J Mol Evol (1985) 21:97–111

Journal of
Molecular Evolution

Comparison of Goose-Type, Chicken-Type, and Phage-Type Lysozymes Illustrates the Changes that Occur in Both Amino Acid Sequence and Three-Dimensional Structure during Evolution

L.H. Weaver[1], M.G. Grütter[1]*, S.J. Remington[1], T.M. Gray[1], N.W. Isaacs[2], and B.W. Matthews[1]

[1] Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA
[2] St. Vincent's School of Medical Research, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia

Summary. The three-dimensional structure of goose-type lysozyme (GEWL), determined by x-ray crystallography and refined at high resolution, has similarities to the structures of hen (chicken) egg-white lysozyme (HEWL) and bacteriophage T4 lysozyme (T4L). The nature of the structural correspondence suggests that all three classes of lysozyme diverged from a common evolutionary precursor, even though their amino acid sequences appear to be unrelated (Grütter et al. 1983).

In this paper we make detailed comparisons of goose-type, chicken-type, and phage-type lysozymes. The lysozymes have undergone conformational changes at both the global and the local level. As in the globins, there are corresponding α-helices that have rigid-body displacements relative to each other, but in some cases corresponding helices have increased or decreased in length, and in other cases there are helices in one structure that have no counterpart in another.

Independent of the overall structural correspondence among the three lysozyme backbones is another, distinct correspondence between a set of three consecutive α-helices in GEWL and three consecutive α-helices in T4L. This structural correspondence could be due, in part, to a common energetically favorable contact between the first and the third helices.

There are similarities in the active sites of the three lysozymes, but also one striking difference. Glu 73 (GEWL) spatially corresponds to Glu 35 (HEWL) and to Glu 11 (T4L). On the other hand, there are two aspartates in the GEWL active site, Asp 86 and Asp 97, neither of which corresponds exactly to Asp 52 (HEWL) or Asp 20 (T4L). (The discrepancy in the location of the carboxyl groups is about 10 Å for Asp 86 and 4 Å for Asp 97.) This lack of structural correspondence may reflect some differences in the mechanisms of action of the three lysozymes. When the amino acid sequences of the three lysozyme types are aligned according to their structural correspondence, there is still no apparent relationship between the sequences except for possible weak matching in the vicinity of the active sites.

Key words: Protein structure — Lysozyme — Structural correspondence — Divergent evolution

Introduction

It is well known that the amino acid sequences and three-dimensional structures of proteins change during evolution. Given sufficient time, proteins coded for by different genes may change to such an extent that it is no longer obvious they came from the same precursor. A prototype of this is provided by the nucleotide-binding domains of different dehydrogenases (Rossmann et al. 1974). These domains have structural similarity but little if any amino acid sequence homology, and it is not obvious whether the different nucleotide-binding domains evolved from a common precursor or arose independently.

* *Present address:* Biotechnology Department, Pharmaceuticals Division, Ciba-Geigy Ltd., CH-4002 Basel, Switzerland
Offprint requests to: B.W. Matthews

Table 1. α-Carbon coordinates for goose egg-white lysozyme

Residue	X	Y	Z	Residue	X	Y	Z	Residue	X	Y	Z
1 Arg	6.5	33.0	50.3	63 Pro	4.4	27.5	41.2	126 Thr	14.8	43.7	34.7
2 Thr	5.4	32.7	46.7	64 Ala	1.4	27.9	38.9	127 Ile	11.2	44.6	33.8
3 Asp	4.1	36.2	46.1	65 Val	3.6	30.3	36.7	128 Gln	10.7	46.0	37.4
4 Cys	7.2	38.2	45.8	66 Ile	6.4	27.6	36.6	129 Lys	13.6	48.4	36.8
5 Tyr	6.7	39.0	42.2	67 Ala	3.7	25.1	35.6	130 Lys	12.5	49.2	33.3
6 Gly	2.9	39.2	42.0	68 Gly	2.3	27.5	33.0	131 Phe	9.0	50.2	34.2
7 Asn	−0.3	37.1	41.7	69 Ile	5.8	27.9	31.4	132 Pro	8.8	51.2	37.9
8 Val	−1.1	35.5	38.4	70 Ile	6.4	24.1	31.5	133 Ser	5.4	52.9	37.3
9 Asn	−4.9	35.7	39.1	71 Ser	3.1	23.5	29.8	134 Trp	3.8	49.4	36.7
10 Arg	−4.8	39.5	39.0	72 Arg	3.9	26.1	27.1	135 Thr	2.0	47.3	39.3
11 Ile	−3.3	39.7	35.5	73 Glu	7.6	25.2	26.6	136 Lys	3.9	44.3	40.5
12 Asp	−5.4	40.4	32.5	74 Ser	7.5	21.4	26.6	137 Asp	1.5	42.0	38.6
13 Thr	−5.0	37.9	29.8	75 His	3.9	20.2	27.4	138 Gln	2.0	43.9	35.4
14 Thr	−6.3	37.3	26.3	76 Ala	5.4	18.9	30.6	139 Gln	5.7	43.6	35.9
15 Gly	−4.9	33.8	26.4	77 Gly	8.1	16.9	28.8	140 Leu	5.1	39.8	36.3
16 Ala	−3.8	31.7	23.4	78 Lys	5.8	15.1	26.3	141 Lys	3.4	39.6	33.0
17 Ser	−3.5	33.1	19.9	79 Val	7.3	16.7	23.3	142 Gly	6.2	41.5	31.2
18 Cys	−5.2	31.0	17.2	80 Leu	10.9	16.3	24.4	143 Gly	8.8	39.3	32.9
19 Lys	−1.6	30.5	16.0	81 Lys	13.6	13.8	23.3	144 Ile	7.1	36.3	31.4
20 Thr	−1.1	28.6	19.3	82 Asn	15.9	13.0	26.2	145 Ser	7.0	37.9	28.0
21 Ala	−4.5	26.9	19.5	83 Gly	15.1	16.6	27.5	146 Ala	10.7	38.6	28.2
22 Lys	−5.3	25.5	16.1	84 Trp	15.8	18.3	24.2	147 Tyr	11.3	35.0	29.1
23 Pro	−2.0	23.5	15.7	85 Gly	13.3	20.3	22.1	148 Asn	10.4	34.2	25.5
24 Glu	−2.7	22.2	19.2	86 Asp	12.8	19.7	18.3	149 Ala	11.6	37.3	23.6
25 Gly	−5.9	20.9	17.5	87 Arg	15.6	21.9	17.2	150 Gly	14.0	39.3	25.8
26 Leu	−8.5	23.2	18.8	88 Gly	18.1	20.6	19.8	151 Ala	13.4	42.5	27.8
27 Ser	−11.1	25.0	16.8	89 Asn	18.6	24.1	21.5	152 Gly	14.2	44.4	24.7
28 Tyr	−11.3	28.1	19.1	90 Gly	16.2	24.0	24.4	153 Asn	10.9	43.2	23.2
29 Cys	−8.8	30.4	20.7	91 Phe	16.8	21.6	27.3	154 Val	8.9	44.6	26.1
30 Gly	−8.3	32.6	23.7	92 Gly	14.5	20.3	30.0	155 Arg	8.3	48.3	25.2
31 Val	−8.3	32.0	27.5	93 Leu	10.9	21.0	31.2	156 Ser	4.7	48.7	26.5
32 Ser	−10.5	29.0	27.3	94 Met	11.3	24.8	30.9	157 Tyr	2.2	46.9	28.6
33 Ala	−8.2	27.4	24.7	95 Gln	13.5	24.7	27.8	158 Ala	−0.5	46.6	25.8
34 Ser	−5.2	28.0	27.0	96 Val	16.6	26.6	28.8	159 Arg	1.7	45.3	23.0
35 Lys	−7.0	26.4	29.9	97 Asp	18.4	27.6	25.5	160 Met	4.3	43.3	25.0
36 Lys	−8.0	23.4	27.7	98 Lys	21.9	26.1	25.4	161 Asp	3.5	39.9	23.5
37 Ile	−4.4	22.9	26.7	99 Arg	23.0	28.9	23.0	162 Ile	3.7	41.1	19.9
38 Ala	−3.3	23.4	30.3	100 Ser	22.4	31.4	25.9	163 Gly	6.9	39.6	18.5
39 Glu	−5.8	20.7	31.3	101 His	22.9	29.5	29.1	164 Thr	6.9	36.7	21.0
40 Arg	−4.5	18.3	28.6	102 Lys	24.8	26.3	29.8	165 Thr	6.3	33.2	19.5
41 Asp	−1.1	18.6	30.3	103 Pro	22.0	23.9	31.0	166 His	2.8	33.3	18.1
42 Leu	−2.2	18.3	33.9	104 Gln	22.3	22.5	34.5	167 Asp	2.4	36.9	19.5
43 Gln	−1.6	14.6	34.4	105 Gly	20.9	19.2	36.0	168 Asp	0.9	35.6	22.7
44 Ala	2.0	14.9	33.1	106 Thr	18.6	16.4	34.8	169 Tyr	3.5	34.1	24.9
45 Met	2.5	18.1	35.1	107 Trp	16.1	17.5	32.2	170 Ala	2.8	35.7	28.2
46 Asp	1.4	16.2	38.2	108 Asn	13.3	17.0	34.6	171 Asn	−1.0	36.0	27.6
47 Arg	4.6	14.0	38.1	109 Gly	14.7	17.7	38.1	172 Asp	−1.1	32.2	27.1
48 Tyr	6.5	17.1	38.9	110 Glu	15.0	20.4	40.7	173 Val	1.2	31.3	29.9
49 Lys	4.1	19.0	41.1	111 Val	18.3	21.8	39.4	174 Val	−0.7	33.5	32.3
50 Thr	6.1	18.6	44.4	112 His	16.8	22.3	36.0	175 Ala	−4.0	31.8	31.3
51 Ile	9.3	19.8	42.9	113 Ile	13.6	23.9	37.4	176 Arg	−2.5	28.4	31.7
52 Ile	7.5	22.7	41.0	114 Thr	15.7	26.2	39.6	177 Ala	−1.0	29.4	35.1
53 Lys	6.1	23.8	44.3	115 Gln	17.8	27.3	36.7	178 Gln	−4.5	30.5	36.2
54 Lys	9.4	23.2	46.2	116 Gly	14.8	28.0	34.5	179 Tyr	−5.8	26.9	35.4
55 Val	11.2	25.6	43.7	117 Thr	13.0	29.9	37.3	180 Tyr	−3.0	25.4	37.4
56 Gly	8.4	28.2	43.7	118 Thr	16.0	31.9	38.1	181 Lys	−3.5	27.8	40.3
57 Glu	8.5	28.3	47.5	119 Ile	16.3	33.0	34.5	182 Gln	−7.0	26.3	40.7
58 Lys	12.3	28.5	47.5	120 Leu	12.7	34.0	34.7	183 His	−5.5	22.8	40.7
59 Leu	12.5	31.4	45.0	121 Ile	13.1	36.0	38.0	184 Gly	−3.4	23.2	43.8
60 Cys	9.2	33.1	46.1	122 Asn	16.0	37.9	36.4	185 Tyr	−0.3	24.8	42.2
61 Val	7.4	32.9	42.8	123 Phe	13.7	38.9	33.6				
62 Glu	4.1	31.2	42.1	124 Ile	11.0	40.0	36.1				

Coordinates are in angstroms in an orthogonal system with axes parallel to a, b, and c* and origin at the crystallographic origin

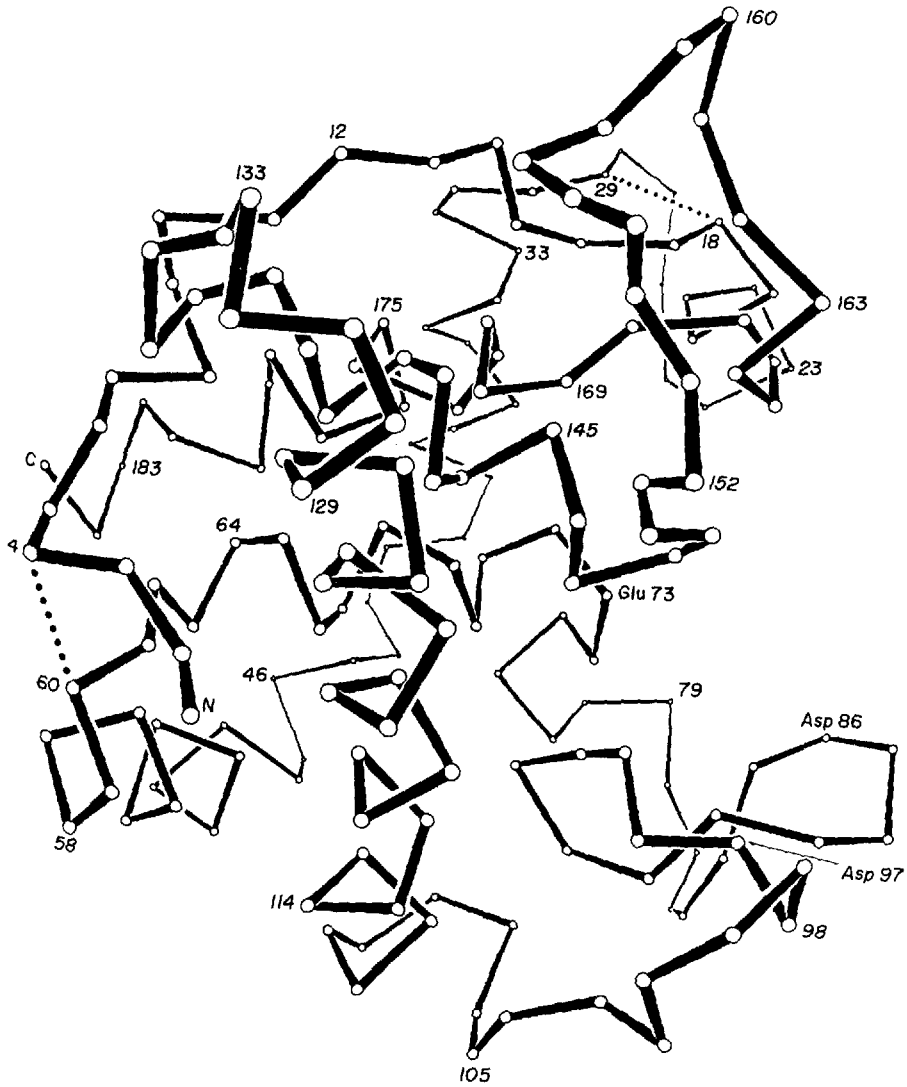

Fig. 1. Backbone conformation of goose egg-white lysozyme (GEWL), with disulfide linkages indicated by dotted lines. The active-site cleft is to the right

In an attempt to distinguish between convergent and divergent evolution of related protein structures, we have determined and compared the structures of a series of lysozymes. The three-dimensional structures of lysozymes from four distinct classes have been reported: (1) chicken-type lysozyme, typified by hen egg-white lysozyme (HEWL) (Blake et al. 1965; Imoto et al. 1972); (2) lysozyme from bacteriophage T4 (T4L) (Matthews and Remington 1974; Remington et al. 1978); (3) goose-type lysozyme from the egg white of Embden Goose (*Anser anser*) (GEWL) (Grütter et al. 1983) this class also includes the lysozyme from Australian Black Swan (Isaacs et al. 1985); and (4) the bacterial lysozyme produced by *Streptomyces erythraeus* (Sarma et al. 1979; Harada et al. 1981). Within a given class the amino acid sequences of the lysozymes are clearly homologous, but there is no obvious sequence correspondence between one class and another (Inouye and Tsugita 1966; Canfield and McMurry 1967; Canfield et al. 1971; Simpson et al. 1980; Schoentgen et al. 1982; Simpson and Morgan, 1983). In spite of the lack of agreement among their amino acid sequences, the three-dimensional structures of GEWL, HEWL, and T4L have elements in common, and in a brief report we argued that the three classes of lysozyme probably diverged from a common, albeit distant, precursor (Grütter et al. 1983).

The availability of accurate coordinates for GEWL has permitted a detailed structural comparison of the three classes of lysozyme. The comparison illustrates the changes that are presumed to have occurred in these enzymes during evolution. In this communication we give additional details of the structure of GEWL and describe its structural relationship to HEWL and T4L.

In collaboration with Dr. V.R. Sarma (unpublished results) we have compared the reported backbone conformation of *Streptomyces erythraeus* ly-

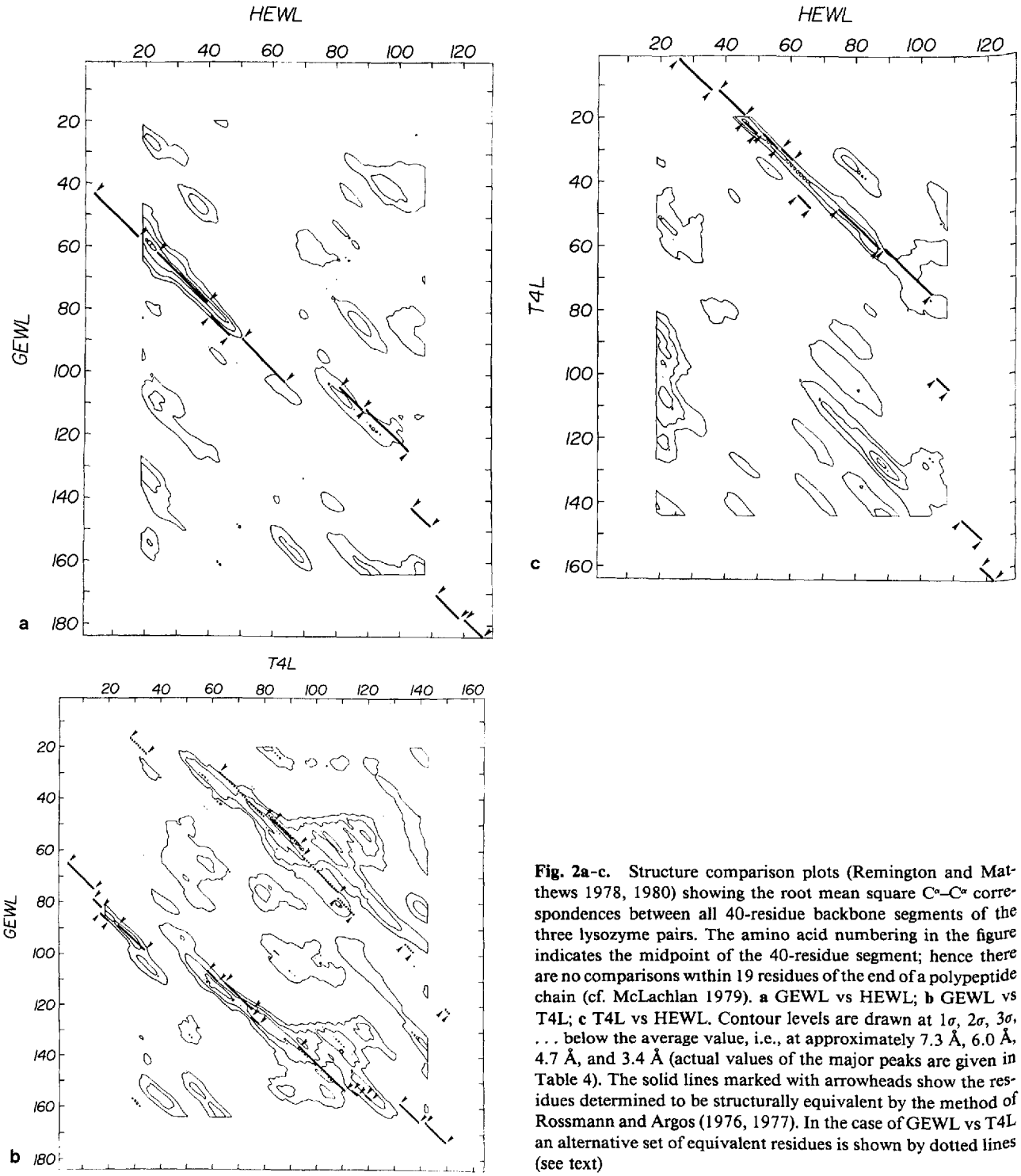

Fig. 2a–c. Structure comparison plots (Remington and Matthews 1978, 1980) showing the root mean square C^{α}–C^{α} correspondences between all 40-residue backbone segments of the three lysozyme pairs. The amino acid numbering in the figure indicates the midpoint of the 40-residue segment; hence there are no comparisons within 19 residues of the end of a polypeptide chain (cf. McLachlan 1979). **a** GEWL vs HEWL; **b** GEWL vs T4L; **c** T4L vs HEWL. Contour levels are drawn at 1σ, 2σ, 3σ, ... below the average value, i.e., at approximately 7.3 Å, 6.0 Å, 4.7 Å, and 3.4 Å (actual values of the major peaks are given in Table 4). The solid lines marked with arrowheads show the residues determined to be structurally equivalent by the method of Rossmann and Argos (1976, 1977). In the case of GEWL vs T4L an alternative set of equivalent residues is shown by dotted lines (see text)

sozyme (Harada et al. 1981) with those of HEWL, T4L, and GEWL, but have not detected any obvious resemblance.

Structure of Goose Lysozyme

The structure of GEWL was initially determined by model building and refined based on 3.2-Å and 2.8-Å resolution electron density maps (Grütter et al. 1983). Data were collected for the native crystals and five isomorphous derivatives by oscillation photography (Rossmann 1979; Schmid et al. 1981). The resolution has since been increased to 1.6 Å and the refinement continued to a crystallographic residual of 19.0% (L.H. Weaver and B.W. Matthews, unpublished results). The C^{α} coordinates are

given in Table 1 and the structure is illustrated in Fig. 1. During refinement, the irregular loop from residues 150 to 170 was reinterpreted, resulting in the "insertion" of four residues and a resultant renumbering of residues within the C-terminal helix. With the exception of this change, the structure is as reported previously (Grütter et al. 1983).

The structure is predominantly helical, with seven α-helices (α_1–α_7) including residues 19–24, 31–45, 48–60, 62–74, 111–131, 136–148, and 170–183. There is also a region within which three extended strands, comprising residues 83–86, 89–92, and 94–98, form a very irregular antiparallel β-sheet.

Structure of Swan Lysozyme

The structure of the goose-type lysozyme from the egg white of the Australian Black Swan *Cygnus atratus* is very similar to that of Embden Goose lysozyme (Isaacs et al. 1985). The C^{α} atoms of swan lysozyme can be superimposed on those of GEWL with a root mean square discrepancy of 2.3 Å. This discrepancy might be due in large part to errors in the coordinates, and the swan lysozyme structure needs to be refined before it can be determined if there are significant differences. Within the limitations of the available data, the backbones of the two goose-type lysozymes are very similar, as is consistent with their homologous amino acid sequences, which differ in only six places (Simpson and Morgan 1983). For the comparison of goose-type lysozyme with the other lysozymes described in the remainder of this paper, we have used the GEWL coordinates.

Comparison of Lysozyme Backbones

The GEWL structure consists of two domains linked in part by a long α-helix (Fig. 1). This overall arrangement is reminiscent of T4L, although the amino-terminal residues of GEWL extend from one domain to the other and thicken the "waist" of the molecule relative to the phage enzyme. The occurrence of three β-strands in the "lower" domain, together with a predominantly α-helical "upper" domain (Fig. 1), suggests similarity with both HEWL and T4L.

To quantitate the relationship among the backbones of GEWL, T4L, and HEWL, we used the methods of both Remington and Matthews (1978, 1980) and Rossmann and Argos (1976, 1977).

The Remington–Matthews approach will detect structural correspondence between any segment of one structure and any segment of the other structure. Structure comparison plots for GEWL vs HEWL, GEWL vs T4L, and HEWL vs T4L calculated with a probe length of 40 residues are shown in Fig. 2, and the best segment comparisons for probe lengths of 40 and 20 residues are summarized in Table 2.

Table 2. Best segment–segment backbone correspondences between lysozyme structures

Segments compared		$R_{C\alpha}$ (Å)	$\overline{R_{C\alpha}}$ (Å)	σ (Å)	$\frac{\overline{R_{C\alpha}} - R_{C\alpha}}{\sigma}$
40-Residue segment					
GEWL	HEWL				
40–79*	3–42*	3.3	8.66	1.19	4.5
22–61	77–116	5.1			3.0
103–142	3–42	5.1			3.0
GEWL	T4L				
31–70	65–104	3.9	8.29	1.52	2.9
40–79	88–127	3.9			2.9
108–147*	63–102*	4.1			2.8
T4L	HEWL				
108–147	69–108	3.7	8.60	1.35	3.6
90–129	4–43	4.5			3.0
74–113	3–42	4.6			3.0
20-Residue segment					
GEWL	HEWL				
57–76*	20–39*	1.0	5.69	1.23	3.8
134–153	88–107	1.8			3.2
165–184	85–104	1.9			3.1
GEWL	T4L				
110–129*	60–79*	0.4	5.38	1.19	4.2
130–149*	88–107*	0.8			3.8
128–147	32–51	1.7			3.1
T4L	HEWL				
92–111	88–107	1.6	5.58	1.22	3.3
136–155	80–99	1.9			3.0

$R_{C\alpha}$ is the root mean square discrepancy between the listed α-carbon segments, $\overline{R_{C\alpha}}$ is the mean value of $R_{C\alpha}$ for all possible segment alignments, σ is the standard deviation of $R_{C\alpha}$, and $(\overline{R_{C\alpha}} - R_{C\alpha})/\sigma$ is the number of standard deviations by which the quoted value of $R_{C\alpha}$ is less than the mean value (cf. Remington and Matthews 1980). Segment alignments marked with asterisks correspond to the overall lysozyme correspondences given in Table 5 and Figs. 2 and 3

In the Rossmann–Argos method, two structures are compared by means of a rigid-body superposition designed to maximize the number of "equivalences" (i.e., spatially corresponding α-carbon atoms) in the two molecules. The results of using this method for comparison of the different lysozymes are summarized in Tables 3 and 4 and Figs. 3 and 4. Table 4 includes results for several other protein comparisons. Protein pairs chosen for inclusion in the table were, as far as possible, those for which the amino acid sequences are known and the correspondence between the sequences is poor or nonexistent. The protein pairs are ranked according to the fraction of residues in the smaller protein (Mol-

Table 3. Overall structural correspondence between lysozymes

HEWL	$d_{H,G}$	GEWL	$d_{G,T}$	T4L	$d_{T,H}$	HEWL
		Δ				
Lys1		Asp41				
Val2		Leu42				
Phe3	8.0	Gln43				
Gly4	5.2	Ala44				
Arg5	3.9	Met45				
Cys6	5.0	Asp46				
Glu7	6.0	Arg47				
Leu8	3.5	Tyr48				
Ala9	4.0	Lys49				
Ala10	3.6	Thr50				
Ala11	2.3	Ile51				
Met12	2.9	Ile52				
Lys13	3.2	Lys53				
Arg14	2.3	Lys54				
His15	2.4	Val55				
Gly16	4.5	Gly56				
Leu17	4.9	Glu57				
Δ		Lys58				
Δ		Leu59				
Δ		Cys60				
Tyr23	4.7	Val61				
Ser24	2.8	Glu62				Δ
Leu25	2.5	Pro63		Met1	4.4	Leu25
Gly26	1.3	Ala64		Asn2	3.5	Gly26
Asn27	1.3	Val65	2.5	Ile3	2.8	Asn27
Trp28	1.3	Ile66	1.5	Phe4	2.1	Trp28
Val29	1.2	Ala67	2.8	Glu5	2.4	Val29
Cys30	1.5	Gly68	2.6	Met6	2.7	Cys30
Ala31	1.4	Ile69	1.6	Leu7	1.8	Ala31
Ala32	1.1	Ile70	2.5	Arg8	1.7	Ala32
Lys33	2.1	Ser71	3.1	Ile9	2.4	Lys33
Phe34	2.7	Arg72	2.3	Asp10	2.4	Phe34
Glu35	1.9	Glu73	2.0	Glu11	1.6	Glu35
						Ser36
						Asn37
Ser36	2.0	Ser74	3.1	Gly12	4.6	Phe38
Asn37	2.7	His75	4.9	Leu13	4.4	Asn39
Phe38	2.7	Ala76				
Asn39	3.0	Gly77				
		Lys78				
Thr40		Val79	5.8	Arg14	3.0	Thr40
		Leu80	5.7	Leu15	4.0	Gln41
		Lys81	3.4	Lys16	6.6	Ala42
Gln41	4.7	Asn82				
Ala42	3.0	Gly83				
Thr43	1.4	Trp84	1.5	Ile17	4.9	Thr43
Asn44	0.9	Gly85	1.7	Tyr18	3.3	Asn44
Arg45	3.7	Asp86	1.9	Lys19	2.5	Arg45
Asn46	3.4	Arg87		Asp20		
Δ		Gly88		Thr21		
Δ				Glu22	4.0	Asn46
Δ				Gly23	4.7	Thr47
Thr51	0.8	Asn89	3.8	Tyr24	5.9	Asp48
Asp52	0.6	Gly90	4.2	Tyr25	6.0	Gly49
						Ser50
						Thr51
Tyr53	0.7	Phe91	3.4	Thr26	1.2	Asp52
Gly54	0.6	Gly92	3.0	Ile27	1.2	Tyr53
Ile55	1.1	Leu93	2.5	Gly28	2.8	Gly54
Leu56	1.0	Met94	1.3	Ile29	3.8	Ile55
						Leu56
Gln57	0.9	Gln95	1.9	Gly30	1.7	Gln57

Table 3. Continued

HEWL	$d_{H,G}$	GEWL	$d_{G,T}$	T4L	$d_{T,H}$	HEWL
Ile58	0.9	Val96	1.7	His31	0.8	Ile58
Asn59	0.6	Asp97	1.6	Leu32	1.6	Asn59
Ser60	0.8	Lys98	2.0	Leu33	1.3	Ser60
Arg61	0.7	Arg99		Thr34		Arg61
Trp62	0.4	Ser100		Lys35		Trp62
Trp63	0.9	His101		Δ		
Cys64	1.4	Lys102		Leu46	6.8	Trp63
Δ				Asp47	5.6	Cys64
Δ				Lys48	4.7	Asn65
Δ						Δ
Δ		Pro103	3.3	Ala49	4.2	Leu75
Ala82	3.9	Gln104	3.8	Ile50	4.7	Cys76
				Gly51	3.3	Asn77
				Arg52	5.3	Ile78
				Asn53	5.8	Pro79
				Cys54	6.3	Cys80
				Asn55	6.1	Ser81
Leu83	3.7	Gly105		Gly56	7.2	Ala82
Leu84	3.0	Thr106		Val57	6.3	Leu83
Ser85	4.3	Trp107	3.6	Ile58	5.0	Leu84
Ser86	4.3	Asn108	3.8	Thr59	3.5	Ser85
Asp87	2.7	Gly109	3.8	Lys60	5.0	Ser86
						Asp87
						Ile88
Ile88	1.7	Glu110	4.2	Asp61	1.7	Thr89
Thr89				Glu62	1.3	Ala90
				Ala63	1.5	Ser91
				Glu64	1.6	Val92
Ala90	2.2	Val111	2.2	Lys65	1.4	Asn93
Ser91	1.9	His112	1.9	Leu66	1.5	Cys94
Val92	3.2	Ile113	2.5	Phe67	1.4	Ala95
Asn93	3.2	Thr114	1.9	Asn68	1.5	Lys96
Cys94	1.9	Gln115	1.9	Gln69	1.2	Lys97
Ala95	2.2	Gly116	2.6	Asp70	1.3	Ile98
Lys96	3.9	Thr117	2.3	Val71	1.7	Val99
Lys97	3.7	Thr118	1.8	Asp72	1.7	Ser100
Ile98	1.8	Ile119	2.9	Ala73	2.2	Asp101
Val99	1.9	Leu120	2.9	Ala74	4.4	Gly102
Ser100	3.9	Ile121	2.5	Val75	4.2	Asn103
Asp101	4.0	Asn122				Gly104
Gly102	3.7	Phe123				
Asn103	6.0	Ile124				
		Lys125	4.5	Arg76		
		Thr126	4.5	Gly77		
		Ile127	4.2	Ile78		
				Δ		
		Gln128	3.4	Lys85		
		Lys129	3.8	Pro86		
		Lys130	4.4	Val87		
		Phe131	4.8	Tyr88		
		Pro132	3.6	Asp89		
		Ser133	3.0	Ser90		
		Trp134	2.8	Leu91		
		Thr135	2.9	Asp92		
		Lys136	3.8	Ala93		
		Asp137	3.0	Val94		
		Gln138	2.2	Arg95		
		Gln139	2.8	Arg96		
		Leu140	2.3	Cys97		
		Lys141	1.2	Ala98		
		Gly142	1.3	Leu99		
Gly104	1.9	Gly143	1.8	Ile100		
Met105	2.0	Ile144	1.2	Asn101	4.2	Met105

Table 3. Continued

HEWL	$d_{H,G}$	GEWL	$d_{G,T}$	T4L	$d_{T,H}$	HEWL
Ala107	3.7	Ala146	1.8	Val103	3.8	Ala107
Trp108	3.9	Tyr147	1.5	Phe104	2.6	Trp108
Val109	3.9	Asn148	1.4	Gln105		Val109
		Ala149	1.8	Met106		
		Gly150	3.7	Gly107		
		Ala151	3.7	Glu108		
		Gly152	3.6	Thr109		
		Asn153	4.0	Gly110		
				Val111		
				Ala112		
				Gly113		
		Val154	3.7	Phe114		
		Arg155	2.9	Thr115		
				Asn116		
				Ser117		
				Leu118		
		Ser156	4.8	Arg119		
		Tyr157	5.3	Met120		
		Ala158		Δ		
		Arg159	3.4	Asn132		
		Met160	3.9	Leu133		
		Asp161	5.7	Ala134		
		Ile162	4.7	Lys135		
		Gly163	3.3	Ser136		
		Thr164	6.1	Arg137		
		Thr165	4.2	Trp138		
		His166	4.1	Tyr139		
				Asn140		
				Gln141		
		Asp167	4.9	Thr142		
Ala110		Asp168	4.5	Pro143		Ala110
Trp111		Tyr169	6.9	Asn144		Trp111
Arg112	1.7	Ala170	5.9	Arg145		Arg112
Asn113	4.8	Asn171	5.2	Ala146	3.4	Asn113
Arg114	1.8	Asp172	5.3	Lys147	4.7	Arg114
Cys115	1.6	Val173	5.6	Arg148	4.7	Cys115
Lys116	2.6	Val174		Val149	4.0	Lys116
Gly117	5.2	Ala175		Ile150	6.2	Gly117
Thr118	6.2	Arg176		Thr151	4.6	Thr118
Asp119				Ala160	5.9	Asp119
Val120	2.8	Ala177		Tyr161	4.9	Val120
Gln121	3.9	Gln178		Lys162	6.9	Gln121
Ala122	3.0	Tyr179		Asn163	5.7	Ala122
Trp123	0.6	Tyr180		Leu164	6.3	Trp123
Ile124	3.4	Lys181				Δ
Arg125	2.4	Gln182				
Gly126	4.2	His183				
Cys127	2.8	Gly184				
		Tyr185				

The three lysozyme backbones are aligned according to the method of Rossmann and Argos (1976, 1977). Distances in angstroms between "equivalent" α-carbon atoms are denoted by $d_{x,y}$, e.g., $d_{H,G}$ for HEWL and GEWL. Triangles indicate that residues are deleted. Amino acids that are the same in two lysozymes have single underlines; a double underline indicates identity in all three lysozyme structures

ecule 2) that have structurally equivalent counterparts in the larger protein. However, different methods of defining "equivalences" were used in some cases, so the ranking should be taken only as a rough guide.

The relationships between the different lysozymes given in Table 3 are similar, but not identical, to those given previously (Rossmann and Argos 1976; Matthews et al. 1981a, b; Grütter et al. 1983). The differences are due in part to the refinement of the respective coordinates, but also are characteristic of the structure comparison procedure. The Rossmann–Argos program employs a complicated algorithm that, in our hands, rarely converges to a well-defined optimum structural correspondence. Rather, the program usually oscillates between a number of alternative structural alignments that are similar in their main features but can differ substantially in detail. The user has to make a somewhat arbitrary choice among these alternatives. There is another problem that occurs when one is trying to obtain an internally consistent alignment of three (or more) structures, as is the case here. For example, in Table 3 Asn 27 of HEWL aligns with Val 65 of GEWL, which aligns with Ile 3 of T4L, which in turn aligns with Asn 27 of HEWL. However, Ala 90 of HEWL aligns with Val 111 of GEWL, which aligns with Lys 65 of T4L, which aligns with Asn 93 (not Ala 90) of GEWL. There are other such inconsistencies, often corresponding to alternative alignments in which two α-helices are "out of register" by one turn. The structural alignments shown in Table 3 were chosen as a compromise between maximizing the number of "equivalences" between the respective lysozymes and achieving internal consistency.

The transformations that relate the three lysozyme structures are given below. The respective lysozyme coordinates are in angstroms in orthogonal Cartesian coordinate systems: (X_H, Y_H, Z_H) for HEWL (Imoto et al. 1972) and (X_T, Y_T, Z_T) for T4L (Remington et al. 1977; L.H. Weaver, T. Gray, and B.W. Matthews, unpublished results). The GEWL coordinates (X_G, Y_G, Z_G) are in an orthogonal system parallel to the crystallographic a, b, and c* axes, with the origin coincident with the origin of the unit cell.

(a) HEWL and GEWL:

$$(1)\quad \begin{cases} 0.1470X_H - 0.6142Y_H + 0.7754Z_H + 9.29 = X_G \\ -0.2673X_H - 0.7301Y_H + 0.6290Z_H - 0.99 = Y_G \\ -0.9523X_H - 0.2997Y_H - 0.0568Z_H + 39.33 = Z_G \end{cases}$$

(b) T4L and GEWL:

$$(2)\quad \begin{cases} -0.3174X_T + 0.6352Y_T + 0.7041Z_T + 19.00 = X_G \\ -0.8425X_T + 0.1520Y_T - 0.5168Z_T + 70.02 = Y_G \\ -0.4353X_T - 0.7572Y_T + 0.4869Z_T + 34.32 = Z_G \end{cases}$$

(c) HEWL and T4L:

$$(3)\quad \begin{cases} 0.5401X_H - 0.5097Y_H - 0.6697Z_H + 64.83 = X_T \\ 0.7950X_H + 0.0479Y_H + 0.6047Z_H - 23.12 = Y_T \\ -0.2761X_H - 0.8590Y_H + 0.4311Z_H + 25.42 = Z_T \end{cases}$$

We will first discuss the relationships among the backbone conformations of the three lysozyme types,

Table 4. Structural comparisons of proteins with weak or nonexistent sequence matching[a]

Proteins[b]		Number of residues		Number of equivalences	Percentage of equivalences		Rms C^{α}–C^{α} (Å)[c]	MBC/C[d]	Reference[e]
Mol. 1	Mol. 2	Mol. 1	Mol. 2		Mol. 1	Mol. 2			
Rhod.(1)	Rhod.(2)	142	135	117	82	87	2.0	1.27	1
GEWL	HEWL	185	129	94	51	73	2.8	1.38	2
TIM	KDPG Ald.	248	225	150	60	67	3.0	1.43	3
GPD(NAD)	LDH(NAD)	148	144	96	65	67	2.9	1.24	4, 5
Hb(β)	Cyt.b_5	146	86	58	40	67	3.9	1.29	5, 6, 7
TIM	PK(A)	247	216	143	58	66	3.2	—	3, 8
CON A	TBSV(P)	237	110	68	29	62	3.4	—	9
SBMV(A)	STNV	187	173	104	56	60	3.7	1.36	10, 11
T4L	HEWL	164	129	76	46	59	3.9	1.45	5, 2
GEWL	T4L	185	164	92	50	56	3.5	1.45	2
λRep.(DNA)	Cro	92	66	36	39	55	3.1	1.24	12
Cyt.b_5	Cyt.b_{551}	86	82	41	48	50	4.9	—	6
CAP(DNA)	Cro	73	66	31	42	47	3.1	1.52	13
SOD	V_L	150	110	51	34	46	2.7	—	14
HBH(FAD)	GR(FAD)	164	161	69	42	43	1.2	—	15
P.Peps.(1)	P.Peps.(2)	181	150	61	34	41	2.0	—	16
GEWL(A)	T4L(A)	185	164	52	28	32	2.7	1.35	2
LDH(NAD)	Flavod.	144	138	39	27	28	2.4	1.23	17

[a] This compilation is taken in part from Matthews and Rossmann (1985)

[b] Mol., molecule; Rhod.(1) and Rhod.(2), first and second domains of rhodanese; GEWL, goose egg-white lysozyme; HEWL, hen (chicken) egg-white lysozyme; TIM, triosephosphate isomerase; KDPG Ald., 2-keto-3-deoxy-6-phosphogluconate aldolase; GPD(NAD), nucleotide adenine dinucleotide (NAD)-binding domain of glyceraldehyde-3-phosphate dehydrogenase; LDH(NAD), NAD-binding domain of lactate dehydrogenase; Hb(β), β-chain of horse hemoglobin; Cyt.b_5, cytochrome b_5; PK(A), domain A of pyruvate kinase; Con A, concanavalin A; TBSV(P), P-domain of tomato bushy stunt virus; SBMV(A), A-subunit of southern bean mosaic virus; STNV, satellite tobacco necrosis virus; T4L, bacteriophage T4 lysozyme; λRep.(DNA), DNA-binding domain of λ repressor protein; Cro, cro repressor protein; Cyt.b_{551}, cytochrome b_{551}; CAP(DNA), DNA-binding domain of catabolite gene activator protein; SOD, superoxide dismutase; V_L, variable domain of immunoglobulin light chain; HBH(FAD), flavin adenine dinucleotide (FAD)-binding domain of *p*-hydroxybenzoate hydroxylase; GR(FAD), FAD-binding domain of glutathione reductase; P.Peps.(1) and P.Peps.(2), first and second domains of penicillopepsin; GEWL(A) and T4L(A), alternative alignment of GEWL and T4L (Figs. 2 and 5); Flavod., flavodoxin

[c] Root mean square distance between equivalent α-carbons of molecules

[d] Minimum base change per codon for the residues that are structurally equivalent in the structures being compared

[e] References are as follows: (1) Ploegman et al. (1978); (2) this work; (3) Lebioda et al. (1982); (4) Rossmann et al. (1974); (5) Rossmann and Argos (1976); (6) Argos and Rossmann (1979); (7) Rossmann and Argos (1975); (8) Levine et al. (1978); (9) Argos et al. (1980); (10) Liljas et al. (1982); (11) Rossmann et al. (1983); (12) Ohlendorf et al. (1983); (13) Steitz et al. (1982); (14) Richardson et al. (1976); (15) Wierenga et al. (1983); (16) Tang et al. (1978); (17) Rao and Rossmann (1973)

and then consider the correspondence among the respective active sites. Next we will ask whether the structural alignment reveals any underlying amino acid homology, and finally we will consider the possible evolutionary relationship among the three lysozymes.

Backbone Conformational Relationships

The percentage of residues that are "equivalent" in any pair of the three lysozymes considered here is reasonably high relative to other examples that have been reported (Table 4). However, there are many differences in conformation at the local level. This is seen most clearly in Figs. 2 and 3 and Table 2. In general, the equivalent residues do not fall into long, uninterrupted stretches, but tend to be broken up into segments of about 5 to 15 residues. Also, the equivalent residues determined by the method of Rossmann and Argos are not necessarily those that have the best local agreement (as measured using probe lengths of 20 and 40 residues). One could imagine a protein to be made up of more-or-less rigid substructures or subdomains that would tend to be conserved during evolution. Such a hypothesis is not supported by the observed relationship among the structures described here. If the three lysozymes diverged from a common evolutionary precursor, then this process has not precisely conserved local substructures, except at the level of individual α-helices and β-sheets. Rather, the lysozymes have undergone conformational changes at both the global and local levels.

From an analysis of the globins, Lesk and Chothia (1980) concluded that the packing of α-helices tends to be conserved during evolution, but that the he-

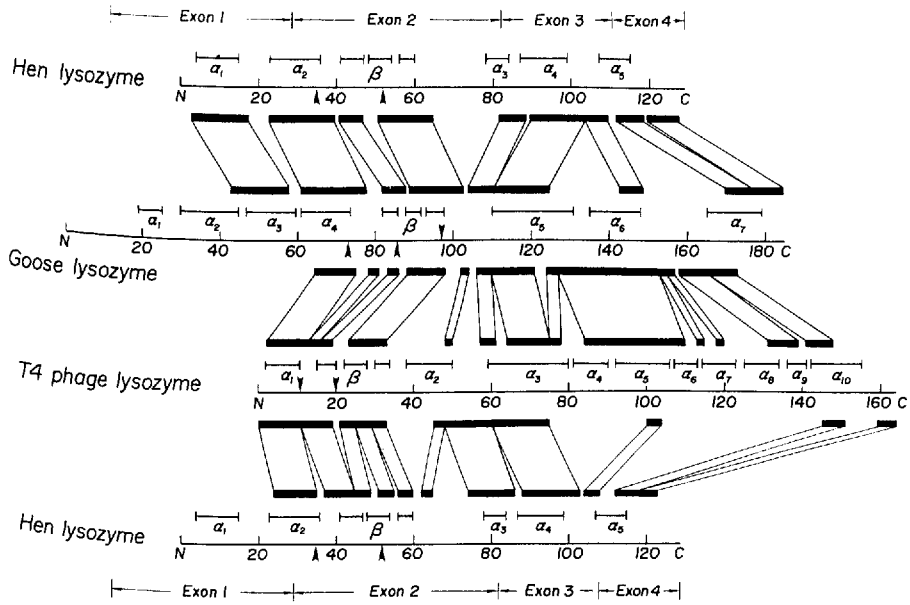

Fig. 3. Structural correspondence between goose, hen, and T4 phage lysozymes. The connected solid bars indicate parts of the polypeptide backbones that structurally correspond when the lysozymes are compared in pairs by the method of Rossmann and Argos (see text). The locations of the α-helices and β-strands are shown, as are the exons of HEWL. The arrows show the locations of residues that may be involved in catalysis, viz. Glu 35 and Asp 52 of HEWL; Glu 11 and Asp 20 of T4L; and Glu 73, Asp 86, and Asp 97 of GEWL

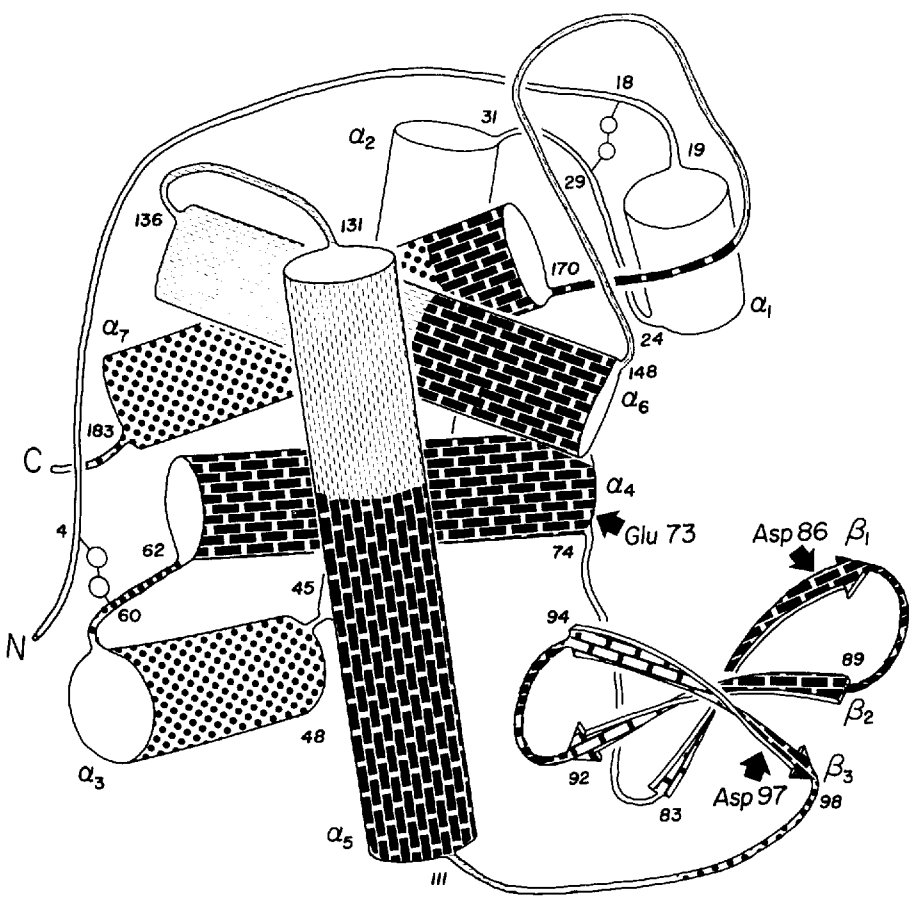

Fig. 4. Simplified drawing of GEWL showing those parts of the structure that are common to hen and to phage lysozymes. Parts common to all three lysozymes are indicated by a brick-like pattern, parts common to GEWL and HEWL are dotted, parts common to GEWL and T4L are dashed, and parts that occur only in GEWL are shown as open areas. [Reprinted, with permission, from Grütter et al. (1983); copyright © 1983, Macmillan Journals, Ltd.]

lices undergo rigid-body displacements relative to each other, with the connecting regions having relatively variable conformations. To some extent we see this tendency in the three lysozyme structures, in that there are corresponding α-helices that have rigid-body displacements, but in other cases corresponding helices have increased or decreased in length or helices in one structure have no counterparts in another (Figs. 3 and 4).

The "overall" alignment of GEWL and T4L has 92 equivalent α-carbons with a root mean square discrepancy of 3.5 Å, and aligns α_4-β_1-β_2-β_3-α_5-α_6 of

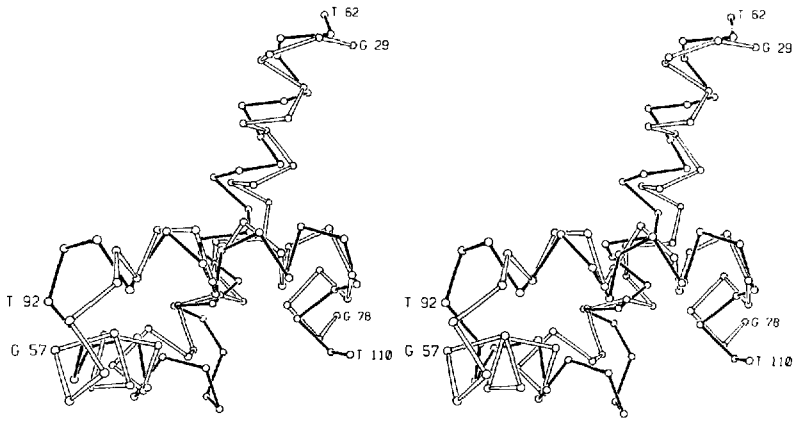

Fig. 5. Structural correspondence of three consecutive α-helices in GEWL and T4L. GEWL backbone shown open and labeled "G," T4L backbone drawn solid and labeled "T." (This structural correspondence is not related to the overall alignment of the two structures shown in Fig. 4)

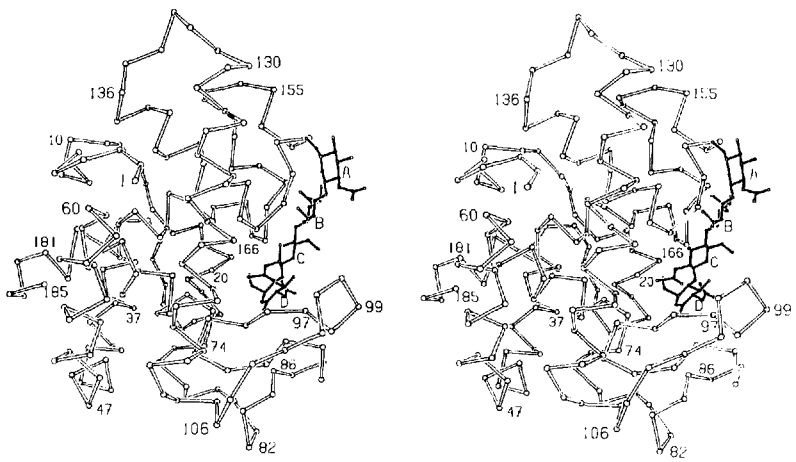

Fig. 6. Stereo drawing of the GEWL backbone, with the coordinates of a tetrasaccharide bound to HEWL (Ford et al. 1974) transformed, using Eqs. (1), into the GEWL coordinate system

GEWL with α_1-β_1-β_2-β_3-α_3-α_5 of T4L. The comparison plot of GEWL with T4L with a probe length of 40 residues (Fig. 2b) includes the expected extended minima corresponding to this alignment. However, there is another extended minimum suggesting a different correspondence between parts of these two molecules. Indeed, this is the case. There are 52 α-carbons of GEWL and T4L that are equivalent with a root mean square discrepancy of 2.7 Å. These equivalences are included in Fig. 2 and predominantly include the superposition of the second, third, and fourth α-helices of GEWL on the third, fourth, and fifth helices of T4L. Within these three helices, 40 consecutive residues superimpose within 2.47 Å (Fig. 5).

This remarkable agreement between the α_3-α_4-α_5 helices of T4L and the α_2-α_3-α_4 helices of GEWL seems to have arisen because it is an energetically favorable packing arrangement determined, in part, by the helix–helix interactions where the first and third helices cross. Chothia et al. (1977) have shown that such helix–helix contacts are normally restricted to three classes, characterized by angles between the helices of about $-82°$, $-60°$, and $+19°$. Indeed, the helix–helix angle in Fig. 5, as defined by Chothia et al. (1977), is $-70°$, a common packing angle. Since the same part of T4L (approximately residues 60–110) has structural similarity to two parts of GEWL (approximately residues 30–80 and 110–160) (Fig. 2), it might be expected that the two parts of GEWL correspond. Indeed, the α_5-strand-α_6 unit and the α_2-α_3-α_4 helices of GEWL can be approximately superimposed, but the overall correspondence is less precise (53 residues equivalent within 4.3 Å).

Active Site Relationships

In a previous comparison of the active sites of T4L and HEWL we found close similarity in the arrangement of key elements involved in substrate binding and in catalysis (Matthews et al. 1981a,b).

In order to explore the overall relationship between the active site of GEWL and those of the other lysozymes, we used Eqs. (1) to transform the coordinates of a tetrasaccharide bound to HEWL (Ford et al. 1974) into the GEWL coordinate system; the result is shown in Fig. 6. The saccharide occupies the active site region in a very reasonable manner. Recently, the binding of the trisaccharide (N-acetylglucosamine)$_3$ to GEWL has been determined experimentally from a difference Fourier map at

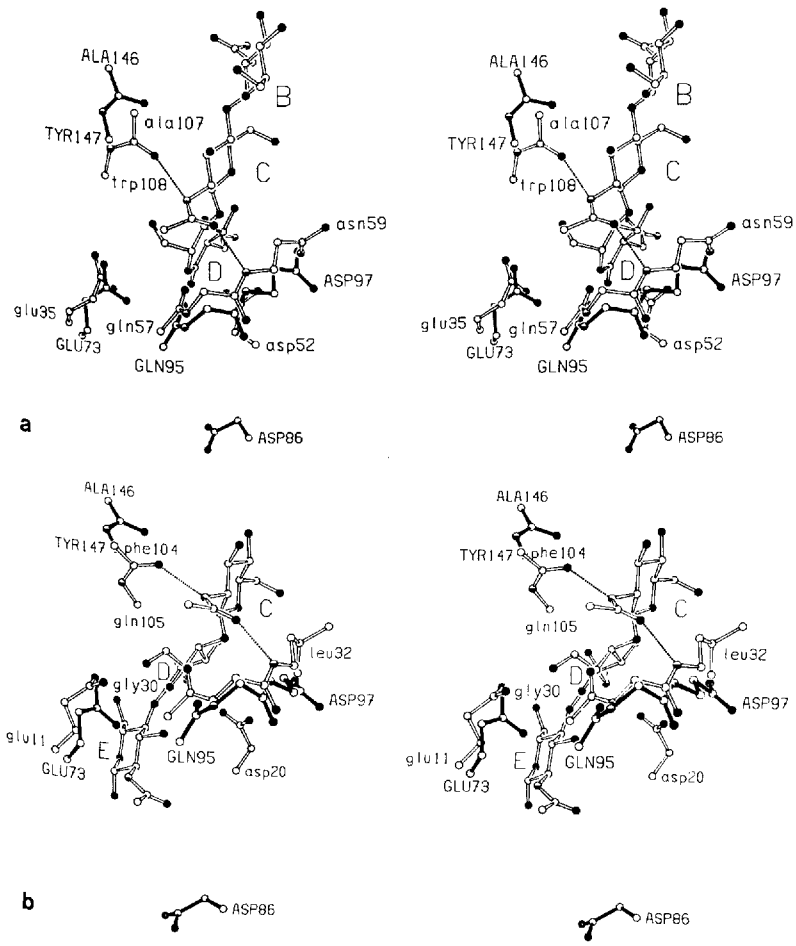

Fig. 7a, b. **a** Potential substrate-binding and catalytic residues in the active site of GEWL superimposed on the active site residues and trisaccharide for HEWL. The two sets of coordinates were superimposed to minimize the discrepancy between GEWL (Glu 73; 95–97 backbone; 146–147 backbone) and HEWL (Glu 35; 57–59 backbone; 107–108 backbone) (20 atoms with root mean square discrepancy of 1.47 Å). GEWL shown with solid bonds and upper-case labels; HEWL, with open bonds and lower-case labels. Oxygen atoms are drawn solid, nitrogens half solid, and carbons open. Observed hydrogen bonds for HEWL are drawn dotted. **b** Active site residues of GEWL superimposed on T4L active site with a trisaccharide as presumed to occupy the C-D-E subsites (Anderson et al. 1981). Superposition based on the best alignment of GEWL (Glu 73; 95–97 backbone; 146–147 backbone) and T4L (Glu 11; 30–32 backbone; 104–105 backbone) (20 atoms with root mean square discrepancy of 1.97 Å). GEWL drawn with solid bonds and upper-case labels. T4L has open bonds and lower-case labels. Symbols for atoms and hydrogen bonds as in a

2.8-Å resolution (L.H. Weaver and B.W. Matthews, unpublished results). This trisaccharide occupies a position close to that of the B, C, and D sugars shown in Fig. 6. The observed saccharide alignment is parallel to and in the same plane as that shown in Fig. 6, but has a bodily translation of about 2.5 Å toward the A subsite. A detailed description will be reported following high-resolution refinement of the GEWL–trisaccharide complex. The results to date confirm the location of the presumed GEWL active site and its alignment relative to the active sites of HEWL and T4L.

We looked for residues in the GEWL active site that might correspond to those elements already shown to be present in both HEWL and T4L. These include the two acidic residues thought to be important in catalysis [viz. Glu 35 and Asp 52 of HEWL (Ford et al. 1974) and Glu 11 and Asp 20 of T4L (Anderson et al. 1981)], as well as parts of the protein backbone that form hydrogen bonds to the N-acetyl group of the saccharide bound in subsite C (Matthews et al. 1981a, b).

In the alignments of the backbones of the three lysozymes, Glu 73 of GEWL coincides with both Glu 35 of HEWL and Glu 11 of T4L (Table 3). Also, there are segments of the GEWL backbone in the active site region that are similar to those seen in T4L and HEWL. However, the counterpart in GEWL to Asp 52 (HEWL) and Asp 20 (T4L) is not obvious. One possibility is Asp 86 (GEWL), which is located in the first β-sheet strand, a position analogous to that of Asp 20 of T4L (Grütter et al. 1983; Isaacs et al. 1985). However, in three dimensions the superposition of Asp 86 (GEWL) on Asp 20 (T4L) is poor, with the respective carboxyl groups about 11 Å apart (Fig. 7b). Similarly, the carboxyl at Asp 86 (GEWL) is about 10 Å from the carboxyl of Asp 52 (HEWL) (Fig. 7a). A second candidate for the "catalytic aspartate" in GEWL is Asp 97. In the respective active site superpositions (Fig. 7), Asp 97 (GEWL) is in better correspondence with Asp 52 (HEWL) and Asp 20 (T4L) than is Asp 86. On the other hand, the superposition is not perfect; the respective carboxyl groups are still about 4 Å apart.

This lack of a clear counterpart in GEWL to the active site aspartates in HEWL and T4L leaves open several possibilities. For example, one or more of the lysozymes may undergo substantial conformational change during catalysis as a result of which the respective aspartates are brought into compa-

Table 5. Active site sequence alignment based on lysozyme structural correspondence[a]

		73							80				84	
GEWL	Arg	Glu	Ser	His	Ala	Gly	Lys	Val	Leu	Lys	Asn	Gly	Trp	
		35						40					43	
HEWL	Phe	Glu	Ser	Asn	Phe	Asn	—	Thr	—	—	Gln	Ala	Thr	
		11											17	
T4L	Asp	Glu	Gly	Leu	Arg	—	—	—	Leu	Lys	—	—	Ile	
		33							40					
EG-SC	Asn	Glu	Ser	Cys	Ala	Glu	Phe	Gly	Asn	—	Gln	—	—	
		86				90							97	
GEWL	Gly	Asp	Arg	Gly	Asn	Gly	Phe	Gly	Leu	Met	Gln	Val	Asp	Lys
		44							50				52	
HEWL	—	Asn	Arg	—	Asn	Thr	Asp	Gly	Ser	—	—	Thr	Asp	Tyr
		18											20	
T4L	—	Tyr	Lys	—	—	—	—	—	—	—	—	—	Asp	Thr
					42								50	
EG-SC	—	—	—	—	Asn	Ile	Pro	Gly	Val	Lys	Asn	Thr	Asp	Tyr

[a] EG-SC is an endo-β-1,4-glucanase from *Schizophyllum commune* (Yaguchi et al. 1983). Underlined residues are those shared by two or more sequences

rable stereochemical positions (although there is no crystallographic evidence for large conformational change in any of the lysozyme–saccharide complexes for which structures have been determined). A second, less likely possibility is that the catalytic activity of GEWL does not require a counterpart to the presumptive "catalytic" aspartates seen in HEWL and T4L. It should be kept in mind that although all three lysozymes cleave the same β-1,4 glycosidic bond, they differ in their specificities toward saccharides with different substituents (Arnheim et al. 1973; Kleppe et al. 1981) and their reactivities toward epoxypropyl β-glycoside inhibitors (Sharon et al. 1974). Studies of the binding of saccharides to GEWL are in progress and may clarify these uncertainties.

Amino Acid Sequences

It is generally accepted that there is no statistically significant overall correspondence between the amino acid sequences of GEWL, HEWL, and T4L. Our own analysis (not shown) by the method of Fitch (1966) supports this finding. However, it is still possible that limited sequence correspondence may exist in, for example, the respective active sites or other limited segments. Jollès and coworkers (Jollès et al. 1981; Schoentgen et al. 1982) have proposed sequence matching between two segments of ostrich lysozyme (a goose-type enzyme) and HEWL. Their proposed alignment for one segment is, in part, consistent with the observed structural correspondence; the other is not.

In terms of the overall structural alignments (Table 3), only two residues are chemically identical in all three lysozymes. The first is the glycine of Gly 54 (HEWL), Gly 92 (GEWL), and Gly 28 (T4L). The second is the glutamate of Glu 35 (HEWL), Glu 73 (GEWL), and Glu 11 (T4L), i.e., the "catalytic" glutamate. Strikingly, in two of the three molecules this glutamate is followed by a serine, and after a few more residues there is a matching Leu–Lys dipeptide in two of the lysozymes. In the case of HEWL and GEWL, there are 95 amino acids in structural correspondence (Table 3), 11 of which are chemically identical. Of these, six (Lys 13, Glu 35, Ser 36, Gly 54, Gly 104, and Ala 107 of HEWL) are among the 40 positions previously found to be invariant in all known chicken-type lysozymes [e.g., see Jollès et al. (1984) and references therein]. In some but not all cases there are obvious similarities between the roles played by these matching residues. For example, Lys 13 of HEWL makes a salt bridge with the carboxyl terminus of the protein (Imoto et al. 1972), and its counterpart residue Lys 53 interacts with the C-terminus of GEWL.

Table 5 shows the active site sequences of GEWL, HEWL, and T4L aligned on the basis of structural correspondence. From Arg 72 to Asp 86 of GEWL the three sequences are aligned strictly according to structural equivalence. From this point on the alignment is based on the assumption that Asp 97 (GEWL) corresponds to Asp 52 (HEWL) and Asp 20 (T4L), although this is very speculative (see above). Also included is a sequence segment for a fungal cellulase proposed to be related to HEWL (Yaguchi et al. 1983), although in this case also the sequence matching is weak.

The limited sequence identity among the different lysozymes is consistent with divergent evolution

from a common but distant precursor, but the level of agreement is so weak that it can hardly be taken as evidence for divergent rather than convergent evolution.

Exon Boundaries

Gilbert (1978) has proposed that the separation of eukaryotic genes into introns and exons might facilitate the evolution of new proteins, and Blake (1978) has suggested that exons might correspond to folded substructures that could be combined together to yield such new protein structures. Evidence both for (Bernard et al. 1978; Craik et al. 1980, 1981) and against (Stein et al. 1980; Quinto et al. 1982) this idea has been presented.

Jung et al. (1980) have shown that the gene of hen lysozyme contains four exons (Fig. 3) and have proposed that each exon corresponds to a functional and, to some extent, structural unit of HEWL. In the previous structural comparison of HEWL and T4L (Matthews et al. 1981a, b), it was noted that the region of best agreement between these two lysozymes corresponded to the second and third exons (Fig. 2), leading Blake and coworkers to argue that the comparison supported the view that recombination within introns can rearrange functional exonic regions into new patterns in new protein products (Artymiuk et al. 1981).

The availability of the GEWL structure will permit additional tests of this hypothesis. One prediction is that if exons do define structural and functional units, their boundaries ought to occur at structurally corresponding positions in HEWL and GEWL. Based on the present structural comparison (Fig. 4), the expected boundaries of exons in GEWL (if they occur) would be at about residues Ile 66, Trp 107, and/or Tyr 147. In addition, Gō (1983) has predicted that the gene for chicken-type lysozymes has or had an additional intron in the region corresponding to residues 53–57 of HEWL. On the basis of the structural comparison of HEWL and GEWL, Gō's prediction would anticipate an intron in the GEWL gene at a position corresponding to residues 91–95.

Inspection of the lysozyme structures does not provide compelling evidence either in support of or against the idea that exon units are conserved in all three molecules. On one hand, large parts of exons 2 and 3 of HEWL are well conserved in GEWL and T4L (Fig. 3). At the same time, there are large deletions and insertions as well, although it could be argued that these differences among the three lysozymes tend to occur at positions close to the exon boundaries. It could also be argued that T4L is completely missing the structural unit corresponding to exon 1 of HEWL, and that GEWL does not have a corresponding domain (Fig. 3). On the other hand, part of the amino-terminal segment of GEWL is very extended and stretches across the molecular surface rather than forming a distinct structural entity. We note that each of the exon boundaries in HEWL occurs within an α-helix, not an obvious way to divide the structure into distinct structural units (see also Craik et al. 1982). Thus, the relation among the three lysozyme structures neither obviously supports nor disproves the notion that the HEWL exons correspond to distinct structural and/or functional units that are conserved during evolution.

Lysozyme Evolution

Many examples exist of protein domains that have similarities in their three-dimensional structures. If there is no corresponding agreement among the amino acid sequences, the origin of the structural correspondence is uncertain—it could have arisen from either divergent or convergent evolution (e.g., see Rossmann et al. 1974; Matthews 1977; Matthews et al. 1981a, b; Matthews and Rossmann 1984). It is reasonable to argue that close overall structural correspondence between two proteins is, in general, a good indication that they have evolved from the same precursor. For example, the close agreement between the two domains of rhodanese (Ploegman et al. 1978) (Table 4) indicates that they probably arose by gene duplication, even though they have essentially no sequence matching. [The use of structural correspondence as evidence for divergent vs convergent evolution is less compelling in cases where the secondary structure is very repetitive, as in the triose phosphate isomerase "α/β barrel" (Banner et al. 1975; Levine et al. 1978).] The structural equivalences among the three lysozyme structures are compared with a number of other reported structural comparisons in Table 4. The GEWL vs HEWL comparison in particular ranks very high, and is suggestive of divergent evolution. Perhaps more compelling is the nature of the structural correspondence among all three lysozymes (Grütter et al. 1983). GEWL and HEWL have in common parts that have no counterparts in T4L. Conversely, GEWL and T4L have common structural elements that do not occur in HEWL. This pattern of structural similarity could easily have arisen through divergent evolution from a common precursor, but would not be expected to have resulted from independent events during evolution.

The differences that do exist between the lysozymes are of the sort that might be expected for distantly related proteins. With more closely related families of proteins, such as the globins or the cytochromes, there is considerable variability in ami-

no acid sequence, but the three-dimensional structures are remarkably conserved. This does not mean that three-dimensional structure changes hardly at all during evolution. In the case of the three lysozymes, part of the core is conserved but there are substantial changes in other local substructures. These changes might reflect different packing requirements dictated by alterations in the amino acid sequences, but could also be associated with altered specificities toward the substrate or toward other macromolecules. In particular, studies of mutant lysozymes suggest that the C-terminal domain of T4L, which is absent in HEWL, enhances the specificity of the enzyme toward the cell walls of *Escherichia coli* (Grütter and Matthews 1982).

Acknowledgments. We are grateful to Drs. E. Zuckerkandl and A.C. Wilson for helpful comments on the manuscript. This work was supported in part by grants from the NIH (GM 21967, GM 20066), the NSF (PCM 8312151), the M.J. Murdock Charitable Trust (to B.W.M.), and the National Health and Medical Research Council of Australia (to N.W.I.).

References

Anderson WF, Grütter MG, Remington SJ, Matthews BW (1981) Crystallographic determination of the mode of binding of oligosaccharides to T4 bacteriophage lysozyme: implications for the mechanism of catalysis. J Mol Biol 147:523–543

Argos P, Rossmann MG (1979) Structural comparisons of heme-binding proteins. Biochemistry 18:4951–4960

Argos P, Tsukihara T, Rossmann MG (1980) A structural comparison of concanavalin A and tomato bushy stunt virus. J Mol Evol 15:169–179

Arnheim N, Inouye M, Law L, Laudin A (1973) Chemical studies on the enzymatic specificity of goose egg white lysozyme. J Biol Chem 248:233–236

Artymiuk PJ, Blake CCF, Sippel AE (1981) Genes pieced together—exons delineate homologous structures of diverged lysozymes. Nature 290:287–288

Banner DW, Bloomer AC, Petsko GA, Phillips DC, Pogson CI, Wilson IA, Corran PH, Furth AI, Milman JD, Offord RE, Priddle J, Waley SC (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution. Nature 255:609–614

Bernard O, Hozumi N, Tonegawa S (1978) Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell 15:1133–1144

Blake CCF (1978) Do genes-in-pieces imply proteins-in-pieces? Nature 273:267

Blake CCF, Koenig DF, Mair GA, North ACT, Phillips DC, Sarma VR (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. Nature 206:757–761

Canfield RE, McMurry S (1967) Purification and characterization of a lysozyme from goose egg-white. Biochem Biophys Res Commun 26:38–42

Canfield RD, Kammesman S, Sobel JH, Morgan FJ (1971) Primary structure of lysozymes from man and goose. Nature (New Biol) 232:16–17

Chothia C, Levitt M, Richardson D (1977) Structure of proteins: packing of α-helices and pleated sheets. Proc Natl Acad Sci USA 74:4130–4134

Craik CS, Buchman SR, Beychok S (1980) Intron–exon splice junctions map at protein surfaces. Proc Natl Acad Sci USA 77:1384–1388

Craik CS, Buchman SR, Beychok S (1981) O_2 binding properties of the product of the central exon of β-globin gene. Nature 291:87–90

Craik CS, Sprang S, Fletterick R, Rutter WJ (1982) Intron-exon splice junctions map at protein surfaces. Nature 299: 180–182

Fitch WM (1966) An improved method of testing for evolutionary homology. J Mol Biol 16:9–16

Ford LO, Johnson LN, Machin PA, Phillips DC, Tjian R (1974) Crystal structure of a lysozyme–tetrasaccharide lactone complex. J Mol Biol 88:349–371

Gilbert W (1978) Why genes in pieces? Nature 271:501

Gō M (1983) Molecular structural units, exons, and function of chicken lysozyme. Proc Natl Acad Sci USA 80:1964–1968

Grütter MG, Matthews BW (1982) Amino acid substitutions far from the active site of bacteriophage T4 lysozyme reduce catalytic activity and suggest that the C-terminal lobe of the enzyme participates in substrate binding. J Mol Biol 154:525–535

Grütter MG, Weaver LH, Matthews BW (1983) Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? Nature 303:828–831

Harada H, Sarma R, Kakudo M, Hara S, Ikenaka T (1981) The three-dimensional structure of the lysozyme produced by *Streptomyces erythraeus*. J Biol Chem 256:11600–11602

Imoto I, Johnson LN, North ACT, Phillips DC, Rupley J (1972) Vertebrate lysozymes. In: Boyer P (ed) The enzymes, vol 7, 3rd ed. Academic Press, New York, pp 665–868

Inouye M, Tsugita A (1966) The amino acid sequence of T4 bacteriophage lysozyme. J Mol Biol 22:193–196

Isaacs NW, Machin KJ, Masakuni M (1985) The three-dimensional structure of the goose-type lysozyme from the egg-white of the black swan, *Cygnus atratus*. Australian J Biol Sci, in press

Jollès J, Schoentgen F, Jollès P (1981) Les lysozymes de type different ont-ils un precurseur commun? C R Hebd Seances Acad Sci (Paris) 292:891–892

Jollès P, Schoentgen F, Jollès J, Dobson DE, Prager EM, Wilson AC (1984) Stomach lysozymes of ruminants II. Amino acid sequence of cow lysozyme 2 and immunological comparisons with other lysozymes. J Biol Chem 259:11617–11625

Jung A, Sippel AE, Grez M, Schultz G (1980) Exons encode functional and structural units of chicken lysozyme. Proc Natl Acad Sci USA 77:5759–5763

Kleppe G, Vasstrand E, Jensen HB (1981) The specificity requirements of bacteriophage T4 lysozyme. Involvement of *N*-acetamido groups. Eur J Biochem 119:589–593

Lebioda L, Hatada M, Tulinsky A, Mavridis IM (1982) Comparison of the folding of 2-keto-3-deoxy-6-phosphogluconate aldolase, triosephosphate isomerase and pyruvate kinase. Implications in molecular evolution. J Mol Biol 162:445–458

Lesk AM, Chothia C (1980) How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. J Mol Biol 136: 225–270

Levine M, Muirhead H, Stammers DK, Stuart DI (1978) Structure of pyruvate kinase and similarities with other enzymes: possible implications for protein taxonomy and evolution. Nature 271:626–630

Liljas L, Unge T, Fridborg K, Jones TA, Lovgren S, Skoglund U, Strandberg B (1982) Structure of satellite tobacco necrosis virus at 3.0Å resolution. J Mol Biol 159:93–108

Matthews BW (1977) X-ray structure of proteins. In: Neurath H, Hill RL (eds) The proteins, vol 3, 3rd ed. Academic Press, New York, pp 403–590

Matthews BW, Remington SJ (1974) The three-dimensional structure of the lysozyme from bacteriophage T4. Proc Natl Acad Sci USA 71:4178–4182

Matthews BW, Rossmann MG (1985) Comparison of protein structures. Methods Enzymol, in press

Matthews BW, Grütter MG, Anderson WF, Remington SJ (1981a) Common precursor of lysozymes from hen egg-white and bacteriophage T4. Nature 290:334–335

Matthews BW, Remington SJ, Grütter MG, Anderson WF (1981b) Relation btween hen egg-white lysozyme and bacteriophage T4 lysozyme: evolutional implications. J Mol Biol 147:545–558

McLachlan AD (1979) Gene duplications in the structural evolution of chymotrypsin. J Mol Biol 128:49–79

Ohlendorf DH, Anderson WF, Lewis M, Pabo CO, Matthews BW (1983) Comparison of the structures of cro and λ repressor proteins from bacteriophage λ. J Mol Biol 164:757–769

Ploegman JH, Drent G, Kalk KH, Hol WGJ (1978) Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. J Mol Biol 123:557–594

Quinto C, Quiroga M, Swain WF, Nikovits WC Jr, Standring DN, Pictet R, Valenzuela P, Rutter WJ (1982) Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of the gene. Proc Natl Acad Sci USA 79:31–35

Rao ST, Rossmann MG (1973) Comparison of super-secondary structures in proteins. J Mol Biol 76:241–256

Remington SJ, Matthews BW (1978) A general method to assess the similarity of protein structures with applications to T4 bacteriophage. Proc Natl Acad Sci USA 75:2180–2184

Remington SJ, Matthews BW (1980) A systematic approach to the comparison of protein structures. J Mol Biol 140:77–99

Remington SJ, Ten Eyck LF, Matthews BW (1977) Atomic coordinates for T4 phage lysozyme. Biochem Biophys Res Commun 75:265–269

Remington SJ, Anderson WF, Owen J, Ten Eyck LF, Grainger CT, Matthews BW (1978) The structure of the lysozyme from bacteriophage T4: an electron density map at 2.4 Å resolution. J Mol Biol 118:81–91

Richardson JS, Richardson DC, Thomas KA, Silverton EW, Davies DR (1976) Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit. J Mol Biol 102:221–235

Rossmann MG (1979) Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. J Appl Crystallogr 12:225–238

Rossmann MG, Argos P (1975) A comparison of the heme binding pocket in globins and cytochrome b_5. J Biol Chem 250:7525–7532

Rossmann MG, Argos P (1976) Exploring structural homology of proteins. J Mol Biol 105:75–96

Rossmann MG, Argos P (1977) The taxonomy of protein structure. J Mol Biol 109:99–129

Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of a nucleotide-binding protein. Nature 250:194–199

Rossmann MG, Abad-Zapatero C, Murthy MRN, Liljas L, Jones TA, Strandberg B (1983) Structural comparisons of some small spherical plant viruses. J Mol Biol 165:711–736

Sarma R, Harada S, Tanaka N, Kakudo M, Hara S, Ikenaka T (1979) Structure of *Streptomyces erythraeus* lysozyme at 6 Å resolution. J Biochem (Tokyo) 86:1765–1771

Schmid MF, Weaver LH, Holmes MA, Grütter MG, Ohlendorf DH, Reynolds RA, Remington SJ, Matthews BW (1981) An oscillation data collection system for high-resolution protein crystallography. Acta Crystallogr A37:701–710

Schoentgen F, Jollés J, Jollés P (1982) Complete amino acid sequence of ostrich (*Struthio camelus*) egg-white lysozyme, a goose-type lysozyme. Eur J Biochem 123:489–497

Sharon N, Eshdat Y, Maoz I, Bernstein Y, Prager EM, Wilson AC (1974) Comparative studies of the active site region of lysozymes from eleven different sources. Isr J Chem 12:591–603

Simpson RJ, Morgan FJ (1983) Complete amino acid sequence of Embden Goose (*Anser anser*) egg-white lysozyme. Biochim Biophys Acta 744:349–351

Simpson RJ, Begg GS, Dorow DS, Morgan FJ (1980) Complete amino acid sequence of the goose-type lysozyme from the egg white of the black swan. Biochemistry 19:1814–1819

Stein IP, Catterall JF, Kristo P, Means AR, O'Malley BW (1980) Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell 21:681–687

Steitz TA, Ohlendorf DH, McKay DB, Anderson WF, Matthews BW (1982) Structural similarity in the DNA-binding domains of catabolite gene activator and *cro* repressor proteins. Proc Natl Acad Sci USA 79:3097–3100

Tang JJN, James MNG, Hsu IN, Jenkins JA, Blundell TA (1978) Structural evidence for gene duplication in the evolution of the acid proteases. Nature 271:618–621

Wierenga RK, Drenth J, Schultz GE (1983) Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of *p*-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. J Mol Biol 167:725–739

Yaguchi M, Roy C, Rollin CF, Paice MG, Jurasek L (1983) A fungal cellulase shows sequence homology with the active site of hen egg-white lysozyme. Biochem Biophys Res Commun 116:408–411

Received June 11, 1984/Revised October 9, 1984