DETERMINATION OF THE AMINO ACID SEQUENCE OF APOVITELLENIN I FROM DUCK'S EGG YOLK USING AN IMPROVED SEQUENATOR PROCEDURE: A COMPARISON WITH OTHER AVIAN SPECIES

A. S. INGLIS
Division of Protein Chemistry, CSIRO, 343 Royal Parade, Parkville, Victoria 3052

and

R. W. BURLEY
Division of Food Research, CSIRO, P.O. Box 52, North Ryde, N.S.W. 2113, Australia

Received 29 November 1976

1. Introduction

As a prerequisite for a study of protein–lipid interactions in avian egg yolk, the apoproteins from the major constituent of yolk – a lipoprotein with a low density (approx. 0.95 g/ml) and a high proportion of lipid (approx. 85%) – are being examined. The apoprotein of lowest molecular weight, apovitellenin I, has been isolated in a pure state from the egg-yolk lipoprotein of the emu (*Dromaeus novaehollandiae*) [1], where it is at least 60% of the total apoprotein, and the hen (*Gallus domesticus*), where it is a lower proportion of the total [2]. Amino acid sequences of these proteins [3,4] show interesting differences, including rare substitutions (e.g. Cys for Trp). Apovitellenin I has since been isolated from eggs of other species and we report here a third sequence, that of apovitellenin I of the duck (*Anas platyrhynchos*).

Sequence determination of these proteins by the overlapping of enzymically released peptides was complicated by the insolubility of some of the peptides and, in emu's apovitellenin I, because all available proteolytic enzymes hydrolysed between residues 22 and 23 (Tyr–Val) [3]. The well-established automatic sequencing procedure of Edman and Begg [5] overcame this latter problem and gave the first 35 residues [3]. This method has, however, several disadvantages when determining larger sequences; e.g., labile or polar thiazolinones (Ser, Cys, Arg, His) give low yields that are further decreased by the HCl conversion [6] and lead to problems of identification after repeated cycles. In addition, heptafluorobutyric acid is not readily removed after cleavage and residual acid may cause loss of hydrophobic residues during extraction of the thiazolinones [7,8]. Some of these problems have now been solved by the use of the more volatile pentafluoropropionic acid for cleavage and a mixture of 1,2-dichloroethane and benzene to extract the thiazolinones. Consequently it has been possible to derive useful information for up to 73 cycles.

2. Materials and methods

2.1. *Materials*

Eggs were from commercial Peking ducks. The major high-lipid lipoprotein was isolated from the yolk and the lipid removed essentially as described for the preparation of apovitellenin I from the yolk of hen's eggs [2]. Molecular weight of the protein, from sedimentation equilibrium measurements, was found to be 9500. Amino acid analyses indicated that the molecule contained the following residues: Asx 5.9, Thr 7.1, Ser 3.0, Glx 8.4, Pro 3.0, Gly 2.9,

Ala 6.8, Val 6.8, Met 1.8, Ile 5.7, Leu 9.1, Tyr 4.0, Phe 3.1, Lys 6.1, Trp 1.9, Arg 6.1 and 5 amide groups [9].

Three cyanogen bromide fragments were obtained from duck's apovitellenin I by a procedure adapted from Mills et al. [10]. They were separated chromatographically on a column (60 × 2.2 cm) of Sephadex G-75 in 70% (v/v) formic acid. The C-terminal peptide, which had no homoserine, contained the following amino acids and molar ratios: Lys 3.2, Asx 0.9, Thr 0.8, Glx 0.8, Gly 1.0, Val 1.0, Ile 0.9, Leu 1.8, Tyr 1.6, Trp 0.8.

Sequenator reagents, ethyl acetate, 1-chlorobutane, 1,2-dichloroethane, *n*-propanol, and pentafluoropropionic acid were purified according to Edman and Begg [5]. Benzene (B.D.H., A.R.) was purified by crystallization (Edman, private communication). Commercial Quadrol (Wyandotte, USA) with a low aldehyde content, phenylisothiocyanate (Fluka, puriss), and heptane (B.D.H. laboratory reagent) were not purified further.

2.2. *Sequenator procedure*

Sequenator analyses of the protein (0.8 μmole) were carried out as described previously for peptides (procedure 3) [11] with the following changes:

(1) Quadrol (0.5 M) was used for coupling.

(2) The cup was stopped during addition of phenylisothiocyanate.

(3) Benzene extraction was extended from 140–240 s.

(4) Extraction was with ethyl acetate at stage 11 until residue 63 but thereafter with 1,2-dichloroethane-benzene (3 : 1).

(5) Cleavage was effected with pentafluoropropionic acid, Stage 16, 225 s.

(6) Thiazolinones were extracted with 1,2-dichloroethane-benzene (3 : 1) and the time was extended from 140–200 s to collect the same volume of extractant.

(7) All samples from the sequenator were dried in a stream of nitrogen and redissolved in 100 μl of 1,2-dichloroethane and 1 μl of pentafluoropropionic acid containing dithiothreitol (5 mg/ml). Appropriate volumes were loaded onto a thin-layer plate and heat converted to the phenylthiohydantoin prior to chromatography [12]. When desirable the remainder of each sample was dried, hydrolysed in hydriodic acid (0.2 ml, in vacuo, 140°C, 24 h) and analysed with a Beckman 120C amino acid analyzer fitted with a scale expander. From residue 63 onwards the total sample was hydrolysed for identification. For analysis of the 14-residue cyanogen bromide fragment (0.5 μmole), succinylated polyornithine (5 mg) was added to the cup. Cleavage was with pentafluoropropionic acid as for the protein procedure.

2.3. *Carboxypeptidase digestion*

Digestion of the cyanogen bromide fragments (residues 30–68 and 69–82) with carboxypeptidase A (Worthington) was carried out in 0.1 M ammonium bicarbonate at 25°C for periods of 1 and 16 h. After treatment with 10% trichloracetic acid followed by centrifugation the supernatant was removed, freeze dried and analyzed for amino acids.

3. Results and discussion

The increased yields obtained on extraction of the polar amino acid derivatives with dichloroethane–benzene instead of chlorobutane are illustrated in fig.1. It shows the scans of the thin-layer chromatography plates from two sequenator runs for residues 4, 5 and 6 of duck's apovitellenin I, i.e., for Phe, Glu and Arg. The arginine derivative was barely detectable after chlorobutane extraction but was easily identified on the thin-layer plate after dichloroethane–benzene extraction. The extra peak (in the lower trace) ahead of the glutamic acid phenylthiohydantoin is also a glutamic acid derivative.

Table 1 gives the sequence of duck's apovitellenin I and, for comparison, those of the emu and hen [3,4]. The duck sequence was derived from sequenator analyses of the whole protein (to residue 73) and of the overlapping C-terminal cyanogen bromide fragment (69–82). This information was supplemented by carboxypeptidase digestion of the C-terminal and the preceding fragment. Up to 62 the residues were identified directly from thin layer plates, and confirmed where desirable by amino acid analyses, although serine was assigned to 58 in the absence of evidence for other residues. From 63–66 the evidence was largely from amino acid analyses. Because only 4 of the 5 amide groups had been established elsewhere, either residue 65 or 66 was necessarily an amide. The sequence Glu–Gln was thus chosen on the grounds of

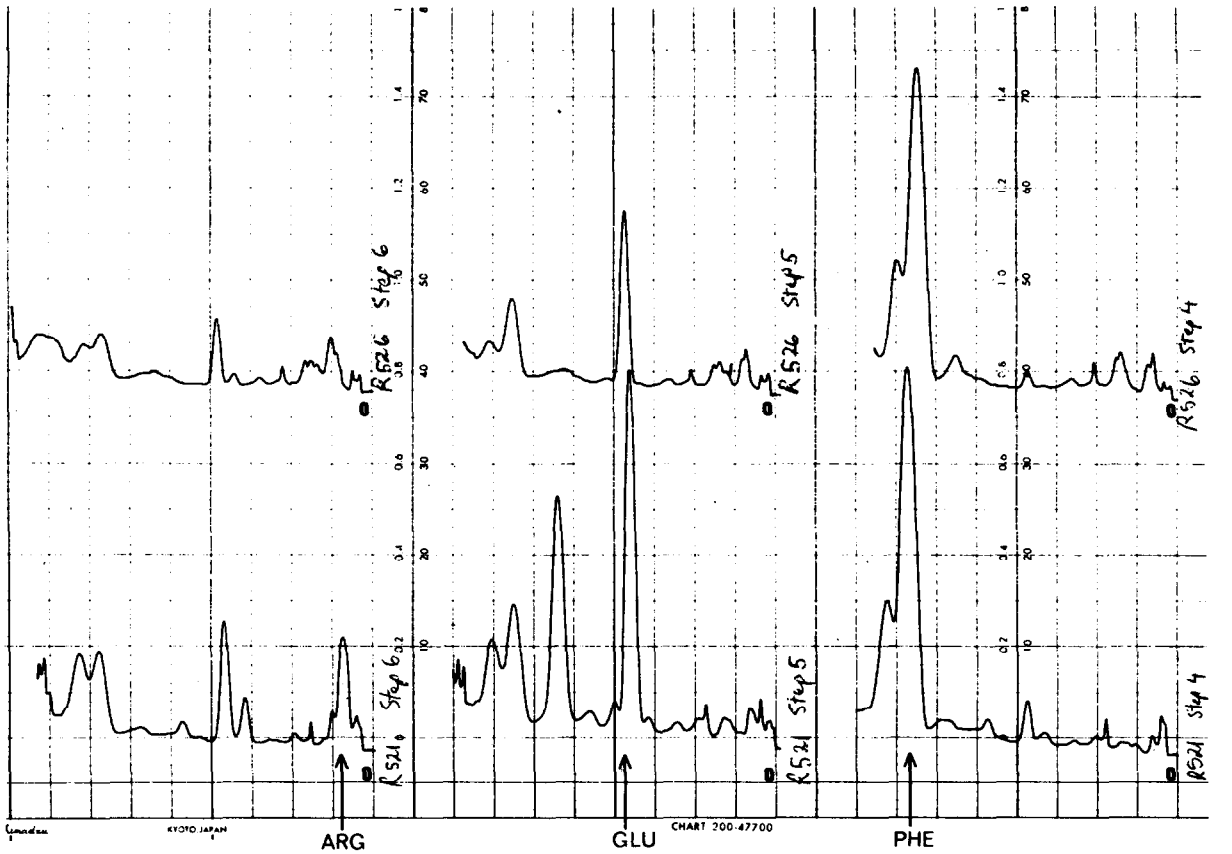

Fig.1. Traces from a Shimadzu TLC scanner after heat conversion and thin layer chromatography [12] of Arg, Glu and Phe derivatives obtained from 2 sequenator runs. The positions of the phenylthiohydantoins are indicated and illustrate the increased yields on extraction of the thiazolinones with 1,2-dichloroethane–benzene (3 : 1) (lower traces) instead of 1, chlorobutane (upper traces). 0 = loading position.

homology with the apovitellenins of the emu and the hen.

The above determination of the sequence of duck's apovitellenin was facilitated by modifying the procedure of Edman and Begg [5] to include a more efficient solvent (dichloroethane–benzene) for extraction and a more volatile acid (pentafluoropropionic acid) for cleavage. Identification of residues was also helped by heat conversion of thiazolinones to thiohydantoins [12]. The dichloroethane–benzene extraction has also been effective in improving yields of acidic residues during the sequencing of polar peptides with the peptide procedure [11]. As illustrated (fig.1) the combined procedures are useful for TLC identification of basic residues as well as the less polar residues in proteins. The phenylthiohydantoin of arginine is usually difficult to determine after more than 30 degradation cycles [13] but at residue 55 arginine was clearly identified from amino acid analysis data.

In sequencing the C-terminal cyanogen bromide fragment (14 residues) succinylated polyornithine [14] was added because we have found it assists in holding small amounts of hydrophobic peptides in the cup. Our results are apparently similar to those of Rochat et al. [15] who recommend parvalbumin as a carrier for peptides and so suggest that the success of their method depends on elimination of ethyl acetate extraction with the dimethylbenzylamine procedure [16].

The amino acid sequence of duck's apovitellenin I

Table 1
Amino acid sequences of apovitellenin I from duck, emu and hen egg yolks

	1				5						10
Duck	Lys	Ser	Ile	Phe	Glu	Arg	Asp		Arg	Arg	Asp
Emu	Lys	Ser	Ile	Phe	Glu	Arg	Asp	Asn	Arg	Arg	Asp
Hen	Lys	Ser	Ile	Ile	Asp	Arg	Glu		Arg	Arg	Asp
	11				15						20
Duck	Trp	Leu	Val	Ile	Pro	Asp	Ala		Ile	Ala	Ala
Emu	Trp	Leu	Val	Ile	Pro	Asp	Ala		Val	Ala	Ala
Hen	Trp	Leu	Val	Ile	Pro	Asp	Ala		Ala	Ala	Ala
	21				25						30
Duck	Tyr	Ile	Tyr	Glu	Thr	Val	Asn		Lys	Met	Ser
Emu	Tyr	Val	Tyr	Glu	Thr	Val	Asn		Lys	Met	Phe
Hen	Tyr	Ile	Tyr	Glu	Ala	Val	Asn		Lys	Val	Ser
	31				35						40
Duck	Pro	Arg	Val	Gly	Gln	Phe	Leu		Ala	Asp	Ala
Emu	Pro	Lys	Val	Gly	Gln	Phe	Leu		Ala	Asp	Ala
Hen	Pro	Arg	Ala	Gly	Gln	Phe	Leu		Leu	Asp	Val
	41				45						50
Duck	Ala	Gln	Thr	Pro	Val	Val	Val		Gly	Thr	Arg
Emu	Ala	Gln	Ile	Pro	Val	Ile	Val		Gly	Thr	Arg
Hen	Ser	Gln	Thr	Thr	Val	Val	Ile		Gly	Ser	Arg
	51				55						60
Duck	Thr	Phe	Leu	Ile	Arg	Glu	Thr		Ser	Lys	Leu
Emu	Asn	Phe	Leu	Ile	Arg	Glu	Thr		Ser	Lys	Leu
Hen	Asn	Phe	Leu	Ile	Asn	Glu	Thr		Ala	Arg	Leu
	61				65						70
Duck	Thr	Leu	Leu	Ala	Glu	Gln	Leu		Met	Glu	Lys
Emu	Ser	Ile	Leu	Ala	Glu	Gln	Met		Met	Glu	Lys
Hen	Thr	Lys	Leu	Ala	Glu	Gln	Met		Leu	Glu	Lys
	71				75						80
Duck	Ile	Lys	Asn	Leu	Trp	Tyr	Thr		Lys	Val	Leu
Emu	Val	Lys	Thr	Leu	Trp	Asn	Thr		Lys	Val	Leu
Hen	Ile	Lys	Asn	Leu	Cys	Tyr	Thr		Lys	Val	Leu
	81	82									
Duck	Gly	Tyr									
Emu	Gly	Tyr	Tyr								
Hen	Gly	Tyr									

(←→) Denotes residues identified by sequenator for duck's apovitellenin I (1–73) and for the C-terminal cyanogen bromide peptide (69–81). (←– – –) From carboxypeptidase A digestion of cyanogen bromide fragments.

(table 1) is similar to those from the hen and the emu. The hen and duck proteins have only 82 residues, both proteins lacking the Asn at residue 18 and the C-terminal tyrosine of the emu protein. In duck's apovitellenin I there are 69 residues identical to those of the emu's, the remaining 13 being the result of conservative changes that could arise from a single base change in the codon. These proteins are clearly homologous although the rate of mutation appears to be high, about 13/100 residues/10^8 years assuming that divergence took place 1.3×10^8 years ago [17]. When duck's apovitellenin I is compared with hen's there are 21 changed residues. Of these, 4 require 2 base changes (18, Ile for Ala; 38, Ala for Leu; 55, Arg for Asn; 62, Leu for Lys). Possibly the ancestors of these species diverged at an earlier date. Alternatively, duck's and hen's apovitellenin I may not be truly orthologous as defined by Fitch [18]. Nevertheless, the changes are in general conservative and suggest that for their role in the lipoprotein the apovitellenins from the avian species require a reasonably specific structural arrangement.

Acknowledgements

We are very grateful to Mr M. R. Rubira for excellent technical assistance and to Dr C. M. Roxburgh and Mr D. T. W. McMahon for the amino acid acid analyses.

References

[1] Burley, R. W. (1973) Biochemistry 12, 1464–1470.
[2] Burley, R. W. (1975) Aust. J. Biol Sci. 28, 121–132.
[3] Dopheide, T. A. A. and Inglis, A. S. (1974) Aust. J. Biol. Sci. 27, 15–21.
[4] Dopheide, T. A. A. and Inglis, A. S. (1976) Aust. J. Biol. Sci. 29, in press.
[5] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[6] Niall, H. (1973) in: Methods in Enzymology; Enzyme Structure Part D (Hirs, C. H. W. and Timasheff, S. N. eds) Vol. 27, pp. 942–1010, Academic Press, New York.
[7] O'Donnell, I. J. and Inglis, A. S. (1974) Aust. J. Biol. Sci. 27, 369–382.
[8] Crewther, W. G., Inglis, A. S., Nicholls, P. W., O'Donnell, I. J., Roxburgh, C. M. and Sparrow, L. G. (1974) Proc. Aust. Biochem. Soc. 7, 2.
[9] Inglis, A. S., Roxburgh, C. M. and Takayanagi, H. (1974) Anal. Biochem. 61, 25–31.
[10] Mills, J. B., Howard, S. C., Scapa, S. and Wihelmi, A. E. (1970) J. Biol. Chem. 245, 3407–3415.
[11] Crewther, W. G. and Inglis, A. S. (1975) Anal. Biochem. 68, 572–595.
[12] Inglis, A. S. (1976) J. Chromatogr. 123, 482–485.
[13] Wittmann-Liebold, B. (1973) Hoppe-Seyler's Z. Physiol. Chem. 354, 1415–1431.
[14] Silver, J. and Hood, L. E. (1974) Anal. Biochem. 60, 235–292.
[15] Rochat, H., Bechis, G., Kopeyan, C., Gregoire, J. and Rietschoten, J. V. (1976) FEBS Lett. 64, 404–408.
[16] Hermodson, M. A., Ericcson, L. H., Titani, K., Neurath, H. and Walsh, K. A. (1972) Biocehmistry 11, 4493–4502.
[17] Fisher, J. (1967) in: The Fossil Record (Harland, W. B. and Holland, C. H. eds) pp. 133–154, The Geological Society, London.
[18] Fitch, W. M. (1973) Ann. Rev. Gen. 7, 343–380.