

Amino Acid Sequence and Immunological Properties of Chachalaca Egg White Lysozyme*

JACQUELINE JOLLÈS¹, FRANÇOISE SCHOENTGEN¹, and PIERRE JOLLÈS¹
ELLEN M. PRAGER² and ALLAN C. WILSON²

¹Laboratory of Proteins, University of Paris V, 45 rue des Saints-Pères,
F-75270 Paris Cedex 05, France

²Department of Biochemistry, University of California, Berkeley,
Cal. 94720, USA

Received December 16, 1975

Summary. The amino acid sequence of lysozyme c from chachalaca egg white was determined. Like other bird lysozymes c, that of the chachalaca has 129 amino acid residues. It differs from other avian lysozymes c by 27 to 31 amino acid substitutions as well as by being devoid of phenylalanine. It contains substitutions at 9 positions which are invariant in the other 7 bird lysozymes of known sequence. Although the chachalaca is classified zoologically in the order Galliformes, which includes chickens and other pheasant-like birds, its lysozyme differs more from those of pheasant-like birds than do the lysozymes c of ducks. Phylogenetic analysis of the sequence comparisons confirms that the lineage leading to chachalaca lysozyme c separated from that leading to other galliform lysozymes c before the duck lysozyme c lineage did. This indicates a contrast between protein evolution and evolution at the organismal level. Immunological comparison of chachalaca lysozyme c with other lysozymes of known sequence provides further support for the proposal that immunological cross-reactivity is strongly dependent on degree of sequence resemblance among bird lysozymes.

Key words: Lysozyme Evolution/Phylogenetic Analysis/Birds/Chachalaca/Sequence-Immunology Correlation/Micro-Complement Fixation

Although the lysozyme of chachalaca egg white is a typical lysozyme c (Prager et al., 1974c) as regards enzymic and physicochemical properties (Prager & Wilson, 1971b; Sharon

*103rd communication on lysozymes from the Laboratory of P. Jollès. Supported in part by grants from C.N.R.S. (ER 102), I.N.S.E.R.M. (Groupe de recherche U-116), N.S.F. (GB-42028X), and N.I.H. (GM-21509).

et al., 1974), it is divergent to an unexpectedly large degree from other bird lysozymes c in antigenic properties (Prager & Wilson, 1971b,c). The chachalaca is classified on the basis of anatomy and way of life in the order Galliformes, along with the chicken and other pheasant-like and quail-like birds (Wetmore, 1960; Hudson et al., 1966). Yet its lysozyme c differs more from that of other galliform birds than do the lysozymes c¹ of ducks, according to immunological tests (Prager & Wilson, 1971b,c).

Extensive immunological comparisons, made chiefly with the micro-complement fixation technique, had indicated a strong correlation between amino acid sequence resemblance and immunological cross-reactivity among lysozymes (Prager & Wilson, 1971b,c; Prager et al., 1972) as well as among other proteins such as azurins (Champion et al., 1975). The sequence-immunology correlation could be explained by proposing that the immunological effects of amino acid substitutions which have been incorporated into homologous naturally-occurring proteins are generally small, roughly equal, and approximately additive (Wilson & Prager, 1974). This explanation makes use of the X-ray observations that protein conformation is strongly conserved despite many amino acid substitutions and that protein evolution is primarily a surface phenomenon (Wilson & Prager, 1974; Champion et al., 1975). It is also possible that proteins such as bird lysozymes and bacterial azurins have many rather than just a few antigenic determinants on their surfaces (Wilson & Prager, 1974; Champion et al., 1975). However, some immunochemists, especially Atassi (1975), doubt that immunological cross reactivity could be dependent predominantly on simply the number of amino acid differences between two proteins. Such workers have found it difficult to reconcile the above findings with the belief that immunological cross-reactivity is best understood by considering the nature, location, and conformational effect of the substitutions.

Thus it was of particular importance to determine the complete chachalaca lysozyme sequence in order to find out whether the chachalaca enzyme indeed differed from the other avian lysozymes by as much or nearly as much as the sequence-immunology correlation predicted or whether, alternatively, the sequence difference was significantly smaller than predicted but that a particular substitution or substitutions had led to an unusual decrease in immunological cross-reactivity. If the first possibility were found to be true, as we predicted, we

¹Since this communication deals exclusively with lysozyme c and not at all with lysozyme g (Prager et al., 1974c), we shall henceforth simply use the term "lysozyme" to indicate the c-type lysozyme.

would have the opportunity of extending the comparison of sequence resemblance and immunological cross-reactivity beyond the range previously observed among bird lysozymes (Prager & Wilson, 1971b,c; Prager et al., 1972; Wilson & Prager, 1974; Champion et al., 1975).

Further impetus to determine the chachalaca lysozyme sequence was provided by the desire to investigate the relationship between molecular evolution and organismal evolution. Recent investigations (Wilson et al., 1974a,b; Prager & Wilson, 1975; King & Wilson, 1975) suggested that these types of evolution often proceeded independently of one another. We were therefore interested in verifying whether, at the molecular level, the chachalaca had diverged considerably from the pheasant-like birds even though on the basis of anatomy and behavior it is classified in the order Galliformes (Wetmore, 1960; Hudson et al., 1966).

MATERIALS AND METHODS

Eggs and Lysozymes. Freshly laid chachalaca (*Ortalis vetula*) eggs were obtained from Frank Wied, Realitos, Texas, and the egg whites were processed and stored as described (Prager & Wilson, 1971b). The chachalaca egg white lysozyme (E.C. 3.2.1.17) was purified according to a previously reported procedure (Prager & Wilson, 1971b). Purified phasianoid and duck lysozymes used in the comparative immunological experiments were those described (Prager & Wilson, 1971a,b, 1972). The source of guinea fowl lysozyme for micro-complement fixation was whole egg white from the collection of A.C. Wilson.

Enzymes and Reagents. Trypsin (E.C. 3.4.4.4), chymotrypsin (E.C. 3.4.4.5), and carboxypeptidases A and B (E.C. 3.4.2.1/2) were purchased from Worthington. Sephadex G-25 fine was obtained from Pharmacia and Dowex 1×2 (200-400 mesh) from Bio-Rad. All other reagents (analytical grade) were obtained from Merck or Prolabo, except mercaptoethanol (Koch-Light) and those used for the Sequencer, which were purchased from Socosi (F-94100, St. Maur).

Antisera and Immunological Methods. Four Dutch Belted rabbits were each given an initial injection of 1 mg of chachalaca lysozyme in Freund's supplemented complete adjuvant (Prager & Wilson, 1971b). A series of intravenous boosts (Prager & Wilson, 1971b) - 0.5 mg of lysozyme per rabbit per injection - was given at 3 and 6 months after the initial injection. The antisera obtained by cardiac puncture after 6 months were pooled as described (Prager & Wilson, 1971b; Champion et al., 1974); the micro-complement fixation titer of the pool was 20,000. The antiserum pools to the chicken, bobwhite quail, Japanese quail, turkey,

ring-necked pheasant, duck A and duck B lysozymes were those from the sixth bleeding described by Prager & Wilson (1971b). Quantitative micro-complement fixation was carried out according to Champion et al. (1974).

Criteria of Lysozyme Purity. The chachalaca enzyme was tested for purity by starch gel electrophoresis at pH 5.3, 7.0, and 11.9, by immunoelectrophoresis, and by sedimentation velocity as described previously (Prager & Wilson, 1971b). SDS (sodium dodecyl sulfate) polyacrylamide gel electrophoresis was done at pH 8.8 according to methods described (Laemmli, 1970; Swank & Munkres, 1971). Sedimentation velocity and SDS gel analysis were also used, prior to initiating sequencing work, to probe for major structural, i.e., size, differences between chachalac lysozyme and the other homologous bird lysozymes.

Reduction, Alkylation, and Enzymatic Digestion. Reduction and alkylation of the chachalaca lysozyme were performed as previously described for guinea fowl egg white lysozyme (Jollès et al., 1972). Tryptic and chymotryptic hydrolyses as well as the separation of the tryptic and chymotryptic peptides were carried out according to Jollès et al. (1972).

Determination of the Sequence. The amino acid compositions of the whole lysozyme and of the peptides were established with a Technicon Autoanalyzer after total hydrolysis (6 M HCl, 110°, for 18, 48, or 72 hrs, under vacuum). The tryptophan content was determined according to Spies & Chambers (1949). The structures were established by automated Edman degradation (Edman & Begg, 1967) in a Socosi Sequencer, Model PS-100, by the Quadrol or dimethylallylamine buffer method. The lysine-containing peptides were modified with 4-sulfophenylisothiocyanate according to Braunitzer et al. (1970); the N-terminal amino acid was identified on an aliquot before modification by the dansylation procedure. The thiazolinones were converted into phenylthiohydantoin (PTH)-amino acids, and these latter were characterized by thin-layer chromatography (chloroform-methanol, 90:10, v/v; pure chloroform), by gas-liquid chromatography (Beckman CG 45 chromatograph), or with an amino acid Autoanalyzer after regeneration of the free amino acid (6 M HCl in the presence of mercaptoethanol, 24 hrs, 150°, under vacuum). Other, more classical, procedures such as digestion by carboxypeptidase have been described elsewhere (Jollès & Jollès, 1969).

RESULTS

Lysozyme Purity

The chachalaca lysozyme was pure by all criteria applied. It was indistinguishable from the chicken enzyme according to sed

Table 1

Amino acid composition of chachalaca egg white lysozyme^a

Amino acid	Time of hydrolysis			Nearest integer
	18 hrs	48 hrs	72 hrs	
Aspartic acid	18.84	18.50	18.36	19
Threonine	6.45	6.12	5.75	7
Serine	9.45	7.88	7.57	10
Glutamic acid	3.88	3.79	3.65	4
Proline	1.95	2.06	2.10	2
Glycine	10.56	10.36	9.80	11
Alanine	11.63	11.20	11.71	12
Valine	4.73	4.90	5.05	5
Half-cystine	7.29	7.88	5.90	8
Methionine	2.37	3.00	3.10	3
Isoleucine	6.79	6.94	7.04	7
Leucine	6.50	6.60	7.00	7
Tyrosine	6.90	6.83	6.85	7
Phenylalanine	0	0	0	0
Tryptophan	6			6
Lysine	8.60	8.20	7.78	9
Histidine	1.70	1.80	1.80	2
Arginine	9.50	9.20	8.72	10
Total				129

^aThe composition is given in terms of residues per molecule. The values for threonine and serine were extrapolated to zero time of hydrolysis, while those for valine and isoleucine were taken from the 72-hr hydrolysate. Half-cystine was determined as S-carboxymethylcysteine, while tryptophan was determined as described by Spies & Chambers (1949).

mentation velocity analysis. By SDS gel electrophoresis it had a mobility of 0.88 relative to chicken and turkey lysozymes. However, we did not consider this evidence of a significant size difference from the other homologous lysozymes since human lysozyme, which has a molecular weight almost identical to chicken lysozyme (Jollès & Jollès, 1971; Canfield et al., 1971), had a relative mobility of 0.91 under the same conditions and since, particularly in this low molecular weight region, mobility may reflect other differences besides molecular weight (Neville, 1971; Tung & Knight, 1972).

Amino Acid Composition

The amino acid composition was established after different hydrolysis times. Values for serine and threonine were extra-

polated to zero time of hydrolysis, whereas those for valine and isoleucine were taken from the 72-hr hydrolysate. The composition is shown in Table 1.

N- and C-Terminal Amino Acids and Sequences

By the dansylation procedure, lysine was determined as the unique N-terminal amino acid; leucine was the unique C-terminal amino acid liberated by the action of carboxypeptidase A (24 hrs, 0.1 M NH_4HCO_3 , yield 53%).

The N-terminal sequence (44 amino acids) was established with the Sequencer (see Figure 1). The different tryptic peptides contained in this sequence were again characterized when the tryptic digest was studied.

Tryptic Peptides

After chromatography of the tryptic hydrolysate on Dowex 1 $\times$ 2, 13 peptides (T-1 - T-13) were characterized. In Table 2, their R_f , mobility (m) value at pH 6.5, and yields are indicated. Twelve peptides account for all the amino acids present in the molecule (Table 3), since peptide T-10 is composed of T-5 + T-1. Six tryptic peptides contained more than one basic amino acid. This result might be due to the relatively short digestion time of the protein with trypsin (7 hrs) and also, in part, to the presence of acidic amino acids next to the basic residues as well as of 2 basic residues side by side.

The structures of all the peptides except T-3 were established with the Sequencer. Peptides T-1, T-4, T-5, T-8, and T-11 were contained in the N-terminal sequence determined directly on the reduced and alkylated undigested lysozyme. The C-terminal sequence of the long tryptic peptide T-7 was established thanks to the N-terminal sequence of the chymotryptic peptide C-8 (Table 4). Peptide T-3 was too short to be sequenced using the automatic Sequencer; the chymotryptic peptide C-5 (Table 4), which contained T-3, allowed establishment of its structure. Peptide T-9 was the unique tryptic peptide devoid of a basic C-terminal residue: it thus constituted the C-terminal sequence of the lysozyme. During all these structural determinations each PTH-amino acid residue was determined by 2 different methods.

Alignment of the Tryptic Peptides

The chymotryptic digest was submitted to gel filtration on Sephadex G-25; 14 fractions (C-1 - C-14) were collected. The different peptides contained in these fractions were purified

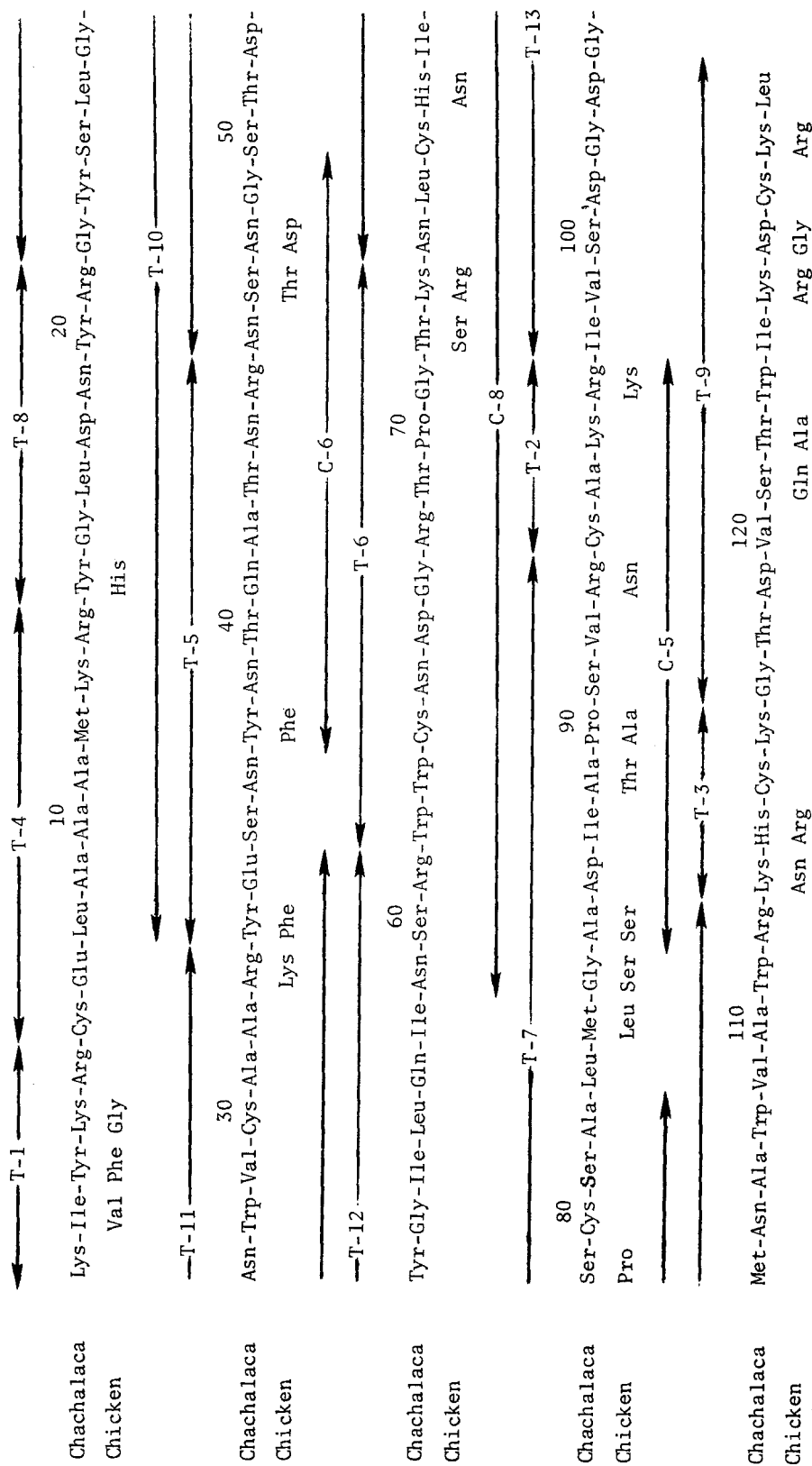


Fig. 1

Primary structure of chachalaca egg white lysozyme. The tryptic and chymotryptic peptides which were isolated and studied are indicated. The first 44 amino acids were determined directly on the reduced, alkylated lysozyme with a Sequencer and characterized again when the tryptic peptides were studied. The sequence of the chicken enzyme is given at those positions where the 2 lysozymes differ

Table 2

R_f and mobility m (at pH 6.5) values, yields, and chemical structure of the tryptic peptides of reduced, alkylated chachalaca lysozyme

Peptide	R_f^a	b^b m	Yield(%)	Structure ^c
T-1	0.40	+0.80	34	Lys-Ile-Tyr-Lys-Arg
T-2	0.12	+0.45	31	SCMCys-Ala-Lys-Arg
T-3	0.08	+0.50	17	Lys-His-SCMCys-Lys
T-4	0.16	0	29	SCMCys-Glu-Leu-Ala-Ala-Met-Lys-Arg
T-5	0.21	0	25	Tyr-Glu-Ser-Asn-Tyr-Asn-Thr-Gln-Ala-Thr-Asn-Arg
T-6	0.35	0	23	Trp-Trp-SCMCys-Asn-Asp-Gly-Arg-Thr-Pro-Gly-Thr-Lys
T-7	0.58	-0.18	32	Asn-Leu-SCMCys-His-Ile-Ser-SCMCys-Ser-Ala-Leu-Met-Gly-Ala-Asp-Ile-Ala-Pro-Ser-Val-Arg
T-8	0.63	0	27	Tyr-Gly-Leu-Asp-Asn-Tyr-Arg
T-9	0.40	-0.13	30	Gly-Thr-Asp-Val-Ser-Thr-Trp-Ile-Lys-Asp-SCMCys-Lys-Leu
T-10	0.20	0	4	T-5 + T-12
T-11	0.53	0	30	Gly-Tyr-Ser-Leu-Gly-Asn-Trp-Val-SCMCys-Ala-Ala-Arg
T-12	0.22	0	6	Asn-Ser-Asn-Gly-Ser-Thr-Asp-Tyr-Gly-Ile-Leu-Gln-Ile-Asn-Ser-Arg
T-13	0.54	-0.35	5	Ile-Val-Ser-Asp-Gly-Asp-Gly-Met-Asn-Ala-Trp-Val-Ala-Trp-Arg

^a Solvent: n-butanol-pyridine-acetic acid water (15:10:3:12, by volume); Whatman No. 1 filter paper as support.

^b m = 0 for Gly; m = +1 for Arg; m = -1 for CySO_3H .

^c SCMCys stands for S-carboxymethylcysteine.

Table 3

Composition of tryptic peptides from reduced, alkylated chachalaca lysozyme. The number of residues per peptide is given in parentheses to the nearest integer. Values are based on hydrolysis for 18 hrs

Amino acid	Tryptic peptide number ^a													Total
	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-11	T-12	T-13		
Asp					2.62(3)	2.00(2)	2.00(2)	1.95(2)	2.05(2)	1.05(1)	3.60(4)	3.00(3)	19	
Thr					1.57(2)	1.51(2)			1.74(2)		1.05(1)		7	
Ser					0.98(1)		2.35(3)		0.95(1)	0.98(1)	2.52(3)	0.96(1)	10	
Glu				0.65(1)	1.80(2)						0.70(1)		4	
Pro						1.00(1)	1.05(1)						2	
Gly						1.75(2)	1.05(1)	1.00(1)	0.88(1)	1.66(2)	1.88(2)	1.90(2)	11	
Ala		1.00(1)		2.30(3) ^b	1.00(1)		2.60(3)			1.64(2)		1.57(2)	12	
Val							0.91(1)		0.74(1)	0.71(1) ^b		1.43(2) ^b	5	
Cys/2 ^c		0.62(1)	0.57(1)	0.59(1)		0.75(1)	1.30(2)		0.52(1)	0.47(1)			8	
Met				0.53(1)			0.57(1)					0.40(1)	3	
Ile	0.57(1)						1.75(2) ^b		0.86(1)		1.63(2) ^b	0.85(1)	7	
Leu				0.66(1)			1.70(2) ^b	0.94(1)	0.88(1)	0.89(1)	0.58(1)		7	
Tyr	0.60(1)				1.37(2)			1.44(2)		0.61(1)	0.54(1)		7	
Phe													0	
Trp ^d						(2)			(1)	(1)		(2)	6	
Lys	1.60(2)	1.00(1)	2.00(1)	1.00(1)		0.81(1)			1.65(2)				9	
His			0.94(1)				0.75(1)						2	
Arg	1.00(1)	0.90(1)		0.70(1)	0.76(1)	0.82(1)	0.93(1)	0.96(1)		0.75(1)	0.73(1)	1.00(1)	10	
Total	5	4	4	9	12	12	20	7	13	12	16	15	129	

^a T-10 is missing as it is composed of T-5 + T-12.

^b Based on hydrolysis for 72 hrs.

^c Determined as S-carboxymethylcysteine.

^d Identified by the Ehrlich reaction (Dalglish, 1952).

Table 4

R_f and mobility m (at pH 6.5) values of 3 chymotryptic overlap peptides between tryptic sequences

Peptide	R_f^a	m^a	Structure ^a
C-5	0.17	+0.33	Arg-Lys-His-SCMCys-Lys-Gly-Thr-Asp-Val-Ser-Thr-Trp
C-6	0.12	0	SCMCys-Asn-Asp-Gly-Arg-Thr-Pro-Gly-Thr-Lys-Asn-Leu
C-8	0.30	-0.20	Gly-Ala-Asp-Ile-Ala-Pro-Ser-Val-Arg-SCMCys-Ala-Lys-Arg-Ile-Val-Ser-Asp-Gly-Asp-Gly-Met-Asn-Ala-Trp

^aSee footnotes to Table 2.

Table 5

Immunological distances of avian lysozymes relative to chachalaca lysozyme

Lysozyme X	Immunological distance		
	Measured with anti-chachalaca lysozyme	Measured with anti-lysozyme X ^a	Average
Chicken	95	159	127
Bobwhite quail	99	156	128
Japanese quail	110	147	129
Turkey	123	173	148
Ring-necked pheasant	100	112	106
Guinea fowl	94	b	-
Duck A	158	169	164
Duck B	164	165	165
Duck C	175	b	-
Duck II	164	b	-
Duck III	156	b	-

^aFrom Prager & Wilson (1971b).

^bNot done because antisera to these lysozymes were not available.

by preparative paper electrophoresis (Jollès & Jollès, 1969; Jollès et al., 1972). Only the chymotryptic peptides containing basic amino acids were studied. In particular, 3 of them (C-5, C-6, and C-8 - see Table 4) were necessary to reconstruct the complete amino acid sequence of chachalaca lysozyme.

Peptide T-10 allowed us to join the long N-terminal sequence of the enzyme (T-1 + T-4 + T-8 + T-11 + T-5) and peptide T-12. The alignment of T-6 + T-7, T-7 + T-2 + T-13, and T-13 + T-3 + T-9 was demonstrated by analysis of the chymotryptic peptides C-6, C-8, and C-5, respectively. The complete sequence as well as the evidence for the overlaps is presented in Figure 1.

Table 5 presents the results of micro-complement fixation experiments comparing chachalaca lysozyme with 6 gallinaceous and 5 duck lysozymes using antiserum to the chachalaca enzyme. The results of reciprocal comparisons are included where available. The percent standard deviation (Champion et al., 1975) of the reciprocal tests is 17%, compared with 19% calculated for all available reciprocal comparisons of bird lysozymes (Prager & Wilson, 1971b; Champion et al., 1974, 1975; Prager et al., 1974b; Jollès et al., this work).

Table 6 summarizes sequence and immunological comparisons for those 8 bird lysozymes of known amino acid sequence, and Figure 2 plots immunological distance (y) versus percent sequence difference (x) for these comparisons. There is an approximately linear relationship between x and y which does not differ greatly from that previously reported for several proteins (Champion et al., 1975). The correlation coefficient estimated from the data in Figure 2 is 0.92.

DISCUSSION

Sequence Aspects

Inspection of the chachalaca amino acid composition and sequence discloses the absence of phenylalanine in this molecule. Lysozymes from certain other species have been found to be totally devoid of one amino acid: all 5 duck lysozymes (Hermann & Jollès, 1970; Hermann et al., 1971; Prager and Wilson, 1972) lack histidine, while baboon lysozyme (Hermann et al., 1973) lacks methionine. There are, moreover, substitutions at 9 positions - residues 2, 38, 47, 48, 79, 86, 89, 126, and 128 - in the chachalaca which are invariant in all other bird lysozymes which had heretofore been sequenced.

Table 6 has summarized the number of amino acid changes in the chachalaca lysozyme relative to the other bird lysozymes. The distribution of these replacements as regards degree of exposure on the molecule's surface is presented in Table 7. With the exception of bobwhite quail and guinea fowl (which share 3 internal substitutions, at position 40, 55, and 91), 96-100% of the substitutions in the chachalaca relative to the other birds are external or surface, and, indeed, two-thirds of them are fully exposed. Thus each substitution, by virtue of its occurrence in an exposed position, would have the potential for influencing immunological reactivity (Prager & Wilson, 1971b,c; Wilson & Prager, 1974). Most of the substitutions relative to chicken lysozyme are concentrated in 2 regions: the N-C region (consisting of residues 1-27 and

Table 6

Sequence and immunological comparison of bird lysozymes^a

Species compared	Chicken	BWQ	JQ	Turkey	GF	Duck II	Duck III	Cha
Chicken		4	6	7	10	22	23	27
Bobwhite quail	6.6		10	11	8	23	24	31
Japanese quail	22	33		10	15	25	26	30
Turkey	18	29	35		16	22	23	28
Guinea fowl	26	28	69	48		29	30	28
Duck II	68	70	118	70	127		6	30
Duck III	64	68	132	74	-	26		31
Chachalaca	127	128	129	148	94	167	156	

^a The number of amino acid differences between any 2 lysozymes appears in the upper right-hand section of the matrix, while the immunological distance appears in the lower left-hand section. The sequence data are from the following references: chicken, Jollès et al. (1963) and Canfield (1963); bobwhite quail, Prager et al. (1972); Japanese quail, Kaneda et al. (1969); turkey, LaRue & Speck (1970); guinea fowl, Jollès et al. (1972); Duck II, Hermann & Jollès (1970); Duck III, Hermann et al. (1971); chachalaca, Jollès et al., this work. The immunological distance values are averages of reciprocal measurements except those in *italics*, which are unidirectional measurements due to the unavailability of antisera to guinea fowl and Duck III. Many of these values have already been presented (Prager & Wilson, 1971b; Prager et al., 1972, 1974b; Champion et al., 1974, 1975; Wilson & Prager, 1974). As previously, we have combined immunological measurements made with anti-duck A with sequence data for Duck II in view of the extreme similarity of these 2 enzymes (Prager & Wilson, 1971b, 1972; Wilson & Prager, 1974). BWQ, bobwhite quail; JQ, Japanese quail; GF, guinea fowl; Cha, chachalaca.

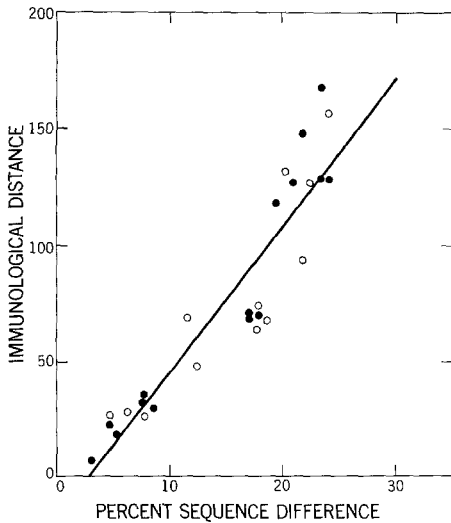


Fig.2

Immunological distance versus percent sequence difference for bird lysozymes of known sequence. Filled and open circles represent reciprocal and unidirectional measurements, respectively. The data are from Table 6. The straight line of best fit calculated by the method of least squares is $y=6.3x-17.9$, where x = percent sequence difference and y = immunological distance. The correlation coefficient is 0.92

Table 7

Location of amino acid substitutions of chachalaca lysozyme relative to other bird lysozymes^a

Lysozyme compared	Number of positions having amino acid replacements			
	Total	External	Surface	Internal
Chicken	27	19	8	0
Bobwhite quail	31	20	8	3
Japanese quail	30	23	7	0
Turkey	28	20	7	1
Guinea fowl	28	17	7	4
Duck II	30	19	10	1
Duck III	31	21	9	1

^a The information concerning degree of exposure of amino acid side chains in chicken lysozyme is taken from Browne et al. (1969) and Lee & Richards (1971), and it is assumed that the other bird lysozymes have the same conformation as chicken lysozyme and that the amino acid side chains replacing those in the chicken have the same accessibility as those they replaced. The total number of exposed, surface, and internal positions are 55, 47, and 27, respectively. 47 positions among all bird lysozymes have undergone substitutions; of these, 30, 12, and 5 are external, surface, and internal, respectively.

122-127, with a disulfide linkage between Cys₆ and Cys₁₂₇ (Fujio et al., 1968, 1971)) and in the area 72-93 (which includes parts of the loop - residues 60-83 (Fainaru et al., 1974) - and of the second loop - residues 76-94 (Fujio et al., 1974)). Eight of the 27 substitutions occur in the N-C region, with 3 more in the immediately adjacent 113-121 area; 10 substitutions occur among residues 72-93. These substitutions would be expected to be of particular importance immunologically (see below).

The electrophoretic behavior of chachalaca lysozyme, reported earlier (Prager & Wilson, 1971b), can also be explained in terms of the sequence. Using the same assumptions as previously (Prager & Wilson, 1971b), we calculate a net charge of +12 at pH 5.3, compared to +9 for chicken lysozyme; the mobility of 117 compared to chicken taken as 100 agrees with the charges calculated. At pH 11.9, the chachalaca would carry a charge of -8, compared to -3 for the chicken. This is consistent with our observation (Prager & Wilson, 1971b) that the chachalaca enzyme moves toward the anode ($R_f = -77$), even though the chicken enzyme still moves toward the cathode (R_f taken as 100) due to electrical endosmosis.

It is noteworthy that 6 of the 9 lysines and 4 of the 10 arginines in chachalaca lysozyme occur next to charged groups

Table 8
Amino acid differences and minimal mutation distances among animal lysozymes^a

Species compared	Chicken	BWQ	JQ	Turkey	GF	Duck	Duck	Cha	Baboon	Humar
						II	III			
Chicken		4	6	7	10	22	23	27	50	52
Bobwhite quail	4		10	11	8	23	24	31	51	53
Japanese quail	7	11		10	15	25	26	30	52	54
Turkey	7	11	10		16	22	23	28	53	54
Guinea fowl	10	8	16	16		29	30	28	54	57
Duck II	27	28	29	27	33		6	30	50	53
Duck III	27	28	29	27	33	7		31	49	52
Chachalaca	30	34	32	32	32	36	36		50	55
Baboon	69	70	69	72	72	68	66	70		14
Human	70	71	72	72	75	69	67	75	14	

^aThe number of amino acid differences between any 2 lysozymes is given in the upper right-hand section of the matrix, while the minimal mutation distances appear in the lower left-hand section. The sources of the avian sequences are indicated in Table 6. The baboon sequence is from Hermann et al. (1973), while that of the human enzyme is from Jollès & Jollès (1971, 1972), Canfield et al. (1971), and Thomsen et al. (1972). The deletion of one amino acid in bird lysozymes relative to mammalian lysozymes has been counted as an amino acid difference but has been ignored in the computation of minimal mutation distances. BWQ, JQ, GF, and Cha as in Table 6.

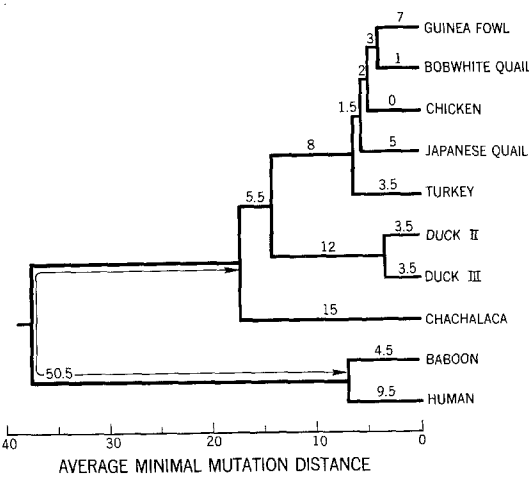


Fig.3
Phylogeny of animal lysozymes constructed from minimal mutation distances (Table 8) according to the Farris method (Farris, 1972). The sequence of rat lysozyme was also considered in constructing the phylogeny. The details of the rat sequence determination are to be reported elsewhere (White & Wilson, in preparation). Since the rat enzyme differs slightly more from the avian lysozymes than do the primate enzymes, the placement of the 10 lysozymes in this figure would not be altered were the complete phylogeny to be shown.

The percent standard deviation of the tree (Fitch & Margoliash, 1967) is 9.7%

(either acidic or basic), in which areas tryptic cleavage would be slowed appreciably. Thus the short (7-hrs) digestion time ultimately proved particularly advantageous, since we were able

to isolate larger tryptic peptides which are easier to sequence with the automatic Sequencer and which at the same time facilitate alignment by overlap of the chymotryptic peptides.

Phylogenetic Implications

Table 8 compiles the amino acid differences and minimal mutation distances for those 10 animal lysozymes whose sequences have been reported. From the minimal mutation distances the phylogeny in Figure 3 was constructed. We note immediately that this phylogeny places the chachalaca outside the ducks relative to the guinea fowl, bobwhite quail, chicken, Japanese quail, and turkey. That is, the chachalaca by a phylogenetic analysis of its lysozyme differs more from the other 5 members of the order Galliformes than does the duck, which is placed in the order Anseriformes. Though at the organismal level ducks are so different in anatomy, physiology, and behavior that they are placed in a different order (Wetmore, 1960), at the level of the lysozyme molecule they are more like the phasianoid birds than the chachalaca is.

To examine in somewhat greater detail the phylogenetic relationship of the bird lysozymes now sequenced, we have used the method of Doolittle & Blombäck (1964; Prager et al., 1972) to determine the amino acid changes which probably occurred along each lineage of the phylogeny of Figure 3. The avian portion of the phylogeny with the specific amino acid changes indicated is shown in Figure 4. The results fit the Farris tree of Figure 3 well both qualitatively and quantitatively. Six changes (at positions 2, 38, 79, 86, 114, and 125) were found along the common lineage leading to the other 6 birds after the chachalaca had diverged off.

In view of the conflict of the lysozyme data with the organismal data (Wetmore, 1960; Hudson, et al., 1966), it is important to examine and derive phylogenies for molecules other than lysozyme. In that way one could distinguish between the following 2 possibilities: 1. at the molecular level the chachalaca is indeed rather different from most other gallinaceous birds or 2. the lysozyme difference is unusually great, perhaps due to a gene duplication event. The immunological information thus far available on transferrin, albumin, and ovalbumin (Prager et al., 1974a)² suggests that the former alternative is more likely. In view of the recent evidence pointing out the independence of molecular and organismal evolution (Wilson et al., 1974a,b; Prager & Wilson, 1975; King & Wilson, 1975), determination of significant differences

² Prager & Wilson, unpublished observations.

Fig. 4

between the chachalaca and the other gallinaceous birds at the molecular level would not automatically lead to exclusion of the chachalaca from the order Galliformes. Only if classification were to be based on molecular criteria or criteria of divergence times would such a reclassification be strongly called for.

Immunological Aspects

The sequence data bear out both qualitatively and quantitatively the initial immunological predictions: chachalaca lysozyme c is quite different from all other bird lysozymes c of known sequence and in fact is more different from the other gallinaceous lysozymes than are the duck lysozymes. Moreover, the correlation coefficient of 0.92 calculated for the data in Figure 2 indicates (Waugh, 1943) that over 80% of the variation in immunological distance between pairs of bird lysozymes is ascribable to the *number* of amino acid substitutions. The remainder of the variation would presumably be due to such factors as the nature, location, and conformational effect of the substitutions.

Of particular interest (Table 5) are the unusually weak reactions of chachalaca lysozyme with antisera elicited by the chicken, bobwhite quail, Japanese quail, and turkey enzymes. It has been reported that over 40% of the antibodies in rabbit antisera to chicken lysozyme may be directed toward the N-C region (Fujio et al., 1968, 1971) and that 5-10% of anti-chicken lysozyme antibodies may be directed toward the first

Fig.4

Bird phylogeny based on lysozyme sequence, with amino acid changes postulated to have occurred along each lineage determined as described (Doolittle & Blombäck, 1964; Prager et al., 1972). Minimal mutation distances were used. A possible ancestral lysozyme sequence was determined (Prager, unpublished) by making use of all bird lysozyme sequences as well as those of the baboon, human, and rat (see Figure 3). In addition, in those cases in which the lysozyme data alone did not suggest an unambiguous choice of amino acid at a particular position in the ancestral sequence, the sequences of human, bovine, and guinea pig α -lactalbumin and the partial sequence of kangaroo α -lactalbumin (summarized by Hill et al., 1974) were taken into account to facilitate deciding on the most likely amino acid at a given position in the ancestral lysozyme. The α -lactalbumin sequences were consulted since lysozyme and α -lactalbumin are considered to have diverged from a common ancestor (e.g., Hill et al., 1974), and it is not unreasonable to assume this divergence of the 2 molecules may have occurred prior to the divergence of mammals and birds. Amino acid changes along each lineage between the possible ancestral sequence and the present day sequences were chosen so as to minimize the total number of mutational events

loop (Arnon & Sela, 1969) and a comparable amount toward the second loop (Fujio et al., 1974). Since 18 of the 27 substitutions in the chachalaca relative to the chicken are in these regions, the weak cross-reactivity may in part be explained by these observations on the distribution of the various populations of antibodies to chicken lysozyme. The reciprocal cross-reactions, measured with anti-chachalaca lysozyme, in contrast, are much stronger. These observations would be consistent with the idea that regions other than the N-C and the loop may elicit more of the antibodies produced when the chachalaca enzyme is injected into rabbits.

The chachalaca versus duck reactions are equally weak in both directions. This could be explained by one or more of the following: 1. more external plus surface changes have occurred in the chachalaca relative to the ducks than relative to any other species except Japanese quail (Table 7); 2. 11 or 12, not just 10, changes occur in the region 72-93; 3. 4 or 5 substitutions occur among residues 57-71, a region in which the chicken and chachalaca enzymes are identical; 4. since the duck enzymes are much less "chicken-like" in sequence than are the other phasianoid lysozymes, their surfaces as a whole would be expected to have a more different array of determinants from the chicken and consequently to exhibit different immunological behavior.

Concluding Remarks

Determination of the chachalaca lysozyme sequence has been valuable from several viewpoints. The 9 substitutions found in the chachalaca enzyme but identical in all other bird lysozymes so far sequenced provide new information on the substitutions compatible with lysozyme function. These data, moreover, allowed construction of a molecular phylogeny with a chachalaca molecule of known sequence. The sequence data, furthermore, permitted extension of the sequence-immunology correlation and suggested possible explanations for the immunological observations in terms of changes in particular determinant areas. Knowledge of the sequence should facilitate design of future immunological experiments such as those involving peptide inhibition in attempts to elucidate what specific regions of the lysozyme molecule have elicited antibody production and how much change in a determinant still permits immunological cross-reactivity.

It should be pointed out, finally, that chachalaca lysozyme is the most different lysozyme one can readily expect to find in egg white of birds (Prager et al., 1974c). We have now examined the 3 major groups of birds known to contain lysozyme c in their egg whites: the phasianoid birds, the ducks, and

the chachalaca. This work suggests the outer limits of variation among bird egg white lysozymes c.

Acknowledgements. The excellent technical assistance of Mr. Ly Quan Le and Mrs. M. Berger is gratefully acknowledged by the authors. We thank T. White for permitting us to make use of the rat lysozyme sequence data. In addition, we thank J. Brodrick and D. Herold for running SDS gels and S. Carlson for advice on the construction of the Farris tree. Lastly, we thank N. Sharon for encouraging the collaboration between the Jollès and Wilson laboratories.

REFERENCES

- Arnon, R., Sela, M. (1969). *Proc.Natl.Acad.Sci.USA* 62, 163
- Atassi, M.Z. (1975). *Immunochemistry* 12, 423
- Braunitzer, G., Schrank, B., Ruhfus, A. (1970). *Z.Physiol.Chem.* 351, 1589
- Browne, W.J., North, A.C.T., Phillips, D.C., Brew, K., Vanaman, T.C., Hill, R.L. (1969). *J.Mol.Biol.* 42, 65
- Canfield, R.E. (1963). *J.Biol.Chem.* 238, 2698
- Canfield, R.E., Kammerman, S., Sobel, J.H., Morgan, F.J. (1971). *Nature New Biol.* 232, 16
- Champion, A.B., Prager, E.M., Wachter, D., Wilson, A.C. (1974). Micro-complement fixation. In: *Biochemical and immunological taxonomy of animals*, C.A. Wright, ed., p. 397. London: Academic Press
- Champion, A.B., Soderberg, K.L., Wilson, A.C., Ambler, R.P. (1975). *J.Mol.Evol.* 5, 291
- Dalgliesh, C.E. (1952). *Biochem.J.* 52, 3
- Doolittle, R.F., Blombäck, B. (1964). *Nature* 202, 147
- Edman, P., Begg, G. (1967). *Eur.J.Biochem.* 1, 80
- Fainaru, M., Wilson, A.C., Arnon, R. (1974). *J.Mol.Biol.* 84, 635
- Farris, J.S. (1972). *Am.Natur.* 106, 645
- Fitch, W.M., Margoliash, E. (1967). *Science* 155, 279
- Fujio, H., Imanishi, M., Nishioka, K., Amano, T. (1968). *Biken J.* 11, 207
- Fujio, H., Martin, R.E., Ha, Y.-M., Sakato, N., Amano, T. (1974). *Biken J.* 17, 73
- Fujio, H., Sakato, N., Amano, T. (1971). *Biken J.* 14, 395
- Hermann, J., Jollès, J. (1970). *Biochim.Biophys.Acta* 200, 178
- Hermann, J., Jollès, J., Buss, D.H., Jollès, P. (1973). *J.Mol.Biol.* 79, 587
- Hermann, J., Jollès, J., Jollès, P. (1971). *Eur.J.Biochem.* 24, 12
- Hill, R.L., Steinman, H.M., Brew, K. (1974). Comparison of the structures of alpha-lactalbumin and lysozyme. In: *Lysozyme*, E.F. Osseman, R.E. Canfield, S. Beychok, eds., p. 55. New York: Academic Press
- Hudson, G.E., Parker, R.A., Vanden Berge, J., Lanzillotti, P.J. (1966). *Amer.Midl.Nat.* 76, 1
- Jollès, J., Jauregui-Adell, J., Bernier, I., Jollès, P. (1963). *Biochim. Biophys.Acta* 78, 668
- Jollès, J., Jollès, P. (1969). *Helv.Chim.Acta* 52, 2671

- Jollès, J., Jollès, P. (1971). *Helv.Chim.Acta* 54, 2668
- Jollès, J., Jollès, P. (1972). *Fed.Eur.Biochem.Soc.Lett.* 22, 31
- Jollès, J., Van Leemputten, E., Mouton, A., Jollès, P. (1972). *Biochim. Biophys.Acta* 257, 497
- Kaneda, M., Kato, I., Tominaga, N., Titani, K., Narita, K. (1969). *J.Biochem.(Tokyo)* 66, 747
- King, M.-C., Wilson, A.C. (1975). *Science* 188, 107
- Laemmli, U.K. (1970). *Nature* 227, 680
- LaRue, J.N., Speck, J.C., Jr. (1970). *J.Biol.Chem.* 245, 1985
- Lee, B., Richards, F.M. (1971). *J.Mol.Biol.* 55, 379
- Neville, D.M., Jr. (1971). *J.Biol.Chem.* 246, 6328
- Prager, E.M., Arnheim, N., Mross, G.A., Wilson, A.C. (1972). *J.Biol.Chem.* 247, 2905
- Prager, E.M., Brush, A.H., Nolan, R.A., Nakanishi, M., Wilson, A.C. (1974a). *J.Mol.Evol.* 3, 243
- Prager, E.M., Fainaru, M., Wilson, A.C., Arnon, R. (1974b). *Immunochemistry* 11, 153
- Prager, E.M., Wilson, A.C., Arnheim, N. (1974c). *J.Biol.Chem.* 249, 7295
- Prager, E.M., Wilson, A.C. (1971a). *J.Biol.Chem.* 246, 523
- Prager, E.M., Wilson, A.C. (1971b). *J.Biol.Chem.* 246, 5978
- Prager, E.M., Wilson, A.C. (1971c). *J.Biol.Chem.* 246, 7010
- Prager, E.M., Wilson, A.C. (1972). *Biochem.Genet.* 7, 269
- Prager, E.M., Wilson, A.C. (1975). *Proc.Natl.Acad.Sci.USA* 72, 200
- Sharon, N., Eshdat, Y., Maoz, I., Bernstein, Y., Prager, E.M., Wilson, A.C. (1974). *Israel J.Chem.* 12, 591
- Spies, J.R., Chambers, D.C. (1949). *Anal.Chem.* 21, 1249
- Swank, R.T., Munkres, K.D. (1971). *Anal.Biochem.* 39, 462
- Thomsen, J., Lund, E.H., Kristiansen, K., Brunfeldt, K., Malmquist, J. (1972). *Fed.Eur.Biochem.Soc.Lett.* 22, 34
- Tung, J.-S., Knight, C.A. (1972). *Anal.Biochem.* 48, 153
- Waugh, A.E. (1943). *Elements of statistical methods*. New York and London: McGraw-Hill Book Co., Inc.
- Whetmore, A. (1960). *Smithson.Misc.Collect.* 139, 1
- Wilson, A.C., Maxson, L.R., Sarich, V.M. (1974a). *Proc.Natl.Acad.Sci.USA* 71, 2843
- Wilson, A.C., Prager, E.M. (1974). *Antigenic comparison of animal lysozymes*. In: *Lysozyme*, E.F. Osserman, R.E. Canfield, S. Beychok, eds., p. 127. New York: Academic Press
- Wilson, A.C., Sarich, V.M., Maxson, L.R. (1974b). *Proc.Natl.Acad.Sci.USA* 71, 3028