Complete Amino Acid Sequence of the Goose-Type Lysozyme from the Egg White of the Black Swan†

Richard J. Simpson,* Geoffrey S. Begg, Donna S. Dorow, and Francis J. Morgan

ABSTRACT: The complete amino acid sequence of the goose-type lysozyme from the egg white of the Australian black swan (*Cygnus atratus*) has been determined largely from analysis of CNBr fragments and tryptic peptides derived from the citraconylated protein. There are 185 amino acid residues in the polypeptide chain; the molecular weight calculated from the sequence analysis is 20 400. The amino acid sequence of black swan goose-type lysozyme is Arg-Thr-Asp-Cys-Tyr-Gly-Asn-Val-Asn-Arg-Ile-Asp-Thr-Thr-Gly-Ala-Ser-Cys-Lys-Thr-Ala-Lys-Pro-Glu-Gly-Leu-Ser-Tyr-Cys-Gly-Val-Pro-Ala-Ser-Lys-Thr-Ile-Ala-Glu-Arg-Asp-Leu-Lys-Ala-Met-Asp-Arg-Tyr-Lys-Thr-Ile-Ile-Lys-Lys-Val-Gly-Glu-Lys-Leu-Cys-Val-Glu-Pro-Ala-Val-Ile-Ala-Gly-Ile-Ile-Ser-Arg-Glu-Ser-His-Ala-Gly-Lys-Val-Leu-Lys-Asn-Gly-Trp-Gly-Asp-Arg-Gly-Asn-Gly-Phe-Gly-Leu-Met-Gln-Val-Asp-Lys-Arg-Ser-His-Lys-Pro-Gln-Gly-Thr-Trp-Asn-Gly-Glu-Val-His-Ile-Thr-Gln-Gly-Thr-Thr-Ile-Leu-Thr-Asp-Phe-Ile-Lys-Arg-Ile-Gln-Lys-Lys-Phe-Pro-Ser-Trp-Thr-Lys-Asp-Gln-Gln-Leu-Lys-Gly-Gly-Ile-Ser-Ala-Tyr-Asn-Ala-Gly-Ala-Gly-Asn-Val-Arg-Ser-Tyr-Ala-Arg-Met-Asp-Ile-Gly-Thr-Thr-His-Asp-Asp-Tyr-Ala-Asn-Asp-Val-Val-Ala-Arg-Ala-Gln-Tyr-Tyr-Lys-Gln-His-Gly-Tyr. There is no sequence homology between the goose-type lysozyme and the other avian egg white lysozyme, chick-type lysozyme; apparently, these related enzymes are products of convergent evolution from originally distinct ancestral genes.

Two distinct forms of lysozyme (*N*-acetylmuramide glycanohydrolase, EC 3.2.1.17) occur in comparable amounts in the egg white of the black swan, *Cygnus atratus* (Arnheim & Steller, 1970). One of these, designated "chick type", is homologous with the lysozyme found in the egg white of the domestic chicken and many other species. The other form of black swan lysozyme, designated "goose type" (Arnheim et al., 1973), is very similar to the enzyme found in the egg white of the Embden goose (Canfield & McMurry, 1967; Dianoux & Jollès, 1967) and of the ostrich (Jollès et al., 1977). Immunological studies have shown that the goose-type lysozyme has a much broader taxonomic distribution than has the chick-type lysozyme in bird egg white (Prager et al., 1974). Only the black swan and the Canada goose, both species of the order anseriformes, have been shown to have both goose-type and chick-type lysozymes in their egg white. Chick-type lysozymes, which all have a molecular weight of ~14 000, exhibit a close similarity in amino acid sequence [see Dayhoff (1972)], immunological properties, and the structure of their active site (Sharon et al., 1974). The goose-type lysozymes have a molecular weight of ~21 000 and do not cross-react immunologically with the chick-type lysozymes (Arnheim et al., 1973). The partial amino-terminal sequences of the black swan goose-type lysozyme (Arnheim et al., 1973), Embden goose lysozyme (Canfield et al., 1971), and the ostrich lysozyme (Jollès et al., 1977) indicate that these goose-type lysozymes are structurally very similar but very different from all of the chick-type lysozymes. This paper describes the complete amino acid sequence of the black swan goose-type lysozyme (SELg)[1] which confirms these preliminary findings.

Experimental Procedure

Materials. Trypsin (code TRTPCK-9HA) and carboxypeptidase A (code 10ADFP) were purchased from Worthington. Cyanogen bromide was purchased from Ajax Chemicals (Australia). Iodoacetic acid and 2-iodosobenzoic acid were from Fluka. Hydroxylamine hydrochloride was obtained from British Drug Houses, and guanidine hydrochloride was from Heico Inc. Citraconic anhydride was from Pierce Chemical Co. Hexadimethrine bromide (polybrene) was obtained from Sigma. Iodo[2-^{14}C]acetic acid (57 mCi/mmol) was purchased from the Radiochemical Centre (Amersham, U.K.).

Methods. Lysozyme activity was measured by lysis of a suspension of killed *Micrococcus lysodeikticus* cells (Parry et al., 1965). A unit of enzyme activity is arbitrarily defined as the increment in percent transmission at 540 nm from 30 s to 3 min.

Column chromatography, S-^{14}C-labeled carboxymethylation, citraconylation, cyanogen bromide cleavage, enzymatic digestions, and amino acid analyses were performed according to standard procedures, which have been described elsewhere (Begg et al., 1978). Additional selective chemical cleavages of proteins were carried out at asparaginyl–glycine bonds with hydroxylamine (Bornstein & Balian, 1977) and at tryptophanyl bonds with 2-iodosobenzoic acid (Mahoney & Hermodson, 1979).

Peptide mixtures resulting from enzymatic and chemical cleavage were invariably fractionated initially on Sephadex G-50 (superfine). High-voltage paper electrophoresis (pH 2.1, 0.61 M formic acid) of pooled chromatographic fractions, using portions estimated to contain 100–200 nmol of peptide or peptides, was employed routinely on Whatman No. 3 chromatography paper to ascertain the purity of peptide fractions and to isolate peptides in quantities sufficient for compositional and sequence analysis. Electrophoresis was carried out for 2 h at 4.5 kV (48 V cm^{-1}), and for preparative purposes peptides were eluted with 1 M acetic acid; peptides were located by sequential staining of guide strips with cadmium–ninhydrin reagent and Ehrlich's reagent (Easley et al.,

† From St. Vincent's School of Medical Research, Fitzroy, Melbourne, Victoria 3065, Australia. *Received October 12, 1979.* Work in the laboratory is supported in part by a grant from the National Health and Medical Research Council of Australia; assistance from the Utah Foundation is gratefully acknowledged.

[1] Abbreviations used: SELg, black swan egg lysozyme (goose type); SCM, *S*-carboxymethyl; Pth, phenylthiohydantoin; CNBr, cyanogen bromide; $NaDodSO_4$, sodium dodecyl sulfate.

0006-2960/80/0419-1814$01.00/0

Table I: Amino Acid Compositions of Fragments Produced by CNBr Cleavage of *S*-Carboxymethyl-Black Swan Lysozyme (Goose Type)

amino acid	CNBr fragments[a]					SCM-SELg[b] residues/mol
	CN1	CN2	CN3	CN4	total	
Lys	3.7 (4)	5.7 (6)	7.0 (7)	1.1 (1)	(18)	17.8
His		1.0 (1)	2.0 (2)	1.9 (2)	(5)	4.7
Arg	3.0 (3)	3.0 (3)	3.9 (4)	1.0 (1)	(11)	10.9
Asp	4.8 (5)	3.9 (4)	5.9 (6)	5.0 (5)	(20)	19.9
Thr[c]	4.8 (5)	1.0 (1)	5.6 (6)	2.2 (2)	(14)	13.1
Ser[d]	2.8 (3)	1.9 (2)	3.8 (4)		(9)	8.9
Glu	1.9 (2)	3.0 (3)	7.0 (7)	2.2 (2)	(14)	14.6
Pro	2.0 (2)	1.0 (1)	1.9 (2)		(5)	5.2
Gly	3.9 (4)	7.5 (8)	7.0 (7)	2.1 (2)	(21)	20.2
Ala	4.9 (5)	2.8 (3)	3.9 (4)	3.2 (3)	(15)	14.7
Cys[e]	2.6 (3)	0.7 (1)			(4)	4.3
Val	2.0 (2)	3.7[g] (4)	3.1 (3)	1.5 (2)	(11)	10.4
Met[f]	0.7 (1)	0.7 (1)	0.8 (1)		(3)	2.7
Ile	1.9 (2)	4.2[g] (5)	4.7[g] (5)	0.9 (1)	(13)	12.6
Leu	2.0 (2)	2.8 (3)	2.0 (2)		(7)	7.2
Tyr	2.0 (2)	1.0 (1)	2.1 (2)	3.9 (4)	(9)	9.1
Phe		1.0 (1)	2.0 (2)		(3)	3.1
Trp[h]		0.4 (1)	1.6 (2)		(3)	3.1
total residues	45	49	66	25	185	
residue no. in sequence	1–45	46–94	95–160	161–185		
yield (%)	41	30	44	80		

[a] Values given are for 24-h total acid hydrolysis. The values in parentheses are the number of residues determined by sequencing. [b] Arnheim et al. (1973) average or extrapolated values from 24-, 48-, and 72-h hydrolyses except for isoleucine; isoleucine was found to increase to 12.6 residues/mol after 120 h, and the composition has been modified accordingly. [c] Corrected for 5% loss during hydrolysis. [d] Corrected for 10% loss during hydrolysis. [e] Determined as *S*-(carboxymethyl)cysteine. [f] Summation of homoserine plus homoserine lactone. [g] 70-h hydrolysis. [h] Determined by analysis of a sample hydrolyzed in 4 M methanesulfonic acid (Simpson et al., 1976).

1969). Yields of peptide are expressed as percentages of the total mass originally digested; where only a portion of the total digest was subjected to final purification by high-voltage paper electrophoresis, the absolute amount of peptide recovered was multiplied by the appropriate factor in order to calculate the final yield.

Automated amino acid sequence analysis of proteins and peptides was performed essentially as previously described (Begg et al., 1978). With the intact lysozyme a sample of 200 nmol was used and the Pth amino acids were identified qualitatively by thin-layer chromatography. In all other cases a sample of ~20 nmol was degraded together with 5 mg of polybrene as carrier (Klapper et al., 1978) and the Pth derivatives were identified and quantitated by high-performance liquid chromatography (Zimmerman et al., 1977); Pth-His and Pth-Arg were similarly identified, but quantitation was not attempted as the internal standard (Pth-S-MeCys) did not remain in the aqueous phase.

Purification of Black Swan Egg White Lysozyme (Goose Type). The purification of SELg was performed by modification of the previous procedure (Arnheim et al., 1973). Batchwise absorption of lysozyme activity from the egg white to CM-cellulose and subsequent elution with 0.4 M ammonium carbonate (pH 9.46) were performed as previously described. The crude lysozyme solution eluted from the CM-cellulose was adjusted to 37% ammonium sulfate by the addition of solid salt with stirring, and the precipitate that formed was allowed to settle overnight. This step removed inactive proteins, including a basic protein which copurifies with goose-type lysozyme and tends slowly to precipitate during subsequent chromatographic procedures (Dianoux & Jollès, 1967). Lysozyme activity was concentrated by adjusting the 37% ammonium sulfate supernatant to 90% ammonium sulfate by the addition of solid salt. The resulting precipitate which contains all the recoverable lysozyme activity was dissolved in a small volume of 0.1 M ammonium carbonate, pH 9.5, and chromatographed on a column of Sephadex G-50 (fine) in the same buffer (Figure 1). Pool I contains goose-type lysozyme, and pool II contains the chick-type lysozyme. SELg prepared in this fashion is homogeneous by the criteria of ion-exchange chromatography on CM-cellulose, electrophoresis in 10% $NaDodSO_4$–polyacrylamide gels, and a unique NH_2-terminal sequence. The amino acid composition is identical with that previously reported (Table I). However, at 120 h of hydrolysis a value of 12.6 residues of isoleucine (cf. 10.9 at 72 h) is obtained (Table I, footnote *b*) which corresponds more closely to the value obtained from the sequence analysis and from the sum of individual peptides.

Peptide Nomenclature. The following prefixes are used to denote the origin of various peptides: CN, cyanogen bromide cleavage; CT, peptides isolated from a tryptic digest of citraconylated SCM-SELg; HA, hydroxylamine cleavage; W, cleavage at tryptophan residues with 2-iodosobenzoic acid. In the case of CN and CT cleavages, where a complete set of peptides was characterized, the peptides are numbered in the order of their position in the final sequence proposal. In the case of the individual HA and W peptides used to complete the sequence determination, arbitrary numbering has been used.

Results

The important features of the isolation and characterization of the peptides used in the sequence determination follow in this section. Detailed results are given in the supplementary material (see paragraph at end of paper regarding supplementary material).

Cyanogen Bromide Cleavage of S-Carboxymethyl-Black Swan Lysozyme (Goose Type). SELg contains three methionine residues (Table I), and thus four peptides should result from CNBr cleavage of the molecule. Gel filtration on Sephadex G-50 (superfine) of CNBr-treated SCM-SELg resulted in two pools containing the single components CN3 and CN4 and a further pool containing CN1 and CN2 together with contamination from CN3. Ion-exchange chromatography of this fraction on SP-Sephadex C-25 resulted in the isolation of CN1 and CN2 in pure form. The amino acid compositions

Table II: Amino Acid Compositions of Tryptic Peptides of Citraconylated *S*-Carboxymethyl-Black Swan Lysozyme (Goose Type)[a]

amino acids	CT1	CT2	CT3	CT4	CT5	CT6[e]	CT7[e]	CT8[e]	CT9[e]	CT10	CT11	CT12	total
Lys			2.8 (3)	1.0 (1)	3.4 (4)	1.9 (2)	1.0 (1)	2.1 (2)	4.0 (4)	0.1		0.9 (1)	18
His						1.0 (1)		1.6 (2)			0.9 (1)	0.8 (1)	5
Arg	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	0.2	11
Asp		2.9 (3)	1.1 (1)	2.0 (2)	0.4	2.0 (2)	1.9 (2)	2.0 (2)	3.1 (3)	0.1	4.8 (5)	0.3	20
Thr[b]		0.9 (1)	3.7 (4)		1.0 (1)			4.2 (5)	1.0 (1)		1.7 (2)	0.1	14
Ser[c]		0.1	3.0 (3)		1.0 (1)	1.0 (1)		1.0 (1)	1.9 (2)	1.0 (1)		0.3	9
Glu			1.9 (2)		2.0 (2)	0.9 (1)	0.9 (1)	2.8 (3)	3.1 (3)	0.1		1.9 (2)	14
Pro			1.8 (2)		0.8 (1)			0.9 (1)	1.0 (1)				5
Gly		1.1 (1)	2.9 (3)	0.2	2.0 (2)	3.0 (3)	2.8 (3)	3.0 (3)	4.0 (4)	0.2	1.0 (1)	1.2 (1)	21
Ala		0.1	3.8 (4)	1.0 (1)	2.1 (2)	1.0 (1)	0.1	0.3	3.1 (3)	1.0 (1)	2.0 (2)	1.0 (1)	15
Cys[d]		0.8 (1)	1.6 (2)		0.7 (1)								4
Val		0.9 (1)	1.1 (1)		2.3[g] (3)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)		1.9 (2)	0.1	11
Met				0.9 (1)			0.6 (1)				0.8 (1)		3
Ile			2.0 (2)		3.9[g] (5)			2.7 (3)	2.0 (2)		0.9[f] (1)		13
Leu			1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	0.9 (1)	1.0 (1)	1.0 (1)				7
Tyr		0.7 (1)	1.0 (1)		0.9 (1)				1.0 (1)	1.0 (1)	0.9 (1)	3.0 (3)	9
Phe							0.8 (1)	0.9 (1)	1.0 (1)				3
Trp[e]						0.5 (1)		0.7 (1)	0.8 (1)				3
total residues	1	9	30	7	25	15	12	27	29	4	17	9	185
yield (%)	11	12	36	23	9	21	18	12	15	34	13	37	
mobility[h]	1.0	0.57	0.52	0.83	0.66	0.78	0.63			0.57	0.52	0.52	

[a] 24-h hydrolysis unless indicated otherwise. Peptides are numbered in their final order in the sequence. Values in parentheses are the number of residues per peptide from sequence data. [b] Corrected for 5% loss during hydrolysis. [c] Corrected for 10% loss during hydrolysis. [d] Determined as *S*-(carboxymethyl)cysteine. [e] Hydrolyzed in 4 M methanesulfonic acid for 24 h (Simpson et al., 1976). [f] 72-h hydrolysis value. [g] 120-h hydrolysis value. [h] Mobility is given with respect to arginine at pH 2.1.

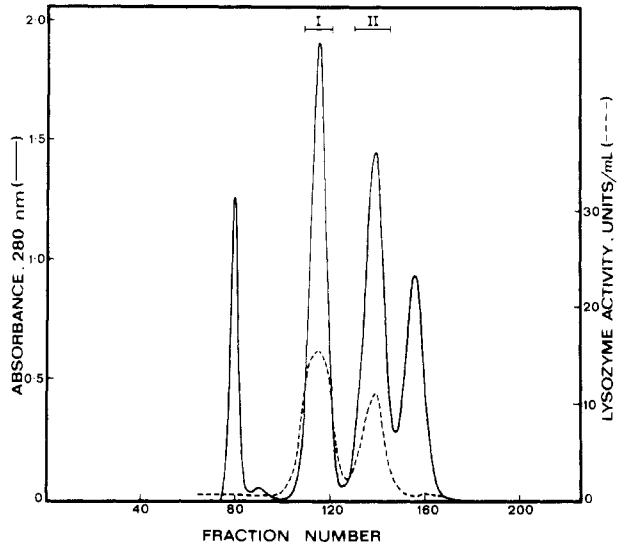

FIGURE 1: Chromatography of black swan egg white lysozyme on Sephadex G-50 fine (2.5 × 180 cm) equilibrated with 0.1 M ammonium carbonate, pH 9.5, at 4 °C. Flow rate was 40 mL/h; fraction size was 4.7 mL. Pool I contains only goose-type lysozyme and pool II only chick-type lysozyme. Sample applied was the crude lysozyme from 700 mL of egg white. Yield: goose-type lysozyme, 40 mg; chick-type lysozyme, 23 mg.

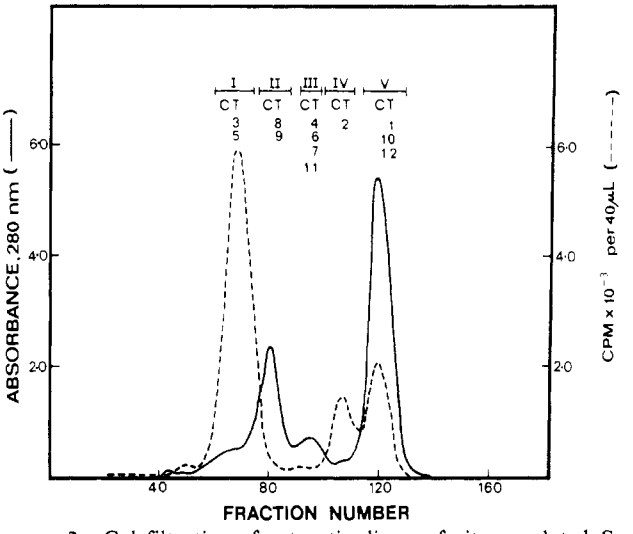

FIGURE 2: Gel filtration of a tryptic digest of citraconylated *S*-[^{14}C]carboxymethyl-black swan lysozyme (goose type) on Sephadex G-50 superfine (1.6 × 143 cm) equilibrated with 0.1 M ammonium bicarbonate at 23 °C. Flow rate was 24 mL/h; fraction size was 2.0 mL. Pooling of fractions is indicated by horizontal bars, and pool labeling is designated by Roman numerals based on their order of elution. Peptides found in each pool have been labeled in accordance with their position in the sequence proposed in Figure 3.

of the four peptides (CN1, CN2, CN3, and CN4), the sum of which accounts for the intact lysozyme molecule, are given in Table I.

Tryptic Digestion of Citraconylated S-Carboxymethyl-Black Swan Lysozyme (Goose Type). SELg contains 11 arginine residues (Table I), and hence 12 peptides are expected from tryptic cleavage of citraconylated SCM-SELg. The elution profile obtained by gel filtration on Sephadex G-50 (superfine) of a tryptic digest of 72 mg (3.25 µmol) of citraconylated SCM-SELg is shown in Figure 2. With the exception of peptides CT8 and CT9 (pool II), all other CT peptides were purified by subsequent preparative high-voltage paper electrophoresis (pH 2.1) of the pooled fractions. Following lyophilization, pool II (Figure 2) was suspended in 0.01 M potassium phosphate, pH 7.4. Peptide CT8 was insoluble in this buffer and was recovered by centrifugation and repeated washings. The soluble material was chromatographed on CM-cellulose in 0.01 M potassium phosphate (pH 7.4) by using a KCl gradient to 0.4 M, resulting in the isolation of peptide CT9. Amino acid compositions of the 12 peptides CT1–12, the sum of which accounts for the intact black swan egg lysozyme (goose type), are given in Table II.

Cleavage of S-Carboxymethyl-Black Swan Lysozyme (Goose Type) at Tryptophan Residues. SELg contains three tryptophan residues as potential cleavage sites for 2-iodosobenzoic acid (Mahoney & Hermodson, 1979). Gel filtration of the 2-iodosobenzoic acid reaction products of SCM-SELg resulted in the isolation of a fraction containing a single component (W2) corresponding to residues 135–185 of the final sequence proposal and a smaller molecular weight fraction

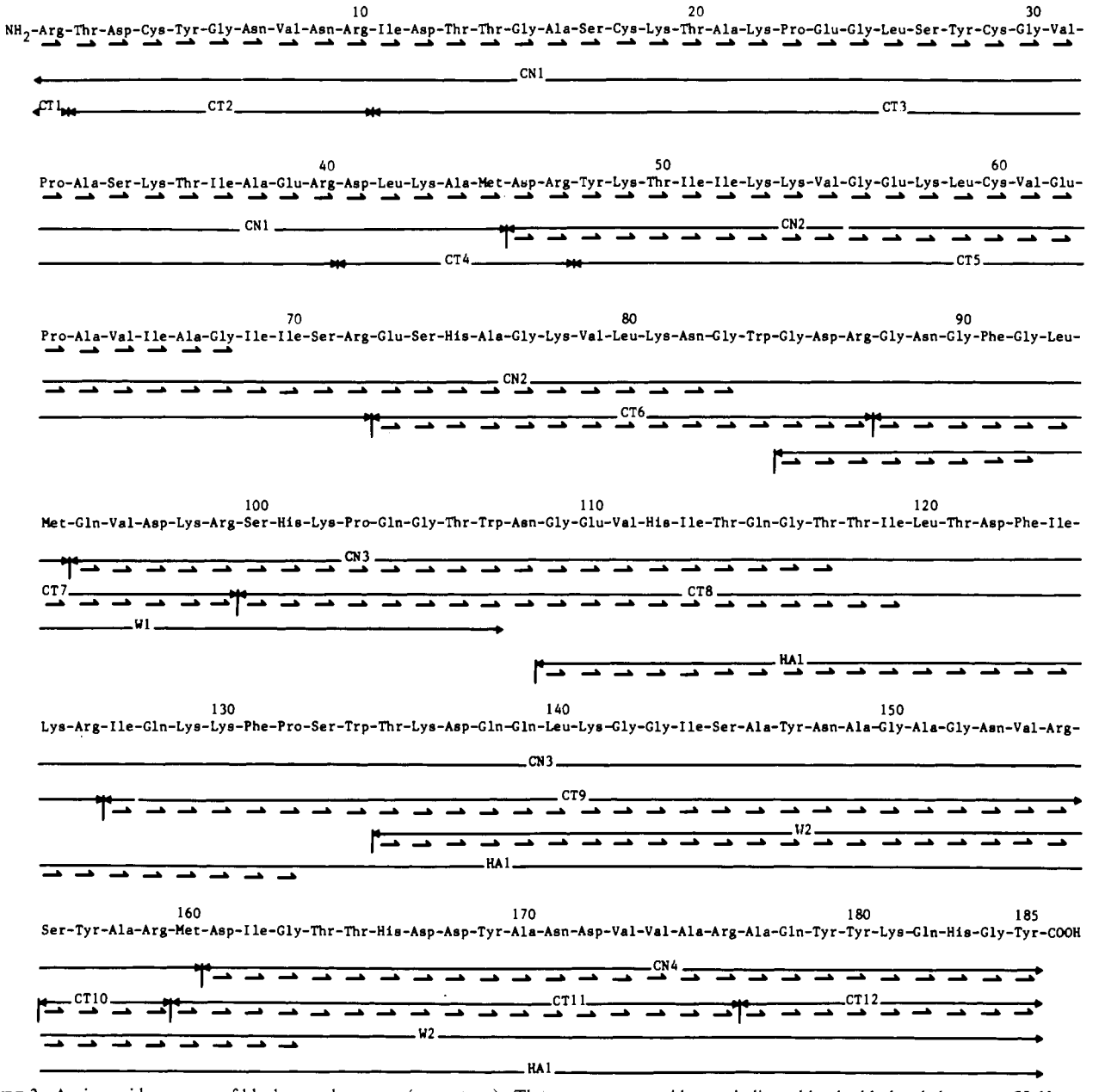

FIGURE 3: Amino acid sequence of black swan lysozyme (goose type). The component peptides are indicated by double-headed arrows. Half-arrows pointing to the right indicate sequences obtained by using the sequenator. Peptide nomenclature is given under Experimental Procedure. Subdigestions of CT6 and CT9, used to complete the sequence of these two peptides (see text), have been omitted for clarity of presentation.

from which peptide W1 (residues 85–107) was readily purified by high-voltage paper electrophoresis. The final yields of W1 and W2 were 22 and 13%, respectively. These two peptides were used in the amino acid sequence determination; other peptide fractions were not subjected to further study.

Hydroxylamine Cleavage of S-Carboxymethyl-Black Swan Lysozyme (Goose Type). A hydroxylamine-generated peptide (HA1) corresponding to residues 109–185 was isolated as a single component by gel filtration on Sephadex G-50 of the hydroxylamine-treated SCM-SELg. Other hydroxylamine-generated peptides were recovered from other fractions of this chromatography and identified but were not used in the subsequent sequence determination.

Amino Acid Sequence Determination. The majority of the sequence was determined by automated degradation of the intact molecule and peptides produced from cyanogen bromide cleavage and tryptic digestion of citraconylated SCM-SELg. The small portion of the sequence not determined from these peptides and the information necessary to order the tryptic peptides were derived from peptides HA1, W1, and W2, described in the preceding two paragraphs. The proposal for the complete amino acid sequence of the swan lysozyme (goose type) and a summary of the strategy used in determining the sequence are given in Figure 3.

Degradation of the intact chain established the sequence to residue 68 and provided the overlap between CN1 and CN2. CN4 was assigned as the COOH-terminal cyanogen bromide fragment as it did not contain homoserine. Sequence analysis of CN2 extended the sequence to residue 83, with an overlap into CT6. The sequences of CT7, CT10, CT11, and CT12 were completely determined by continuous automated degradation. Partial sequences of 11 and 26 residues, respectively, were obtained for CT6 and CT9. These peptides were redigested with trypsin following deblocking, and the peptide mixtures were fractionated by high-voltage paper electrophoresis; in each case the COOH-terminal arginine-containing peptide was sequenced to complete the characterization of the parent peptide. In the case of CT8, a partial amino-terminal

sequence, encompassing the first 20 residues, was obtained. These data provided evidence for the unaligned amino acid sequence of the lysozyme molecule with the exception of the region encompassing residues 120–126.

Sequence analysis of CN3 confirmed the sequence from residues 95 to 117 and provided the overlap between CT7 and CT8. A sequence of 24 steps on the hydroxylamine peptide HA1 provided the remaining sequence from residues 120 to 126, together with the ordering of CT8 and CT9. Sequence data from peptide W1 allowed the ordering of CT6 and CT7 and those from W2 the ordering of CT9, CT10, and CT11. The remaining tryptic peptide, CT12, which did not contain arginine was assigned as the COOH-terminal peptide. The sequence of W2 also gave confirmatory evidence of the alignment of CN3 and CN4, while the sequence of CN4 in turn confirmed the alignment of CT11 and CT12. Peptides CT1–5 were not subjected to sequence determination; they were aligned on the basis of the sequence of the intact molecule and of CN2, and their individual compositions supported the sequence data in this region.

When intact SCM-SELg was digested with carboxypeptidase A with an enzyme/substrate ratio of 1:55 at 35 °C, tyrosine only was released at digestion times up to 1 h; a yield of 0.8 residue of tyrosine per mol of SCM-SELg was attained after 15 min of digestion. Likewise, carboxypeptidase A digestion of peptide CN4 at an enzyme/substrate ratio of 1:100 at 37 °C released only tyrosine at digestion times up to 2 h, reaching a yield of 0.9 residue/mol of peptide after 40 min of digestion. These data confirm the assignment of tyrosine as the COOH-terminal residue as evidenced by automatic Edman degradation of CT12 and CN4.

Discussion

It is now known that at least two radically distinct forms of lysozyme occur in the egg white of birds; one form (goose-type lysozyme) is typified by the lysozyme found in the egg white of the Embden goose (Dianoux & Jollès, 1967; Canfield & McMurry, 1967) and the other form (chick-type lysozyme) is typified by the enzyme found in the egg white of the domestic hen (Canfield, 1963). These two forms of lysozyme bear no structural or immunological resemblance and, although they both display classical muramidase activity, there are significant differences in their enzymic specificity and inhibition properties (Canfield et al., 1971; Jollès et al., 1968; Arnheim & Steller, 1970). In contrast to most avian species, which possess only one or the other of these lysozymes in their egg white, the Australian black swan, *C. atratus*, and the Canada goose have a mixture of both chick-type lysozyme and goose-type lysozyme in their egg white (Arnheim, 1974). A large body of work has been devoted to structural analysis of the chick-type lysozyme, including amino acid sequence determination [see Dayhoff (1972)] and X-ray crystallography of the enzyme from several species (Phillips, 1974; Jollès et al., 1974).

For some time the chick-type enzyme was considered to be the most common form of lysozyme in the egg white of avian species; however, recent immunological studies have demonstrated that goose-type lysozyme is more widely distributed, being detected in nine different orders of birds compared with only two orders of birds for chick-type lysozyme (Prager et al., 1974).

The results presented in this communication permit a proposal for the primary structure of goose-type lysozyme from the egg white of the Australian black swan (Figure 3). There are 185 amino acid residues in the polypeptide chain, and the amino acid composition, determined from the sequence, is in close agreement with that obtained from quantitative amino acid analysis of the protein (Arnheim et al., 1973). One feature of the sequence (Figure 3) is the uneven distribution of amino acid residues in the polypeptide chain; all 4 half-cystine residues are situated in the NH_2-terminal half of the chain, and there is a high predominance of aromatic amino acids in the COOH-terminal half. Another feature is the unusually high incidence of paired amino acids (total of 12) in the sequence, 8 of which are situated in the COOH-terminal half of the molecule. The molecular weight of SELg, calculated from the sequence analysis, is 20 400, a value which is in good agreement with that obtained by sodium dodecyl sulfate gel electrophoresis (Arnheim et al., 1973).

The most salient feature of the sequence of SELg presented here (Figure 3) is that it displays no structural similarity to that of the chick-type lysozyme from a variety of species (Dayhoff, 1972). Increasingly, gene duplication is invoked as a mode of evolution of similar groups of polypeptides (Dayhoff & Barker, 1972); the functionally related egg white lysozymes are really a striking exception as they appear to be products of convergent evolution from originally distinct ancestral genes. Furthermore, SELg displays no sequence homology with the other egg white proteins ovalbumin (McReynolds et al., 1978), ovomucoid (Kato, 1978), and transferrin (Dayhoff, 1972). No sequence homology was evidenced between SELg and lysozymes from such diverse species as chalaropsis (Hash, 1974), bacteriophage T4 (Tsugita & Inouye, 1968), mouse (Riblet, 1974), and man (Canfield et al., 1971).[2]

An inspection of the sequence using the helical and β-sheet conformational parameters, P_α and P_β, of Chou & Fasman (1978) revealed that approximately 11% of the amino acid residues might be involved in helix formation and 28% in β-sheet formation. There were three possible helical segments predicted (residues 57–62, 73–81, and 136–141) and eight β-sheet segments (residues 1–5, 11–15, 65–69, 92–96, 104–108, 113–124, 160–167, and 175–180) with a high predominance of β sheet (five segments including two long stretches) in the COOH-terminal half of the molecule. As the goose-type lysozyme from the black swan is functionally related to, yet structurally different from, the chick-type enzyme and a number of other lysozymes, a comparison of their three-dimensional structures may reveal general principles governing their mode of catalysis, in the manner described for α-chymotrypsin and subtilisin (Robertus et al., 1972) and more recently for the kinases (Anderson et al., 1979). Preliminary X-ray crystallographic studies on the black swan goose-type lysozyme (Masakuni et al., 1979) reveal that the space group of this protein ($P2_1$ with one molecule of protein in the asymmetric unit; cell parameters a = 46.2 Å, b = 65.1 Å, c = 38.7 Å, β = 110°) is significantly different from that of the chick-type lysozyme (Phillips, 1974) and lysozyme from bacteriophage T4 (Matthews & Remington, 1974).

Despite several recent studies (Royal et al., 1979; Nguyen-Huu et al., 1979), the structural organization of egg white protein genes is not yet clear. However, there is some evidence that ovalbumin, ovomucoid, and chick-type lysozyme are located in different genomic regions (Nguyen-Huu et al., 1979), possibly subject to a common regulatory molecule. The se-

[2] A comparison of the sequence of black swan goose-type lysozyme with 1279 amino acid sequences by computer search of a data collection, using the mutation data scoring matrix with no added bias, has shown no evidence of any significant relationship between the goose-type lysozyme sequence and the other sequenced proteins in the collection (Dayhoff et al., 1979).

quence studies presented here should facilitate identification and localization in the genome of the DNA sequences coding for the goose-type lysozyme. This, in turn, should improve our understanding of the mode of expression and the regulation by steroid hormones of this group of functionally related genes.

Immunological studies have demonstrated that the chicken and duck, which have no goose-type lysozyme in their egg white (Arnheim et al., 1973), nevertheless possess this enzyme in their bone marrow (Hindenburg et al., 1974; Arnheim, 1974). It seems probable that many, if not all, birds possess genes for both goose-type and chick-type lysozyme but that their regulation is under separate control in different tissues. This provides a potentially valuable system for the study of differential gene expression.

Acknowledgments

We thank the Fisheries and Wildlife Department of Victoria for assistance in the collection of black swan eggs and Dr. Hugh Niall for helpful discussion.

Supplementary Material Available

Elution profiles for column purification of (a) black swan lysozymes, (b) various CNBr fragments, (c) hydroxylamine cleavage fragments, and (d) iodosobenzoic acid cleavage fragments, a representative high-performance liquid chromatogram of a standard mixture of Pth amino acids, and tables presenting the results of automated Edman degradation and a summary of amide assignments (20 pages). Ordering information is given on any current masthead page.

References

Anderson, C. M., Zucker, F. H., & Steitz, T. A. (1979) *Science 204*, 375–380.

Arnheim, N. (1974) in *Lysozyme* (Osserman, E. F., Canfield, R. E., & Beychok, S., Eds.) pp 153–161, Academic Press, New York.

Arnheim, N., & Steller, R. (1970) *Arch. Biochem. Biophys. 141*, 656–661.

Arnheim, N., Hindenburg, A., Begg, G. S., & Morgan, F. J. (1973) *J. Biol. Chem. 248*, 8036–8042.

Begg, G. S., Pepper, D. S., Chesterman, C. N., & Morgan, F. J. (1978) *Biochemistry 17*, 1739–1744.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132–145.

Canfield, R. E. (1963) *J. Biol. Chem. 238*, 2698–2707.

Canfield, R. E., & McMurry, S. (1967) *Biochem. Biophys. Res. Commun. 26*, 38–42.

Canfield, R. E., Kammerman, S., Sobel, J. H., & Morgan, F. J. (1971) *Nature (London), New Biol. 232*, 16–17.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, Washington, D.C.

Dayhoff, M. O., & Barker, W. C. (1972) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) pp 41–46, National Biomedical Research Foundation, Washington, D.C.

Dayhoff, M. O., Hunt, L. T., Barker, W. C., Schwartz, R. M., & Blomquist, M. C. (1979) *Protein Sequence Data File*, National Biomedical Research Foundation, Washington, D.C.

Dianoux, A.-C., & Jollès, P. (1967) *Biochim. Biophys. Acta 133*, 472–479.

Easley, C. W., Zegers, B. J. M., & De Vijlder, M. (1969) *Biochim. Biophys. Acta 175*, 211–213.

Hash, J. H. (1974) in *Lysozyme* (Osserman, E. F., Canfield, R. E., & Beychok, S., Eds.) pp 95–103, Academic Press, New York.

Hindenberg, A., Spitznagel, J., & Arnheim, N. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1653–1657.

Jollès, J., Périn, J.-P., & Jollès, P. (1977) *Mol. Cell. Biochem. 17*, 39–44.

Jollès, P., Saint-Blancard, J., Charlemagne, D., Dianoux, A.-C., Jollès, J., & Le Baron, J. (1968) *Biochim. Biophys. Acta 151*, 532–534.

Jollès, P., Bernier, I., Berthou, J., Charlemagne, D., Faure, A., Hermann, J., Jollès, J., Périn, J.-P., & Saint-Blancard, J. (1974) in *Lysozyme* (Osserman, E. F., Canfield, R. E., & Beychok, S., Eds.) pp 31–54, Academic Press, New York.

Kato, I. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, p 140, National Biomedical Research Foundation, Washington, D.C.

Klapper, D. G., Wilde, C. E., & Capra, J. D. (1978) *Anal. Biochem. 85*, 126–131.

Mahoney, W. C., & Hermodson, M. A. (1979) *Biochemistry 18*, 3810–3814.

Masakuni, M., Simpson, R. J., & Isaacs, N. W. (1979) *J. Mol. Biol. 135*, 313–314.

Matthews, B. W., & Remington, S. J. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 4178–4182.

McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M., & Brownlee, G. G. (1978) *Nature (London) 273*, 723–728.

Nguyen-Huu, M. C., Stratmann, M., Groner, B., Wurtz, T., Land, H., Giesecke, K., Sippel, A. E., & Schütz, G. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 76–80.

Parry, R. M., Jr., Chandan, R. C., & Shahani, K. M. (1965) *Proc. Soc. Exp. Biol. Med. 119*, 384–386.

Phillips, D. C. (1974) in *Lysozyme* (Osserman, E. F., Canfield, R. E., & Beychok, S., Eds.) pp 9–30, Academic Press, New York.

Prager, E. M., Wilson, A. C., & Arnheim, N. (1974) *J. Biol. Chem. 249*, 7295–7297.

Riblet, R. J. (1974) in *Lysozyme* (Osserman, E. F., Canfield, R. E., & Beychok, S., Eds.) pp 89–93, Academic Press, New York.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., & Wilcox, P. E. (1972) *Biochemistry 11*, 2439–2449.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J. L., Le Meur, M., Brégégègre, F., Gannon, F., Le Pennec, J. P., Chambon, P., & Kourilsky, P. (1979) *Nature (London) 279*, 125–132.

Sharon, N., Eshdat, Y., Maoz, I., Bernstein, Y., Prager, E. M., & Wilson, A. C. (1974) *Isr. J. Chem. 12*, 591–603.

Simpson, R. J., Neuberger, M. R., & Liu, T.-Y. (1976) *J. Biol. Chem. 251*, 1936–1940.

Tsugita, A., & Inouye, M. (1968) *J. Mol. Biol. 37*, 201–212.

Zimmerman, C. L., Apella, E., & Pisano, J. J. (1977) *Anal. Biochem. 77*, 569–573.