Tanford, C., Kawahara, K., & Lapanje, S. (1967) *J. Am. Chem. Soc. 89*, 729.
Van Bruggen, E. F. J. (1968) in *Physiology and Biochemistry of Hemocyanins* (Ghiretti, F., Ed.) Academic Press, New York.
Van Holde, K. E., & Cohen, L. B. (1964) *Brookhaven Symp. Biol. No. 17*, 184.
Wood, E. J., & Peacocke, A. R. (1973) *Eur. J. Biochem. 35*, 410.
Yang, J. T. (1961) *Adv. Protein Chem. 16*, 323.

Amino Acid Sequence of California Quail Lysozyme. Effect of Evolutionary Substitutions on the Antigenic Structure of Lysozyme†

Ibrahim M. Ibrahimi,‡ Ellen M. Prager,* Thomas J. White,§ and Allan C. Wilson

ABSTRACT: To examine the effect of amino acid substitutions in lysozyme on the binding of antibodies to lysozyme, we purified lysozyme from the egg whites of California quail and Gambel quail. Tryptic peptides were isolated from digests of the reduced and carboxymethylated lysozymes and subjected to quantitative analysis of their amino acid compositions. The two proteins were identical by this criterion. Each peptide from the California quail lysozyme was then sequenced by quantitative Edman degradation, and the peptides were ordered by homology with other bird lysozymes. California quail lysozyme is most similar in amino acid sequence to bobwhite quail lysozyme, from which it differs by two substitutions: arginine for lysine at position 68 and histidine for glutamine at position 121. California and bobwhite quail lysozymes were antigenically distinct from each other in quantitative microcomplement fixation tests, indicating that substitutions at one or both of these positions can alter the antigenic structure of lysozyme. Yet neither of these positions is among those claimed to account for the precise and entire antigenic structure of lysozyme [Atassi, M. Z., & Lee, C.-L. (1978) *Biochem. J. 171*, 429–434]. Two possible explanations for this discrepancy are discussed.

Although much progress has been made in our understanding of the antigenic structure of globular proteins, this understanding is incomplete in important respects. Lysozyme *c* is an example of a globular protein which has been studied particularly extensively by immunochemists. Especially notable are recent studies by Atassi and co-workers, who claim to have delineated precisely the complete antigenic structure of lysozyme[1] (Atassi & Habeeb, 1977; Atassi & Lee, 1978). According to these authors the number of sites at which antibodies bind to lysozyme is only three, regardless of whether the antisera are produced in rabbits or in goats. The sites comprise residues at the following positions: site I, residues 5, 7, 13, 14, and 125; site II, residues 33, 34, 113, 114, and 116; site III, residues 62, 87, 89, 93, 96, and 97 (Atassi & Lee, 1978). These sites were delineated by studies with chemically modified lysozyme and with peptides isolated from lysozyme or synthesized to simulate certain surface parts of the molecule (Atassi & Lee, 1977, 1978; Lee & Atassi, 1977).

Investigators from other laboratories have also contributed to the knowledge of the antigenic structure of lysozyme. Evidence from these other laboratories is consistent with the possibility that lysozyme has additional antigenic sites (Maron et al., 1972; Arnon et al., 1974; Fainaru et al., 1974; Matthyssens et al., 1974; Fujio et al., 1974; Wilson & Prager, 1974).

We now present further evidence that evolutionary substitutions at positions outside the sites proposed by Atassi and co-workers alter the antigenic structure of lysozyme. This evidence was obtained by sequencing lysozyme from the California quail and comparing it immunologically with other lysozymes of known amino acid sequence, especially the lysozyme of the bobwhite quail.

Materials and Methods

Eggs and Lysozymes. Freshly laid California quail (*Lophortyx californicus*) and Gambel quail (*Lophortyx gambeli*) eggs were obtained from A-1 Game Birds, Ashland, Oregon. The egg whites were separated from the yolks and stored frozen at −10 °C.

Prior to lysozyme purification, individual California quail egg whites were tested for the presence of electrophoretic variants by cellulose acetate electrophoresis at pH 8.6 (Ibrahimi, 1977). No variants were found. Lysozyme was then purified from 1800 mL of pooled egg white, derived from 477 eggs, according to procedures previously described (Arnheim et al., 1969; Prager & Wilson, 1971a). The purity of the lysozyme obtained was tested by electrophoresis in different media at several pH values (Prager & Wilson, 1971a; Ibrahimi, 1977; Prager et al., 1978), by immunodiffusion, and by amino-terminal analysis.

Gambel quail lysozyme was purified from 570 mL of pooled egg white, derived from 133 eggs, in the same manner as was California quail lysozyme, and its purity was evaluated by the same criteria. Gambel quail lysozyme was included in this study because taxonomists (Leopold, 1978) consider the Gambel quail to be the closest living relative of the California quail, and we therefore hoped to obtain lysozymes differing at only one or two positions.

All other lysozymes used in this investigation were those previously purified and characterized (Prager & Wilson, 1971a; Jollès et al., 1976).

† From the Department of Biochemistry, University of California, Berkeley, California 94720. *Received November 29, 1978.* This work was supported in part by grants from the National Science Foundation (DEB74-11866A01) and the National Institutes of Health (GM-21509). I.M.I. was partially supported by a travel grant from the Research Council of the University of Jordan.

‡ On leave from the Department of Biology, University of Jordan, Amman, Jordan.
§ Present address: Cetus Corporation, Berkeley, CA 94710.

[1] Abbreviations used: lysozyme, lysozyme *c*; CM (Cm in figure), carboxymethyl.

0006-2960/79/0418-2736$01.00/0

Enzymes and Reagents. Trypsin (TPCK-treated) and α-chymotrypsin were obtained from Worthington Biochemical Corp. All chemicals used for sequencing were either spectral or sequential grade.

Antisera and Immunological Methods. Antisera to California quail and Gambel quail lysozymes were produced in Dutch Belted rabbits according to the immunization protocol previously described (Prager & Wilson, 1971a; Jollès et al., 1976). Four rabbits were injected with each lysozyme, and the antisera obtained after 6 months of immunization were pooled in inverse proportion to their microcomplement fixation titers (Champion et al., 1974). All other antiserum pools used—to the lysozymes from chicken, bobwhite quail, turkey, Japanese quail, ring-necked pheasant, chachalaca, and duck—were likewise derived after 6 months of immunization and have been previously described (Prager & Wilson, 1971a; Jollès et al., 1976).

Quantitative microcomplement fixation analysis was carried out according to Prager & Wilson (1971a) and Champion et al. (1974). At times the following modifications (Ibrahimi, 1977) were used. All tubes (a maximum of 96) in one experiment were accommodated in one rack, and the cell lysis step was carried out in a shaking water bath. At intervals, samples were withdrawn from the control tubes, and the progress of lysis was followed by monitoring the turbidity spectrophotometrically at 600 nm rather than by looking periodically at the control tubes.

The degree of antigenic difference is given in terms of immunological distance, which is equal to 100 times the log of the factor by which the antiserum concentration must be raised for a heterologous antigen to produce a complement fixation curve whose peak height is equal to that produced by the homologous antigen (the immunogen) (Prager & Wilson, 1971a; Champion et al., 1974). Immunological distance is related to percent cross-reactivity in the quantitative precipitin test by the equation $y = 2.24(100 - z)$, where y = units of immunological distance and z = percent cross-reactivity in the precipitin test (Prager & Wilson, 1971b).

Amino Acid Compositions. Proteins, both native and reduced and alkylated, and peptides were hydrolyzed in 6 M HCl at 110 °C for 24, 48, or 72 h under vacuum, and the amino acid compositions were determined with a Beckman 121 automatic amino acid analyzer. Tryptophan was determined following hydrolysis in *p*-toluenesulfonic acid.

Separation of Tryptic Peptides. The tryptic peptides were chromatographed on a Dowex 50 column as previously described (Canfield, 1963; Prager et al., 1972). High-voltage paper electrophoresis at pH 6.4 and 1.9 (Ibrahimi, 1977) was then used in the event that the column chromatographic peaks contained more than one peptide. Alternatively, the tryptic peptides were separated by peptide mapping (Canfield, 1963; Arnheim et al., 1969; Prager et al., 1972).

Sequential Edman Degradation. The degradations were carried out manually on 0.05–0.2 μmol of each peptide. For all residues the thiazolinone and phenylthiohydantoin derivatives were hydrolyzed to the free amino acids (6 M HCl; 140 °C; 20 h; under vacuum), which were then identified with the automatic amino acid analyzer. The derivatives of aspartic acid and asparagine gave aspartic acid, those of glutamic acid and glutamine gave glutamic acid, those of serine and (carboxymethyl)cysteine gave alanine, that of tryptophan gave alanine and glycine, and that of threonine gave α-aminobutyric acid. The derivatives of the other 12 amino acids yielded only the amino acid actually present in the native molecule. The phenylthiohydantoin derivatives of aspartic acid, asparagine, glutamic acid, glutamine, (carboxymethyl)cysteine, tryptophan, serine, and threonine were identified by thin-layer chromatography on Cheng Chin polyamide sheets (Summers et al., 1973). For serine and threonine the identification was confirmed by dansyl amino-terminal analysis (Hartley, 1970). Occasionally the subtractive Edman procedure was also used to confirm the identification. The C terminus of each peptide was identified by amino acid analysis of the unhydrolyzed peptide after the cleavage of the penultimate residue.

Other Sequencing Procedures. The following additional procedures were used during the course of chemical analysis of the primary structures of California and Gambel quail lysozymes: reduction and carboxymethylation; tryptic and chymotryptic hydrolysis; cyanogen bromide cleavage; peptide mapping. These procedures are described in detail in the supplementary material (see paragraph at the end of this paper regarding supplementary material).

Results

Lysozyme Yields and Purity. California quail lysozyme (2900 mg) was obtained. Upon electrophoresis, the purified lysozyme migrated as a single protein band through cellulose acetate at pH 8.6, through polyacrylamide at pH 4.5, and through starch at pH 5.3 and 11.9; at pH 8.9 on polyacrylamide there was a faint trailing band in the lysozyme region, perhaps due to deamidation. In the Ouchterlony immunodiffusion test, antisera elicited by the pure lysozyme gave only one precipitin line when reacted with whole egg white or the pure immunogen. Only lysine was detected at the amino terminus by thin-layer chromatography of the phenylthiohydantoin derivative and by amino acid analysis of that derivative hydrolyzed to the free amino acid.

Gambel quail lysozyme (740 mg) was obtained. This lysozyme likewise satisfied the electrophoretic and immunological criteria of purity used (with the exception of a very faint extra leading band on polyacrylamide at pH 8.9 in addition to the faint trailing band described above for California quail lysozyme). In all systems used California and Gambel quail lysozymes had an identical electrophoretic mobility. Gambel quail lysozyme likewise yielded only lysine as the amino-terminal residue.

Amino Acid Composition. As shown in Table I, the amino acid compositions indicate that California quail lysozyme differs by a minimum of two amino acid substitutions from bobwhite quail lysozyme. The California quail protein contains one more residue of both arginine and histidine and one less residue of both lysine and glutamate. Gambel quail lysozyme, in turn, is identical with California quail lysozyme in composition (Table I).

Peptide Mapping. Reduced, alkylated lysozyme (50 mg) was digested with trypsin for 12 h, and 1 mg of the digest was subjected to peptide mapping. The peptides were eluted from 25 spots on the map detected by ninhydrin, and their amino acid compositions were determined. The maps for California quail and Gambel quail lysozymes were superimposable, and no difference in the amino acid compositions of the eluted peptides was detected. However, two peptides differed in amino acid composition from their counterparts in bobwhite quail lysozyme (Table II). Peptide T-9 contained one more arginine and one less lysine, while peptide T-16 contained one more histidine and one less glutamate. Since bobwhite quail T-9 contains lysine only in one place, at the C terminus (Prager et al., 1972), the arginine for lysine substitution could already be assigned to position 68. Similarly, since bobwhite quail T-16 yields only one glutamate upon amino acid analysis, corresponding to glutamine in the intact peptide (Prager et

Table I: Amino Acid Composition of California Quail, Gambel Quail, and Bobwhite Quail Lysozymes

amino acid	residues per molecule: California quail[a]	Gambel quail[a]	bobwhite quail[b]
cysteine (CM)[c]	8.03 (8)	8.00	8
aspartic acid	21.19 (21)	20.88	21
threonine	6.97 (7)	6.92	7
serine	9.85 (10)	9.88	10
glutamic acid	3.87 (4)	3.95	5
proline[c]	1.70 (2)	2.22	2
glycine	12.20 (12)	12.22	12
alanine	12.15 (12)	11.73	12
valine	7.31 (7)	7.12	7
methionine	2.25 (2)	2.04	2
isoleucine	5.01 (5)	4.89	5
leucine	7.90 (8)	7.99	8
tyrosine	2.81 (3)	3.19	3
phenylalanine	2.99 (3)	2.94	3
histidine	1.99 (2)	2.20	1
lysine	6.05 (6)	6.34	7
arginine	10.55 (11)	10.61	10
tryptophan[d]	6.07 (6)	5.33	6

[a] The protein was hydrolyzed in duplicate for 24, 48, and 72 h. The values listed are the results of extrapolation to 0 time or 72 h, followed by normalization to chicken lysozyme treated in the same way. The values in parentheses in the California quail column are those obtained from the amino acid sequence. Significant differences in composition from bobwhite quail lysozyme are underlined. [b] Prager et al. (1972). [c] Determined on reduced and carboxymethylated lysozyme. [d] Determined by *p*-toluenesulfonic acid hydrolysis for 48 h.

al., 1972), histidine could be assigned to position 121 in the California quail.

Isolation and Amino Acid Sequence of the Tryptic Peptides of California Quail Lysozyme. A tryptic digest of reduced, carboxymethylated lysozyme (300 mg) was passed over a Dowex 50 column. Aliquots of the peak fractions eluted from this column were tested for purity by high-voltage paper electrophoresis at both pH 6.4 and 1.9. Fractions showing more than one peptide at either pH were further purified by high-voltage preparative electrophoresis (Ibrahimi, 1977). The amino acid compositions and properties of all the pure peptides thus obtained are summarized in Table III. These results confirmed those obtained by peptide mapping. The amino acid sequences of all the tryptic peptides, T-1 to T-18, were determined (Figure 1), with full details of the sequencing given in the supplementary material.

Alignment of the Tryptic Peptides and Reliability of the Sequence. The tryptic peptides were aligned by homology with chicken lysozyme (Canfield, 1963; Jollès et al., 1963) without resorting to overlapping peptides, as shown in Figure 1. The use of homology is the main weakness of the sequence. However, the extensive homology of the two lysozymes and, indeed, of all bird lysozymes sequenced to date (Jollès et al., 1979) leaves little doubt that the proposed alignment is correct. The only sequence difference that could be overlooked by the proposed alignment would be reciprocal replacements that involve whole tryptic peptides. Such a replacement would probably result in a big sequence difference and consequently a big immunological difference. No such large difference was observed when the two lysozymes were compared immunologically (see below). Since both free arginine and free lysine were obtained upon tryptic digestion, the assignment of these residues by homology might be considered least reliable. However, arginine at position 14 is invariant in all lysozymes of known amino acid sequence, including many analyzed with a sequencer and with chymotryptic overlap peptides.

Some difficulty was encountered in sequencing T-9 and T-13 due to high extractability of the shortened peptides into butyl acetate. This difficulty was circumvented by the availability of large quantities of each peptide, which made it possible to repeat the sequencing and to use subtractive Edman analysis.

Finally, we call attention to the unequivocal demonstration of asparagine and not aspartic acid at position 103 in California quail lysozyme. Initial studies (Canfield, 1963; Jollès et al., 1963) indicated the presence of aspartic acid at that position in chicken lysozyme, but it was subsequently reported (Imoto et al., 1972; Phillips, 1974) that asparagine occurred at position 103 in the chicken. Furthermore, reexamination

Table II: Amino Acid Compositions of California and Gambel Quail Lysozyme Tryptic Peptides Differing in Composition from the Homologous Bobwhite Quail Lysozyme Tryptic Peptides

amino acid	peptide[a] T-9: California quail	Gambel quail	bobwhite quail	T-16: California quail	Gambel quail	bobwhite quail
Cys (CM)[b]	0.27 (1)	0.27 (1)	1			
Asp	2.00 (2)	2.00 (2)	2	1.23 (1)	1.45 (1)	1
Thr				0.94 (1)	0.83 (1)	1
Ser						
Glu				(0)	(0)	1
Pro						
Gly[c]	1.40 (1)	1.31 (1)	1	0.97 (1)	1.17 (1)	1
Ala				1.18 (1)	0.78 (1)	1
Val				1.03 (1)	0.58 (1)	1
Met						
Ile				0.92 (1)	0.92 (1)	1
Leu						
Tyr						
Phe						
His				0.67 (1)	0.69 (1)	0
Lys	(0)	(0)	1			
Arg	0.87 (1)	0.57 (1)	0	1.00 (1)	1.00 (1)	1
Trp	ND[d]	ND	2	ND	ND	1
net change[e]	−Lys, +Arg			−Glu, +His		

[a] The values given are molar ratios and the numbers in parentheses are the assumed integral number of residues of each amino acid per peptide. The bobwhite quail lysozyme peptides (integral values) are according to Prager et al. (1972). [b] Low recovery of Cys (CM) from maps. [c] Usually extra Gly is recovered from maps. [d] ND, not determined. [e] California and Gambel quails relative to bobwhite quail.

Table III: Amino Acid Composition and Properties of the Tryptic Peptides of Reduced and Carboxymethylated California Quail Lysozyme

	tryptic peptide[a]																	
amino acid	T-1, T-12	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10	T-11	T-13	T-14	T-15	T-16	T-17	T-18	total
Cys (CM)[b]			0.93 (1)			0.22 (1)			0.44 (1)		1.33 (3)			0.59 (1)		0.76 (1)		8
Asp					1.62 (2)	1.22 (1)	3.48 (3)	4.56 (4)	2.06 (2)	0.27 (0)	4.19 (4)	3.22 (3)	0.84 (1)		1.14 (1)			21
Thr							1.04 (1)	2.20 (2)		0.98 (1)	1.91 (2)				0.72 (1)			7
Ser						0.83 (1)	2.09 (2)	2.08 (2)		1.00 (1)	2.77 (3)	0.99 (1)			0.25 (0)			10
Glu			1.08 (1)				1.87 (2)	1.00 (1)							0.28 (0)			4
Pro										0.92 (1)	1.00 (1)							2
Gly[c]		0.92 (1)			0.92 (1)	1.84 (2)	0.44 (0)	2.14 (2)	1.06 (1)	1.20 (1)	0.60 (0)	2.24 (2)		0.17 (0)	0.90 (1)	0.94 (1)		12
Ala			3.3[4] (3)			2.00 (2)	1.41 (1)	0.38 (0)		0.19 (0)	2.74 (3)	2.00 (2)		0.05 (0)	0.76 (1)			12
Val		0.64 (1)				0.73 (1)		0.71 (1)			1.00 (1)	1.47 (2)			0.76 (1)			7
Met			1.04 (1)									1.03 (1)						2
Ile								1.10 (1)			1.84 (2)	0.46 (1)			0.64 (1)			5
Leu			1.06 (1)		0.87 (1)	0.89 (1)		1.03 (1)			2.77 (3)	0.07 (0)			0.32 (0)		1.00 (1)	8
Tyr					0.70 (1)	0.79 (1)		0.91 (1)							0.32 (0)			3
Phe		0.48 (1)					1.71 (2)											3
His					0.56 (1)										0.71 (1)			2
Lys	1.00 (2)		1.00 (1)			0.79 (1)					0.70 (1)			1.00 (1)				6
Arg		1.00 (1)		1.00 (1)	1.00 (1)	0.16 (0)	1.00 (1)	0.90 (1)	1 00 (1)	1.00 (1)		1.13 (1)	1.00 (1)		1.00 (1)	1.00 (1)		11
Trp[d]						1.24 (1)			0.11 (2)			0.52 (2)			0.21 (1)			6
total	2	4	8	1	7	12	12	16	7	5	23	15	2	2	9	3	1	129
N terminus	Lys	Val	Cys (CM)	Arg	His	Gly	Phe	Asn	Trp	Thr	Asn	Ile	Asn	Cys (CM)	Gly	Gly	Leu	
net charge[e]	+1	+1	−1	+1	ND	0	0	−1	−1	+1	−3	0	+1	0	ND	0	0	
pH[f]	1.9	6.4	1.9	6.4	1.9	1.9			6.4		1.9	6.4	6.4	1.9	1.9	1.9	1.9	

[a] The peptides are numbered serially from the amino terminus of the protein; the numbers in the table are molar ratios and those in parentheses are the assumed integral number of residues of each amino acid per peptide. [b] Cys(CM) is partially destroyed during isolation and hydrolysis. [c] Extra Gly is recovered as a contaminant, especially in T-11. [d] Determined by *p*-toluenesulfonic acid hydrolysis. [e] Calculated from electrophoretic mobility relative to aspartic acid at pH 6.4 (Offord, 1966); ND, not determined. [f] pH at which preparative electrophoresis was done.

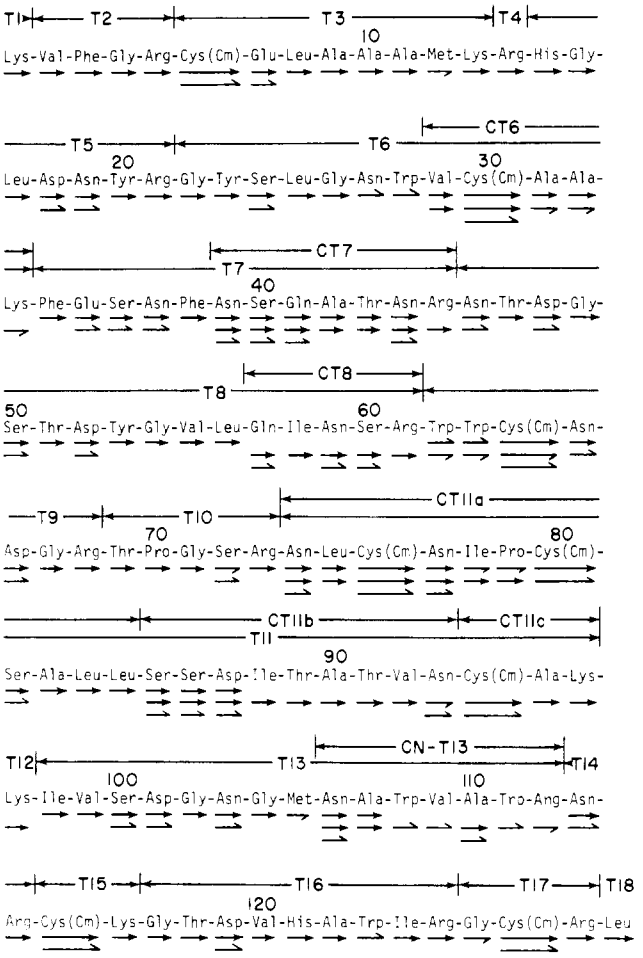

FIGURE 1: Amino acid sequence of California quail lysozyme. The peptides isolated are indicated by double-headed arrows. Residues identified by the quantitative manual Edman (→), manual Edman followed by phenylthiohydantoin thin-layer chromatography (⇢), and subtractive Edman (⟶) methods are indicated below the sequence. Ser and Thr were further identified by the dansyl–Edman method, with Ser [and also Cys(CM)] quantitated as Ala and with Thr quantitated as α-aminobutyric acid. The following abbreviations are used: T, tryptic peptide; CT, chymotryptic peptide isolated from a tryptic peptide; CN-T, a peptide isolated from a cyanogen bromide digest of a tryptic peptide. In cases where subfragments of the tryptic peptides were isolated, the row of symbols immediately below the residue names refers to analyses done on the parent tryptic peptides, and lower row(s) refer to analyses done on the subfragment. The only exceptions are for residues 80 and 81, where both rows refer to T-11 itself, and occasional uses of (⟶) in the third row of symbols to represent use of Edman degradation followed by phenylthiohydantoin thin-layer chromatography on the parent peptide and also on the subfragment. Full details concerning the sequence determination of all peptides, including quantitation, appear in the supplementary material.

(E. M. Prager, unpublished experiments) of the electrophoretic mobility at pH 6.4 of peptide T-13 in the chicken and bobwhite quail strongly suggests asparagine at position 103 in the lysozymes of both species. In view of these findings we consider California quail, chicken, and bobwhite quail lysozymes identical at that position. We suspect (Jollès et al., 1979) that facile deamidation is the cause of early reports of aspartic acid at position 103.

Immunological Cross-Reactivity. The microcomplement fixation assay was used to compare California quail, Gambel quail, and bobwhite quail lysozymes. As shown in Figure 2A, California quail and Gambel quail lysozymes could not be distinguished immunologically. In contrast, California quail and bobwhite quail lysozymes were found to differ antigenically, as had been noted in earlier comparative studies (White, 1976). This is evident from experiments both with antisera to California quail lysozyme (Figure 2A) and with antisera to bobwhite quail lysozyme (Figure 2B). Indeed, every antiserum tested distinguished readily between California and bobwhite quail lysozymes even though these two enzymes differ by only two amino acid substitutions.[2]

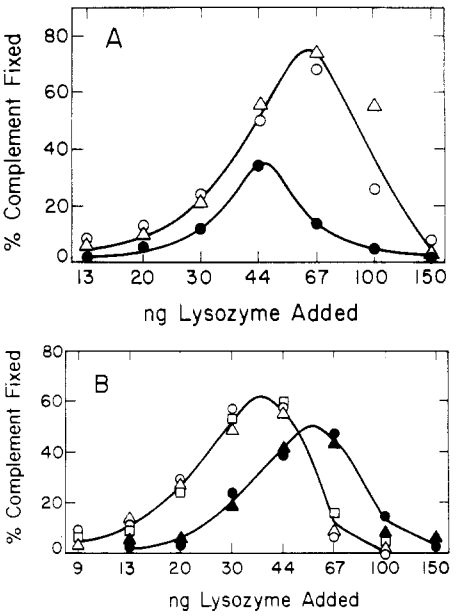

FIGURE 2: Immunological cross-reactivity among quail lysozymes measured by microcomplement fixation analysis. (A) Rabbit antisera prepared against California quail lysozyme tested with pure California quail (△), Gambel quail (○), and bobwhite quail (●) lysozymes. (B) Rabbit antisera prepared against bobwhite quail lysozyme tested with pure bobwhite quail (○, △, □) and California quail (●, ▲) lysozymes. One milliliter of a 1:8000 and of a 1:5300 dilution of the antiserum pool was used for each tube in (A) and (B), respectively. The multiple symbols plotted for each antigen in (B) represent duplicate or triplicate determinations done in the same experiment. In (A) the averages of duplicate determinations are shown.

Antisera to California quail lysozyme were tested also against all avian lysozymes of known sequence. These results and the results of additional comparisons of bird lysozymes of known sequence are summarized in Table IV and Figure 3. As the figure shows, there is a correlation between the degree of immunological difference (expressed in immunological distance units) and the degree of sequence difference.

Twenty-seven of the forty-four pairwise immunological comparisons in Figure 3 were published previously (Jollès et al., 1976). The additional 17 comparisons presented here confirm the correlation seen previously between the immunological distance and the degree of sequence difference. A slightly stronger correlation is evident when one eliminates from the calculation those positions at which the residues are fully buried in the interior of native chicken lysozyme and presumably inaccessible to antibodies (Figure 3B).

Discussion

Of all the distinct lysozymes *c* whose primary structures are known, those of the California quail and bobwhite quail are the most similar. The present work indicates that these two lysozymes differ by only two amino acid substitutions, at positions 68 and 121. The most similar lysozymes *c* previously

[2] The average immunological distance between the two lysozymes is 7.6 units (Table IV), which implies that about 3% of the antibodies are non-cross-reactive (Prager & Wilson, 1971b).

Table IV: Sequence and Immunological Comparison of Bird Lysozymes[a]

species compared	CaQ	chicken	BWQ	JQ	turkey	GF	RNP	duck II	duck III	Cha
California quail		3.1	1.6	7.0	7.0	5.4	10.0	18.6	19.4	23.3
chicken	3.3		3.1	4.7	5.4	7.8	7.7	17.1	17.8	20.9
bobwhite quail	7.6	6.6		7.8	8.5	6.2	10.8	17.8	18.6	24.0
Japanese quail	23	22	33		7.8	11.6	9.3	19.4	20.2	23.3
turkey	10[b]	18	29	35		12.4	7.7	17.1	17.8	21.7
guinea fowl	*18*	*26*	*28*	*69*	*48*		10.0	22.5	23.3	21.7
ring-necked pheasant	26	28	42	66	28	*25*		18.5	19.3	18.5
duck II	72	68	70	118	70	*127*	108		4.7	23.3
duck III	*73*	*64*	*68*	*132*	*74*	–	*101*	*26*		24.0
chachalaca	117	127	128	129	148	*94*	106	167	*156*	

[a] The percent amino acid sequence difference between any two 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 Jollès et al. (1976) and sources listed therein and from Jollès et al. (1979) and this work. The extra amino-terminal glycine in the ring-necked pheasant lysozyme (Jollès et al., 1979) relative to all other lysozymes has been counted as an amino acid difference. The immunological distance values are averages of reciprocal measurements except those in italics, which are unidirectional measurements since antisera to guinea fowl and duck III were not available. Many of these data have already been presented (Prager & Wilson, 1971a; Prager et al., 1974; Jollès et al., 1976, 1979). Immunological data obtained with anti-duck A have been combined with sequence data for duck II in view of the great similarity of these two lysozymes (Prager & Wilson, 1972) and in view of the unavailability of antisera to duck II. The following abbreviations are used: CaQ, California quail; BWQ, bobwhite quail; JQ, Japanese quail; GF, guinea fowl; RNP, ring-necked pheasant; Cha, chachalaca. [b] In the precipitin test (Prager & Wilson, 1971b) turkey and California quail lysozymes were also readily distinguishable: with anti-California quail lysozyme, turkey lysozyme precipitated 93% of the antibodies precipitated by the homologous enzyme (K. M. Helm-Bychowski, personal communication).

known were those of the chicken and bobwhite quail, which differ from each other by four substitutions.

Antigenic Structure. The two quail lysozymes proved useful for testing a hypothesis concerning the antigenic structure of lysozyme. The antigenic structure proposed by Atassi and co-workers (Atassi & Habeeb, 1977; Atassi & Lee, 1978), as mentioned in the introduction, consists of 16 antibody-binding residues arranged in three sites or determinants (Table V). The 16 residues were claimed to account for the entire antigenic structure of lysozyme. Such an antigenic structure would lead one to expect that the two quail lysozymes might be identical antigenically, since neither position 68 nor position 121 is among the 16 key positions (Table V) at which antibody binding is proposed to take place.

Inadequacy of the Three-Determinant Model. The most significant finding presented in this article is that, contrary to the above expectation, the two quail lysozymes are antigenically dissimilar.

The California quail thus joins a growing list of species whose lysozymes differ from one another in amino acid sequence but are identical at the 16 antigenic positions mentioned above. This list now includes the chicken, turkey, Japanese quail, and bobwhite quail along with the California quail. According to the above model, all five of these lysozymes should be antigenically identical, and yet all are readily distinguishable in the microcomplement fixation assay (Table IV). The 10 comparisons among the five lysozymes just listed above involve a total of 72 substitutions at 15 different positions, all outside the proposed antigenic sites. Of further interest is the observation (Table V) that 7 of the 16 proposed antigenic residues are invariant among all bird lysozymes sequenced, which might imply that only 9 positions in the lysozyme molecule account for the full range of immunological differences observed among all bird lysozymes (Table IV and Figure 3).

The average antigenic effect of substitutions outside the proposed 16 antigenic positions was calculated. To do this we restricted our attention to the five aforementioned lysozymes which are identical with one another at the 16 antigenic positions. The mean immunological distance (y_1) among the five lysozymes was divided by the mean number of substitutions (x_1). Thus, we assumed a relationship between y_1 and x_1 of the form shown in eq 1. The mean value of a obtained

$$y_1 = ax_1 \quad (1)$$

Table V: Proposed Antigenic Sites in Lysozyme[a]

position	amino acid in chicken lysozyme	location in chicken lysozyme[b]	variable among birds
	Site I		
5	arginine	external	no
7	glutamic acid	external	no
13	lysine	external	no
14	arginine	external	no
125	arginine	external	yes
	Site II		
33	lysine	surface	yes
34	phenylalanine	surface	yes
113	asparagine	external	yes
114	arginine	surface	yes
116	lysine	surface	yes
	Site III		
62	tryptophan	surface	no
87	aspartic acid	external	no
89	threonine	external	yes
93	asparagine	external	yes
96	lysine	surface	no
97	lysine	external	yes

[a] From Atassi & Lee (1978). [b] From Browne et al. (1969) and Lee & Richards (1971).

was 2.6 immunological distance units per amino acid substitution.

This value was then used to aid in calculating the mean antigenic effect of substitutions within the 16 antigenic positions. This calculation was done with pairs of lysozymes differing by substitutions at one or more of the antigenic positions. Each of these pairs differs both by substitutions at some of the proposed antigenic positions as well as by substitutions elsewhere in lysozyme. Thirty-four such pairs have been compared immunologically. The mean immunological distance among these lysozymes (y) may be related to the two types of substitutions by eq 2,

$$y = ax_1 + bx_2 \quad (2)$$

where x_1 is the number of substitutions outside the antigenic positions and x_2 is the number of substitutions at the antigenic positions. With the a value computed earlier and the observed values of y, x_1, and x_2, we calculated b, the mean immunological distance attributable to an amino acid substitution at an antigenic position. This value, 8.4, is greater than that

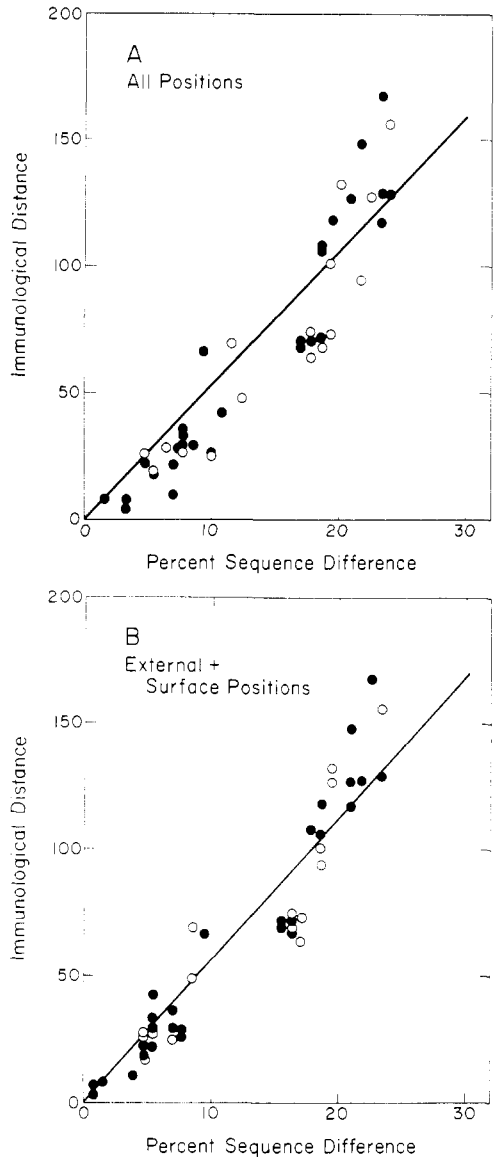

FIGURE 3: Dependence of immunological distance on percent sequence difference for 10 bird lysozymes of known sequence. Filled and open circles represent reciprocal and unidirectional immunological measurements, respectively. The data are from Table IV and include 44 of the 45 possible pairwise comparisons that can be made among the bird lysozymes of known sequence. Each regression line is drawn through the origin by the method of Steel & Torrie (1960). (A) Immunological distance plotted against the percent sequence difference at all positions in lysozyme. The regression line is described by the equation $y = 5.3x$, where x is percent sequence difference and y is immunological distance, and the chi-square (χ^2) value is 221. The straight line of best fit calculated by the method of least squares (not shown) has almost the same χ^2 value (235) and a correlation coefficient (r) between y and x of 0.93. (B) Immunological distance plotted against the percent sequence difference at the 102 or 103 positions in lysozyme where the residues are fully external or lie on the surface of lysozyme [Browne et al. (1969) and Lee & Richards (1971); cf. also Jollès et al. (1976, 1979)]. Differences at positions where the residues are fully internal were omitted, but x was calculated as a percent of the total number of residues in lysozyme. The regression line in Figure 3B is described by $y = 5.7x$, and χ^2 is 133. χ^2 for the line of best fit calculated by the method of least squares (not shown) is 190. The correlation coefficient is 0.95, somewhat stronger than the value of 0.93 obtained when amino acid substitutions at all positions are considered.

computed for substitutions outside the antigenic positions. If, however, in these calculations we neglect positions at which the amino acid residue is fully buried in the interior of lysozymes and thus inaccessible to antibodies, then a is 3.8 and b is 5.9. Thus, the average immunological effects of substitutions outside the proposed antigenic positions are nearly as large as those at antigenic positions.

A corollary of critical importance is that since most evolutionary substitutions lie outside these 16 positions, most of the antigenic difference between any two lysozymes is due to substitutions outside these positions. This follows from the observation that, on the average, the ax_1 term in eq 2 is equal to the bx_2 term when all positions are considered. When internal substitutions are neglected, ax_1 in eq 2 exceeds bx_2 by a factor of 2. The antigenic model of Atassi et al. (Atassi & Habeeb, 1977; Atassi & Lee, 1978) therefore appears to explain only a part—about a third to a half—of the antigenic difference between pairs of lysozymes.[3]

Additional Determinants. To account for the discrepancy just discussed, we consider first the possibility that there are additional antigenic determinants on lysozyme. The loop region of lysozyme, made up of the residues from positions 64 to 80 in the amino acid sequence, has been reported to have a minor determinant. From 5 to 14% of the antibodies in a typical antilysozyme serum bind to this determinant, according to Arnon (1977), who summarizes evidence concerning this determinant. It is notable that position 68, besides being part of the loop, is one of the two positions at which California quail and bobwhite quail lysozymes differ in amino acid sequence. Elsewhere, Ibrahimi et al. (1979) show that the substitution at this position probably makes a major contribution to the antigenic difference between these two lysozymes. Thus, by assuming that lysozyme has four determinants, namely, the loop determinant and the three determinants given in Table VI, one could in principle account for the antigenic difference between California quail and bobwhite quail lysozyme.

This four-determinant model is unsatisfactory, however, because it predicts immunological identity among the following lysozymes: turkey,[4] chicken, California quail, and Japanese quail. By invoking a fifth determinant, in particular the active-site determinant, one can theoretically explain the antigenic difference between the turkey lysozyme and the other three lysozymes. Although this is a minor determinant to which only 5–10% of the antibodies against lysozyme are directed (P. V. Hornbeck and A. C. Wilson, unpublished experiments), turkey lysozyme differs radically from other lysozymes at a key position (position 101, with glycine in place of aspartic acid) in this region of lysozyme (Arnheim et al., 1974). Consistent with this observation, turkey lysozyme appears to bind poorly to antibodies directed against the active site of chicken lysozyme (Maron et al., 1972).

Even a five-determinant model does not explain why chicken, California quail, and Japanese quail lysozymes differ antigenically from one another. Additional positions have been suggested to be antigenic, e.g., positions 20, 23, and 53 (Strosberg et al., 1971), but at these three positions all bird lysozymes are identical. Hence, the known or suspected antigenic positions on lysozyme do not suffice to account for the observed antigenic differences among these three lysozymes. Nevertheless, the antigenic differences among them are significant. With eq 1 we calculate that the average immunological effect of each amino acid substitution by which these lysozymes differ from one another is 2.5 immunological distance units. These immunological effects could be explained

[3] On the average the number of evolutionary substitutions that have occurred outside the proposed determinants exceeds the number of substitutions within these proposed determinants by a factor of 3.5.

[4] Although turkey lysozyme differs from the three other lysozymes listed at position 73 in the loop region, this position is probably not important antigenically (Fainaru et al., 1974).

were one to postulate the existence of still more determinants.

Consistent with the possibility of multiple determinants is the prediction that rabbit and bird lysozymes differ in amino acid sequence by at least 50 amino acid substitutions. This prediction is based on the observation that the mammalian lysozymes sequenced to date differ by an average of 54 (range 49–61) substitutions from bird lysozymes of known sequence (White et al., 1977; Jollès et al., 1979). Thus, if one had a hypothesis (Reichlin, 1975) that the rabbit's own lysozyme served as a screen, one would expect the rabbit to make antibodies to those many regions at which rabbit lysozyme differs from bird lysozyme. Since a typical determinant consists of five amino acid residues (Reichlin, 1975), a 40% sequence difference between two lysozymes makes it likely that any given cluster of five residues will differ by substitutions in the two species. Recently, moreover, the extent to which the immunized animal uses its own protein as a screen has been questioned (Kazim & Atassi, 1977, 1978).

Conformational Considerations. Alternatively, one could consider a completely different type of explanation for the observed immunological effects, such as the conformation hypothesis suggested by Atassi & Habeeb (1977) and White et al. (1978). The conformation hypothesis states that evolutionary substitutions have a high probability of producing subtle but long-range conformational changes in globular proteins and that these changes have significant effects on the complement-fixing properties of the aggregates formed when antibodies react with globular proteins. This hypothesis predicts that when such substitutions occur outside antigenic determinants they produce slight conformational differences affecting those determinants, thereby altering immunological reactivity. With this hypothesis, the existence of only a few determinants needs to be invoked in order to explain antibody interaction with lysozyme and the observed immunological differences among lysozymes.

The possibility that closely related lysozymes differ in conformation has already been suggested by tryptophan fluorescence experiments (Formoso & Forster, 1976) and low-resolution X-ray studies (Bott & Sarma, 1976). Very high resolution X-ray or NMR analysis may be needed to provide a rigorous demonstration of such differences.

Evolutionary Considerations. The finding that the California quail, Gambel quail, and bobwhite quail have extremely similar lysozymes is consistent with the taxonomists' judgment that these species are extremely similar in appearance and way of life, with evidence that these species can hybridize with one another (Johnsgard, 1973) and with evidence that these species have very similar transferrins, albumins, ovalbumins, and several enzymes (E. M. Prager and A. C. Wilson, unpublished experiments). The apparent identity of California and Gambel quail lysozymes agrees with the belief of taxonomists that these two quails are one another's closest living relatives (Leopold, 1978).

However, when the lysozyme data are subjected to phylogenetic analysis there is a surprising result, which is diagramed in Figure 3 of the following paper by Jollès et al. (1979). The guinea fowl lysozyme is as closely related, in branching order, to California quail lysozyme as are California and bobwhite quail lysozymes to each other. This phylogenetic anomaly appears to be the result of convergence or parallelism, a subject which has been discussed by Boulter (1978) and to which we intend to address ourselves in detail elsewhere.

Acknowledgments

We thank M. A. Baker for carrying out polyacrylamide gel electrophoresis, K. M. Helm-Bychowski for conducting precipitin studies with anti-California quail lysozyme, and S. S. Carlson and P. V. Hornbeck for helpful advice and discussion.

Supplementary Material Available

Detailed information on experimental procedures (reagents used, determination of amino acid composition, reduction and alkylation, chemical and enzymatic cleavage, peptide mapping, electrophoresis) utilized during sequencing work; complete details on the sequence determination of all the tryptic peptides of California quail lysozyme; the elution profile of the tryptic digest of reduced, carboxymethylated California quail lysozyme on Dowex 50; a peptide map of the tryptic digest of reduced, alkylated California quail lysozyme, along with a table indicating which peptide(s) each spot contained; and tables giving the compositions of all California and Gambel quail tryptic peptides isolated from maps (17 pages). Ordering information is given on any current masthead page.

References

Arnheim, N., Prager, E. M., & Wilson, A. C. (1969) *J. Biol. Chem. 244*, 2085–2094.

Arnheim, N., Millett, F., & Raftery, M. A. (1974) *Arch. Biochem. Biophys. 165*, 281–287.

Arnon, R. (1977) in *Immunochemistry of Enzymes and Their Antibodies* (Salton, M. J. R., Ed.) pp 1–28, Wiley, New York.

Arnon, R., Teicher, E., & Scheraga, H. A. (1974) *J. Mol. Biol.* 90, 403–407.

Atassi, M. Z., & Habeeb, A. F. S. A. (1977) in *Immunochemistry of Proteins* (Atassi, M. Z., Ed.) Vol. 2, pp 177–264, Plenum Press, New York and London.

Atassi, M. Z., & Lee, C.-L. (1977) *Biochim. Biophys. Acta 495*, 354–368.

Atassi, M. Z., & Lee, C.-L. (1978) *Biochem. J. 171*, 429–434.

Bott, R., & Sarma, R. (1976) *J. Mol. Biol. 106*, 1037–1046.

Boulter, D. (1978) in *Evolution of Protein Molecules* (Matsubara, H., & Yamanaka, T., Eds.) pp 243–250, Japan Scientific Societies Press, Tokyo.

Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, T. C., & Hill, R. L. (1969) *J. Mol. Biol. 42*, 65–86.

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

Champion, A. B., Prager, E. M., Wachter, D., & Wilson, A. C. (1974) in *Biochemical and Immunological Taxonomy of Animals* (Wright, C. A., Ed.) pp 397–416, Academic Press, London.

Fainaru, M., Wilson, A. C., & Arnon, R. (1974) *J. Mol. Biol. 84*, 635–642.

Formoso, C., & Forster, L. S. (1976) *Biochim. Biophys. Acta 427*, 377–386.

Fujio, H., Martin, R. E., Ha, Y.-M., Sakato, N., & Amano, T. (1974) *Biken J. 17*, 73–86.

Hartley, B. S. (1970) *Biochem. J. 119*, 805–822.

Ibrahimi, I. M. (1977) Ph.D. Thesis, University of California, Berkeley, CA.

Ibrahimi, I. M., Eder, J., Prager, E. M., Wilson, A. C., & Arnon, R. (1979) *Immunochemistry* (in press).

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C., & Rupley, J. A. (1972) *Enzymes, 3rd Ed. 7*, 665–868.

Johnsgard, P. A. (1973) *Grouse and Quails of North America*, pp 1–553, University of Nebraska, Lincoln, NE.

Jollès, J., Jauregui-Adell, J., Bernier, I., & Jollès, P. (1963) *Biochim. Biophys. Acta 78*, 668–689.

Jollès, J., Schoentgen, F., Jollès, P., Prager, E. M., & Wilson, A. C. (1976) *J. Mol. Evol. 8*, 59–78.

Jollès, J., Ibrahimi, I. M., Prager, E. M., Schoentgen, F., Jollès, P., & Wilson, A. C. (1979) *Biochemistry* (following paper in this issue).
Kazim, A. L., & Atassi, M. Z. (1977) *Biochim. Biophys. Acta 494*, 277–282.
Kazim, A. L., & Atassi, M. Z. (1978) *Immunochemistry 15*, 67–70.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.
Lee, C.-L., & Atassi, M. Z. (1977) *Biochem. J. 167*, 571–581.
Leopold, A. S. (1978) *The California Quail*, pp 1–281, University of California Press, Berkeley, CA.
Maron, E., Eshdat, Y., & Sharon, N. (1972) *Biochim. Biophys. Acta 278*, 243–249.
Matthyssens, G. E., Giebens, G., & Kanarek, L. (1974) *Eur. J. Biochem. 43*, 353–362.
Offord, R. E. (1966) *Nature (London) 211*, 591–593.
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. (1971a) *J. Biol. Chem. 246*, 5978–5989.
Prager, E. M., & Wilson, A. C. (1971b) *J. Biol. Chem. 246*, 7010–7017.
Prager, E. M., & Wilson, A. C. (1972) *Biochem. Genet. 7*, 269–272.
Prager, E. M., Arnheim, N., Mross, G. A., & Wilson, A. C. (1972) *J. Biol. Chem. 247*, 2905–2916.
Prager, E. M., Fainaru, M., Wilson, A. C., & Arnon, R. (1974) *Immunochemistry 11*, 153–156.
Prager, E. M., Wilson, A. C., Pêrin, J.-P., & Jollès, P. (1978) *Immunochemistry 15*, 577–583.
Reichlin, M. (1975) *Adv. Immunol. 20*, 71–123.
Steel, R. G. D., & Torrie, J. H. (1960) *Principles and Procedures of Statistics—With Special Reference to the Biological Sciences*, pp 1–481, McGraw-Hill, New York.
Strosberg, A. D., Van Hoeck, B., & Kanarek, L. (1971) *Eur. J. Biochem. 19*, 36–41.
Summers, M. R., Smythers, G. W., & Oroszlan, S. (1973) *Anal. Biochem. 53*, 624–628.
White, T. J. (1976) Ph.D. Thesis, University of California, Berkeley, CA.
White, T. J., Ibrahimi, I. M., & Wilson, A. C. (1978) *Nature (London) 274*, 92–94.
Wilson, A. C., & Prager, E. M. (1974) in *Lysozyme* (Osserman, E. F., Canfield, R. E., & Beychok, S., Eds.) pp 127–141, Academic Press, New York.

Amino Acid Sequence of Pheasant Lysozyme. Evolutionary Change Affecting Processing of Prelysozyme†

Jacqueline Jollès, Ibrahim M. Ibrahimi,‡ Ellen M. Prager,* Françoise Schoentgen, Pierre Jollès, and Allan C. Wilson

ABSTRACT: The amino acid sequence of ring-necked pheasant egg white lysozyme *c* was determined by automated Edman degradation of the reduced, alkylated tryptic peptides. Alignment of the tryptic peptides into a single chain containing 130 amino acids was done on the basis of automatic sequencing of the first 35 residues and by homology with chicken egg white lysozyme. Besides differing from the chicken enzyme by 9 amino acid substitutions, ring-necked lysozyme has an extra glycine at the amino terminus. Four other species of pheasant were shown to have lysozymes with a conventional amino terminus, beginning with lysine. The amino termini are now known for a total of 28 lysozymes *c*, and all except the ring-necked pheasant enzyme begin with lysine. To account for the extra residue in the ring-necked pheasant lysozyme, it is suggested that an evolutionary shift in the site of proteolytic cleavage of prelysozyme has occurred. Phylogenetic analysis of 10 avian and 3 mammalian lysozymes of known sequence placed the pheasant at a greater distance from the chicken than the turkey is from the chicken, in contrast to traditional taxonomic placement but consistent with evidence obtained from several other proteins. Immunological distances between ring-necked pheasant lysozyme and other bird lysozymes are consistent with the degree of sequence difference between ring-necked pheasant and other bird lysozymes and fit predictions based on the previously observed correlation between immunological cross-reactivity and sequence difference.

For several years we have been comparing the primary structures of lysozymes from various species. Amino acid sequences are already known for 12 lysozymes *c* (Jollès et al., 1976; White et al., 1977; Ibrahimi, 1977), and partial sequences are available for several others (Riblet, 1974; Morgan & Arnheim, 1974; Jollès et al., 1977; Ibrahimi, 1977). Knowledge of the positions at which sequence variation occurs has enhanced understanding of the structural requirements for lysozyme function. The sequence comparisons have also been valuable for studies of molecular evolution and of antigenic structure (Jollès et al., 1976; White et al., 1977, 1978).

While working out additional lysozyme sequences, which are necessary for in-depth studies of molecular evolution and antigenic structure, we discovered a novel structural feature, namely, an extra amino acid at the amino terminus of a bird

† From the Laboratory of Proteins, University of Paris V, F-75270 Paris Cedex 06, France (J.J., F.S., and P.J.), and the Department of Biochemistry, University of California, Berkeley, California 94720 (I.M.I., E.M.P., and A.C.W.). *Received November 29, 1978.* This work was supported in part by grants to P.J. from CNRS (ER No. 102) and INSERM (Group U-116) and to A.C.W. from the National Science Foundation (DEB74-11866A01) and the National Institutes of Health (GM-21509). I.M.I. was partially supported by a travel grant from the Research Council of the University of Jordan. This work is the 111th communication on lysozymes from the laboratory of P.J.

‡ On leave from the Department of Biology, University of Jordan, Amman, Jordan.

0006-2960/79/0418-2744$01.00/0