Aust. J. Biol. Sci., 1985, **38**, 13–22

Three-dimensional Structure of Goose-type Lysozyme from the Egg White of the Australian Black Swan, *Cygnus atratus*

N. W. Isaacs,[A] *K. J. Machin*[A] *and M. Masakuni*[A,B]

[A] St Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy, Vic. 3065.
[B] Present address: Department of Biochemistry, National Defence Medical College, Tokorozawa, Japan.

Abstract

The egg white of *C. atratus* contains two forms of lysozyme, a 'chick-type' which is similar to that found in the egg white of the domestic hen, and a 'goose-type' similar to that found in the egg white of the Embden goose. The molecular structure of the goose-type lysozyme has been determined at a resolution of a 2·8 Å by X-ray crystallographic analysis. The structure consists of two domains linked by a long stretch of α-helix. In all, there are seven helical segments in the structure. While there is no amino acid sequence homology with either hen egg-white or bacteriophage T4 lysozymes, there are portions of the structure where the folding of the main chain is similar to that found in portions of either hen egg-white lysozyme or T4 lysozyme or both. In particular, there is a consistency of structure in the arrangement of acid groups in the catalytic site.

Gō plots calculated for this structure and for the bacteriophage T4 lysozyme structure show that both have similar 'modules' of structure with boundaries occurring at structurally equivalent positions. Three of the common boundaries are equivalent structurally to three of the four module boundaries observed in Gō plots of hen egg-white lysozyme. The variation in the position of the remaining boundary may be related to differences in substrate binding.

Introduction

'Goose-type' lysozymes differ from 'chick-type' lysozymes in molecular weight, amino acid composition and immunological properties (Arnheim *et al.* 1973). Although both forms of lysozyme catalyse the same reaction, the goose-type lysozyme has a preference for a peptide-bound substrate. In the standard lysozyme assay (lysis of a suspension of killed *Micrococcus lysodeikticus* cells) black swan goose-type lysozyme has a specific activity six to seven times greater than hen egg-white lysozyme. Black swan goose-type lysozyme consists of 185 amino acid residues (M_r = 20 400) and its amino acid sequence (Simpson *et al.* 1980) shows no homology with either hen egg-white or phage T4 lysozymes, but closely resembles the Embden goose egg-white lysozyme, differing in only six of the common 185 residues (Simpson and Morgan 1983). In this paper we describe the three-dimensional structure of the polypeptide chain of goose-type lysozyme of the Australian black swan, *Cygnus atratus*, which has been determined using X-ray crystallographic methods at 2·8 Å resolution.

Materials and Methods

Purification and Crystallization of Lysozyme

Two crystal forms of black swan goose-type egg-white lysozyme (hereafter referred to as SELg) may be grown from NaCl solutions. The monoclinic form (Masakuni *et al.* 1979), space group $P2_1$ with

0004-9417/85/010013$02.00

$a = 46{\cdot}2$ Å, $b = 65{\cdot}1$ Å, $c = 38{\cdot}7$ Å, $\beta = 110°$, is almost isomorphous with the monoclinic form of lysozyme from the egg white of the Embden goose (Grütter *et al.* 1979). These monoclinic crystals of SELg have a tendency to form twins through a twofold rotation about the *c* axis. The second crystal form is orthorhombic, space group $P2_12_12$ with $a = 91{\cdot}8$ Å, $b = 65{\cdot}4$ Å, $c = 38{\cdot}8$ Å, and one molecule in the asymmetric unit.

Table 1. Statistics for heavy-atom derivatives

Derivative	Concn (mM)	Soaking time (days)	Sites	R_{merge} [A]	R_c [B]	$<F_{pH}-F_p>$	$<f_H>$	E [C]
Native	—	—	—	0·14	—	—	—	—
K_2PtCl_4	1	1	3	0·12	0·63 (5 Å)	16·7	12·7	15·5
K_2PtCl_4	10	1	4	0·12	0·68 (3 Å)	26·4	17·1	27·5
$(NH_3)_2Pt(NO_2)_2$	10	5	5	0·12	0·42	37·3	39·4	20·2
K_2PdCl_4	0·5	5	7	0·13	0·67	28·4	17·7	30·0

[A] $R_{merge} = \Sigma|I-\bar{I}|/\Sigma I$ for intensities measured on different films.
[B] $R_c = \Sigma|(F_{pH}-F_p)-f_H|/\Sigma|F_{pH}-F_p|$ for centric reflections, where F_{pH} is the structure factor of the heavy-atom derivative, F_p is the structure factor of the native protein, and f_H is the heavy-atom contribution.
[C] The r.m.s. lack of closure $E = <(F_{pH}-F_p)-f_H>$.

Table 2. Resolution dependence of heavy-atom derivatives

Parameter[A]	Resolution at: 9·71 Å	7·18 Å	5·69 Å	4·72 Å	4·03 Å	3·51 Å	3·12 Å	2·80 Å	All
				K_2PtCl_4 (1 mM)					
$<f_H>$	17·29	14·19	12·01	10·25	—	—	—	—	12·67
E	16·69	15·19	16·04	14·28	—	—	—	—	15·46
				K_2PtCl_4 (10 mM)					
$<f_H>$	26·34	25·03	20·95	19·52	16·48	14·50	12·11	10·70	17·07
E	41·87	41·20	24·87	21·76	24·65	23·35	17·97	25·55	27·54
				$(NH_3)_2Pt(NO_2)_2$ (10 mM)					
$<f_H>$	67·03	61·69	55·51	44·85	39·45	34·46	28·03	24·17	39·41
E	27·47	21·56	22·99	17·97	20·29	18·49	18·32	18·58	20·25
				K_2PdCl_4 (0·5 mM)					
$<f_H>$	31·09	27·26	24·07	20·76	18·39	15·49	11·77	9·39	17·72
E	33·53	27·95	30·32	26·80	34·00	31·98	26·64	25·45	30·08
Figure of merit, $\bar{m}$	0·83	0·82	0·81	0·80	0·73	0·69	0·56	0·43	0·64

[A] Definitions given in Table 1.

Data Collection and Phase Calculation

Diffraction data were collected by rotation methods (Schwager *et al.* 1975) for the native protein (orthorhombic form) and four isomorphous derivatives prepared by soaking the crystals in buffered solutions of heavy-atom compounds. The locations of the heavy-atom binding sites were determined from difference Patterson and difference Fourier syntheses. Heavy-atom refinement and multiple isomorphous replacement phasing (Dickerson *et al.* 1961), including the anomalous dispersion data for the neutral platinum derivative, resulted in an overall figure of merit of 0·64 for 4930 reflections to 2·8 Å resolution. Details of the heavy-atom refinement and phase calculations are given in Tables 1 and 2.

Interpretation of the Electron-density Map

An electron-density map calculated with these phases was interpreted in an optical comparator. Although it clearly showed a number of helices, and the amino acid sequence could be fitted into the long helix (residues 110–132) it was not possible to interpret unambiguously the connections between the helices. At this stage the structure of the Embden goose lysozyme was made available (Grütter *et al.* 1983) which allowed a model with most of the connections, excluding the irregular β-sheet region (residues 76–91), to be obtained. Coordinates obtained for 1125 atoms from 160 residues were regularized (Dodson *et al.* 1976) and used to calculate phases which were then recombined with the isomorphous phase set using Bricogne's (1976) procedure. For the combined phases the average difference was 54° from the isomorphous phases and 33° from the calculated phases. An electron-density map calculated with these phases and plotted onto transparent sheets at a scale of 3 Å per centimetre showed the complete trace of the protein chain and allowed for the reinterpretation of the isomorphous map in the optical comparator. New coordinates were recorded for 733 atoms from 84 of the 185 residues although the first two residues of the *N*-terminal end still could not be positioned. The present crystallographic residual for the model (with regularized coordinates for 183 of the 185 residues) is 0·43 for 2·5 Å data.

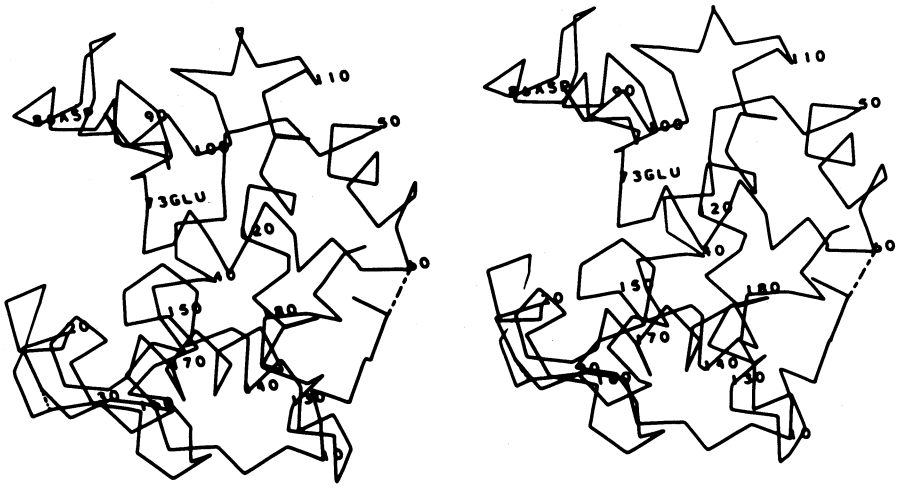

Fig. 1. A stereo diagram of the folding of the backbone chain (residues 3–185) of swan egg-white lysozyme (goose-type).

Results

A trace of the polypeptide chain of SELg is shown in Fig. 1 and a stylized illustration of the topology of the molecule is shown in Fig. 2. There are two disulfide bonds in the structure between residues 4 and 60 and between 18 and 29. The structure consists predominantly of seven α-helices (α_1–α_7) (residues 18–24, 31–43, 48–60, 63–74, 110–132, 136–146, 168–179) with three small strands of irregular antiparallel β-sheet structure (residues 84–86, 88–91, 95–97). The two domains in the structure (Fig. 2) are connected by the long stretch of α-helix (α_5). The left domain (Fig. 2) contains predominantly α-helixes, while the right domain contains the irregular β-sheet. The cleft between these domains contains the acidic residues (Glu 73 and Asp 86) which Schoentgen *et al.* (1982) have suggested could be equivalent to the catalytically active Glu 35 and Asp 52 of the hen egg-white lysozyme. In this structure of SELg, Glu 73 and Asp 86 are located on either side of the cleft with a distance of 10 Å between the carboxyl groups. Distances between equivalent catalytic residues for aligned structures are as follows:

HEWL		SELg		T4L
Glu 35	2·4 Å	Glu 73	1·4 Å	Glu 11
Asp 52	3·0 Å	Asp 86	4·3 Å	Asp 20

Discussion

This structure of SELg is geometrically identical to the reported structure of the closely homologous goose-type lysozyme from the egg white of the Embden goose (GEWL) (Grütter *et al.* 1983). These authors have compared the backbone structure (α carbon positions) of GEWL with the structures of hen egg-white lysozyme (HEWL) and bacteriophage T4 lysozyme (T4L) (Matthews *et al.* 1981). They report r.m.s. discrepancies of 3·2 Å between GEWL and HEWL; 3·2 Å between GEWL and T4L; and 3·8 Å between HEWL and T4L. All α carbons in GEWL have a counterpart in either HEWL or T4L and they argue that the three lysozyme structures diverged from a common ancestor which was either a goose-type or hen-type molecule.

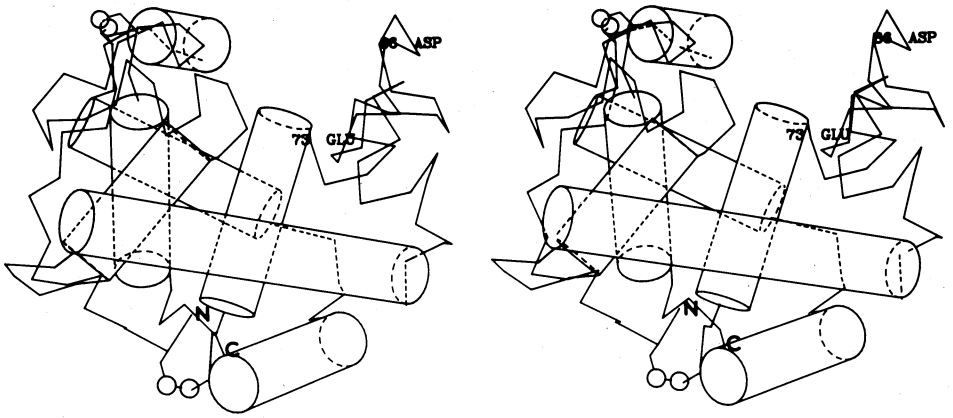

Fig. 2. A stereo diagram of the folding of the backbone chain (residues 3–185) of swan egg-white lysozyme (goose-type). The cylinders represent α-helical sections and the disulfide bridges are shown as circles. This diagram was produced by a computer program written by A. M. Lesk and K. D. Hardman (1982).

Although the accuracy of the SELg structure does not allow for a detailed comparison, it exhibits the same structural similarity with HEWL and T4L. A simple comparison may be made by representing each helix by a vector directed along the helix axis from the *N*- to the *C*-terminal end. Once these vectors are referred to a common centre of mass, the angles between them allow a rotation matrix to be computed to align the vectors. When the axis of the long helix in SELg (residues 110–132) is aligned with the axis of the long helix in T4L (residues 60–79), then it is also found that the helices 63–74 and 136–146 in SELg almost overlap the helices 3–11 and 95–106 in T4L. With this alignment of molecules, the helices 48–60, 63–74 and 110–132 in SELg are also spatially equivalent to the helices 5–15, 24–34, 88–99 respectively in HEWL. Furthermore, the alignment produces a spatial equivalence in the residues of the catalytic site (Glu 35 and Asp 52 in HEWL, Glu 11 and Asp 20 in T4L, Glu 73 and Asp 86 in SELg) which occupy equivalent positions with the glutamic acid always found at the end of an α-helix and the aspartic acid found in a β-sheet region. Table 3 summarizes these data.

Intron–Exon Structure

It has been proposed by Gilbert (1978) that the arrangement of eukaryotic genes into introns and exons could facilitate the evolution of new proteins. This concept was

Table 3. Angles between equivalent α-helices after alignment of SELg, HEWL and T4L structures

SELg and HEWL co-ordinates (from the Protein Data Bank, Bernstein *et al.* 1977) have been transformed into the T4L co-ordinate system [using the matrix given by Matthews *et al.* (1981) to transform HEWL]

HEWL helix	Angle HEWL–SELg	SELg helix	Angle SELg–T4L	T4L helix
5–15	36°	48–60		N.E.[A]
24–34	20°	63–74	27°	3–11
88–99	9°	110–132	1°	60–79
N.E.[A]		136–146	15°	95–106

[A] No equivalence.

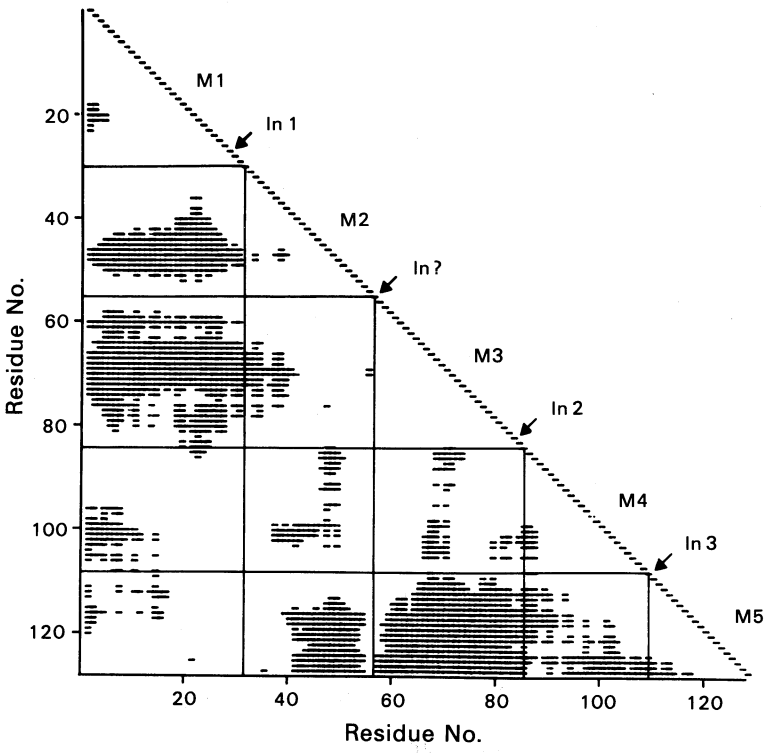

Fig. 3. Modules of chicken egg-white lysozyme (after Gō 1983). The dark regions represent pairs of C_α atoms that are separated by more than 23 Å. The five modules are located by drawing horizontal and vertical lines which avoid the dark regions and intersect on the diagonal. Intron positions of the gene structure are identified by arrows.

extended by Blake (1978) with the suggestion that evolution would be further accelerated if the exons coded for integrally folded subunits of protein structure. Apparent support

for these ideas came when the structure of T4 lysozyme was compared with hen egg-white lysozyme and it was found that the region of best homology (Matthews *et al.* 1981) corresponded with the portion of the HEWL structure encoded by the second and third exons. Jung *et al.* (1980) had previously determined the four exon structure of the chicken lysozyme gene and shown that each exon encoded for a structural and functional domain of the protein. Gō has shown that an analysis of the distribution of the C_α–C_α distances in haemoglobin (Gō 1981) and in chicken lysozyme (Gō 1983) can distinguish subdomains or 'modules' of the structures from which the position of the intron–exon boundaries in the protein sequence can be deduced.

A C_α–C_α distance (or Gō) plot for HEWL is shown in Fig. 3. Horizontal and vertical lines which meet on the diagonal at the intron–exon boundary positions in the sequence (residues 28, 82 and 108) subdivide the map into domains or modules. The average number of residues in each module is 26. In addition there is a fourth point of intersection between residues 53–57 and Gō has postulated the presence of another intron–exon boundary at this point in the ancestral gene of chicken egg-white lysozyme or in the genes of other contemporary lysozymes.

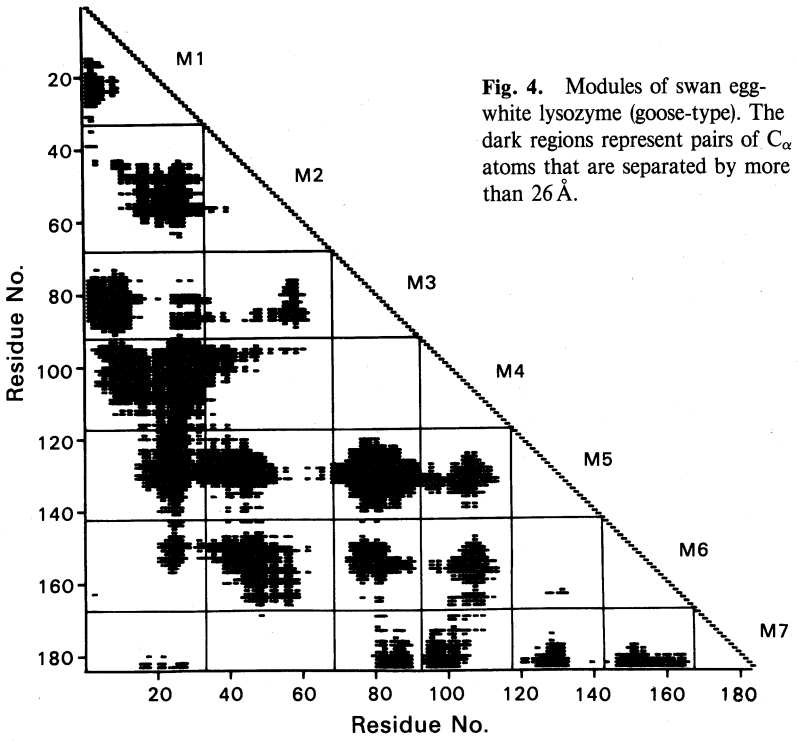

Fig. 4. Modules of swan egg-white lysozyme (goose-type). The dark regions represent pairs of C_α atoms that are separated by more than 26 Å.

We have calculated Gō plots for both SELg (Fig. 4) and T4L (Fig. 5). In the SELg plot there are seven modules with boundaries around residues 33, 68, 92, 117, 142 and 167. For T4L there are six modules with boundaries around residues 7, 30, 73, 105 and 143. In each case the module boundary is located on an α-helix or a β-sheet structure. (It may be surprising that the module boundaries occur within α-helices or β-sheet structures rather than in connecting loops. However, module boundaries by definition

occur at residues whose C_α positions are less than $\simeq 23$ Å from every other C_α in the structure. If the protein structure is seen as a number of helices or sheets packed around the central core, these secondary structures are more likely to have their closest approach points towards their centres, rather than at their extremities). Using the approximate

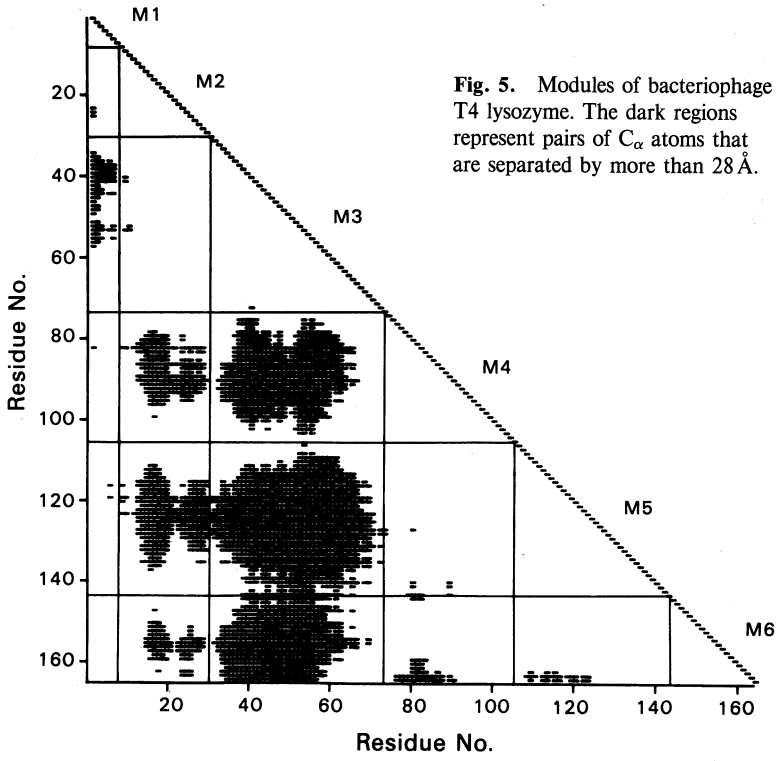

Fig. 5. Modules of bacteriophage T4 lysozyme. The dark regions represent pairs of C_α atoms that are separated by more than 28 Å.

Table 4. Minimum distance between residues at or near module boundaries for aligned structures

HEWL	Distance (Å)	SELg	Distance (Å)	T4L	Distance (Å)	HEWL
28		65	3·1	6	2·4	28
29	1·7	66				
56		91	2·2	28		
57	1·6	92		29	1·4	56
82	18·7	119	2·6	71	18·4	82
105	2·6	147	1·0	104	3·0	105
		172	1·9	147		

alignment of molecules derived from the overlaps of helices as described above, it is found that the residues at or close to the module boundaries in each structure can be closely superimposed; except for the boundary at residue 82 in HEWL, 117 in SELg and 73 in T4L where there is a close superposition only between SELg and T4L (Table 4 and Fig. 6).

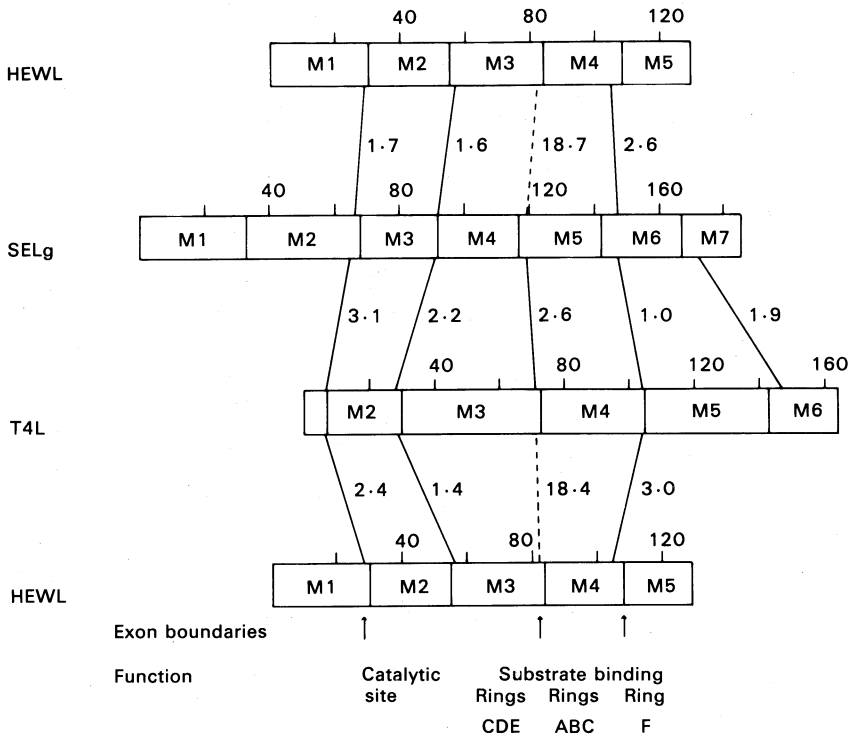

Fig. 6. Comparison of the positions of the module boundaries in the sequences of chicken lysozyme (HEWL), goose-type swan egg lysozyme (SELg) and bacteriophage T4 lysozyme (T4L). The module boundaries are marked with vertical bars and lines between the sequences connect boundaries occurring in spatially equivalent parts of the structures. The distances (in Å) between C_α atoms of the aligned structures are given. The positions of the intron–exon boundaries found in the chicken lysozyme gene are shown.

The data from these plots can be summarized as follows:

(1) There are five structure modules in HEWL with boundaries at or near residues 28, 55, 82, 108.

(2) There are seven structure modules in SELg with boundaries at or near residues 33, 68, 92, 117, 142, 167.

(3) There are six structure modules in T4L with boundaries at or near residues 7, 30, 73, 105, 143.

(4) Three of the four boundaries in HEWL (28, 82, 108) occur at known intron–exon splice junctions.

(5) Between the overlapped SELg and T4L structures there is spatial equivalence for residues at or near all the T4L module boundaries.

(6) Between the overlapped HEWL structure and both SELg and T4L structures there is spatial equivalence for residues at or near the first, second and fourth boundaries, but not for the third.

(7) In each case the module boundary occurs in an α-helical or β-sheet structure.

These results could be taken as support for Gō's suggestion (1981) that module boundaries fall on intron–exon junctions and that a precursor of the HEWL gene contained an intron between residues 53–57. The SELg structure has three module

boundaries occurring at regions which are structurally equivalent to those in HEWL (Table 4 and Fig. 6). An exception is the boundary at residue 82 in HEWL where the corresponding boundary in SELg (residue 119) does not occur at an equivalent part of the structure. The explanation for this difference may be found in Gō's (1983) analysis of the relationship between the module structure and the functional elements of the HEWL structure. No residue in module 1 of HEWL plays an obvious role in substrate binding or in the catalytic mechanism. The boundaries of 29 and 57 in HEWL match well with 66 and 92 in SELg (and the whole of module 2 is similar in structure). This module contains the catalytic residues and the contact sites for rings D, E and F of the substrate. The enzyme cleaves the bond between rings D and E of the substrate. The corresponding module 3 of SELg has the equivalent catalytic residues and a similarity in structure. There is evidence (Grütter *et al.* 1983) that the substrate binds and is cleaved in this region in the homologous GEWL structure.

On the other hand, the boundary at 82 in HEWL does not match the corresponding boundaries in SELg (119) or in T4L (73), and when the module structures are compared (3 and 4 of HEWL with 4 and 5 of SELg) agreement is poor, even though helices 88–99 (HEWL) and 110–132 (SELg) match in part. It could be argued that in this region the two structures have diverged in response to a functional requirement. In HEWL, modules 3 and 4 contain the residues that bind the substrate (rings C, D and E for module 3 and rings A, B and C for module 4). The goose-type lysozyme has a preference for peptide-bound saccharides and the presumptive binding site of the peptide would occur in this region of the structure. It is notable that the structure of T4 lysozyme, which also cleaves peptide-bound substrates is more like that of SELg than HEWL in this region.

Until the gene structure of a goose-type lysozyme is known, the supposed correlation between the intron–exon structure of the gene and the module structure of the protein as detected from Gō plots must remain tenuous. However, this present work does suggest that the Gō plot is very effective in detecting regions of similarity and dissimilarity between functionally related molecules.

Acknowledgments

Permission to collect black swan eggs was granted by the Victorian Department of Fisheries and Wildlife. We thank the staff of that Department for their assistance in collecting them. We are grateful to Dr P. M. Colman, Division of Protein Chemistry, CSIRO, for providing access to microdensitometer facilities and the PROTEIN set of computer programs written by Dr W. Steigemann. The protein diagrams were drawn by Dr J. M. Guss, and the Gō plots were computed by L. Guddat. We particularly thank Dr B. W. Matthews for communicating results of the Embden goose lysozyme structure prior to its publication and acknowledge the referees' constructive comments. This work was supported by a grant from the National Health and Medical Research Council.

References

Arnheim, N., Hindenburg, A., Begg, G. S., and Morgan F. J. (1973). Multiple genes for lysozymes in birds. Studies on black swan egg white lysozymes. *J. Biol. Chem.* **248**, 8036–42.

Bernstein, F. C., *et al.* (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Molec. Biol.* **112**, 535–42.

Blake, C. C. F. (1978). Do genes-in-pieces imply proteins-in-pieces? *Nature (London)* **273**, 267.

Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr.* A **32**, 832–47.

Dickerson, R. E., Kendrew, J. C., and Strandberg, B. (1961). The crystal structure of myoglobin: phase determination to a resolution of 2Å by the method of isomorphous replacement. *Acta Crystallogr.* **14**, 1188–95.

Dodson, E. J., Isaacs, N. W., and Rollett, J. S. (1976). A method for fitting satisfactory models to sets of atomic positions in protein structure refinements. *Acta Crystallogr.* A **32**, 311–15.

Gilbert, W. (1978). Why genes in pieces? *Nature (London)* **271**, 501.

Gō, M. (1981). Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature (London)* **291**, 90–2.

Gō, M. (1983). Modular structural units, exons and function in chicken lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1964–8.

Grütter, M. G., Rine, K. L., and Matthews, B. W. (1979). Crystallographic data for lysozyme from the egg white of Embden goose. *J. Molec. Biol.* **135**, 1029–32.

Grütter, M. G., Weaver, L. H., and Matthews, B. W. (1983). Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? *Nature (London)* **303**, 828–30.

Jung, A., Sippel, A. E., Grez, M., and Schutz, G. (1980). Exons encode functional and structural units of chicken lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5759–63.

Lesk, A. M., and Hardman, K. D. (1982). Computer-generated schematic diagrams of protein structures. *Science (Wash., D.C.)* **216**, 539–40.

Masakuni, M., Simpson, R. J., and Isaacs, N. W. (1979). Preliminary X-ray diffraction studies on the 'goose-type' lysozyme from the egg-white of the black swan *Cygnus atratus*. *J. Molec. Biol.* **135**, 313–14.

Matthews, B. W., Remington, S. J., Grütter, M. G., and Anderson, W. F. (1981). Relation between hen egg white lysozyme and bacteriophage T4 lysozyme: evolutionary implications. *J. Molec. Biol.* **147**, 545–58.

Schoentgen, F., Jolles, J., and Jolles, P. (1982). Complete amino acid sequence of ostrich (*Struthio camelus*) egg-white lysozyme, a goose-type lysozyme. *Eur. J. Biochem.* **123**, 489–97.

Schwager, P., Bartels, K., and Jones, A. J. (1975). Refinement of setting angles in screenless film methods. *J. Appl. Crystallogr.* **8**, 275–80.

Simpson, R. J., Begg, G. S., Dorow, D. S., and Morgan, F. J. (1980). Complete amino acid sequence of the goose-type lysozyme from the egg white of the black swan. *Biochemistry* **19**, 1814–19.

Simpson, R. J., and Morgan, F. J. (1983). Complete amino acid sequence of Embden goose (*Anser anser*) egg-white lysozyme. *Biochim. Biophys. Acta* **744**, 349–51.

Manuscript received 11 July 1984, accepted 9 November 1984