Aust. J. Biol. Sci., 1976, **29**, 175–80

Primary Structure of Apovitellenin I from Hen Egg Yolk and its Comparison with Emu Apovitellenin I

T. A. A. Dopheide and A. S. Inglis

Division of Protein Chemistry, CSIRO, Parkville, Vic. 3052.

Abstract

The amino acid sequence of apovitellenin I from hen egg yolk has been determined using both automatic and manual procedures; it comprises 82 residues. Hen apovitellenin shows considerable homology with emu apovitellenin I which contains 84 residues. Besides two deletions in the sequence, the hen protein differs in 28 positions from the emu protein; 26 of these positions may have arisen from single base changes. The changes are largely conservative ones, which suggest that the structure and function have been preserved despite extensive mutation.

Introduction

Recently we reported the amino acid sequence of the major apoprotein from the very low-density lipoproteins from emu egg yolk (Dopheide and Inglis 1974). Burley (1975) has isolated a similar protein from the eggs of another bird species, the hen, and the determination of the primary sequence of this molecule is described here.

Following a study of the cytochrome sequences in birds, Howard *et al.* (1974) have raised doubts as to the correctness of the classical evolutionary pathway as proposed for these birds. A comparison of the emu and hen egg yolk apoproteins was therefore made to see whether the changes in the amino acid sequences of these easily obtained and conveniently sized proteins could shed light on this question.

Materials and Methods

Hen egg apovitellenin was a gift of Dr R. W. Burley. For sequence studies, the protein was reduced with a tenfold excess (based on 1 μmol half-cystine per 8 mg protein) of dithiothreitol in 6 M guanidine–HCl/0·5 M tris–HCl at pH 8·3, and alkylated with the theoretical quantity (based on total thiol) of iodoacetic acid.

Trypsin (TPCK-treated), pepsin, chymotrypsin and DFP-treated carboxypeptidase A came from Worthington Biochemical Corp.

The reduced and carboxymethylated protein was dialysed thoroughly against *N*-ethylmorpholine acetate buffer (0·1 M, pH 8·0) and then digested with trypsin or chymotrypsin at 1:100 enzyme–substrate ratio (w/w) for 8 h at 37°C.

Digestion with carboxypeptidase A was in 0·1 M NH_4HCO_3 at 25°C, whilst peptic digestions were in 5% formic acid.

Details of peptide purifications, amino acid analyses and protein sequence analyses were similar to those reported earlier (Dopheide and Inglis 1974). In addition, when the presence of tryptophan was suspected, samples were hydrolysed in 3 M *p*-toluenesulphonic acid containing tryptamine (Liu and Chang 1971). In this work the sequenator, operated according to the procedure of Crewther and Inglis (1975), was also used to sequence the first 15 residues of an insoluble, 18-residue-long tryptic peptide (T7). Residues from the sequenator were converted to thiohydantoins by heat (Inglis *et al.* 1975) before thin layer chromatography.

Results

The molecular weight of the reduced, carboxymethylated hen egg yolk apovitellenin was found to be 8000±800 by electrophoresis in sodium dodecyl sulphate on acrylamide gels (Weber and Osborne 1969). The amino acid analysis (Table 1) indicated the presence of 82 residues, from which a molecular weight of 9391 was obtained. The original analysis reported by Burley (1975) has one less residue of both isoleucine and tyrosine.

Table 1. Amino acid composition of apovitellenin I from hen egg yolk

The protein was hydrolysed, in duplicate, for 24, 48 and 72 h, in 3 M *p*-toluenesulphonic acid containing 0·05% tryptamine. Serine and threonine values have been corrected for losses by extrapolation to zero time; valine, isoleucine and leucine values are given as the 72-h values. Total number of residues = 82. Molecular weight obtained from residue weights = 9391

Amino acid	Mol/mol of protein	No. found in sequence	Amino acid	Mol/mol of protein	No. found in sequence
Asp	8·1	8	Ile	6·8	7
Thr	5·0	5	Leu	9·3	9
Ser	3·9	4	Tyr	4·0	4
Glu	8·2	8	Phe	2·1	2
Phe	2·0	2	Lys	6·1	6
Gly	2·8	3	Arg	5·8	6
Ala	8·3	8	Trp	0·9	1
Val	6·8	7	$CysSO_3H$[A]	1·0	1[B]
Met	1·0	1	$MetO_2$[A]	1·0	1[C]

[A] Obtained from duplicate 24-h analyses of a performic acid oxidized protein sample.
[B] Found as Cys(Cm). [C] Found as Met.

Tyrosine was the only amino acid found at the *C*-terminus by selective tritiation (Holcomb *et al.* 1968), and lysine the only *N*-terminal residue by dansylation of the protein (Gray 1967). Tryptic peptides were separated on G25 Sephadex (five coupled columns, each 120 by 0·9 cm) in pyridine–1 M ammonia (70:30, v/v). A fraction, insoluble in the chromatography solvent and isolated before the digest was applied to the column system, was virtually pure (peptide T7) by the criteria of a single *N*-terminus and amino acid analysis. The column eluate was monitored by high-voltage paper electrophoresis at pH 6·5; relevant fractions were pooled and further purified. The amino acid compositions and sequences of the recovered tryptic peptides are shown in Table 2. The peptides T1, T2, T3, T4, T6, T9, T11, T12 and T13 were degraded to completion by the dansyl–Edman method. The insoluble peptide (T7, 18 residues) was partly sequenced with the sequenator (15 residues) using a peptide procedure (Crewther and Inglis 1975); digestion of peptide T7 with thermolysin gave the overlapping *C*-terminal peptide, Val-Ile-Gly-Ser-Arg. The peptide designated T5 in Table 4 was not detected in the Sephadex column eluate. It was not sought further because sequenator analysis of the intact protein unequivocally established the sequence of the first 40 residues and the arrangement of peptides T1–T7; peptide T5 would occupy positions 11–29 in the amino acid sequence, a region almost identical with the corresponding region of the emu protein.

Peptide T8 was degraded until the *C*-terminal Arg was reached. The mobility of the peptide showed that one single carboxyl group was present. This was located at the *C*-terminus of the peptic peptide Leu-Ile-Asx-Glx ($m = -0{\cdot}45$) by stepwise degradation and measurement of the mobility of the residual peptide after each step. The mobility, carrying a negative sign, increased after each step, until after step 3; the remaining Glx (as the major spot) had a mobility of $-0{\cdot}90$, thus establishing the sequence as Leu-Ile-Asn-Glu.

Table 2. Composition and sequence of some tryptic peptides of apovitellenin I from hen egg yolk
Evidence for the structure of these peptides is discussed in the text. Peptide composition is given as mol amino acid/mol peptide

Peptide and mobility	Composition	Sequence
T1	Lys	Lys
$m = 1{\cdot}05$		
T2	$Asp_{1.0}\ Ser_{0.8}\ Ile_{1.6}\ Arg_{0.9}$	Ser-Ile-Ile-Asp-Arg
$m = 0{\cdot}0$		
T3	$Glu_{1.0}\ Arg_{1.0}$	Glu-Arg
$m = 0{\cdot}0$		
T4	Arg	Arg
$m = 1{\cdot}10$		
T6	$Ser_{0.9}\ Pro_{1.0}\ Val_{1.0}\ Arg_{0.9}$	Val-Ser-Pro-Arg
T7	$Asp_{1.2}\ Thr_{1.8}\ Ser_{1.8}\ Glu_{2.1}\ Gly_{1.9}\ Ala_{1.2}\ Val_{2.6}\ Ile_{1.0}\ Leu_{2.2}\ Phe_{1.0}\ Arg_{1.0}\ Lys_{0.2}$	Ala-Gly-Gln-Phe-Leu-Leu-Asp-Val-Ser-Gln-Thr-Thr-Val-Val-Ile-Gly-Ser-Arg
T8	$Asp_{2.0}\ Thr_{1.0}\ Glu_{1.0}\ Ala_{1.0}\ Ile_{1.0}\ Leu_{0.9}\ Phe_{0.9}\ Arg_{0.7}$	Asn-Phe-Leu-Ile-Asn-Glu-Thr-Ala-Arg
$m = 0{\cdot}0$		
T9	$Thr_{0.9}\ Leu_{0.9}\ Lys_{1.0}$	Leu-Thr-Lys
T10	$Glu_{3.0}\ Ala_{1.0}\ Met_{0.8}\ Leu_{2.0}\ Lys_{1.0}$	Leu-Ala-Glu-Gln-Met-Leu-Glu-Lys
$m = -0{\cdot}28$		
T11	$Ile_{1.0}\ Lys_{1.0}$	Ile-Lys
T12	$Cys(Cm)_{0.8}\ Asp_{1.0}\ Thr_{0.9}\ Leu_{1.0}\ Tyr_{1.0}\ Lys_{1.0}$	Asn-Leu-Cys(Cm)-Tyr-Thr-Lys
$m = 0{\cdot}0$		
T13	$Gly_{1.0}\ Val_{0.9}\ Leu_{0.9}\ Tyr_{1.2}$	Val-Leu-Gly-Tyr

Table 3. Composition and sequence of some chymotryptic peptides of apovitellenin I from hen egg yolk
Evidence for the structure of these peptides is discussed in the text. Peptide composition is given as mol amino acid/mol peptide

Peptide and mobility	Composition	Sequence
C1	$Asp_{1.0}\ Ser_{0.8}\ Gly_{0.9}\ Val_{0.8}\ Ile_{0.7}\ Phe_{1.0}\ Arg_{0.9}$	Val-Ile-Gly-Ser-Arg-Asn-Phe
$m = 0{\cdot}31$		
C2	$Asp_{1.2}\ Thr_{2.0}\ Ala_{1.8}\ Glu_{3.0}\ Met_{0.8}\ Ile_{0.8}\ Leu_{3.5}\ Lys_{1.0}\ Arg_{1.0}$	Leu-Ile-Asn-Glu-Thr-Ala-Arg-Leu-Thr-Lys-Leu-Ala-Glu-Gln-Met-(Leu)
$m = 0{\cdot}0$		
C3	$Asp_{1.0}\ Glu_{0.9}\ Ile_{0.9}\ Leu_{1.0}\ Lys_{1.8}$	Glu-Lys-Ile-Lys-Asn-Leu
C4	$Thr_{0.9}\ Val_{0.8}\ Leu_{0.8}\ Lys_{1.0}$	Thr-Lys-Val-Leu

Peptide T10 contained two carboxyl groups, as judged from its mobility. It was degraded until the *C*-terminal Lys was reached. The positions of the two carboxyl groups were obtained from the thermolytic peptides Met-Leu-Glu-Lys ($m = 0{\cdot}0$)

and Leu-Ala-Glu-Gln ($m = -0\cdot49$); the latter peptide was tritiated at the *C*-terminal residue by the method of Holcomb *et al.* (1968), then digested for 1 h with carboxypeptidase A and the [^{3}H]glutamine (at the origin) identified as the only radioactive spot after electrophoresis at pH 6·5.

Table 4. Composition of amino acid sequences of apovitellenin I from hen and from emu (Dopheide and Inglis 1974)

The numbering is that of the emu protein and gaps have been left to maximize the homology. Residues printed in italics are those related via single base changes, residues printed in bold face are those related via multiple changes. Vertical arrows indicate positions of cleavage by trypsin. T, tryptic peptides. C, chymotryptic peptides. ← – – → Residues identified by sequenator

	5 10 15 20
Emu	Lys-Ser-Ile-Phe-Glu-Arg-Asp-Asn-Arg-Arg-Asp-Trp-Leu-Val-Ile-Pro-Asp-Ala-Val-Ala-
Hen	Lys-Ser-Ile- *Ile*-*Asp*-Arg-*Glu* — Arg-Arg-Asp-Trp-Leu-Val-Ile-Pro-Asp-Ala-*Ala*-Ala-
	T1 ↑ T2 ↑ T3 ↑ T4 ↑
	← –
	25 30 35 40
Emu	Ala-Tyr-Val-Tyr-Glu-Thr-Val-Asn-Lys-Met-Phe-Pro-Lys-Val-Gly-Gln-Phe-Leu-Ala-Asp-
Hen	Ala-Tyr- *Ile*-Tyr-Glu-*Ala*-Val-Asn-Lys- *Val* -*Ser* -Pro-*Arg*-*Ala*-Gly-Gln-Phe-Leu-**Leu**-Asp-
	T5 ↑ T6 ↑
	– →
	← – – – – – – – – – – – – – – – –
	45 50 55 60
Emu	Ala-Ala-Gln- Ile-Pro-Val- Ile-Val-Gly-Thr-Arg-Asn-Phe-Leu-Ile-Arg-Glu-Thr-Ser-Lys-
Hen	*Val*-*Ser*-Gln-*Thr*-*Thr*-Val- *Val*- *Ile*-Gly-*Ser*-Arg-Asn-Phe-Leu-Ile-**Asn**-Glu-Thr-*Ala*-*Arg*-
	T7 ↑ T8 ↑
	←——— C1 ———→ ←———————
	– – – – – – – – – – – – – – – – – →
	65 70 75
Emu	Leu-Ser- Ile-Leu-Ala-Glu-Gln-Met-Met-Glu-Lys-Val-Lys-Thr-Leu-Trp-Asn-Thr-Lys-
Hen	Leu-*Thr*-*Lys*-Leu-Ala-Glu-Gln-Met-*Leu* -Glu-Lys- *Ile*-Lys-*Asn*-Leu-*Cys*-*Tyr*-Thr-Lys-
	T9 ↑ T10 ↑ T11 ↑ T12 ↑
	——— C2 ———————→ ←———— C3 ————→ ←————
	80 84
Emu	Val-Leu-Gly-Tyr-Tyr
Hen	Val-Leu-Gly-Tyr —
	T13
	- C4 ——→

The remaining overlaps were constructed from the sequence of some chymotryptic peptides (Table 3). These were separated and purified by the methods used for the tryptic peptides. Peptide C1 was completely degraded. Peptide C2 was labelled with tritium at the *C*-terminal residue (Holcomb *et al.* 1968); on hydrolysis and electrophoresis at pH 1·9 both Met and Leu were found to be labelled. A peptic digest of peptide C2 gave the peptides Thr-Lys-Leu and Thr-Ala-Arg-Leu. Dansyl–Edman degradation of peptide C2 gave the sequence Leu-Ile-Asn-Glu-Thr-Ala-Arg, so that the remainder of peptide C2 could be placed unambiguously. The hetero-

geneity at the *C*-terminus is not surprising in view of the specificity of chymotrypsin. Peptides C3 and C4 were sequenced without complications.

Discussion

The determination of the amino acid sequence of hen egg yolk apovitellenin I shows that it is homologous with apovitellenin I isolated from the low-density lipoprotein fraction of emu egg yolk (Dopheide and Inglis 1974). The molecular weights of around 9000 are similar, and amino acid analyses indicate that there are 82 residues in the hen protein compared with 84 in the emu protein. In the amino acid sequence, however, 28 residues are changed in the hen protein as compared with the emu protein. Most of these changes are conservative (the charge distribution is virtually unchanged) and the overall net charge is unaltered, so that the molecules appear to function in a similar manner.

In Table 4 the two sequences have been aligned to facilitate comparison. It is seen that the first 28 residues and the last 6 residues of the hen protein are little changed compared with the corresponding areas in the emu protein. In these regions of the hen protein there are two deletions, corresponding to residues 8 and 84 of the emu protein, whilst homologous residues Ile, Asp, Glu, Ala, Ile and Ala replace Phe, Glu, Asp, Val, Val and Thr respectively. Of the remaining 22 changes throughout the molecule, 20 could occur with only a single base change, although the changes at residues 76 and 77 (Cys and Tyr for Trp and Asn respectively) are comparatively rare. The altered residues 38 (Ala for Leu) and 56 (Asn for Arg) require two base changes. The highly polar regions (residues 5–10) and the predominantly hydrophobic areas (residues 11–23 and residues 33–48) of the emu protein have been preserved in the hen sequence. The deletion of the *C*-terminal Tyr may have occurred by formation of a terminating triplet from the Tyr triplet via a single base change. The replacement of Trp(76) in the emu protein by Cys in the hen protein could well have structural implications, although Burley (personal communication) has evidence that the protein exists *in vivo* in the reduced state.

The 30 differences between the emu and the hen proteins are equivalent to at least 38 point mutations per 100 amino acid residues. For comparison between the same bird groups, cytochrome C shows only 3 point mutations per 100 residues (Howard *et al.* 1974). Published general mutation rate estimates for cytochrome C, haemoglobin and the immunoglobulins are respectively 3, 14 and 32 point mutations per 100 residues per 10^8 years (McLaughlin and Dayhoff 1972). Applying the higher rate of mutation, the divergence between the ratites and the galliformes could be of the order of 10^8 years, which supports the antiquity of this divergence rather than the proposition of Howard *et al.* (1974) that the ratites are a relatively recent degenerate group. On the other hand, it is possible that the apovitellenins are functionally intact over a wide range of sequence variation; the size and distribution of polar and non-polar residues at least seem conserved here, but the composition of most of the domains is greatly changed. In this case of presumed low constraint, the mutation rate could be much higher, which in turn leads to a more recent divergence date.

No previous quantitative estimates of the divergence exist; fossil evidence seems absent and the rate of protein evolution is largely unknown. A comparison of these apoproteins is the first quantitative evidence that the evolutionary distance is large. Further comparative work could throw additional light on these questions, and also

on the effectiveness of egg-yolk proteins as markers for the molecular evolution of birds.

Acknowledgments

The technical assistance of Mr P. McK. Strike is gratefully acknowledged.

References

Burley, R. W. (1975). *Aust. J. Biol. Sci.* **28**, 121–32.
Crewther, W. C., and Inglis, A. S. (1975). *Anal. Biochem.* **68,** 572–85.
Dopheide, T. A., and Inglis, A. S. (1974). *Aust. J. Biol. Sci.* **27**, 15–21.
Gray, W. R. (1967). *Methods Enzymol.* **11**, 139–51.
Holcomb, G. H., James, S. A., and Ward, D. N. (1968). *Biochemistry* **7**, 1291–6.
Howard, M. L., Joubert, F. J., and Strydom, D. J. (1974). *Comp. Biochem. Physiol.* **48**B, 75–85.
Inglis, A. S., Nicholls, P. W., and Strike, P. McK. (1975). *J. Chromatogr.* **107**, 73–80.
Liu, T. Y., and Chang, Y. H. (1971). *J. Biol. Chem.* **246**, 2842–8.
McLaughlin, P. J., and Dayhoff, M. O. (1972). 'Atlas of Protein Sequence and Structure'. Vol. 5. pp. 47–52. (National Biomedical Research Foundation: Maryland.)
Weber, K., and Osborne, M. (1969). *J. Biol. Chem.* **244**, 4406–12.

Manuscript received 28 October 1975