derived by digestion with a number of proteolytic enzymes. The molecular weight of the polypeptide chain of ovalbumin is 42699.

Ovalbumin has four sites of postsynthetic modification; in addition to the acetylated N terminus, the carbohydrate moiety is located at Asn-292, and the two phosphorylated serines are at residues 68 and 344. The 'signal sequence' of ovalbumin is between residues 234 and 252. The heptapeptide released during the conversion of ovalbumin to plakalbumin by subtilisin digestion corresponds to residues 346–352. The hen ovalbumin polymorphism characterised by an Asn→Asp replacement results from a mutation at residue 311.

The amino acid sequence of ovalbumin deduced from these amino acid sequence studies is in complete agreement with the sequence of mRNA determined by McReynolds et al. [*Nature (Lond.)* 273, 723–728 (1978)].

Ovalbumin is the most abundant protein in egg white, and is readily purified and crystallised in gram quantities [1]. Its function is not clear and remains a matter for speculation. Ovalbumin is a monomer with a molecular weight of 45000 as determined by physical methods [2]. Its amino acid composition [3,4] is unremarkable, although the thiol and disulphide contents have proved difficult to establish, perhaps due to disulphide interchange. It is now apparent that there are four cysteines and one cystine [5], and peptides containing these residues have been purified and sequenced [6]. The N terminus of the protein is blocked by an acetyl group [7]. The C-terminal sequence has been shown to be Cys-Val-Ser-Pro-COOH [5,8].

Treatment of ovalbumin with subtilisin under suitable conditions causes the formation of a new crystalline form of ovalbumin, plakalbumin [9]. A heptapeptide, Glu-Ala-Gly-Val-Asp-Ala-Ala is released by this treatment [10]. The resulting two chains of plakalbumin can be separated by gel filtration in urea, and the smaller, 33-residue chain corresponding to the C-terminal portion of ovalbumin has been sequenced [11].

Ovalbumin has three sites of postsynthetic modification in addition to the N-terminal acetyl group. The glycoprotein has a single carbohydrate moiety comprising a core of two *N*-acetyl-glucosamine and four mannose units with variable numbers of additional residues of the same sugars (e.g. [12]). The carbohydrate moiety is attached to an asparagine residue, and the sequence of a glycopeptide [13] shows it to contain the Asn-X-Thr 'recognition site'. Hen ovalbumin is known to contain two phosphoryl groups [14] that are attached to serine residues [15]. Milstein [16] has purified two phospho-

serine peptides by a diagonal electrophoresis method and has determined the amino acid sequences of these peptides. Amino acid sequence studies on phosphoserine peptides from ovalbumins from a number of species indicate that the 'recognition site' for phosphorylation is Ser-X-Glu [17].

A genetic variant of hen ovalbumin characterised by a change in electrophoretic mobility [18] has been shown to result from an Asn→Asp replacement [19].

The synthesis of ovalbumin by oviduct tissue has proved to be a suitable system for studying factors affecting the rate of protein synthesis [20], including the influence of steroid hormones [21]. Interest in this system has led to the isolation [22,23] and sequence determination of ovalbumin mRNA [24]. The isolation of the mRNA has in turn led to the description of 'split genes' in hen DNA [25]. Ovalbumin mRNA has been translated in a cell-free system using conditions that prevented acetylation of the N terminus [26], and the amino acid sequence of the N-terminal 35 residues has been determined by automatic Edman degradation. These experiments also showed that ovalbumin does not have a transient hydrophobic leader sequence. It has subsequently been shown that ovalbumin appears to have an internal 'signal sequence' [27].

The present paper describes the sequence determination of the complete ovalbumin molecule, and allows all the previous work to be drawn together in the context of a single structure.

MATERIALS AND METHODS

Materials

Ovalbumin was purified from single egg whites by chromatography on CM-cellulose followed by precipitation by 60% saturation with $(\text{NH}_4)_2\text{SO}_4$. Details of these procedures and the source of the egg white have been reported elsewhere [5]. The ovalbumin was either oxidised by performic acid [16] or reduced and carboxymethylated [5].

Alkaline phosphatase, chymotrypsin, pepsin and trypsin were obtained from Worthington Biochemical Corporation (Freehold, NJ, USA). Subtilisin (protease VII) was from

Abbreviations. Dns, dansyl, 5-dimethylaminonaphthalene-1-sulphonyl; Cm-Cys, *S*-carboxymethylcysteine; Cys(O_3H), cysteic acid; Met(O_2), methionine sulphone; $m_{6.5}$, m_2 , electrophoretic mobilities at pH 6.5 and pH 2, respectively.

Enzymes. Trypsin (EC 3.4.21.4); chymotrypsin (EC 3.4.21.1); pepsin (EC 3.4.23.1); thermolysin (EC 3.4.24.4); subtilisin (EC 3.4.21.14); serine carboxypeptidase (or carboxypeptidase Y) (EC 3.4.16.1); staphylococcal serine proteinase (EC 3.4.21.19); alkaline phosphatase (EC 3.1.3.1).

Sigma Chemical Company (St Louis, MO, USA). Thermolysin was a gift from Dr R. P. Ambler. Staphylococcal serine proteinase was from Miles Laboratories Ltd (Stoke Poges, Slough, UK). Carboxypeptidase Y, phenylisothiocyanate, triethylamine and trifluoroacetic acid were from Pierce Chemical Co. (Rockford, IL, USA) or from Rathburn Chemicals (Walkerburn, Peebleshire, UK). Cyanogen bromide was supplied by Eastman Kodak Co. (Rochester, NY, USA). Iodo[2-¹⁴C]acetic acid (40–60 Ci/mol) was from the Radiochemical Centre (Amersham, UK). All other chemicals were from BDH Chemicals (Poole, Dorset, UK) and were Analar or Aristar grade. Analar pyridine was redistilled after refluxing with ninhydrin (2 g/l) for 1 h.

Overall Strategy and Methods of Cleavage

Fragments and smaller peptides of ovalbumin were generated as indicated in Fig. 1 and were purified by gel filtration, ion-exchange chromatography and electrophoresis. Cyanogen bromide cleavage was done at room temperature (15–20 °C) for 24 h in 70% (v/v) formic acid at a 50-fold molar excess of CNBr over methionine residues, and the resulting fragments purified on Sephadex G-50 (Fig. 2).

The ovalbumin or its cyanogen bromide fragments were digested at a concentration of approximately 2 mg/ml with enzyme (1:50 by weight) at 37 °C for 12–16 h in 1% (w/v) NH₄HCO₃, pH 7.8 for trypsin, chymotrypsin, subtilisin and staphylococcal serine proteinase, and in 0.01 M HCl for pepsin. Thermolysin digestions were done in CH₃COONH₄ (0.2 M with respect to acetate), pH 8.5, containing 0.05 M CaCl₂ [28]. Carboxypeptidase Y digests were done according to Kuhn et al. [29]. Subdigestions of peptides were performed using 20–30 µg enzyme/100 nmol substrate at a concentration of 100–200 nmol/ml for 1–16 h at 37 °C. After these enzymic treatments the peptide mixtures were fractionated or freeze-dried immediately. Dephosphorylation of peptides using alkaline phosphatase was done on paper strips according to Milstein [16].

Peptide Purification

Peptide mixtures were fractionated by gel filtration using Sephadex G-25 or by cation-exchange chromatography using Technicon P resin. For the latter method, the freeze-dried peptides from the enzyme digests or from the gel filtration of the CNBr fragments (Fig. 2) were dissolved in 2 ml 0.2 M pyridine and the pH adjusted to 2–3 with 98% (w/w) formic acid. The sample was applied to a column (0.6 × 150 cm) that had been equilibrated with 0.2 M pyridine acetate, pH 3.1. Elution was at 40 °C at a flow rate of 48 ml/h with a gradient of pyridine and acetic acid using a nine-chambered Technicon Autograd (see Fig. 4). Approximately 10% of the effluent was

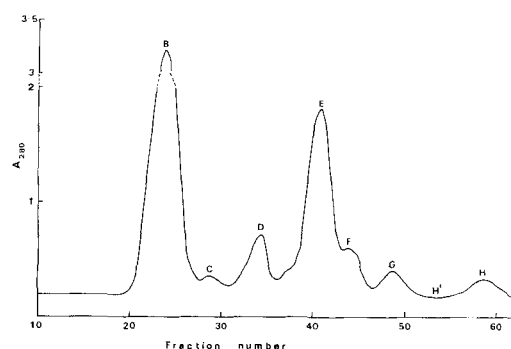


Fig. 2. Separation of CNBr fragments on Sephadex G-50. Reduced, carboxymethylated ovalbumin (1 µmol, 45 mg) that had been cleaved with CNBr (see Materials and Methods) was loaded onto a column of Sephadex G-50 (2 × 130 cm) run in 50% formic acid at room temperature with a hydrostatic head of 50 cm (flow rate approx. 8 ml/h). Fraction size was 2 ml.

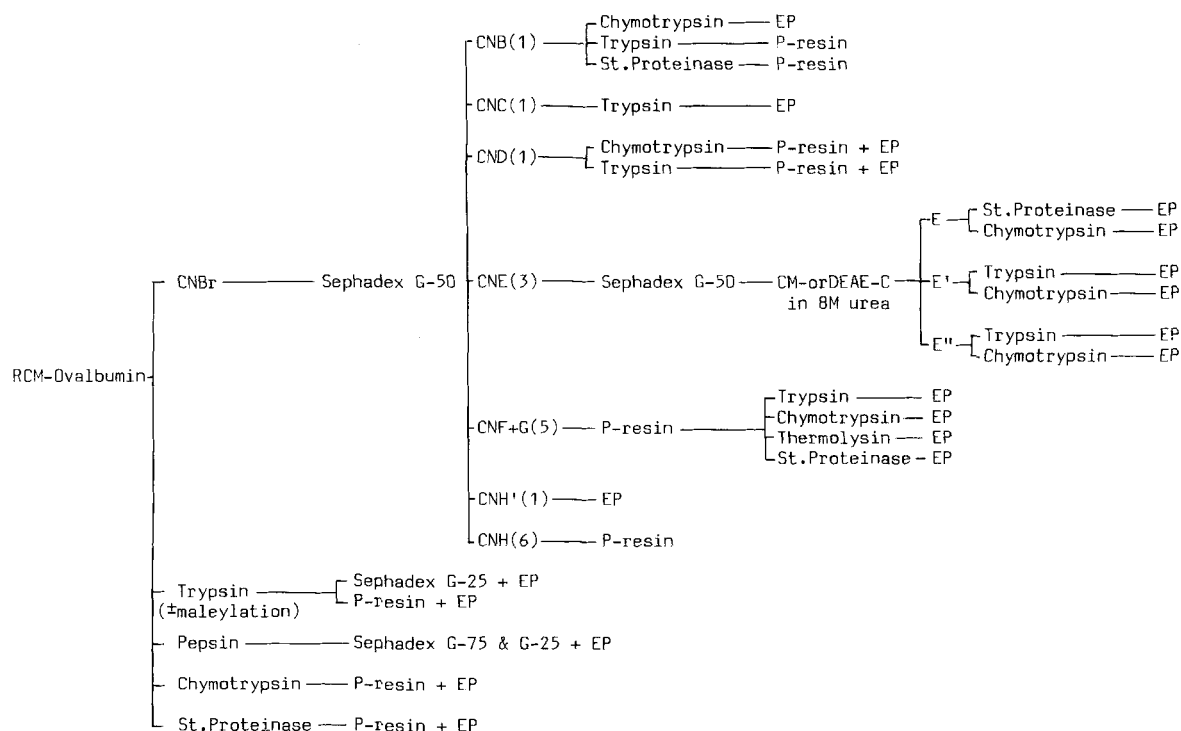


Fig. 1. General scheme showing methods of cleavage and separation of peptides. Details of methods are given in the text. RCM = reduced and carboxymethylated, EP signifies electrophoresis. The numbers in brackets indicate the number of peptides purified from the pool.

diverted for continuous detection of peptides with ninhydrin reagent, and the remainder was collected in 2.5-ml fractions. After completion of the run, the resin was removed from the column and washed with two 25-ml samples of ammonia solution (specific gravity 0.88)/pyridine (1:1, v/v). The washes were freeze-dried and examined for the presence of peptides after high-voltage paper electrophoresis.

Further purification of peptides was carried out when necessary by high-voltage electrophoresis at pH 6.5, 3.5 or 2 [4,17]. The voltage gradient was 60 V/cm. Peptides were detected by staining with cadmium/ninhydrin reagent [30] or chlorine stain [31]. At pH 6.5, peptide mobility ($m_{6.5}$) was expressed relative to aspartic acid, with valine as a neutral marker. At pH 2 mobility (m_2) was measured relative to valine with dansyl-OH as a neutral marker.

The amide content of a peptide was determined by correlating its electrophoretic mobility at pH 6.5 with its molecular weight [32]. If the amide position was still ambiguous, the electrophoretic mobility of the peptide was determined after the appropriate number of rounds of Edman degradation.

The following specific stains were used for the initial characterisation of peptides: Ehrlich stain for tryptophan [33], Pauly stain for histidine and tyrosine [34] and Sakaguchi stain for arginine [35].

Peptides were eluted from paper strips with 1% acetic acid for basic peptides, or with 1% NH_4OH for acidic and neutral peptides.

Peptide Nomenclature

Upper case letters indicate that the peptide was derived from digestion by: C, chymotrypsin; CN, CNBr; P, pepsin; S, subtilisin; St, staphylococcal serine proteinase; Th, thermolysin; T, trypsin. Numerals or upper case letters which immediately follow these letters refer to the peak number from the ion-exchange or gel filtration fractionation. Lower case letters refer to the electrophoretic mobility at pH 6.5, with 'a' being the most acidic. Numerals refer to electrophoretic mobility at pH 2 with '1' being the most acidic.

Amino Acid Analysis

Hydrolysis was done in evacuated tubes with 0.5 ml 18% HCl (Aristar HCl diluted with an equal volume of water) at 110°C usually for 20 h, although occasionally for 72–96 h to check for the presence of peptide bonds resistant to hydrolysis. Quantitative analyses were done using a Locarte amino acid analyser with a single column eluted by a 3.5-h sequence of four stepwise buffer changes. Tryptophan was estimated by reaction with Ehrlich reagent [33].

Amino Acid Sequence Determination

The amino acid sequences were determined using the dansyl-Edman method as described by Hartley [36], except that 2% (v/v) triethylamine was incorporated in the 50% (v/v) pyridine. The dansylated amino acids were identified by two-dimensional thin-layer chromatography [37] on polyamide plates (5 × 5 cm). Dns-Pro was determined after hydrolysis of the dansyl-peptide for 4 h instead of the usual 16 h. Dns-Arg, Dns-His and ϵ -Dns-Lys were separated by using ethanol/0.05 M Na_3PO_4 (1:3, v/v) [36] after the first three solvents. Confirmation of the sequence results was sometimes obtained using subtractive Edman determinations.

RESULTS AND DISCUSSION

Cyanogen Bromide Fragments

Ovalbumin has 16 methionine residues [5]; Fig. 2 shows the elution profile of the CNBr fragments on Sephadex G-50. Calibration of the column with suitable standard proteins (IgG, myoglobin, cytochrome *c* and myoglobin CNBr fragments) allowed the molecular weights of the fragments to be estimated. The amino acid compositions of all 17 CNBr fragments and their positions in the sequence are shown in Table 1. In all cases there is reasonable agreement between the compositions determined by amino acid analysis and the numbers of each residue shown by sequence determination.

Peak B corresponds to one fragment (CNB) with an approximate molecular weight of 13500. Trypsin, chymotrypsin and staphylococcal serine proteinase digestions of CNB were done and the resulting peptides separated by ion-exchange chromatography using Technicon P resin. Fragment CNC was further purified by rerunning the appropriate fractions from peak C through the Sephadex G-50 column. CNC has a molecular weight of approximately 9000 and the amino acid analysis does not contain homoserine, suggesting that the fragment is derived from the C terminus of ovalbumin. A trypsin subdigest of CNC was done and the peptides separated by electrophoresis. Fragment CND was also further purified by rerunning through the Sephadex G-50 column. Its amino acid composition shows the presence of methionine, and it is now apparent that the fragment corresponds to fragments CNE and CNF20; the Met-Glu peptide bond between residues 273 and 274 is partially resistant to CNBr cleavage. CND was treated with chymotrypsin and staphylococcal serine proteinase, and the resulting peptides separated by ion-exchange chromatography using Technicon P resin.

Peak E contains three fragments (CNE, CNE' and CNE'') that were purified by chromatography on CM-cellulose in 7.9 M urea (Fig. 3). The breakthrough peak contained CNE and a carbohydrate-containing fragment (also present in peak F) that were separated by chromatography on DEAE-cellulose in urea. Fragment CNE was digested with chymotrypsin and staphylococcal serine proteinase, CNE' and CNE'' with trypsin and chymotrypsin; the resulting peptides were separated by ion-exchange chromatography on Technicon P resin and by electrophoresis.

Fractions corresponding to peaks F and G were pooled and the fragments purified by ion-exchange chromatography on Technicon P resin. Fragments CNF1, CNF10, CNF13, CNF19 and CNF20 were all isolated by this procedure. Subdigests of these fragments were done using trypsin, chymotrypsin, thermolysin or staphylococcal serine proteinase, and the peptides separated by electrophoresis.

Peak H was pooled, and the fragments CNH3, CNH4, CNH10, CNH17 and CNH18 purified by ion-exchange chromatography on Technicon P resin. CNH10 was a mixture of two dipeptides, Ala-Hse and Ser-Hse.

Fragment CNH'1 was purified from the fractions between peaks G and H by electrophoresis at pH 2. This fragment does not absorb at 280 nm and is ninhydrin-negative although it does give a positive reaction with chlorine stain. It corresponds to the N-terminal eight residues of ovalbumin.

Proteolytic Digests of Whole Ovalbumin

Peptides derived from proteolytic digests of whole ovalbumin were used to complement information from the CNBr

Table 1. Amino acid composition of CNBr fragments. The results are expressed as mol / mol fragment, and are uncorrected for destruction of Ser or Thr, or for incomplete cleavage next to hydrophobic residues. (-) means insignificant amount. Values in brackets are those found in the sequence. The results for ovalbumin were taken from Fothergill & Fothergill [5]

	Ovalbumin	Sum of CNBr fragments	CNB	CNC	CNF	CNE ¹	CNE ^{II}	CNF1	CNF10
Pro	6.1(6)	5.4(6)	2.1(2)	1.8(2)	-	1.5(2)	-	-	-
Asx	51.7(51)	29.1(51)	11.6(15)	6.5(6)	3.1(3)	3.3(3)	3.5(3)	1.1(1)	-
Thr	15.0(15)	15.2(15)	5.2(2)	2.3(2)	1.7(2)	-	1.0(1)	1.0(1)	1.0(1)
Ser	36.9(38)	33.4(38)	11.2(13)	9.9(12)	3.3(4)	-	-	1.1(1)	1.0(1)
His	- (16)	14.0(16)	0.9(1)	-	0.7(1)	0.9(1)	0.6(1)	1.0(1)	0.9(1)
Glx	50.9(48)	49.3(48)	16.7(17)	9.5(8)	5.8(6)	3.3(3)	3.5(3)	1.8(2)	1.1(1)
Pro	15.5(14)	14.9(14)	5.9(6)	2.5(2)	1.3(1)	1.3(1)	-	-	1.0(1)
Gly	18.5(19)	21.7(19)	8.3(7)	7.0(6)	1.1(1)	-	1.1(1)	-	1.0(1)
Ala	54.3(55)	56.5(55)	10.3(8)	12.1(14)	-	2.5(2)	3.3(3)	-	1.0(1)
Val	29.6(31)	27.7(31)	9.4(10)	6.4(8)	2.5(3)	1.7(2)	2.3(3)	1.1(1)	-
Ile	13.3	-	-	-	-	-	-	-	-
Ile	24.6(25)	21.9(25)	8.5(9)	5.3(6)	1.4(2)	2.6(3)	0.8(1)	-	0.8(1)
Leu	32.1(32)	51.8(32)	12.2(12)	4.8(4)	4.9(6)	1.2(1)	1.9(2)	2.0(2)	1.8(2)
Tyr	9.8(10)	10.3(10)	6.0(6)	0.8(0)	-	0.9(1)	-	0.9(1)	-
Pro	20.0(20)	21.6(20)	5.9(4)	5.3(6)	1.2(1)	3.7(4)	2.2(2)	-	1.0(1)
His	8.0(7)	5.7(7)	0 (1)	4.0(4)	-	1.7(2)	-	-	-
Lys	19.1(20)	20.7(20)	5.6(5)	2.5(2)	0.9(1)	1.8(2)	2.4(3)	1.0(1)	0.8(1)
Arg	16.7(15)	16.7(15)	9.1(8)	3.7(3)	-	-	-	-	-
Trp	3.0(3)	3.0(3)	1.0(1)	-	+ (1)	-	+ (1)	-	-
Sum	385 (385)	376.9(385)	129.9(132)	84.4(85)	28.9(32)	26.4(27)	23.6(24)	11.0(11)	11.4(12)
N terminus	As-Gly		Val	Gly	Leu	Glu	Val	Glu	Lys
Sequence position			41-172	301-385	242-273	9-35	173-196	288-298	228-239

a

	CNE ¹	CNE ^{II}	CNE ^{III}	CNF1	CNF3	CNF4	CNF10	CNF17	CNF18
-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
1.0(1)	-	-	-	-	-	-	-	-	-
1.1(1)	1.0(1)	-	2.0(2)	0.8(1)	-	1.0(1)	1.0(1)	-	-
1.0(1)	0.9(1)	0.7(1)	0.5(1)	1.3(1)	1.0(1)	1.9(2)	0.8(1)	0.9(1)	-
3.3(4)	1.2(1)	2.0(2)	-	-	-	-	1.1(1)	-	-
1.9(2)	-	1.0(1)	-	-	-	-	-	-	-
-	1.1(1)	-	2.1(2)	-	-	-	-	-	-
-	1.0(1)	-	2.0(2)	2.2(2)	-	1.2(1)	0.9(1)	-	-
1.9(2)	1.1(1)	1.3(1)	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
-	0.9(1)	0.8(1)	0.8(1)	-	-	-	-	-	-
-	1.0(1)	1.0(1)	-	1.0(1)	-	-	-	-	-
-	1.0(1)	0.7(1)	-	-	-	-	-	-	-
1.2(1)	1.1(1)	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
1.5(1)	-	2.1(2)	-	-	-	-	1.0(1)	1.1(1)	-
0.9(1)	0.9(1)	2.1(2)	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-
14.8(14)	11.2(11)	11.7(12)	7.4(8)	5.3(5)	1.0(1)	4.1 ⁽²⁾ (2)	4.8(5)	2.0(2)	-
Pro	Tyr	Glu	As-Gly	Ser	Ile	Ser Ala	Ala	Lys	
192-210	212-222	275-285	1-8	36-60	211	240-241 299-300	223-227	296-287	

b

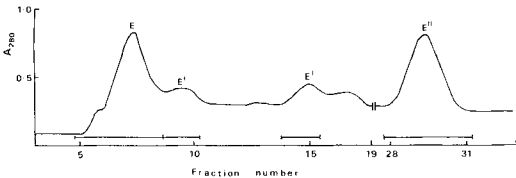


Fig. 3. Separation of fragments CNE, CNE' and CNE'' on CM-cellulose. Peak E from the CNBr fragment separation (Fig. 2) was freeze-dried, and dissolved in 1.5 ml of acetate buffer (0.05 M, pH 5.26), 7.9 M urea. The sample was applied to a column of CM-cellulose (0.9 x 32 cm) run in the same buffer at room temperature. The flow rate was 10 ml/h, and 4-ml fractions were collected. Fractions 5-8, 9-10, 14-15 and 28-31 were each pooled and desalted into 50% formic acid using a Sephadex G-25 column. Pools 9-10 and 14-15 showed the same amino acid analysis, and the separation was presumably due to the different charged forms of homoserine

fragments and to provide evidence for overlaps. Four enzymes were used: chymotrypsin, pepsin, staphylococcal serine proteinase and trypsin. Approximately 1-6 μmol ovalbumin was digested by each of the proteolytic enzymes, and Fig. 4 shows the elution profile for the fractionation of the chymotryptic peptides by ion-exchange chromatography on Technicon P resin. Comparable results were obtained from the staphylococcal serine proteinase and trypsin digests. In addition, a pepsin digest and a second trypsin digest were fractionated on Sephadex G-25 followed by electrophoresis.

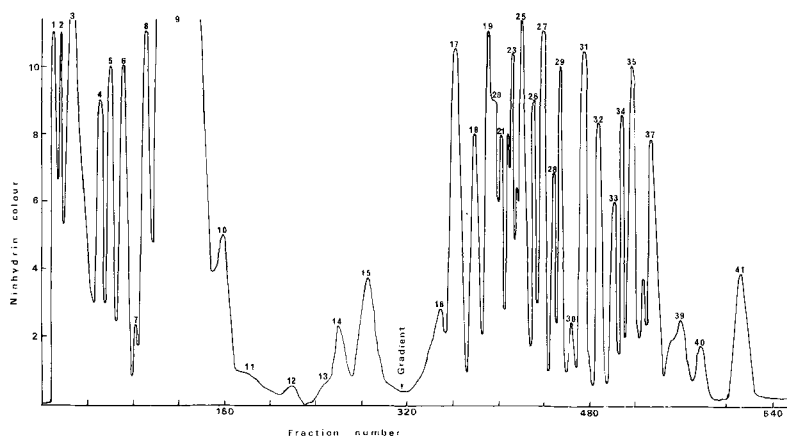


Fig. 4. Elution profile of peptides from chymotrypsin digest of ovalbumin on Technicon *P* resin. About 6 μ mol of reduced, carboxymethylated ovalbumin that had been digested with chymotrypsin was dissolved in about 2 ml 0.2 M pyridine, and the pH adjusted to 3 with formic acid. The sample was loaded onto a column (0.9 $\times$ 150 cm) of Technicon type *P* chromobead resin maintained at 40 $^{\circ}$ C with a water jacket. The flow rate was 36 ml/h and approximately 10% of the effluent was mixed with ninhydrin reagent to detect peptides. The following three buffers were used: (A) 0.2 M pyridine, pH 3.1 (16.1 ml pyridine + 280 ml acetic acid/l); (B) 2.0 M pyridine, pH 5.1 (161 ml pyridine + 143 ml acetic acid/l); (C) 2.0 M pyridine pH 6.5 (161 ml pyridine + 4 ml acetic acid/l). The column was eluted with buffer A until tube 300 when a gradient formed by a nine-chambered Technicon Autograd was used. The Autograd chambers were filled as follows: (1) 162 ml A; (2) 162 ml A; (3) 99 ml A + 27 ml B + 36 ml water; (4) 9 ml A + 81 ml B + 68 ml water; (5) 157 ml B; (6) 157 ml B; (7) 117 ml B + 40 ml C; (8) 157 ml C; (9) 157 ml C.

Sequence Determination

The complete sequence is shown in Fig. 5; the CNBr fragments, the major peptides derived by proteolytic digestion and the results of corresponding sequential degradations are indicated. Amino acid compositions and electrophoretic mobilities of important peptides are given in Table 2. Evidence for assignment of amides is shown in Table 3.

Abundant evidence was obtained to establish unambiguously virtually all of the sequence. For the majority of the residues at least two different lines of evidence were used and in some cases many more. For two overlaps, however, because of the nature of the sequence and the specificity of CNBr cleavage and that of the proteolytic enzymes, relatively limited information was available. The evidence used to establish these overlaps will be described below. In addition, those regions for which sequences had previously been published were sometimes only confirmed by peptide compositions in this study.

The largest CNBr fragment, CNB (together with the five-residue fragment CNH3 shown to be N-terminal to CNB) is the only fragment with inadequate overlap information. All the other fragments have been overlapped into two sections. One section has been unambiguously assigned to the N terminus and the other to the C terminus. CNB and CNH3 must therefore link these two sections. This interpretation was confirmed when the sequence of ovalbumin mRNA [24] became available in the final stages of the amino acid sequence determination.

There is no sequence evidence from this study for the Met-Met sequence at residues 210-211 although free homoserine was detected (CNH4) from the CNBr digest. Amino acid sequence work on turkey ovalbumin, however, has led to the isolation of a tryptic peptide with the following sequence: Val-Thr-Glu-Gln-Glu(Ser,Lys)Pro-Val-Gln-Met-Met-Tyr(Gln,Ile,Gly,Leu)Phe-Lys [38]. This peptide is obviously homologous with residues 200-218 of hen ovalbumin, and supports the proposed sequence. Once again, the mRNA sequence provided confirmation.

The Structure of Ovalbumin

Although ovalbumin has been studied over many years with an extensive array of physico-chemical techniques,

relatively little is known of its secondary and tertiary structures and no X-ray crystallographic work is available. The determination of the amino acid sequence, however, provides indirectly additional information concerning the topology of the molecule.

The single carbohydrate moiety of ovalbumin is attached to Asn-292 (and not residue 10 as previously alleged [39]) since fragment CNF1 is a glycopeptide and since the tryptic peptide purified by Furka and Sebestyen [13] clearly corresponds to residues 291-322. The Asn-Leu-Thr 'recognition site' is probably situated on the surface of ovalbumin to allow access by the glycosyltransferase(s) and to permit a suitable hydrophilic environment for the carbohydrate. It is interesting to note that ovalbumin has a second potential 'recognition site' for glycosylation: Asn-Leu-Ser (311-313). It is thus an example of an unusual type of glycoprotein containing two sites of the Asn-X-Ser/Thr type, where only one of the sites is glycosylated (see Hunt and Dayhoff [40]) and is in marked contrast to the situation in pig ribonuclease where all three of such sites carry carbohydrate [41]. Moreover, it is unlikely that the second non-glycosylated site in ovalbumin is buried because this site contributes to the antigenic determinants. The genetic polymorphism of ovalbumin that has been shown to involve an Asn $\rightarrow$ Asp replacement [19] is due to a mutation at Asn-311. The two forms of ovalbumin differing only at this residue can be distinguished immunochemically using quantitative precipitin assays [19].

Comparing the sequences of the phosphoserine peptides determined by Milstein [16] with the complete sequence shows the phosphoryl groups to be attached to Ser-68 and Ser-344. As in the case of the glycosylation site, these sites of post-synthetic modification are probably situated on the surface of the ovalbumin molecule. From a comparison of sixteen phosphoserine peptide sequences from ovalbumins from eight species it is notable that a glutamic acid residue always occurs two residues toward the C terminus from the phosphorylated serine [17]. Caseins also have phosphorylated serines, and in α_1 casein and β casein the phosphoserines are always followed in the $n+2$ position by glutamic acid or another phosphoserine [42,43]. It is noteworthy that all Ser-X-Glu (or phosphoserine) sequences in the caseins are phosphorylated, whereas there are three such sequences in ovalbumin that are not: Ser-Met-Glu (7-9), Ser-Ala-Glu

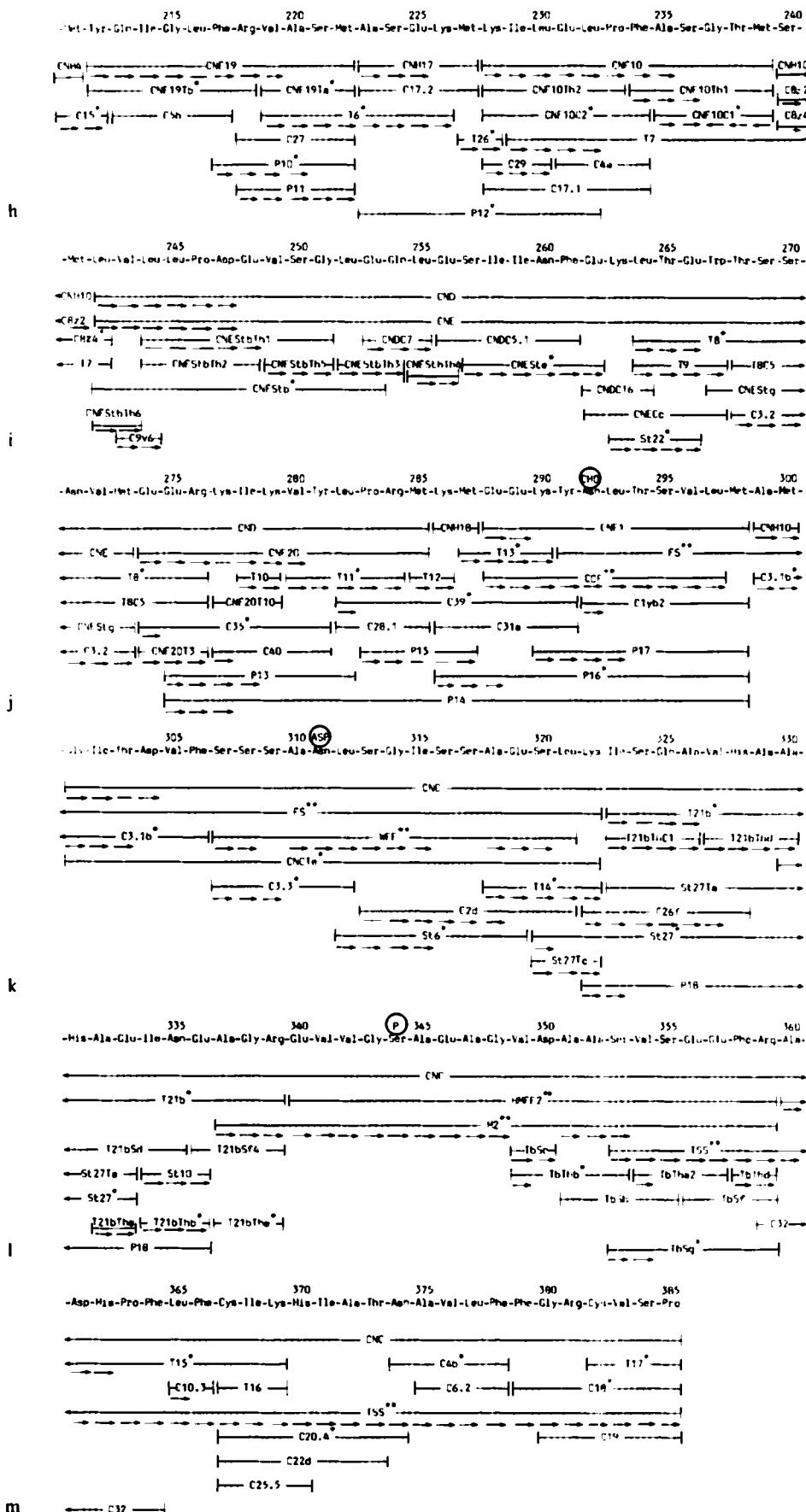


Fig. 5. Amino acid sequence of ovalbumin. Residues were identified after Edman degradation (—) or by carboxypeptidase Y treatment (---). Rounds of degradation for which no residue was identified are indicated (---). Peptide nomenclature and details of the methods are given in the text. The amino acid compositions and peptide mobilities of those peptides indicated by (*) are given in Table 2. Some previously published sequences are indicated by (**) and are taken from the following references: PGW [26], HMEF1 [17], M1 [16], CCF [13], FS [46], WFF [19], M2 [16], HMEF2 [17], and TSS [11]. The positions of the phosphoserines are indicated by P in a circle; the carbohydrate by CHO in a circle, and the genetic variant by ASP in a circle.

(317–319) and Ser-Glu-Glu (355–357). In these cases it is unlikely that the 'middle' residue has prevented phosphorylation since phosphoserine-344 is followed by an alanine (as it is in all eight ovalbumins) and also peptides from Chinese goose and Aylesbury duck ovalbumins are both phosphorylated at Ser-Met-Glu sequences [17].

Residues 346–352 probably occupy a particularly exposed position since only these residues are released by digestion of ovalbumin with subtilisin under suitable conditions [10]. The region is very close to the phosphoserine-344. This portion of the sequence may also be implicated in intermolecular interactions in the crystal since release of the hepta-

peptide causes ovalbumin to form plate-like crystals instead of the usual needles [9].

Thompson and Fisher [6] have sequenced peptides containing the four thiols and one disulphide, and it is now apparent that the residues involved in the disulphide bridge are Cys-73 and Cys-120, although there appears to be the possibility of disulphide interchange with Cys-382 [5]. It may be that disulphide interchange is involved in the conversion of the newly-synthesized 'R-ovalbumin' to the more stable 'S-ovalbumin' on storage [44].

Ovalbumin has been shown to contain an internal 'signal sequence' involved in membrane translocation, which is

Table 7. Data for peptides important in establishing the sequence. Experimental details and peptide nomenclature are given in the text. Values are molar ratios, and are uncorrected for destruction or partial hydrolysis. (–) means insignificant amount in the composition, or not done for mobility and N-terminal determination. (—) means no N-terminal residue detected.

Peptide	CNH ¹ 1Th1	CNH ¹ 1Th2	C21	CNF ¹ 14a	St17	121a	St32	C25 .2	CNF ¹ C3	C6.3	C9v2	P3	C21 .1	CNH ¹ 112	CNH ¹ 110a	CNH ¹ 110b
m _{6,5}	—	—	—	–0.45	–0.47	0	+0.29	—	—	—	0	—	—	+0.50	0	+0.40
m ₂	0	+0.65	0	+0.76	—	+1.25	—	+0.82	—	+0.87	+0.69	+1.04	+0.99	+1.44	+1.26	+1.70
Amino acid																
Cys-Cys	—	—	—	—	0.9	—	—	—	0.9	—	—	—	—	—	—	—
Asx	—	—	—	1.3	1.2	—	1.2	2.1	—	—	—	1.6	2.1	—	1.1	1.2
Thr	—	—	—	—	—	—	—	—	—	—	—	1.8	2.1	—	0.9	0.7
Ser	0.9	0.9	1.8	—	—	—	—	—	0.9	—	—	1.1	1.1	—	1.0	—
His	—	1.5	—	—	—	—	—	—	0.8	—	—	—	—	—	—	—
Glx	—	—	1.2	1.4	1.4	1.2	1.4	1.1	—	—	—	1.0	1.1	—	—	1.2
Pro	—	—	—	—	—	—	—	—	1.1	—	—	—	—	—	—	—
Gly	1.1	1.1	2.0	—	—	—	—	—	—	—	—	0.9	1.0	1.2	—	—
Ala	—	2.0	1.8	—	—	—	1.1	1.1	1.1	1.1	1.0	1.0	1.1	1.0	—	—
Val	—	—	—	1.0	0.9	—	0.9	—	—	—	1.1	1.0	—	—	—	—
Met	—	—	0.6	—	—	—	—	—	—	—	0.9	—	—	—	—	—
Ile	—	0.9	1.1	—	—	—	—	0.9	1.9	—	—	1.1	1.0	—	—	—
Leu	—	—	—	—	—	0.8	1.0	—	—	1.0	—	1.0	0.6	0.9	—	0.9
Tyr	—	—	—	—	—	—	—	1.0	—	—	1.0	0.8	—	—	—	—
Phe	—	—	1.1	2.4	2.6	—	—	0.9	—	—	—	—	—	—	—	—
His	—	—	—	—	—	—	1.6	0.9	—	—	—	—	—	—	—	—
Lys	—	—	—	0.9	0.9	1.0	0.9	—	—	—	—	1.0	0.9	0.9	—	1.0
Arg	—	—	—	—	—	—	—	—	—	—	—	1.0	1.1	—	1.1	—
Trp	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Sum	2	6	10	8	8	3	8	8	6	3	4	14	12	4	4	5
a N terminus	—	Ile	—	—	Phe	Glx	Leu	His	Cys	Ser	Ala	Val	Leu	Leu	Asx	Thr

C41	CNB 118	C1v2 9a	C1v2 5b	P4	C20 12b	12 9e2	C9v3	C925	CNB 114e	C31b	CNB 114b	St24	CNB 118.2	St16	CNB 119
+0.69	+0.51	+0.66	+0.47	+0.76	0	–0.20	0	0	+0.41	+0.38	–0.20	0	—	–0.27	0
+1.60	+1.44	—	—	+1.07	+0.95	+0.55	+0.87	+0.92	+1.21	+1.19	+0.98	+0.90	+0.69	—	+1.13
—	—	—	0.3	—	—	—	—	—	—	—	—	—	1.3	1.3	—
—	—	2.3	1.2	0.8	1.1	3.5	2.9	—	—	—	—	—	—	—	—
—	—	—	0.9	—	—	1.0	1.2	—	—	—	—	—	—	—	—
—	—	3.1	1.0	2.0	1.9	3.0	—	1.0	1.9	1.1	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	1.3	1.1	—	—	1.3	1.1	—	—	—	2.4	1.4	2.2	2.3	1.1
—	—	1.1	—	—	—	1.1	1.1	—	—	—	—	1.9	1.9	—	—
—	—	1.7	1.1	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	0.7	—	—	1.1	—	—	1.1	0.9	1.1	—	—	—	—
1.9	1.1	—	1.0	1.2	—	1.3	1.0	—	—	—	—	—	0.9	1.0	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	0.8	—	—	1.0	2.0	0.9	—	—	—	—	0.9	1.0	—	—
—	—	0.8	—	1.2	1.9	2.1	—	—	1.0	1.0	0.6	1.8	2.1	1.0	1.0
2.0	—	0.7	—	—	—	0.9	—	1.0	—	0.8	0.9	0.9	1.8	0.6	1.0
—	—	—	—	1.2	—	—	—	—	—	—	—	—	—	—	—
1.9	—	0.6	—	—	—	1.1	1.0	—	—	—	—	—	0.7	0.8	—
1.0	1.0	—	—	—	1.1	0.6	—	—	1.0	1.1	1.0	1.0	—	—	0.9
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5	5	11	8	7	7	20	10	2	5	5	6	7	12	7	4
Lys	Val	—	Ala	Val	Ser	Asx	Asx	Ser	—	Ala	Leu	Arg	Tyr	Tyr	Glx

b

C11	T4	C21	C31	St13	C11T	P8	C11T	C11T	C11T	C11T	T3	C37	C11T	St29	T5
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-	-0.22	-	-	-0.25	0	0	-	-0.42	-	-	+0.37	+0.67	-	0	-0.20
+0.80	+0.65	+0.45	-	-	-	+0.69	+0.56	-	-	-	+1.33	+1.22	+1.05	+1.49	+1.02
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.1	1.8	1.2	2.3	1.1	1.2	1.9	1.1	1.0	1.2	-	-	-	-	1.1	1.9
-	1.5	1.1	1.1	-	0.9	0.9	-	1.0	-	-	-	-	-	-	1.1
-	-	-	1.8	1.1	0.8	2.7	1.0	1.1	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.1	1.1	1.0	2.8	1.2	1.1	2.0	1.2	0.9	-	-	-	-	1.1	1.2	2.2
0.9	1.2	-	-	-	-	0.8	1.0	-	-	-	-	-	-	-	1.0
2.0	1.7	-	1.1	-	1.0	-	-	-	-	-	-	1.1	1.0	-	-
-	2.7	2.9	-	-	-	0.8	-	-	-	1.1	1.1	-	-	1.0	2.0
-	-	-	1.0	0.9	-	1.2	0.7	1.0	1.9	0.6	1.0	-	-	-	-
-	-	-	-	-	-	0.6	-	-	-	-	-	-	-	-	1.0
1.0	1.0	-	2.3	0.8	0.8	0.9	-	-	-	0.5	0.9	-	-	-	-
1.0	1.3	0.9	1.1	0.9	-	1.1	1.1	-	0.9	-	-	1.0	0.9	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.8	0.9	-	-	-	-	-	-	-	-	0.9	1.1	-	-	0.8	1.9
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0.9	1.0	0.9	0.8	-	1.0	0.9	-	-	-	-	1.0	0.9	1.0	1.9	0.9
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	16	10	16	7	8	17	6	5	4	4	5	4	5	6	13
-	Gly	-	Glx	Leu	Ser	Ile	-	Val	Val	Ala	Ala	Lys	Gly	Lys	Ala

St11	CNF	CNF	CNF	C15	CNF	P10	CNF	T6	P12	T26	CNF	C11	C11T	CNF	C11
13	13Tb	13T1	13T2	13T3	13T4	13T5	13T6	13T7	13T8	13T9	13T10	13T11	13T12	13T13	13T14
-0.23	+0.10	-	+0.37	0	+0.30	+0.34	-0.07	0	0	-	-	-	-	-0.57	-0.26
-	-	+0.76	-	+0.80	-	+1.07	-	+0.97	+1.26	-	+1.90	+0.73	+0.82	+0.31	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.1	-	-	-	-	-	-	-	-	-	-	-	-	-	1.2	1.2
1.9	0.9	0.9	-	-	-	-	-	-	-	-	-	1.0	-	-	-
-	-	1.0	1.0	-	-	0.7	1.1	1.9	1.1	-	-	1.1	1.0	1.0	0.9
-	-	0.8	0.7	-	-	-	1.4	-	-	-	-	1.0	-	-	-
2.1	1.4	3.6	1.3	-	1.1	-	-	1.3	2.2	-	1.2	-	-	2.9	1.0
1.1	0.9	1.0	1.0	-	-	-	-	-	-	-	1.2	-	-	1.1	-
-	-	-	-	-	1.0	-	-	-	-	-	-	1.0	-	0.9	-
0.9	-	-	-	-	-	0.9	1.0	2.2	1.0	-	-	1.0	-	-	-
1.0	1.0	1.7	0.9	-	-	1.1	0.9	1.1	-	-	-	-	-	0.9	-
0.9	-	-	-	0.9	-	0.9	-	1.1	1.0	1.0	-	-	0.9	-	-
-	-	-	-	-	0.9	-	-	1.1	-	0.8	-	-	-	-	1.6
-	-	-	-	-	1.0	-	-	-	1.8	-	1.8	-	1.1	3.7	-
-	-	-	-	1.2	0.7	-	-	-	-	-	-	-	-	-	-
0.9	0.9	-	-	-	1.0	0.8	-	-	-	-	0.8	-	-	-	0.9
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	0.9	0.8	-	-	-	-	1.0	1.9	1.0	0.9	-	-	-	-
1.0	0.9	-	-	-	0.9	1.2	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	6	11	6	2	7	6	4	8	10	2	7	5	3	15	6
Asx	-	-	Ser	Met	-	Phe	-	Val	-	Met	-	Ala	Ser	-	Ser

d

St72	T8	C35	T11	C39	P16	T13	C3.1	CNC	C3.3	St6	T14	St27	T21b	T21b	T21b
0	-0.40	+0.34	+0.37	+0.36	0	-0.44	-0.28	-0.14	0	-0.31	0	+0.23	-	-0.43	+0.55
+1.30	+0.75	+1.51	+1.22	+1.33	+0.73	+1.27	+0.49	-	+0.63	-	+1.38	-	+1.04	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	1.2	-	-	-	0.9	-	1.2	2.1	1.1	-	-	-	1.0	0.9	-
1.0	1.7	-	-	-	1.1	-	1.0	1.0	-	-	-	-	-	-	-
-	1.9	-	-	-	1.0	-	-	5.7	2.2	3.0	1.0	2.1	1.0	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.2	3.1	2.7	-	1.8	2.1	2.2	-	1.3	-	1.2	1.2	2.2	3.0	1.1	-
-	-	-	0.8	1.0	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	1.2	2.7	-	1.2	-	-	1.1	-	1.0
-	-	-	-	-	-	-	1.1	2.2	1.0	1.1	1.0	3.6	4.8	-	0.9
-	1.0	1.0	0.9	-	1.1	-	0.8	0.9	-	-	-	1.0	0.8	-	-
-	0.6	-	-	1.6	1.6	0.9	1.0	-	-	-	-	-	-	-	-
-	-	0.9	-	-	-	-	0.9	1.8	-	0.7	-	0.9	2.0	1.0	-
0.9	1.1	-	0.9	1.1	2.0	-	-	1.9	0.9	0.8	0.9	1.1	-	-	-
-	-	1.0	1.0	0.9	1.0	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	0.9	1.0	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	1.6	2.2	-	-
0.9	-	1.9	-	2.2	1.6	1.0	-	1.0	-	-	1.0	1.1	-	-	-
-	1.1	1.1	1.2	1.0	-	-	-	-	-	-	-	1.3	-	1.1	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	13	8	5	10	13	4	8	22	6	8	5	14	17	3	3
Lys	Leu	Glx	Val	Leu	Lys	Met	Ala	-	Ser	Leu	Ala	Ser	Ile	Ile	-

c

f

Table 3. Evidence for amide assignments

Residue	Evidence
Glu-9 Asp-13	Peptide CNE'T4a $m_{6.5} = -0.45$, therefore two acids.
Glu-17	Peptide T21a is neutral at pH 6.5, therefore one acid.
Asn-24 Glu-25 Asn-26	Residue 25 is Glu since it is a site of staph. proteinase cleavage. The other two residues must be amides since peptide C20.2a $m_{6.5} = -0.05$.
Asp-47	Peptide CNBT10a is neutral at pH 6.5, therefore one acid.
Gln-52 Asn-54	Peptide CNBT10b $m_{6.5} = +0.40$, therefore two amides.
Asp-60 Asp-67	From Milstein [16].
Glu-70 Gln-72 Asn-78	From Henderson et al. [17].
Asp-85	Peptide C20.2b is nearly neutral at pH 6.5, therefore one acid.
Asn-88 Gln-89	Subpeptide C9v3Th2 with sequence Asx-Glx is neutral at pH 6.5, therefore two amides.
Asn-94 Asp-95	Subpeptides C9v3Th3 with sequence Ile-Thr-Lys-Pro-Asx-Asx- is neutral at pH 6.5, therefore one acid, one amide. After 5 rounds of Edman degradation $m_{6.5}$ was the same as Asp, therefore 6th residue is Asp.
Glu-108 Glu-109	Peptide CNBT14b $m_{6.5} = -0.20$, therefore two acids.
Glu-116	Peptide St24 is neutral at pH 6.5, therefore one acid.
Gln-119 Glu-123	Peptide St16 $m_{6.5} = -0.27$, therefore one amide, one acid. Final residue is Glu since it is a site of staph. proteinase cleavage.
Glu-130 Asn-133 Gln-135 Asp-139 Gln-140	Peptide T4 $m_{6.5} = -0.12$, therefore two acids and three amides. Subpeptide T4Cc (Gly, Gly, Leu, Glx, Pro, Ile, Asx, Phe) $m_{6.5} = -0.28$, therefore one acid, one amide. Automatic sequence determination of peptide C19.1 showed that residue 130 is Glu and 133 is Asn. Residue 135 is Glu since peptide C8z1Th1 $m_{6.5} = 0$ due to cyclised glutamine. Both peptides C8z1Th1 and C8z1 showed no change in composition after two rounds of Edman degradation. Peptide C8z1Th3 after 3 rounds of Edman degradation was neutral at pH 6.5, therefore residue 139 is acid and 140 is amide (residue 143 is acid).
Glu-143	Peptide C8z1Ta $m_{6.5} = -0.57$, therefore one acid.
Asn-146	Peptide C14 is neutral at pH 6.5, therefore one amide.
Glu-150	Site of staph. proteinase cleavage to give peptides St 13 and CNBT8Std.
Gln-152 Asn-154	Peptide C2a $m_{6.5} = -0.31$, therefore one acid, two amides. Acid is Glu 150.
Asn-159 Gln-162	Peptide CNBT2.3 is neutral at pH 6.5, therefore two amides.
Asp-167 Gln-169	Peptide CNBC1z3Tha $m_{6.5} = -0.42$, therefore one acid, one amide. The amide assignment is from McReynolds et al. [24].
Asn-176	Peptide CNE"C2 is neutral at pH 6.5, therefore one amide.
Glu-185	Peptide CNE"T4 is neutral at pH 6.5, therefore one acid.
Asp-190 Glu-191	Peptide St 29 is neutral at pH 6.5, therefore two acids.
Asp-192 Gln-194	Peptide CNE"Ta $m_{6.5} = -0.42$, therefore three acids, one amide. Amide is Gln-194 since peptide C8z.2 (Thr, Glx, Ala, Met, Pro, Phe) is neutral at pH 6.5.
Glu-202	Peptide CNF13 Stb is neutral at pH 6.5, therefore one acid.
Gln-203 Glu-204	Peptide CNF13Sta $m_{6.5} = -0.53$, therefore one acid, one amide. Acid is second residue due to staph. proteinase specificity.
Gln-209	Peptide CNF13Stc $m_{6.5} = +0.37$, therefore one amide.
Gln-213	Peptide CNF19Tb $m_{6.5} = +0.30$, therefore one amide.
Glu-225	Peptide CNH17 is neutral at pH 6.5, therefore one acid.
Glu-231	Peptide CNF10 is neutral at pH 6.5, therefore one acid.
Asp-247 Glu-248	Peptide CNESbtTh1 $m_{6.5} = -0.62$, therefore two acids.
Glu-253 Gln-254	Peptide CNDC7 $m_{6.5} = -0.49$, therefore one acid, one amide. After one round of degradation, peptide was neutral at pH 6.5, therefore second residue is amide.
Glu-256	Peptide CNESbtTh4 $m_{6.5} = -0.55$, therefore one acid.
Asn-260 Glu-262	Peptide CNESb $m_{6.5} = -0.26$, therefore one acid, one amide. Acid is residue 262 due to staph. proteinase specificity. Also peptide CNDC16 is neutral at pH 6.5.
Glu-266	Peptide T9 $m_{6.5} = -0.37$, therefore one acid.
Asn-271	Peptide C3.2 is neutral at pH 6.5, therefore one amide.
Glu-274 Glu-275	Peptide CNF2013 $m_{6.5} = -0.37$, therefore two acids.
Glu-288 Glu-289	Peptide T13 $m_{6.5} = -0.44$, therefore two acids.
Asn-292	Peptide C1yb2 is neutral at pH 6.5, therefore one amide.

Residue	Evidence
Asp-304	Peptide C3.1b $m_{6.5} = -0.28$, therefore one acid.
Asn-311	Peptide C3.3 is neutral at pH 6.5 therefore one amide.
Glu-319	Peptide T14 is neutral at pH 6.5, therefore one acid.
Gln-325	Peptide T21bThc1 is neutral at pH 6.5, therefore one amide.
Glu-333	Peptide T21bTha $m_{6.5} = -0.60$, therefore one acid.
Asn-335 Glu-336	Peptide St10 $m_{6.5} = -0.51$, therefore one acid, one amide. Acid is residue 336 due to staph. proteinase specificity.
Glu-340 Glu-346	From Milstein [16].
Asp-350	From Henderson et al. [17].
Glu-356 Glu-357 Asp-361 Asn-374	From Thompson et al. [11].

masked or buried in native ovalbumin [27]. This sequence is located between residues 234 and 252.

It has recently been pointed out by Hunt and Dayhoff [45] that the amino acid sequence of ovalbumin, antithrombin III and α_1 -proteinase inhibitor are significantly homologous. This relationship indicates that these three proteins have diverged from a common, albeit ancient, ancestor and reopens the question concerning the still enigmatic function of ovalbumin. Perhaps like antithrombin III and α_1 -proteinase inhibitor, ovalbumin may function as a proteinase inhibitor. If so, ovalbumin will be one of the few proteins whose function has been discovered from a knowledge of its amino acid sequence.

We would like to thank Dr Richard Ambler for many helpful discussions on the strategy and tactics of protein sequencing, Mrs Jean Bathgate for excellent technical assistance, the Medical Research Council and the University of Aberdeen for financial support, and the Science Research Council for a studentship (ADN).

REFERENCES

- Sorensen, S. P. L. & Hoyrup, M. (1915-1917) *Comp. Rend. Trav. Lab. Carlsberg*, 29, 12-67.
- Warner, R. C. (1954) in *The Proteins* (Neurath, H. & Bailey, K., eds) vol. 2, pp. 443-485, Academic Press, London.
- Tristram, G. R. & Smith, R. H. (1963) *Adv. Protein Chem.* 18, 227-318.
- Fothergill, L. A. & Fothergill, J. E. (1970) *Eur. J. Biochem.* 17, 529-532.
- Fothergill, L. A. & Fothergill, J. E. (1970) *Biochem. J.* 116, 555-561.
- Thompson, E. O. P. & Fisher, W. K. (1978) *Aust. J. Biol. Sci.* 31, 433-442.
- Narita, K. & Ishii, J. (1962) *J. Biochem. (Tokyo)* 52, 367-373.
- Niu, C. I. & Fraenkel-Conrat, H. (1955) *J. Am. Chem. Soc.* 77, 5882-5885.
- Linderstrom-Lang, K. & Ottesen, M. (1949) *Comp. Rend. Trav. Lab. Carlsberg*, 26, 403-442.
- Ottesen, M. (1958) *Comp. Rend. Trav. Lab. Carlsberg*, 30, 211-270.
- Thompson, E. O. P., Sleight, R. W. & Smith, M. B. (1971) *Aust. J. Biol. Sci.* 24, 525-534.
- Tai, T., Yamashita, K., Ito, S. & Kobata, A. (1977) *J. Biol. Chem.* 252, 6687-6694.
- Cunningham, L. W., Clouse, R. W. & Ford, J. D. (1963) *Biochim. Biophys. Acta*, 78, 379-381.
- Perlmann, G. E. (1952) *J. Gen. Physiol.* 35, 711-726.
- Flavin, M. (1954) *J. Biol. Chem.* 210, 771-784.
- Milstein, C. P. (1968) *Biochem. J.* 110, 127-134.
- Henderson, J. Y., Moir, A. J. G., Fothergill, L. A. & Fothergill, J. E. (1981) *Eur. J. Biochem.* 114, 439-450.
- Lush, I. E. (1964) *Genet. Res.* 5, 257-268.

19. Wiseman, R. L., Fothergill, J. E. & Fothergill, L. A. (1972) *Biochem. J.* 127, 775–780.
20. Palmiter, R. D. (1975) *Cell*, 4, 189–197.
21. Palmiter, R. D., Moore, P. B. & Mulvihill, E. R. (1976) *Cell*, 8, 557–572.
22. Means, A. R., Comstock, J. P., Rosenfeld, G. C. & O'Malley, B. W. (1972) *Proc. Natl Acad. Sci. USA*, 69, 1146–1150.
23. Haines, M. E., Carey, N. H. & Palmiter, R. D. (1974) *Eur. J. Biochem.* 43, 549–560.
24. McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M. & Brownlee, G. G. (1978) *Nature (Lond.)* 273, 723–728.
25. Breathnach, R., Mandel, J. L. & Chambon, P. (1977) *Nature (Lond.)* 270, 314–319.
26. Palmiter, R. D., Gagnon, J. & Walsh, K. A. (1978) *Proc. Natl Acad. Sci. USA*, 75, 94–98.
27. Lingappa, V. R., Lingappa, J. R. & Blobel, G. (1979) *Nature (Lond.)* 281, 117–121.
28. Ambler, R. P. & Meadway, R. J. (1968) *Biochem. J.* 108, 893–895.
29. Kuhn, R. W., Walsh, K. A. & Neurath, H. (1974) *Biochemistry*, 13, 3871–3877.
30. Heilmann, J., Barrolier, J. & Watzke, E. (1957) *Hoppe-Seyler's Z. Physiol. Chem.* 309, 219–220.
31. Rydon, H. N. & Smith, P. W. G. (1953) *Nature (Lond.)* 169, 922–923.
32. Offord, R. E. (1966) *Nature (Lond.)* 211, 591–593.
33. Smith, I. (1953) *Nature (Lond.)* 171, 43–44.
34. Offord, R. E. (1969) in *Data for Biochemical Research* (Dawson, R. M. C., Elliot, D. C., Elliot, W. H. & Jones, K. M., eds) 2nd edn, p. 530. Clarendon Press, Oxford.
35. Sakaguchi, S. (1950) *J. Biochem. (Tokyo)* 37, 231–236.
36. Hartley, B. S. (1970) *Biochem. J.* 119, 805–822.
37. Woods, K. R. & Wang, K. T. (1967) *Biochim. Biophys. Acta*, 133, 369–370.
38. Moir, A. J. G. (1973) Ph.D. Thesis, University of Aberdeen.
39. Kiely, M. L., McKnight, G. S. & Schimke, R. T. (1976) *J. Biol. Chem.* 251, 5490–5495.
40. Hunt, L. T. & Dayhoff, M. O. (1970) *Biochem. Biophys. Res. Commun.* 39, 757–765.
41. Jackson, R. L. & Hirs, C. H. W. (1970) *J. Biol. Chem.* 245, 624–636.
42. Mercier, J.-C., Grosclaude, F. & Ribadeau-Dumas, B. (1971) *Eur. J. Biochem.* 23, 41–51.
43. Brignon, G., Ribadeau-Dumas, B., Grosclaude, F. & Mercier, J.-C. (1971) *Eur. J. Biochem.* 22, 179–185.
44. Smith, M. B. (1964) *Aust. J. Biol. Sci.* 17, 261–270.
45. Hunt, L. T. & Dayhoff, M. O. (1980) *Biochem. Biophys. Res. Commun.* 95, 864–871.
46. Furka, A. & Sebestyen, F. (1970) *Acta Biochim. Biophys. Acad. Sci. Hung.* 5, 131–136.

A. D. Nisbet, Department of Pathology, University of Aberdeen, Medical School Buildings, Foresterhill, Aberdeen, Great Britain, AB9 2ZD

R. H. Saundry, Department of Haematology, Louis Jenner Laboratories, St Thomas' Hospital and Medical School, London, Great Britain SE1 7EH

A. J. G. Moir, Department of Biochemistry, University of Birmingham Medical School, P.O. Box 363, Birmingham, Great Britain, B15 2TT

L. A. Fothergill and J. E. Fothergill, Department of Biochemistry, University of Aberdeen, Marischal College, Aberdeen, Great Britain, AB9 1AS