

Crystal Structure of the Lysozyme from Bacteriophage Lambda and its Relationship with V and C-type Lysozymes

Christine Evrard^{1,2}, Jacques Fastrez² and Jean-Paul Declercq¹

¹Laboratoire de Chimie Physique et de Cristallographie and ²Laboratoire de Biochimie Physique et des Biopolymères Université Catholique de Louvain, Place L. Pasteur, 1 B1348, Louvain-la-Neuve Belgium

Abstract

Like other lysozymes, the bacteriophage lambda lysozyme is involved in the digestion of bacterial walls. This enzyme is remarkable in that its mechanism of action is different from the classical lysozyme's mechanism. From the point of view of protein evolution, it shows features of lysozymes from different classes.

The crystal structure of the enzyme in which all tryptophan residues have been replaced by aza-tryptophan has been solved by X-ray crystallography at 2.3 Å using a combination of multiple isomorphous replacement, non-crystallographic symmetry averaging and density modification techniques. There are three molecules in the asymmetric unit. The characteristic structural elements of lysozymes are conserved: each molecule is organized in two domains connected by a helix and the essential catalytic residue (Glu19) is located in the depth of a cleft between the two domains. This cleft shows an open conformation in two of the independent molecules, while access to the cavity is much more restricted in the last one. A structural alignment with T4 lysozyme and hen egg white lysozyme allows us to superpose about 60 C^α atoms with a rms distance close to 2 Å. The best alignments concern the helix preceding the catalytic residue, some parts of the β sheets and the helix joining the two domains. The results of sequence alignments with the V and C lysozymes, in which weak local similarities had been detected, are compared with the structural results.

Keywords : aza-tryptophan ; evolution ; lysozyme ; structure ; transglycosylase

Abbreviations used : λL, lambda lysozyme ; NCS, non-crystallographic symmetry ; HEWL, hen egg white lysozyme ; T4L, T4 lysozyme.

INTRODUCTION

The bacteriophage lambda lysozyme (λL) is a small protein of 158 amino acid residues (Imada & Tsugita, 1971). Like other lysozymes, it cleaves glycoside bonds between the C1 of N-acetyl muramic acids (NAM) and C4 of N-acetyl glucosamines (NAG) of the peptidoglycan of the bacterial walls. However, its mechanism has been shown to be different from that of the other lysozymes; it is not a hydrolase but a transglycosylase: the hydroxyl function OH on the C6 of the muramic acid is the nucleophile on cleavage of the muropeptide instead of the more common water molecule (Taylor *et al.*, 1975; Bienkowska-Szewczyk *et al.*, 1981).

Lysozymes are widespread in nature. They are produced by plants and higher organisms as a first mechanism of defence against bacterial invaders. They are used by bacterial viruses to facilitate infection or to release virions at the end of an infection cycle (for recent reviews on lysozymes see Jollès, 1996). Efforts have been devoted to the understanding of the relationships between lysozymes of different origins by comparison of their sequences and tertiary structures. Several classes of lysozymes have been defined on the basis of sequence alignments; the best known are the C-type (chicken-type), G-type (goose-type) and V-type (viral type) lysozymes. No statistically significant sequence similarities are detected between members of these classes but interesting structural similarities have been discovered (Rossmann & Argos, 1976; Remington & Matthews, 1978; Matthews *et al.*, 1981; Weaver *et al.*, 1985a). These comparisons have led to the proposal that the genes encoding these enzymes have evolved from a remote common ancestor and these lysozymes have become a classical example of divergent evolution. More recently, other enzymes have been shown to share the same fold. The soluble lytic

transglycosylase (Slt70) of *Escherichia coli*, a 70 kDa enzyme, contains a catalytic domain showing significant structural similarity with the T4 lysozyme (Thunnissen *et al.*, 1994). The recently determined structures of barley chitinase and of *Streptomyces* N174 chitosanase appear to be quite similar to those of goose lysozyme and T4 lysozyme, respectively. On the basis of these comparisons of structures, the lysozyme superfamily was tentatively divided into two families: a bacterial family and an eukaryotic family (Monzingo *et al.*, 1996). Lysozyme-like structures have also been proposed for virulence factors in bacterial pathogens of plants and animals (Mushegian *et al.*, 1996).

The *E. coli* phage λ is closely related to the *Salmonella typhimurium* phage P22 (Susskind & Botstein, 1978; Wiggins & Hilliker, 1985). The P22 lysozyme sequence is easily aligned with that of the T4 lysozyme (T4L) and its structure has been modeled on the T4L pattern (Weaver *et al.*, 1985b). This led us to analyse the evolutionary relationship between λ L and V-type lysozymes using multiple sequence alignments, secondary structure predictions and pattern recognition methods. The alignments with the V-type sequences have allowed us to detect an essential catalytic residue, Glu19, whose mutation into Gln completely abolished the activity. Alignments with C-type sequences were also attempted; they revealed the presence in λ L of a sequence fragment highly similar to that constituting the β -sheet in these lysozymes and a possible repetition of the β -sheet motif. They also suggested a role for a second residue, Asp34; its mutation into Asn decreased the activity significantly. However, global alignments could not be constructed and a full structural model could not be generated (Jespersen *et al.*, 1992; Fastrez, 1996).

Here, we report the three-dimensional structure of λ L, we compare this structure with those of T4L and hen egg white lysozyme (HEWL) and we interpret the partial sequence alignments developed earlier to analyse the evolutionary relationships between these lysozymes.

RESULTS AND DISCUSSION

Structure determination and quality of the model

As described in the preliminary paper (Evrard *et al.*, 1997), we were unable to grow crystals suitable for X-ray analysis from the unmodified bacteriophage lambda lysozyme. Expression, purification and crystallization of the mutant in which all tryptophan residues have been replaced by aza-tryptophan have been described elsewhere (Soumilion & Fastrez, 1992; Soumilion *et al.*, 1995; Evrard *et al.*, 1997).

The structure was solved by the multiple isomorphous replacement method including averaging over the non-crystallographic symmetry (NCS) and density modification techniques (see Table 1 and Materials and Methods). A total of four heavy atom derivatives were used in phase determination with data extending to 3.0 Å resolution. Three molecules are present in the asymmetric unit. NCS restraints were applied up to the end of the refinement except in some regions of the chain where very significant differences appeared in the electron density. The structure has been refined to 2.3 Å spacings with an *R*-factor of 0.1635. No interpretable electron density is present for the last four residues (155 to 158) of each molecule. The current model contains 462 (3 × 154) residues, 121 water molecules and three isopropanol molecules. The quality of the final model structure was estimated with the program PROCHECK (Laskowski *et al.*, 1993). The root mean square (rms) deviations from ideality of this model are reported in Table 2. The upper limit of the mean coordinate error of all atoms assessed from Luzzati plots (Luzzati, 1952) is 0.22 Å (see Table 2). Analysis of the main-chain torsion angles shows that 91.2% of the non-glycine residues are located in the most favoured regions of the Ramachandran plot, and the remaining 8.8% in the additional allowed regions (Figure 1). Coordinates and structure factors have been deposited with the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) under accession numbers 1AM7 and R1AM7SF, respectively.

Overall structure

λ L is a globular protein of approximate dimensions 40 Å × 32 Å × 32 Å. The folded enzyme clearly shows features from C and V-type lysozymes (see following paragraphs). The crystal structure of λ L contains three independent molecules, referred to here as A, B and C. Comparison of these molecules reveals the high similarity of molecules A and C whereas a significant conformational change in regions from residues 51 to 60 and 128 to 141 is observed for molecule B. The overall structures of molecules A and B are illustrated as ribbon diagrams in Figure 2. Least-square superposition of the three molecules was carried out with the program X-plor, using as guides the C $^{\alpha}$ atoms of the residues subjected to NCS restraints. The overall rms deviations between the C $^{\alpha}$ of the complete structural fold are 2.75, 0.28 and 2.72 Å for the superposition of molecules A with B, A with C and B with C, respectively. Comparison between molecules B and C will not be detailed further because of the

very similar fold of molecules A and C. Rms deviations for backbone atoms and for side-chain atoms are illustrated in Figure 3. Comparison of molecules A and B indicates the relatively high mobility of residue regions 51 to 60 and 128 to 141. The large rms deviation value associated to these parts of the enzyme is ascribed to the conformational difference in these residue regions. When these residues are excluded, the rms deviation value between the C α of molecules A and B decreases to 0.25 Å.

Figure 1. Ramachandran plot of the refined structure (program PROCHECK, Laskowski et al, 1993). The three independent molecules are superposed on the same diagram. Glycine and non-glycine residues are designated by triangles and squares, respectively. The most favourable regions are the darkest; the favourable, generously allowed and forbidden regions are progressively lighter. 91.2% of the non-glycine residues are in the most favoured regions.

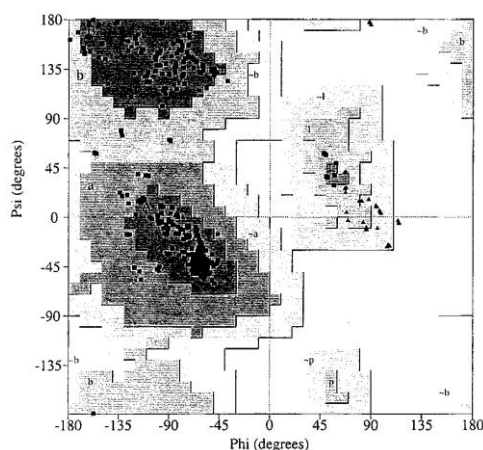


Figure 2. Stereoscopic view of two of the three independent molecules (A and B) illustrated as ribbon diagrams. The orientations of the molecules are exactly the same as in Figure 6 on which the positions of some residues are labelled. The side-chain of the catalytic residue Glu9 is represented by sticks, at the C-terminal end of helix 1. Domain I (see the text) is at the bottom of the views and domain II at the top. The two domains are separated by a vertical helix (H3). This Figure was produced using the program O (Jones et al., 1991).



The 154 observed amino acid residues of λ L are organized in two domains connected by a helix. We define here domain I at the "bottom" of the longest molecular axis, and domain II at its "top". There are six helices, six β strands forming two β sheets in molecules A and C, and only four β strands forming one β sheet in molecule B. A topological diagram showing the arrangement of the secondary structural elements in comparison with those of T4L and HEWL is presented in Figure 4. Helix 1 is a regular α helix. Its N-terminal part starts in domain II and the helix extends to the interface between the two domains. The C-terminal residue of this helix (Glu19) has been identified by Jespers *et al.* (1992) as the essential catalytic residue of λ L. It is located in the depth of a cleft between the two domains, with its side-chain protruding inside the cavity. Domain I starts after this first helix. It is constituted by a β sheet structure and a second helix (H2). A long and complicated loop (14 residues) makes the junction between Glu19 and the first β strand (β 1) beginning with residue Asp34. This residue had been proposed by Jespers *et al.* (1992) as a candidate for the second catalytic residue. It is located in the external part of the first domain, while Glu19 is inside the cleft and the interaction of a substrate molecule with the two acidic side-chains seems almost impossible. This β 1 strand is the second strand of a twisted four-strand antiparallel β sheet composed successively of strands β 2, β 1, β 5 and β 6. These four strands are very short (three to five residues) and interconnected by short turns, except between β 2 and β 5. This is where the first large difference appears between molecules A and C on one side and molecule B on the other side. In molecules A and C, a long loop is followed by a short β ladder (β 3, β 4).

Table 1. Crystallization, data collection and phase determination statistics

	Native 1	Pt1	Pt2	Hg1	Native2	Hg2
<i>Crystallization</i>						
Buffer	Citrate	Citrate	Citrate	Citrate	Phosphate	Phosphate
Soaking concentration (mM)		5.0	1.0	1.5		1.0
Soaking time (days)		2	2	2	2	2
<i>Data collection</i>						
Beamline	X31	BW7B	BW7B	X11	BW7A	BW7A
Wavelength (Å)	1.000	0.859	0.859	0.912	0.900	0.900
Resolution (Å)	2.3	2.8	2.5	2.5	2.4	2.5
Measured reflections	104,257	60,359	77,184	129,443	99,406	61,690
Unique reflections	21,455	12,235	17,046	17,210	18,888	13,500
R_{sym}^a	0.090	0.074	0.102	0.136	0.082	0.084
$\langle I/\sigma(I) \rangle$ in the highest resolution shell	3.3	2.1	2.6	2.8	5.8	2.3
Completeness (%)	98.8	98.9	98.9	99.5	98.2	78.7
R_{iso}^b		0.223	0.205	0.215		0.135
<i>MIR phasing (20 to 3.0 Å)</i>						
No. of sites		7	3	1		3
Total occupancy		2.1	0.6	0.3		0.5
Phasing power ^c						
centric		0.81	0.46	0.34		0.76
acentric		1.14	0.54	0.41		0.95
R_{cullis}^d centric		0.77	0.92	0.95		0.80
acentric		0.82	0.95	0.97		0.87
anomalous		0.95	0.99	0.99		0.99

Pt1, sodium tetrachloroplatinate; Pt2, platinum terpyridine chloride; Hg1, mersalyl acid; Hg2, methyl mercury chloride.

^a $R_{\text{sym}} = \sum |I(h)_i - \langle I(h) \rangle| / \sum I(h)_i$.

^b $R_{\text{iso}} = \sum |F_{\text{PH}} - F_{\text{P}}| / \sum F_{\text{P}}$.

^c $\langle F_{\text{H}} / |F_{\text{PH}} - |F_{\text{P}} + F_{\text{H}}|| \rangle$.

^d $R_{\text{cullis}} = \langle |F_{\text{PH}} - |F_{\text{P}} + F_{\text{H}}|| \rangle / \langle |F_{\text{PH}} - F_{\text{P}}| \rangle$.

Table 2. Results of the refinement

<i>Bulk solvent correction</i>		
Solvent density	0.34 e Å ⁻³	
Solvent probe radius	0.25 Å	
Temperature factor (B)	50 Å ²	
	Reflections with	
Resol.-range (Å)	$I \geq \sigma(I)$	R-value

2.3-infinity	20,814	0.1635
Upper limit of the estimated error on atomic positions (Luzzati, 1952) (Å)		0.22
RMS deviations from ideality (Engh & Huber, 1991)		
Bonds (Å)	0.009	
Angles (°)	1.69	
Improper dihedrals (°)	1.27	
Target standard deviations between <i>B</i> factors (Å ²)		
of NCS equivalent atoms		2.0
of bonded atoms:	Back-bone	1.5
	Side-chain	2.0
of atoms connected by an angle:	Back-bone	2.0
	Side-chain	2.5

It extends to the outside of domain I and the loop between the two strands (residues 56 to 58) forms the lower lip of the active site cavity, in an open conformation. In molecule B, the β ladder is disrupted by a water molecule hydrogen bonded to Val53 N, Ser61 N and Ser61 O. This water molecule throws apart the two strands, replacing the β ladder by a long loop. The outer part of the loop is folded in such a way that it partially closes the access to the cavity. After the β sheet structure, the domain I is terminated by a slightly distorted α helix (H2). Helix 3 is parallel to the long molecular axis and makes the connection between the two domains. It is immediately followed by helices 4 and 5, which are approximately antiparallel to each other and perpendicular to helix 3. These three helices are regular α helices. The long loop joining helices 5 and 6 includes the second region where a large difference appears between the molecules. This loop constitutes the upper lip of the large mouth characterizing the active site of λ L. In molecules A and C, this loop adopts an open conformation allowing free access to the active site. In molecule B, the N-terminal end of H6 comes unwound on about one helix turn and the loop extends across the active site cleft limiting the access to the cavity. The two lips are separated by about 23 Å in the open conformation and by 10 Å in molecule B. The two different conformations, open for molecules A and C and closed for molecule B, are well illustrated by a comparison of the molecular surfaces (Figure 5) produced by the program GRASP (Nicholls *et al.*, 1993).

For each of the independent molecules, electron density is found in the active site cleft. This density is consistent with the presence of an isopropanol molecule whose oxygen atom interacts with the backbone nitrogen of Gly38. The isopropanol molecule might play a role in the stabilization of the overall structure of λ L.

A number of intramolecular interactions between sequentially distant residues are observed. Among them, it is worth noting a short hydrogen bond between the side-chain ($O^{\delta 2}$) of the catalytic residue (Glu19), at the C-terminal end of helix H1 and the side-chain of Ser126 (O^{γ}) in the loop between helices 5 and 6 (2.53 Å, 2.49 Å, 2.54 Å). Salt bridges also exist between Arg8 (H1) and Asp109 (H4), between Arg50 (loop in the β structure) and Asp75(H2), between Arg66 (turn connecting $\beta 5$ and $\beta 6$) and Asp13(H1), between Arg66 and Asp22 (loop between H1 and $\beta 1$), between Arg114 (H5) and Asp140 (H6).

Figure 3. Root mean square deviations as a function of residue number, between molecules A and B (top), and between molecules A and C (bottom). The thick lines are for the backbone atoms, the thin lines for the remaining atoms.

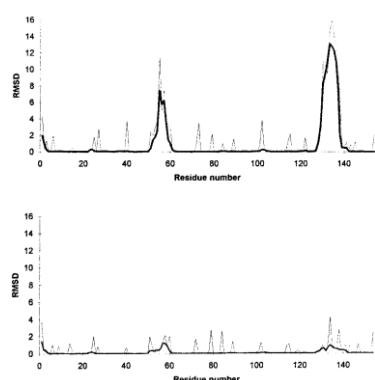


Figure 4. Topological diagram comparing the arrangement of the secondary structural elements in HEWL, λ L and T4L. The helices are represented as grey rectangles and the β strands as white arrows. The beginning and the end of the secondary structural elements are labelled. In molecule B of λ L, helix H6 begins at residue 139, and the β ladder (β 3, β 4) is replaced by a long loop. The asterisk (*) indicates the position of the catalytic residue.

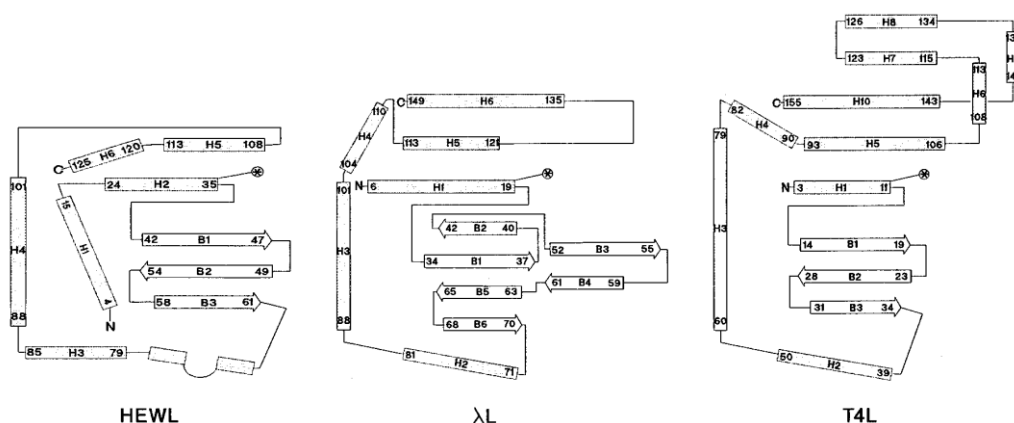
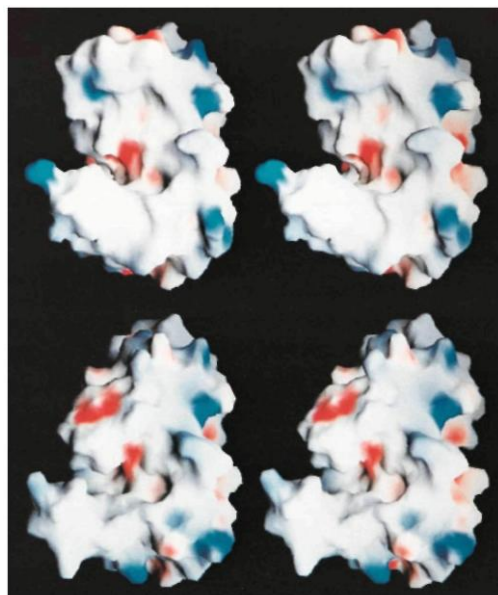


Figure 5. Stereoscopic view showing a comparison of the molecular surfaces of molecule A (top) and molecule B (bottom) in the same orientation. The surface is coloured according to the local electrostatic potential, ranging from blue (the most positive region) to red (the most negative). Domains I and II (see the text) are at the bottom and the top of the views, respectively. The position of the active site corresponds to the red spot in the depth of the cleft between the two domains. This cleft shows an open and a closed conformation in molecule A and B, respectively (Figure produced using the program GRASP, Nicholls et al, 1993).



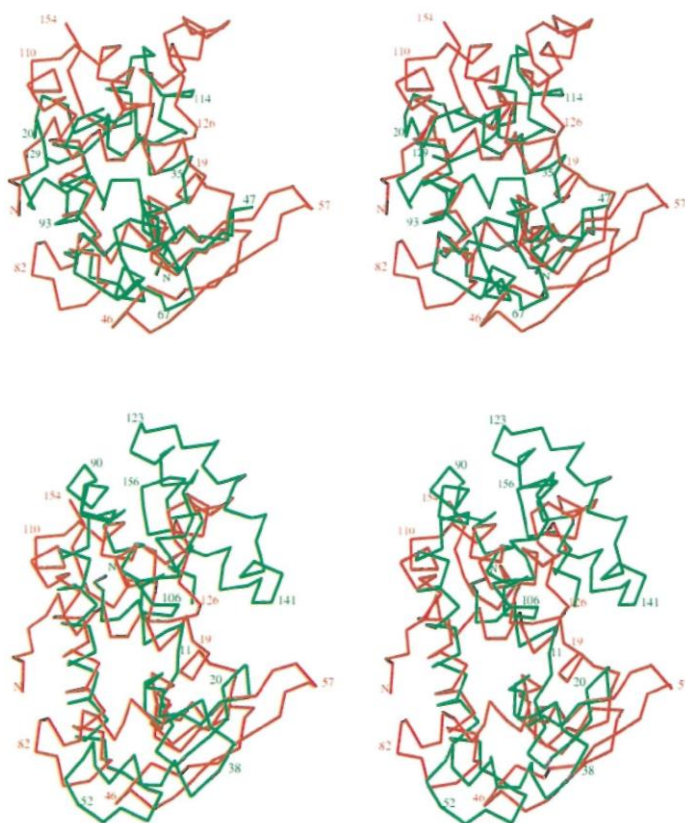
Packing of the molecules

The three independent molecules of the asymmetric unit are nearly perfectly stacked on top of each other in the direction of the *b* crystallographic axis. A succession of three independent molecules in this direction corresponds exactly to the length of one unit cell. A kind of infinite column of molecules (...-A-B-C-A-B-C-...) appears thus in that direction.

In the two other directions, the (*a,c*) face could be separated in four equivalent rectangles (nearly squares) by tracing two perpendicular lines joining the centres of the opposite edges of the face, and one column of

molecules could be projected on each of these rectangles. The intermolecular contacts between the columns are much looser than inside one column. The interactions between the three independent molecules in one column are very different from each other. As described above, there are two mobile loops (residues 55 to 60 and 129 to 134) guarding the gate of the deep cleft in which the catalytic residue (Glu19) is located. This gate in molecule B is much more closed than in molecules A and C. In the interaction between these latter two molecules, the two clefts interpenetrate each other in a very symmetrical way. The interactions of molecule B with the two other molecules are more superficial.

Figure 6. Stereoscopic view of the superposition of the C $^{\alpha}$ -trace of molecule A with HEWL (top) and T4L (bottom). λ L is red, HEWL and T4L are green (Figure produced using the program O, Jones *et al.*, 1991).



Structural comparison of lambda lysozyme (λ L) with T4 lysozyme (T4L) and hen egg white lysozyme (HEWL)

In their comparison of G-type, C-type and V-type lysozymes, Weaver *et al.* (1985a) emphasized the conservation of essential structural elements. The active site was shown to be always located in a crevice between two domains connected by a long helix (the third helix of T4L, H3, or the fourth helix of HEWL, H4). A catalytically essential glutamic acid residue was invariably located in the groove between the two domains on the carboxyl side of a helix (H1 in T4L and H2 in HEWL), which is immediately followed by an irregular β sheet located in domain I. Domain II was made up of less well conserved helices.

In view of results reported previously in which the possible relationships between λ L and the V and C-type lysozymes were analysed, we examined our models for similarities with the structure of T4L (Weaver & Matthews, 1987) and HEWL (Diamond *et al.*, 1974) using respectively the data sets 3LZM and 6LYZ available in the Protein Data Bank. The best conserved motif in the three structures, consisting of the two helices labelled H1 and H3 in λ L, clearly appeared to correspond respectively to H1 and H3 in T4L or to H2 and H4 in HEWL. Although the two helices do not show the same length, they are characterized by similar relative orientations and displacements. A β -sheet structure invariably appears between these two helices. A more detailed structural

comparison was thus initiated. We first fitted manually these secondary elements, between λ L and T4L on one side and between λ L and HEWL on the other side using the program O (Jones *et al.*, 1991). No attempts were made to produce some particular alignment between T4L and HEWL, since such comparisons had already been developed (Rossmann & Argos, 1976; Remington & Matthews, 1978; Matthews *et al.*, 1981; Weaver *et al.*, 1985a). This initial step was followed by an automatic improvement of the superposition (lsq_imp in O): in an iterative process, all pairs of C $^{\alpha}$ atoms belonging to a fragment of at least three residues, with a distance ≤ 3.8 Å were included in the least squares fit of the two molecules. Using molecule A of λ L, it was possible to align 54 residues of T4L with a rms distance of 2.047 Å between the C $^{\alpha}$ atoms, and 62 residues of HEWL, with a rms distance of 1.848 Å. A better agreement appears between molecule B of λ L and T4L: 58 residues with a rms distance of 2.103 Å, while the results remain approximately unmodified for HEWL: 61 residues, rms = 1.803 Å. Figure 6 shows stereoscopic views of the superpositions of the C $^{\alpha}$ -traces of molecule A and HEWL and of molecule A and T4L. The details of the structural alignments are given in Scheme I. It was later checked that a completely automatic structural alignment using the "brute_force" available in the program LSQMAN (Kleywegt & Jones, 1997) provides very comparable alignments. The superposition of H1 of λ L with H1 of T4L and H2 of HEWL make coincide the essential catalytic residues of T4L and HEWL (Glu11 and Glu35, respectively) with residue Glu19 of λ L, at the end of this helix, confirming the results obtained by Jespers *et al.* (1992). This helix is followed by a long loop, itself followed by six short β strands: strands β 2, β 1, β 5, β 6 form a distorted antiparallel β sheet, while in molecules A and C only, β 3 and β 4 form a β ladder, replaced by a long loop in molecule B. A part of the first β sheet (β 1, β 5, β 6) coincides exactly with the central part of the three stranded β sheet 2 described in HEWL. The two β strands (β 4, β 5) of λ L are nearly collinear and coincide with one of the β strands of T4L and β 6 of λ L also partially corresponds to the next and last strand of the β sheet of T4L. In λ L, the β region is followed by the H2 helix, which is also observed in T4L and is labelled H3 in HEWL. However, none of these helices is superimposed. More surprisingly, there seems to be a good superposition between H2 of λ L and two short fragments (60 to 64 and 74 to 77) of the sequence of HEWL, separated by a loop composed of residues 65 to 73. These two short fragments form two turns more or less comparable to two successive turns of a helix: hydrogen bonds are observed between O61...N73, O62...N75 and O63...N76, and the ϕ and ψ conformational angles of residues 60 to 63 [(-83°, -7°) (-79°, -28°) (-126°, -42°) (-84°, -30°)] and of residues 75 to 76 [(-62°, -23°) (-85°, -15°)] are acceptable for a helical conformation. Of course, such short fragments were not described as helices in the original structure of HEWL, but the present superposition suggests the insertion of a loop in the middle of a short helix. Interestingly, the first of these short turns in HEWL (residues 60 to 63) is characterized by the sequence SRWW perfectly identical to that of the superposed residues 71 to 74 in λ L. In HEWL, Blake *et al.* (1967) have shown from the analysis of complexes between the enzyme and substrate analogues or inhibitors that Trp62 and Trp63 play an important role in substrate binding. The replacement of Trp62 by an oxindolealanine and the simultaneous loss of enzymatic activity confirmed this result (Blake *et al.*, 1981). The spatial coincidence of the C $^{\alpha}$ atoms of these important residues in HEWL and in λ L allows us to suppose that they play the same role in the two enzymes. In HEWL, the four side-chains are accessible in the cleft containing the catalytic residues. In λ L, the side-chains of Ser71, Arg72 and Trp73 are closely superposed on the corresponding side-chains (60 to 62) of HEWL while the side-chain of Trp64 adopts a completely different orientation and becomes completely inaccessible. The helices H3 of λ L and of T4L as well as H4 of HEWL show a very good superposition from their N-terminal side. Their lengths are very comparable in λ L and HEWL while H3 of T4L is longer by about two complete turns. In both λ L and T4L, H3 is immediately followed by helix H4, nearly perpendicular to H3, but the large difference of length of the two H3 helices has the consequence that the two H4 helices are more or less parallel, but separated by about 14 Å. In spite of this large spatial distance, the two H5 helices of λ L and T4L are again nicely superimposed, while no corresponding helix appears in HEWL. The similarities in the C-terminal part are much fewer and do not usually correspond to comparable secondary structures, except for the end of H6 of λ L which corresponds to a short helical fragment (H6) of HEWL. In molecule B of λ L only, a short fragment (136 to 139) of the very flexible loop coincides spatially with a fragment of helix H10 of T4L, but their secondary structures are not similar.

Scheme I. Amino acid sequences of superposed structural elements of λ L, T4L and HEWL. Residues located in β sheets and in helices are singly and doubly underlined, respectively. Residues 17, 73, 74 and 124 of λ L, noted with the one-letter code W (Trp) correspond in this structure to aza-tryptophan ($C^{\epsilon 2}$ replaced by $N^{\epsilon 2}$). * β 4 is present in molecules A and C only. **The superposition of the fragment 136 to 139 of λ L with 146 to 149 of T4L is only observed for molecule B of λ L.

```

      3      11      22      30 31 34      58      62
T4L  IFEMLRIDE.....EGYTTIGIG HLLT.....ITKDE
      9      H1      19      31  β1 37      59 β4* β5      β6      H2 75      76 79      86      90
λL  KAFLDMLAWSE...HGYDVIV...LKSTGAGRYQLLSRWWD AYRK...FSPKS
      25      35      38      44      51      64 74 77      88
HEWL LGNWVCAAKFE...FNTQATN.....TDYGILOINSRWWC...NLCN....ITA

      63      75      93      105      146 149
T4L  AEKLFNODVDAAV...AVRRCALINMVFO.....AKRV.....
      91      H3      103      112 H5      128      136 ** 139      146 H6 150
λL  QDAVALQOIKERG...DIROAIDRCSNIWASLP...EHKA...FKEAG
      91      102      107      112      120 124
HEWL SVNCAKKIVSDG.....AWVAWR.....VOAWI

```

Scheme II. Local sequence similarities between the V and C-type lysozymes and λ L. When multiple alignments were run with the phage or the C-type lysozymes, only the sequences of T4L or of HEWL are shown; the β strands and the α -helices are singly or doubly underlined and numbered. Vertical lines and colons point to alignment of identical or similar residues, respectively. The number of standard deviations above the similarity scores of random alignments are given in parentheses. A local similarity in 3D-profile alignments between λ L and HEWL is also shown (c); it implies a sequence alignment between T4L and HEWL which agrees with a reported structural alignment (Weaver et al, 1985a or Taylor & Orengo, 1989).

(a)

$$\begin{array}{cccccccccccc} & & 1 & & & H1 & & & 11 & & & \\ T4L : & M & N & I & F & E & M & L & R & I & D & E & G & L & R \\ & : & : & | & | & & & & | & | & & & & & \\ \lambda_L : & K & A & F & L & D & M & L & A & W & S & E & G & T & D \\ & 9 & & & & & H1 & & & & & 19 & & \end{array} \quad (3.3\sigma)$$

(b)

$\begin{array}{ccccccccccccccccccccccccccccc} & & 36 & & 40 & & \beta_1 & & & & 50 & \beta_2 & & & & \beta_3 & & 61 \\ \text{HEWL : } & S & N & F & N & T & Q & A & T & N & R & N & T & - & D & G & S & T & D & Y & G & I & L & Q & I & N & S & R & W & W \\ & | & : & & & & & & & & & & & & & & | & | & | & | & | & | & | & | & | & | & | & | & | & | \\ \lambda_L : & S & D & H & P & R & K & L & V & T & L & N & P & K & L & K & S & T & G & A & G & R & Y & Q & L & S & R & W & W \\ & 46 & & & & 50 & & \beta_3 & & & & & & & \beta_4 & & \beta_5 & & & & \beta_6 & & 72 \end{array}$
 (5.9σ/6.8σ)

(the gap in the HEWL sequence arises because a residue is present here in the sequences of many C-type [vsoszymes]).

(c)

T4L : ITKDEAEKLFNQDVD
 Seq. align. : 88 H3 101 (2.7σ)
 λL : GLKDFSPKSQDAVALQOIKER G A L P M I D R
 3D-profile : 88 H4 101 (3.0σ)
 HEWL : A L L S D I T A S V N C A K K I V S D G N G M N A W V A

(residues contributing positively to the 3D-profile alignment are in bold).

(d)

$\begin{array}{cccccccccccccccc} & & & 89 & & & & H5 & & & & 101 & & & & \\ T4L : & D & S & L & D & \underline{A} & \underline{V} & \underline{R} & \underline{R} & \underline{C} & \underline{A} & \underline{L} & \underline{I} & \underline{N} & \underline{M} & \underline{V} \\ & | & & & & | & : & | & | & & & & | & & & \\ \lambda L : & D & I & R & Q & A & I & D & R & C & S & N & I & W & A & S \\ & 112 & & & & & & H5 & & & & & & 124 & & \end{array} \quad (3.5\sigma)$

(e)

$$\begin{array}{cccccccccccccccc}
 & & 139 & & & & H10 & & & & 151 & & & & & \\
 T4L : & Y & N & Q & T & \underline{P} & \underline{N} & \underline{R} & \underline{A} & \underline{K} & \underline{R} & \underline{V} & \underline{I} & \underline{T} & \underline{T} & \underline{F} & \underline{R} & \underline{T} \\
 & | & & | & & & : & | & & : & | & & | & : & & & & (4.5\sigma) \\
 \lambda L : & Y & G & Q & \underline{F} & \underline{E} & \underline{H} & \underline{K} & \underline{A} & \underline{D} & \underline{S} & \underline{L} & \underline{I} & \underline{A} & \underline{K} & \underline{F} & \underline{K} & \underline{A} \\
 & & 132 & & & & H6 & & & & 141 & & & & & & &
 \end{array}$$

Comparison of local sequences similarities with V or C lysozymes and structural alignments

Comparisons between the sequences of enzymes belonging to different classes of the lysozyme superfamily (C-type, G-type and V-type lysozymes, barley chitinase and *Streptomyces* chitosanase), have detected very few similarities and no alignment could be realized. Our attempts to align the sequence of λ L with those of the V or C-type lysozymes did not permit us to construct a global alignment either, however, local similarities were detected which led us to identify the most essential catalytic residue. The significance of these partial alignments can now be re-examined in the light of the structural information.

The sequence comparisons were done by a modified dot plot method described earlier (Argos, 1987; Jespers *et al.*, 1992): searches were run for aligned sequence fragments of N residues ($11 < N < 35$) whose similarity score (SSc; Dayhoff scoring scheme, Dayhoff *et al.*, 1983) was a few standard deviations (σ) above the mean of random alignments. Comparison matrices were established between the sequence of λ L and those of T4L or of prealigned phage lysozyme (Garvey *et al.*, 1986; Fastrez, 1996), then between λ L and 18 prealigned C-type lysozymes (Nitta & Sugai, 1989). Particular attention was given to fragments implying catalytic residues or conserved structural elements. The significant results going from the amino to the carboxyl terminal of λ L are presented in Scheme II and are commented on below. Global alignments were also attempted using the 3Dprofile method of Eisenberg (Bowie *et al.*, 1991) but without success; here a local 3D-profile similarity was observed between the sequences connecting the two domains in λ L or in HEWL; it is also shown in Scheme II.

In the amino terminal of λ L, a 14 residues fragment containing helix H1 is found to align with the sequence supporting helix HI of T4L or of pre-aligned phage lysozymes; the similarity score (SSc) is 3.3σ above the mean of random alignments (Scheme II(a)). This alignment corresponds to the structural one given in Scheme I.

Several fragments of the λ L sequence can be aligned with the sequences corresponding to the β -sheet of the other phage lysozymes (Jespers *et al.*, 1992) but none of these alignments corresponds to the structural alignment. In the comparison with the C-type lysozymes, a 29 residues fragment was found to align with the region corresponding to the HEWL β -sheet (5.9σ above the random mean, Scheme II(b)). The region of highest similarity, from λ L residues 61 to 74 (6.8σ above the mean, shaded in Scheme II(b)), is also very well aligned structurally: $\beta 5$ and $\beta 6$ of λ L are aligned with $\beta 2$ and $\beta 3$ of HEWL. However, although running in the same direction (Figure 4), $\beta 3$ of λ L is not super-posable to $\beta 1$ of HEWL. This may arise from the insertion of a large sequence fragment and a probable reorganization of the λ L β -sheet. The structural alignment of the SRWW tetrapeptide has been discussed in the previous section.

The sequence of the long helix connecting the two domains (H3) is relatively poorly conserved in the V-type lysozymes (Fastrez, 1996). In the multiple alignment of the sequence of λ L with the phage lysozymes, a 15 residues sequence fragment of weak similarity with the amino terminal of helix H3 was detected (2.7σ above the mean, shaded alignment in Scheme II(c)). No significant sequence similarity was observed with the C-type lysozymes but a 29 residues fragment indicative of local structural similarity was found in the 3D profile alignment with HEWL (3.0σ above the mean, score per residue = 0.16, at the lower limit of confidence of 3D profiles as defined by Lüthy *et al.*, 1992). The combined alignments of λ L with V and C-type lysozymes imply a local alignment between the last two which appears to correspond to the structural alignments reported by Weaver *et al.* (1985a) or Taylor & Orengo (1989).

A 15 residues fragment is found in the λ L and T4L sequence alignment (3.5σ above the mean, Scheme II(d)) which correspond approximately with the structural alignment of helices H5; it aligns C120 of λ L with C97 of T4L. A disulphide bridge has been constructed between C97 and a cysteine introduced by mutation of Ile3 (Perry & Wetzel, 1984). The proximity between λ L-C120 and λ L-F11, implied by the local alignment, is indeed observed; it emphasises a similar spatial relationship between the helices H1 and H5 in both enzymes, although H5 is shorter in λ L. In HEWL, a disulphide bridge is also observed between a cysteine in the middle of helix H2 (structurally aligned with T4L-H1) and the end of the short helix H5; this bridge may point to a conserved spatial relationship that existed in the presumed common ancestor.

In the region of helix H10 of T4L, a highly significant alignment is found on 17 residues (4.5σ above the mean, Scheme II(e)); it suggests that the helices λ L-H6 and T4L-H10 are related. The orientations of these helices *versus* H1 and H5, however, are not identical; this may result from a rigid body movement in λ L as a consequence of the deletion of helices H6 to H9.

These local sequence similarities had led us to build a crude and partial model of λ L which guided our experiments before structure determination (Jespers *et al.*, 1992; P. Soumilion & J.F., unpublished results). The

hypothesis that helices equivalent to H1, H3, H5 and H10 of T4L were present in λ L (H3 being shorter in this enzyme), that there was some kind of duplication of the β sheet and a miniaturization of the C-terminal domain is supported by the determination of the structure.

CONCLUSION

The agreement between local sequence alignments and structural alignments described above supports the hypothesis that λ L is evolutionarily more closely related to the other phage lysozymes; however, the C-terminal domain of λ L is rather small; in this respect, λ L appears to be different from the other members of the bacterial family (Monzingo *et al.*, 1996). Two features of λ L connect it to the eukaryotic and particularly the C-type family: a shortening of the helix connecting the two domains associated with a decrease in the size of the C-terminal domain and the presence in the β -sheet of a sequence fragment highly similar to the most conserved part of the sequences of the C lysozymes. The origin of this sequence fragment and the mechanism of its insertion into the β -sheet of λ L remain mysterious.

As far as the mechanism is concerned, the structural data confirm the role of Glu19 as an essential catalytic residue. Its location in the active site is quite similar to that of the corresponding glutamic acid in other lysozymes. There is apparently no second catalytic residue in the active site: Asp34, which had previously been envisaged to play a significant role in catalysis, appears to be too far from Glu19 to be involved and no other residue with a suitable side-chain is detected in the active site cavity. The necessity of a second residue to assist in lysozyme catalysis had been questioned earlier (Hardy & Poteete, 1991; Weaver *et al.*, 1995). At this stage, the origin of the change in mechanism between λ L and the other lysozymes remains unexplained. This question and the determination of the possible role of the conformational change in the amino terminal of helix H6 will have to await further experimental work.

MATERIALS AND METHODS

Crystallization and data collection

In spite of numerous attempts, we were unable to grow crystals suitable for X-ray analysis from the unmodified bacteriophage lambda lysozyme (Evrard *et al.*, 1997). The mutant in which all tryptophan residues have been replaced by azatryptophan was prepared as previously described (Soumilion *et al.*, 1995). Crystals were grown by sitting drop vapour diffusion of the protein solution (20 mg ml⁻¹ in 50 mM NaH₂PO₄/Na₂HPO₄ (pH 6.7), 0.02% (w/v) NaN₃) mixed 50:50 with a well solution of 20% (w/v) PEG4000, 15% (v/v) 2-propanol, 0.1 M sodium citrate (pH 5.3), as previously reported (Evrard *et al.*, 1997). The crystals are orthorhombic, space group P2₁2₁2₁ with cell dimensions $a = 73.01$ Å, $b = 78.80$ Å, $c = 82.31$ Å. The crystal used for the collection of the native data set (native 1) was grown under microgravity conditions during the flight STS-67 of the space shuttle.

If we assume that two or three molecules are present in the asymmetric unit, the calculation of V_M (Matthews, 1968) leads to values of 3.4 Å³ dalton⁻¹ and 2.2 Å³ dalton⁻¹, respectively, with associated solvent content of 62% and 44%. Both V_M values are consistent with those obtained for globular proteins though the situation in which the asymmetric unit contains three molecules seems more probable. As reported in the preliminary paper (Evrard *et al.*, 1997), all attempts to solve the structure by molecular replacement were unsuccessful, suggesting the need for multiple isomorphous replacement.

Three heavy atom derivatives were prepared by soaking the crystals in the standard mother liquor containing the heavy atom: sodium tetrachloroplatinate (Pt1), platinum terpyridine chloride (Pt2) and mersalyl acid (Hg1). The fourth heavy atom derivative, methyl mercury chloride (Hg2), could not be dissolved in the mother liquor due to the presence of the citrate buffer. A modified mother liquor, in which citrate was replaced by phosphate, was used for soaking the crystals and a second native data set (native2) was also collected from a crystal soaked in this liquor. All the data were collected at room temperature using a synchrotron radiation source and a MAR Research imaging plate detector at the EMBL Hamburg outstation. The details are summarized in Table 1. All the measurements were indexed and integrated using the program DENZO (Otwinowski, 1993) and then merged with the program SCALEPACK (Otwinowski, 1993). The program SCALEIT from the CCP4 package (Collaborative Computing Project, Number 4, 1994) was used to scale all the data sets to native1 data and 5% of the reflections were flagged for use in free-R calculations. In the comparison of native1 and native2 data, a value of R_{iso} (defined in Table 1) = 0.152 was computed showing that the two native data sets were not sufficiently isomorphous to allow the use of Hg2 derivative with native1. For this reason, the isomorphous differences between the Hg2 data and the native2 data were computed and a new pseudo derivative was created by adding

these differences to the native data. The R_{iso} value and all the MIR phasing values concerning the Hg2 compound in Table 1 refer to this pseudo derivative. It is important to note that all the attempts to solve the structure without this additional pseudo derivative were unsuccessful.

Structure determination

Self-rotation searches were attempted on the native data looking for non-crystallographic symmetry, with the programs X-PLOR (Brünger, 1992a) and AMoRe (Navaza, 1987). A systematic search through all rotation possibilities was conducted using several ranges of resolution. All these attempts were unsuccessful. Moreover no pure translation peak was found in the native Patterson function.

An examination of the isomorphous differences as a function of resolution led us to believe that some lack of isomorphism could occur usually for data beyond 3.0 Å resolution so that all the MIR calculations were restricted to a resolution of 20 to 3.0 Å.

Pt1 was the first available derivative and a difference Patterson computed and interpreted by the program SHELXS86 (Sheldrick, 1985) allowed us to identify three heavy atom sites. The coordinates and occupancies of these sites were refined with the program MLPHARE (Otwinowski, 1991). The resulting SIR phases were used to look for additional sites and for sites in other derivatives by the analysis of difference Fourier syntheses. New sites were progressively incorporated and their agreement with the corresponding difference Patterson functions was checked by the program VECREF (Tickle, 1991) from the CCP4 package. Finally, all the derivatives were refined with the programme MLPHARE, including the anomalous dispersion data (Table 1). Since all the anomalous occupations were positive, it was concluded that the handedness of the structure was correct. The figure of merit was 0.50 for all reflections between 20.0 and 3.0 Å. The resulting phases were improved by solvent flattening and histogram fitting implemented in the program DM (Cowtan, 1994). Although some secondary elements were visible in the resulting electron density map, the chain tracing was not possible at this stage. However, the distribution of the electron density clearly suggested that the asymmetric unit should contain three independent protein molecules. The localization of the heavy atom sites in the neighbourhood of each of these three supposed molecules was carefully examined, looking for some similar pattern. A relationship between molecule 1 and 2 was found using for each molecule two platinum atoms of the Pt1 compound and one mercury atom of the Hg2 compound. The same two platinum atoms from Pt1 and one from Pt2 allowed us to find the relationship between molecules 1 and 3. The two transformations correspond approximately to nearly perpendicular 2-fold axes, one with an inclination of 11 degrees with respect to the *c* direction, the other one 25 degrees with respect to the *a* direction. This can explain the lack of success of the self rotation function when looking for non-crystallographic symmetry: the directions of the non-crystallographic 2-fold axes are close to the crystallographic 2-fold screw axes, but not close enough to produce a pure translation peak in the native Patterson function.

After skeletonization and superposition of the three parts of the map containing the molecules, a mask corresponding to one molecule was created from the superposed bones, using the program MAMA (Kleywegt & Jones, 1994), allowing us to improve the transformation matrices between the three parts of the map (correlations = 0.59 and 0.63). The program DM (Cowtan, 1994) was run again, but applying the non-crystallographic symmetry. The resulting averaged map was very nice and allowed us to trace the chain for three fragments, representing 87 residues and to unambiguously assign parts of the known sequence (Sanger *et al.*, 1982) to these fragments by means of the "slider-guess" option in O. After a number of recyclings, including structure factor calculations (program SFALL), phase combination (program SIGMAA, Read, 1986), mask modification (program MAMA, Kleywegt & Jones, 1994) and visual examination, 103 residues were identified ($R = 0.515$, $R_{free} = 0.551$, Brünger, 1992b) and refined with strict NCS using the slowcool protocol of X-PLOR ($R = 0.408$, $R_{free} = 0.463$; Brünger, 1992a). After tracing about 20 more residues, it clearly appeared that for some parts of the molecule the strict NCS had to be relaxed. NCS restraints with a weight of 300 kcal mole⁻¹ Å⁻² were then applied except for residues 1 to 3, 23 to 25, 51 to 60, 102 to 104 and 128 to 141; for these residues, the chain was traced or retraced independently for the three molecules. The three chains appeared then uninterrupted from residue 1 to residue 154. In all the subsequent refinements, the available data, with $I \geq \sigma(I)$ between infinity and 2.3 Å were included and a bulk solvent correction was applied (Table 2). Using the program X-PLOR (Brünger, 1992a), several rounds of positional refinement alternating with individual B-factor refinement and manual rebuilding steps were applied. The side-chains of 21 residues, mainly on the surface of the molecules, were also removed from the NCS restraints and two alternate conformations had to be applied to the side-chain of Arg78 in molecules A and C. A total of 118 water molecules as well as three isopropanol molecules (each bound to one protein molecule) were progressively included. It was never possible to trace residues 155 to 158 in any of the three molecules. The electron density is well defined in all parts of the three molecules, with a slightly

lower quality in the region 131 to 134 of molecule A and 56 to 59 of molecule C. Electron density is missing for the extremities of some long side-chains at the surface of the molecules, principally for some lysine residues. At the end of the refinement, the programs OOPS (Kleywegt & Jones, 1995) and PROCHECK (Laskowski *et al.*, 1993) were used to correct some minor local problems. The final R-indices for observed data between infinity and 2.3 Å are $R = 0.1627$ and $R_{\text{free}} = 0.2134$. It has been checked that some additional release of the NCS restraints could allow us to reach a lower R-value but the R_{free} -value remained unmodified, indicating an overfitting of the experimental data and consequently this additional release was not applied. During the very final cycles of refinement, the R_{free} flagged reflections were included and the final R-value for all available reflections with $\Delta\sigma(I)$ between infinity and 2.3 Å is 0.1635. These results are summarized in Table 2.

Acknowledgements

The authors are indebted to the Fonds National de la Recherche Scientifique (Belgium) and to the Fonds de Développement Scientifique (Université Catholique de Louvain) for financial support. C.E. thanks the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture for a fellowship. We thank the European Union for support of the work at EMBL Hamburg through the HCMP to Large Installations Project, contract no. CHGE-CT93-0040. We are grateful to the members of the scientific staff at the EMBL outstation for their help during data collection and processing.

References

- Argos, P. (1987). A sensitive procedure to compare amino acid sequences. *J. Mol. Biol.* **193**, 385-396.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer based archival file for macromolecular structures, *J. Mol. Biol.* **112**, 535-542.
- Bienkowska-Szewczyk, K., Lipinska, B. & Taylor, A. (1981). The R gene product of bacteriophage λ is a murein transglycosylase. *Mol. Gen. Genet.* **184**, 111-114.
- Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. Roy. Soc. ser. B.* **167**, 378-388.
- Blake, C. C. F., Cassels, R., Dobson, C. M., Poulsen, F. M., Williams, R. J. P. & Wilson, K. S. (1981). Structure and binding properties of hen lysozyme modified at tryptophan 62. *J. Mol. Biol.* **147**, 73-95.
- Bowie, J. U., Lüthy, R. & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, **253**, 164-170.
- Brünger, A. T. (1992a). *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
- Brünger, A. T. (1992b). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.
- Collaborative Computing Project Number 4 (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760-763.
- Cowan, K. (1994). DM, an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newsletter Protein Crystallog.* **31**, 24-28.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983). Established homologies in protein sequences. *Methods Enzymol.* **91**, 524-545.
- Diamond, R., Phillips, D. C., Blake, C. C. F. & North, A. C. T. (1974). Real-space refinement of the structure of hen egg white lysozyme. *J. Mol. Biol.* **82**, 371-391.
- Engh, R. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392-400.
- Evrard, C., Declercq, J.-P. & Fastrez, J. (1997). Crystallization and preliminary X-ray analysis of bacteriophage lambda lysozyme in which all tryptophans have been replaced by aza-tryptophans. *Acta Crystallog. sect. D*, **53**, 217-219.
- Fastrez, J. (1996). Phage lysozymes. In *Lysozymes: Model Enzymes in Biochemistry and Biology* (Jolies, P., éd.), pp. 35-64, Birkhäuser Verlag, Basel.
- Garvey, K. J., Saedi, M. S. & Ito, J. (1986). Nucleotide sequence of *Bacillus* phage ϕ 29 genes 14 and 15: homology of gene 15 with other phage lysozymes. *Nucl. Acids Res.* **14**, 10001-10008.

- Hardy, L. W. & Poteete, A. R. (1991). Reexamination of the role of Asp20 in catalysis by bacteriophage T4 lysozyme. *Biochemistry*, **30**, 9457-9463.
- Imada, M. & Tsugita, A. (1971). Amino-acid sequence of λ phage endolysin. *Nature New Biol.* **233**, 230-231.
- Jespers, L., Sonveaux, E. & Fastrez, J. (1992). Is the bacteriophage lambda lysozyme an evolutionary link or a hybrid between the C and V-type lysozymes? Homology analysis and detection of the catalytic amino acid residues, *J. Mol. Biol.* **228**, 529-538.
- Jollès, P. (1996). *Lysozymes: Model Enzymes in Biochemistry and Biology*, Birkhauser Verlag, Basel.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110-119.
- Kleywegt, G. J. & Jones, T. A. (1994). Halloween. . . masks and bones. In *From First Map to Final Model* (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 59-66, SERC Daresbury Laboratory, UK.
- Kleywegt, G. J. & Jones, T. A. (1995). Efficient rebuilding of protein structures. *Acta Crystallog. sect. D*, **52**, 829-832.
- Kleywegt, G. J. & Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, 525-545.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK. *J. Appl. Crystallog.* **26**, 283-291.
- Lüthy, R., Bowie, J. U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83-85.
- Luzzati, P. V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallog.* **5**, 802-810.
- Matthews, B. W. (1968). Solvent content in protein crystals, *J. Mol. Biol.* **33**, 491-497.
- Matthews, B. W., Remington, S. J., Grütter, M. G. & Anderson, W. F. (1981). Relation between hen egg white lysozyme and bacteriophage T4 lysozyme: evolutionary implications, *J. Mol. Biol.* **147**, 545-558.
- Monzingo, A. F., Marcotte, E. M., Hart, P. J. & Robertus, J. D. (1996). Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. *Nature Struct. Biol.* **3**, 133-140.
- Mushegian, A. R., Fullner, K. J., Koonin, E. V. & Nester, E. W. (1996). A family of lysozyme-like virulence factors in bacterial pathogens of plants and animals. *Proc. Natl Acad. Sci. USA*, **93**, 7321-7326.
- Navaza, J. (1987). On the fast rotation function. *Acta Crystallog. sect. A*, **43**, 645-653.
- Nicholls, A., Bharadway, R. & Honig, B. (1993). GRASP: graphical representation and analysis of surface properties. *Biophys. J.* **64**, 166-170.
- Nitta, K. & Sugai, S. (1989). The evolution of lysozyme and α -lactalbumin. *Eur. J. Biochem.* **182**, 111-118.
- Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom parameters. In *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), pp. 80-86, SERC Daresbury Laboratory, UK.
- Otwinowski, Z. (1993). Oscillation data reduction programs. In *Data Collection and Processing* (Sawyer, L., Isaacs, N. & Bailey, S. W., eds), pp. 56-62, SERC Daresbury Laboratory, UK.
- Perry, L. J. & Wetzel, R. (1984). Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science*, **226**, 555-557.
- Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with error. *Acta Crystallog. sect. A*, **42**, 140-149.
- Remington, S. J. & Matthews, B. W. (1978). A general method to assess similarity of protein structures, applications to T4 bacteriophage lysozyme. *Proc. Natl Acad. Sci. USA*, **75**, 2180-2184.
- Rossmann, M. G. & Argos, P. (1976). Exploring structural homology of proteins, *J. Mol. Biol.* **105**, 75-95.
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. (1982). Nucleotide sequence of bacteriophage λ DNA. *J. Mol. Biol.* **162**, 729-773.
- Sheldrick, G. M. (1985). *SHELXS86. Program for the Solution of Crystal Structures*, University of Göttingen, Germany.
- Soumilion, P. & Fastrez, J. (1992). A large decrease in the heat-shock-induced proteolysis after tryptophan starvation leads to an increased expression of the phage lambda lysozyme in *E. coli*. *Biochem. J.* **286**, 187-191.

Soumillion, P., Jespers, L., Vervoort, J. & Fastrez, J. (1995). Biosynthetic incorporation of 7-azatryptophan into the phage lambda lysozyme: estimation of tryptophan accessibility, effect on enzymatic activity and protein stability. *Protein Eng.* **5**, 451-456.

Susskind, M. M. & Botstein, D. (1978). Molecular genetics of bacteriophage P22. *Microbiol. Rev.* **42**, 385-413.

Taylor, A., Das, B. C. & Van Heijenoort, J. (1975). Bacterial-cell-wall peptidoglycan fragments produced by phage λ or Vi II endolysin and containing 1,6-anhydro-N-acetylmuramic acid. *Eur. J. Biochem.* **53**, 47-54.

Taylor, W. R. & Orengo, C. A. (1989). Protein structure alignment, *J. Mol. Biol.* **208**, 1-22.

Thunnissen, A.-M. W. H., Dijkstra, A. J., Kalk, K. H., Rozeboom, H. J., Engel, H., Keck, W. & Dijkstra, B. W. (1994). Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature*, **367**, 750-753.

Tickle, I. (1991). Refinement of SIR heavy atoms parameters in Patterson *versus* reciprocal space. In *Isomorphous Replacement and Anomalous Scattering* (Wolf, W., Evans, P. R. & Leslie, A. G. W., eds), pp. 87-95, SERC Daresbury Laboratory, UK.

Weaver, L. H. & Matthews, B. W. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Angstroms resolution, *J. Mol. Biol.* **193**, 189-199.

Weaver, L. H., Grütter, M. G., Remington, S. J., Gray, T. M., Isaacs, N. W. & Matthews, B. W. (1985a). 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, *J. Mol. Evol.* **21**, 97-111.

Weaver, L. H., Rennell, D., Poteete, A. R. & Matthews, B. W. (1985b). Structure of phage P22 gene 19 lysozyme inferred from its homology with phage T4 lysozyme. Implication for lysozyme evolution. *J. Mol. Biol.* **184**, 739-741.

Weaver, L. H., Grütter, M. G. & Matthews, B. W. (1995). The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the goose-type lysozymes lack a catalytic aspartate residue, *J. Mol. Biol.* **245**, 54-68.

Wiggins, B. A. & Hilliker, S. (1985). Genetic and DNA mapping of the late regulation and lysis genes of *Salmonella* bacteriophage P22 and coliphage λ . *J. Virol.* **56**, 1030-1033.

Wolf, W., Evans, P. R. & Leslie, A. G. W. (1991). *Isomorphous Replacement and Anomalous Scattering*, CCP4 Daresbury Laboratory Study Weekend, nos. DL/SCI/R32, ISSN 0144-5677, Daresbury Laboratory, Warrington WA4 4AD, UK.