

Crystal Structure of Uncleaved Ovalbumin at 1.95 Å Resolution

Penelope E. Stein^{1†}, Andrew G. W. Leslie², John T. Finch²
and Robin W. Carrell¹

¹*Department of Haematology, University of Cambridge
Hills Road, Cambridge CB2 2QH, U.K.*

²*MRC Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, U.K.*

(Received 14 January 1991; accepted 20 June 1991)

Ovalbumin, the major protein in avian egg-white, is a non-inhibitory member of the serine protease inhibitor (serpin) superfamily. The crystal structure of uncleaved, hen ovalbumin was solved by the molecular replacement method using the structure of plakalbumin, a proteolytically cleaved form of ovalbumin, as a starting model. The final refined model, including four ovalbumin molecules, 678 water molecules and a single metal ion, has a crystallographic *R*-factor of 17.4% for all reflections between 6.0 and 1.95 Å resolution. The root-mean-square deviation from ideal values in bond lengths is 0.02 Å and in bond angles is 2.9°. This is the first crystal structure of a member of the serpin family in an uncleaved form. Surprisingly, the peptide that is homologous to the reactive centre of inhibitory serpins adopts an α -helical conformation. The implications for the mechanism of inhibition of the inhibitory members of the family is discussed.

Keywords: ovalbumin; crystal structure; serpins; reactive centre; molecular replacement

1. Introduction

The serpins (Carrell & Travis, 1985) are a superfamily of more than 20 homologous proteins, found in animals, plants and viruses, that probably evolved from a serine protease inhibitor (Hunt & Dayhoff, 1980; Carrell & Boswell, 1986). They include plasma inhibitors that control enzymes of the coagulation, fibrinolytic, complement and kinin cascades, as well as proteins without any known inhibitory activity such as hormone-binding globulins, angiotensinogen and egg-white ovalbumin.

Loebermann *et al.* (1984) reported the crystal structure of human antitrypsin[‡] that had been proteolytically cleaved at its reactive centre peptide bond (Met358–Ser359) and showed an unexpected separation of the new chain termini by 67 Å

(1 Å = 0.1 nm). The new C terminus forms the fourth strand (s4A§) of the six-stranded β -sheet A that runs parallel to the long axis of the molecule. It was predicted that in intact antitrypsin, this strand would be withdrawn from sheet A to form an external loop, placing the reactive centre peptide bond near the position of Ser359 in the crystal structure. Other serpins, with the exceptions of ovalbumin and angiotensinogen (Bruch *et al.*, 1988; Gettins, 1989; Stein *et al.*, 1989), show evidence for a similar conformational change following reactive centre cleavage (Carrell & Owen, 1985; Bruch *et al.*, 1988; Gettins & Harten, 1988; Pemberton *et al.*, 1988, 1989). There is no reported crystal structure of an inhibitory serpin in its native uncleaved form, either alone or as a complex with its target protease.

Ovalbumin shows no inhibitory activity, but its putative reactive centre (Ala358/Ser359||) is readily identified by sequence alignment with inhibitory

† Present address: Department of Medical Microbiology and Infectious Diseases, 1–41 Medical Sciences Building, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.

‡ Abbreviations used: antitrypsin, α_1 -antitrypsin (α_1 -proteinase inhibitor); serpin, serine protease inhibitor; n.m.r., nuclear magnetic resonance; r.m.s., root-mean-square; s.d., standard deviation; FAST, fast-scanning area-sensitive television detector.

§ Strands of β -sheets are referred to as in the following example: s4A is the fourth strand of β -sheet A.

|| Residue numbering is based on sequence alignment with antitrypsin (Huber & Carrell, 1989). Residues of sequence insertions are given consecutive letters (a, b, c, etc.) with the number of the last aligned residue.

serpins. Subtilisin cleaves six amino acid residues from this site to form the modified protein, plakalbumin (Linderström-Lang, 1952; Satake *et al.*, 1965). The crystal structure of plakalbumin was reported by Wright *et al.* (1990) and has been refined to a crystallographic *R*-factor of 25% for all data to 2.8 Å resolution. The cleaved ends are separated by 27 Å in this structure, but the conformational change seen in cleaved antitrypsin has not taken place. The plakalbumin structure provides a partial model for intact antitrypsin. It supports the prediction by Loebermann *et al.* (1984) that cleavage of antitrypsin is followed by insertion of the new C terminus as an additional strand (s4A) in the centre of a pre-existing five-stranded sheet to make an expanded six-stranded sheet. Wright *et al.* (1990) and Engh *et al.* (1990) have discussed this tertiary structure transformation based on a comparison of the structures of cleaved antitrypsin and plakalbumin. Sequence differences were noted (particularly Arg345 in ovalbumin replacing Thr345 in antitrypsin) that would explain why the C terminus generated by cleavage is not inserted into the A sheet in plakalbumin. We have now determined the structure of uncleaved ovalbumin to provide a model for the intact serpin reactive centre.

Ovalbumin comprises 60 to 65% of the total protein in egg-white (Warner, 1954), but its function is unknown. A dephosphorylated form of ovalbumin found in egg-yolk (Saito & Martin, 1966) may act as an amino acid store for the growing embryo. A possible role of ovalbumin in the transport and storage of metal ions has also been proposed (Taborsky, 1974) and a single strong binding site for several metal ions has been found (Goux & Venkatasubramanian, 1986). Ovalbumin shows no recognized protease inhibitory activity (Long & Williamson, 1980; Ødum, 1987), despite sequence identity of about 30% with antitrypsin and with other functional inhibitors of the serpin family. The Ala-Ser bond at the putative reactive centre suggests specificity for elastase, but ovalbumin acts as a substrate, rather than as an inhibitor, of this enzyme (Wright, 1984).

Ovalbumin is a glycoprotein with a relative molecular mass of 45,000. The amino acid sequence of hen ovalbumin, comprising 385 residues, was deduced from the mRNA sequence by McReynolds *et al.* (1978) and is in complete agreement with the sequences of the purified protein (Nisbet *et al.*, 1981) and the cloned DNA (Woo *et al.*, 1981). The sequence includes four thiol groups with a single disulphide bond between Cys87h and Cys133 (Thompson & Fisher, 1978). The N terminus of the protein is acetylated (Narita & Ishii, 1962). Ovalbumin does not have a classical N-terminal leader sequence, although it is a secretory protein. Instead, the hydrophobic sequence between residues 50 and 66 may act as an internal signal sequence involved in transmembrane location (Robinson *et al.*, 1986).

Heterogeneity in ovalbumin preparations may

arise from several sources. A single heterogeneous carbohydrate chain is covalently linked to the amide nitrogen of Asn298. At least six different ovalbumin glycopeptides have been identified which all share a common core structure: mannose β (1-4) glcNAc β (1-4) glcNAc-Asn298 (Atkinson *et al.*, 1981; Narasimhan *et al.*, 1980; Tai *et al.*, 1975, 1977a,b; Shepherd & Montgomery, 1978; Yamashita *et al.*, 1978).

Heterogeneity in the electrophoretic behaviour of ovalbumin preparations has been attributed to variable phosphorylation at Ser87c and at Ser350 (Milstein, 1968; Nisbet *et al.*, 1981). Three main ovalbumin forms are present with, respectively, two, one and zero phosphate groups per molecule (Linderström-Lang & Ottesen, 1949; Perlmann, 1952), the diphosphorylated form accounting for about 80% of the total (Cann, 1949). Further heterogeneity may be associated with genetic polymorphisms; a Glu→Gln substitution has been reported at residue 295 (Ishihara *et al.*, 1981) and an Asn→Asp substitution at residue 317 (Wiseman *et al.*, 1972). In stored eggs, a proportion of the ovalbumin apparently undergoes a poorly characterized change to form a variant (S-ovalbumin) with increased heat stability (Smith & Back, 1962, 1965).

Microcrystals of uncleaved ovalbumin were first reported nearly 100 years ago (Hopkins & Pinkus, 1898) and the results of a preliminary X-ray diffraction study were described by Miller *et al.* (1983). We have grown larger ovalbumin crystals by reducing the heterogeneity of the protein preparation and report here the structure of uncleaved ovalbumin refined to 1.95 Å resolution.

2. Experimental Methods

(a) Purification of ovalbumin

Ovalbumin was purified from single, newly-laid hens eggs by ammonium sulphate fractionation (Warner, 1954) followed by ion-exchange chromatography (Goux & Venkatasubramanian, 1986). Protein prepared from different eggs was not pooled to avoid heterogeneity arising from genetic polymorphism. Newly-laid eggs were used to avoid possible contribution from the S-ovalbumin variant.

All steps were carried out at room temperature unless otherwise stated. The egg-white was separated from the yolk and added to an equal volume of saturated ammonium sulphate. The mixture was stirred for 30 min, then centrifuged for 30 min at 2000 g. The precipitate was discarded. The supernatant was acidified to pH 4.6 by dropwise addition of 0.5 M-sulphuric acid, incubated for 12 to 24 h at 4°C, then centrifuged at 2000 g for 30 min. The supernatant was discarded. The ovalbumin-rich precipitate was dissolved in distilled water to a volume of 25 ml, dialysed into 10 mM-sodium phosphate buffer (pH 7.0) and loaded onto a 2.5 cm × 40 cm DEAE-cellulose column previously equilibrated with the same buffer. A linear gradient of 10 to 100 mM-sodium phosphate buffer pH 7.0 (1:21 total volume) was applied and the column was eluted at 75 ml/h. Ovalbumin was separated into 3 fractions (A to C; Fig. 1). Goux & Venkatasubramanian (1986) showed, using ³¹P n.m.r., that fractions A and B consist of monophosphorylated

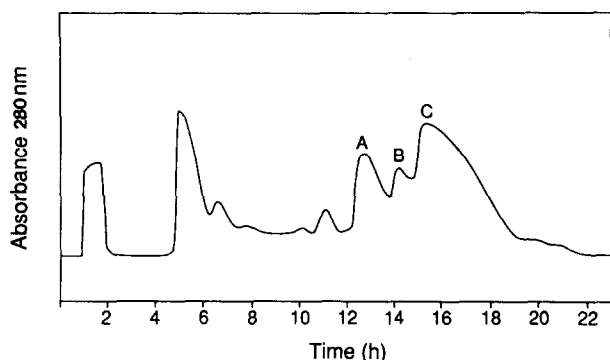


Figure 1. Purification of ovalbumin by ion exchange chromatography on DEAE-cellulose, eluting from 10 to 100 mM-sodium phosphate at pH 7.0. Peaks A and B, monophosphorylated ovalbumin; peak C, diphosphorylated ovalbumin.

ovalbumin (phosphate at Ser87c only), while fraction C is diphosphorylated ovalbumin (phosphates at Ser87c and Ser350). Fraction C was collected and concentrated in an Amicon ultrafiltration cell with a PM10 membrane. Protein concentration was determined from measurements of absorbance at 280 nm, assuming an extinction coefficient of 7.5 for a 1% (w/v) solution of ovalbumin. This preparation migrated as a single compact band, with an apparent relative molecular mass of 43,000, on SDS/polyacrylamide gel electrophoresis under reducing conditions, typically comprising about 96% diphosphorylated and 4% monophosphorylated ovalbumin estimated by isoelectric focusing and densitometry. The yield of diphosphorylated ovalbumin from each egg was approximately 400 mg.

(b) Crystallization

Ovalbumin, purified as described above, was crystallized using the conditions described by Miller *et al.* (1983).

Crystals were grown in sealed glass vials (Cambridge Glass Blowing) at 37°C from freshly prepared protein (25 mg/ml) in 50% saturated ammonium sulphate/50 mM-cacodylate buffer (pH 6.4). Needle-shaped crystals appeared in 1 to 3 months, reaching their maximum size over a further week. The biggest crystals had dimensions of 1.5 mm × 0.5 mm × 0.4 mm, although satellite crystals often grew from one end. Crystals were harvested into a stabilizing medium of 55% saturated ammonium sulphate/45 mM-cacodylate buffer (pH 6.4). The crystals are triclinic (space group *P*1) with 4 molecules in the asymmetric unit and a specific volume, V_m (Matthews, 1968) of 1.95 Å³/Da. The unit cell constants ($a = 62.9$ Å, $b = 84.7$ Å, $c = 71.5$ Å, $\alpha = 87.5^\circ$, $\beta = 104^\circ$, $\gamma = 108.5^\circ$) agree with those reported by Miller *et al.* (1983), although some crystals prepared from different batches of protein had cell dimensions that differed by up to 0.3% (Stein, 1990). Protein from a washed, dissolved crystal ran with the same mobility as intact ovalbumin on SDS/polyacrylamide gel electrophoresis, with 95% diphosphorylated and 5% monophosphorylated forms, estimated by isoelectric focusing and densitometry.

(c) Data collection and processing

Crystals were mounted with their longest (*b*) axis parallel to the length of the capillary. The data were collected at room temperature using 1 or 2 crystals. A dataset to 1.9 Å resolution was collected using radiation of wavelength 1.01 Å on the X31 beamline at the EMBL outstation, DESY, Hamburg with the image plate detector system developed at EMBL by Hendrix and Lentfer. In all, 2 data collection runs were needed because of the limited dynamic range of the detector. In the initial "high-resolution" run (collecting to 1.9 Å), many "low-resolution" reflections were overloaded and a 2nd run, with shorter exposure times, was used to collect low-resolution data (to 2.7 Å). The data were processed using a modified version of the MOSFLM package originally developed by Nyborg & Wonacott (1977). All other data

Table 1
Native ovalbumin data sets

Detector	FAST	Image plate	5-circle
No. crystals	2	1	1
Resolution range (Å)	36 to 2.5	27 to 1.9	30 to 8.0
No. reflections	111,741	352,293	2905
No. independent reflections	46,698	94,362	1281
Completeness of data (%)	99 (to 2.5 Å)	94 (to 1.95 Å)	95 (to 8.5 Å)
$R_{\text{merge}}(\%)^1$	6.9 ² (9.8 at 2.5 Å)	5.2 ^{2,3} (10.0 at 1.95 Å)	4.6 ² (4.8 at 8.5 Å)
Intensities > 3 s.d. (%)	84.8 (69.2 at 2.5 Å)	89.1 (81.7 at 1.95 Å)	98.9 (98.9 at 8.5 Å)

$$R_{\text{merge}} = \frac{\sum \sum |I(h)_j - \langle I(h) \rangle|}{\sum \sum I(h)_j}$$

where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

² Agreement between datasets:

$$R\text{-factor} = \frac{\sum |F_1 - F_2|}{\sum F_1}$$

where F_1 and F_2 are structure factor amplitudes for equivalent reflections of datasets 1 and 2, respectively.

FAST (1) versus image plate (2),

$R\text{-factor} = 8.2\%$ (10 to 2.5 Å)

FAST (1) versus 5-circle (2),

$R\text{-factor} = 6.8\%$ (12 to 8 Å)

³ Overall value for merged high (1.9 Å) and low (2.7 Å) resolution data. R_{merge} for high-resolution data alone is 6.8% and for low-resolution data alone is 4.1%.

were collected using nickel filtered CuK α radiation from a rotating anode source. Data to 2.5 Å resolution were collected with an Enraf-Nonius FAST diffractometer (Arndt & Gilmore, 1979) and the data collection program, MADNES, with integration of reflection intensities using a 3-dimensional profile-fitting method. Low-resolution data (8 to 30 Å) were collected using a 5-circle diffractometer, as some of the low-resolution terms in the FAST data set were poorly determined due to problems associated with the backstop shadow. Subsequent data reduction and scaling, in each case, was performed with the programs ROTAVATA and AGROVATA. These and other programs named in the text are part of the CCP4 package (distributed by SERC Daresbury Laboratory, Daresbury, U.K.). Diffraction data statistics are summarized in Table 1. The FAST data were used for all the molecular replacement calculations, the molecular dynamics refinement and the initial stages of restrained least-squares refinement (PROLSQ). The high-resolution image plate data were used for all subsequent refinement.

(d) Non-crystallographic symmetry

The triclinic unit cell shows pseudo-monoclinic symmetry with molecular packing similar to that in the monoclinic (C2) crystal form of ovalbumin described by Miller *et al.* (1983). The length of the *b* axis in the triclinic cell (84.7 Å) is almost exactly twice that in the monoclinic cell (41.8 Å).

A self-rotation function (Crowther, 1972) was computed using the program POLARRFN. The Patterson function was calculated in the resolution range 8 to 4 Å and the rotation function was integrated within a sphere of radius 20 Å about the origin, applying an overall temperature factor of 20 Å². The rotation function was computed relative to an orthogonal co-ordinate system with *c*, *b** × *c*, *b** along *x*, *y*, *z*, respectively. The parameters of the rotation function were expressed in spherical polar angles (κ , ω , ϕ). The biggest peak in the self-rotation function, besides the origin peak, was at a rotation (κ) of 180° about the axis defined by $\omega = 18.8^\circ$ and $\phi = 277.6^\circ$, in the expected position of the crystallographic *b* axis (Fig. 2).

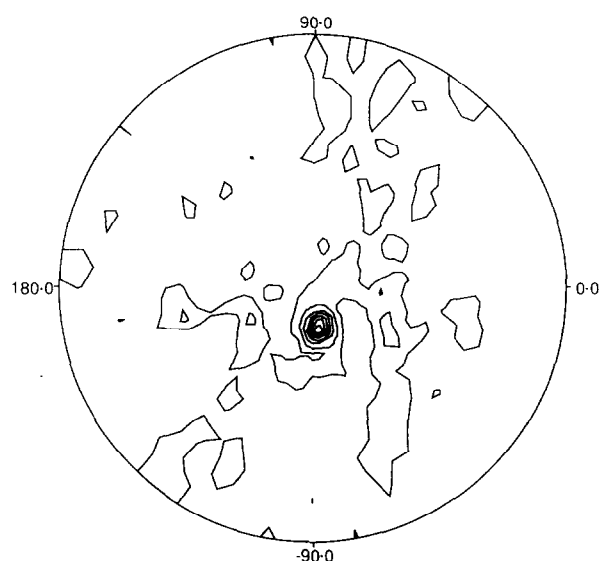


Figure 2. Ovalbumin self-rotation function ($\kappa = 180^\circ$). Resolution 8 to 4 Å; integration radius 20 Å; contoured at $3 \times$ r.m.s. value of map. The peak is in the position of the crystallographic *b* axis.

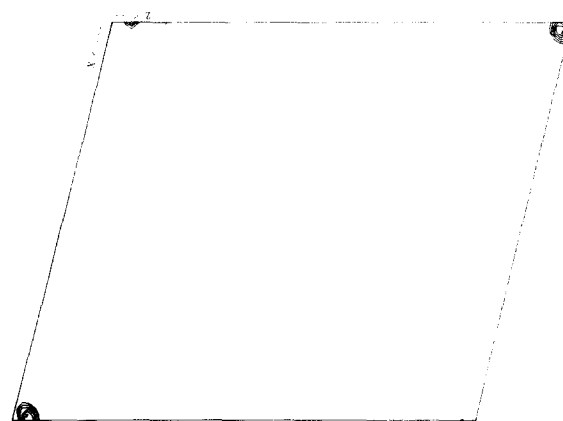


Figure 3. Ovalbumin native Patterson function. Resolution 10 to 4 Å; section 35 of 70 sections through the *b* axis ($x = a$, $z = c$).

The peak height was 80% of the height of the origin peak (3.6 times the r.m.s. value of the map, with the highest noise peak at 1.1 times the r.m.s.). Only a single peak was resolved despite there being 4 molecules in the unit cell. This indicates that the 4 molecules are arranged as 2 similarly orientated pairs, the molecules within each pair being related by an approximate 2-fold symmetry axis parallel to the *b* axis.

A native Patterson function was calculated in the resolution range 10 to 4 Å. The biggest peak, besides the origin peak, represented a vector with a length equal to half that of the *b* axis and a direction that was very close to this axis (fractional co-ordinates 0.024, 0.500, -0.030) (Fig. 3). Only a single peak was resolved. This indicates translational symmetry in the unit cell, consistent with a molecular repeat approximately halfway along the *b* axis (close to a complete lattice translation parallel to the *b* axis in the monoclinic cell). The peak height was only 14% of the height of the origin peak (12.6 times the r.m.s. value of the map), suggesting that the orientations of the molecules related by this *b*/2 translation are slightly different. In the diffraction pattern of the triclinic crystals there is no evidence of the systematic absences (for reflections with *k* odd) that would be expected if the *b*/2 translation were exact.

(e) Molecular replacement with cleaved antitrypsin model

Before the plakalbumin co-ordinates became available, an unsuccessful attempt was made to solve the ovalbumin structure by molecular replacement using cleaved antitrypsin (Brookhaven entry number, 5API) as a model. Cross-rotation functions were calculated, using procedures described in the following section, for several antitrypsin models (all atoms; main-chain and C β atoms only; omitting loops and turns of more than 4 residues; omitting parts of the structure with least sequence homology to ovalbumin). A total of 2 solutions related by the self-rotation matrix were consistently found, with peak heights of approximately 5 times the r.m.s. value of the map. The highest noise peak was approximately 4 times the r.m.s. value. The signal-to-noise ratio was greatest at a resolution of 10 to 6 Å and tended to decrease as the resolution was extended. Translation functions based on these solutions produced no convincing solution.

(f) *Molecular replacement with plakalbumin model*

The orientations and positions of the ovalbumin molecules in the unit cell were successfully determined by Patterson search methods using all atoms of the partially refined model of plakalbumin. The plakalbumin coordinates were kindly given to us by Dr T. Wright and Professor R. Huber.

Normalized structure factor amplitudes were calculated for a model that contained a single plakalbumin molecule in a triclinic cell ($100 \text{ \AA} \times 100 \text{ \AA} \times 100 \text{ \AA}$). The cross-rotation function (Crowther, 1972) was computed using the program ALMN, with data in the resolution range 8 to 4 Å, integrating between radii of 5 Å and 30 Å. An orthogonal co-ordinate system was used for rotation function calculations with a , $c^* \times a$, c^* along x , y , z , respectively. The 2 highest peaks P ($\alpha = 198.5^\circ$, $\beta = 27.4^\circ$, $\gamma = 166.5^\circ$) and Q ($\alpha = 15.6^\circ$, $\beta = 152.3^\circ$, $\gamma = 343.3^\circ$), were 10.6 and 12.2 times the r.m.s. value of the map, respectively, with the highest noise peak at 6.3 times the r.m.s. value (Fig. 4). As expected, the 2 solutions are related to each other by the symmetry operation defined by the self-rotation matrix.

The positions of the molecules in the unit cell were determined by translation searches using the orientations P and Q found in the cross-rotation function. The model comprised 4 plakalbumin molecules (A to D) in the ovalbumin unit cell. Molecule A, in orientation P, was fixed in an arbitrary position defining the origin of the unit cell. Molecule B, related to A by the non-crystallographic 2-fold axis, was generated using orientation Q. The $b/2$ translation vector derived from the native Patterson function was applied to molecules A and B to generate molecules C and D, respectively. The model at this stage did not account for the difference in orientation of the molecules related by the $b/2$ translation (indicated by the size of the peak in the native Patterson function). Partial structure factor amplitudes were calculated for each molecule of the model, then merged and scaled to the observed structure factor amplitudes. A fast Fourier transform based translation function (TFPART), implementing the function of Crowther & Blow (1967), was used to search in turn for the positions of molecules B, C and D, fixing them as they were found. The resolution range was 8 to 4 Å. Self-vectors of the observed Patterson function were omitted by making the assumption that the known and unknown structures have the same self-vector set. Results are shown in Table 2. In the 1st search for molecule B, 2 solution vectors related by the $b/2$ translation appeared as well-defined peaks. Molecule B was translated by the vector corresponding to the highest peak before proceeding to the next search. In the search for molecule C, the highest peak was at the origin, with a smaller $b/2$ related peak, indicating that molecule C was already correctly positioned in the model. The final translational search, for molecule D, had the lowest signal-to-noise ratio for reasons that are not clear. Molecule D was moved by the translation vector corresponding to the highest peak.

(g) *Model building and refinement*

The initial model, of 4 plakalbumin molecules with the orientations found in the cross-rotation function and positions determined by the translation functions, had an R -factor of 49.6% for all data to 6 Å resolution. These positions and orientations were optimized using the program CORELS (Sussman *et al.*, 1977). A total of 3 of the molecules (B, C and D) were refined as individual, constrained ('rigid') groups moving with 6 degrees of

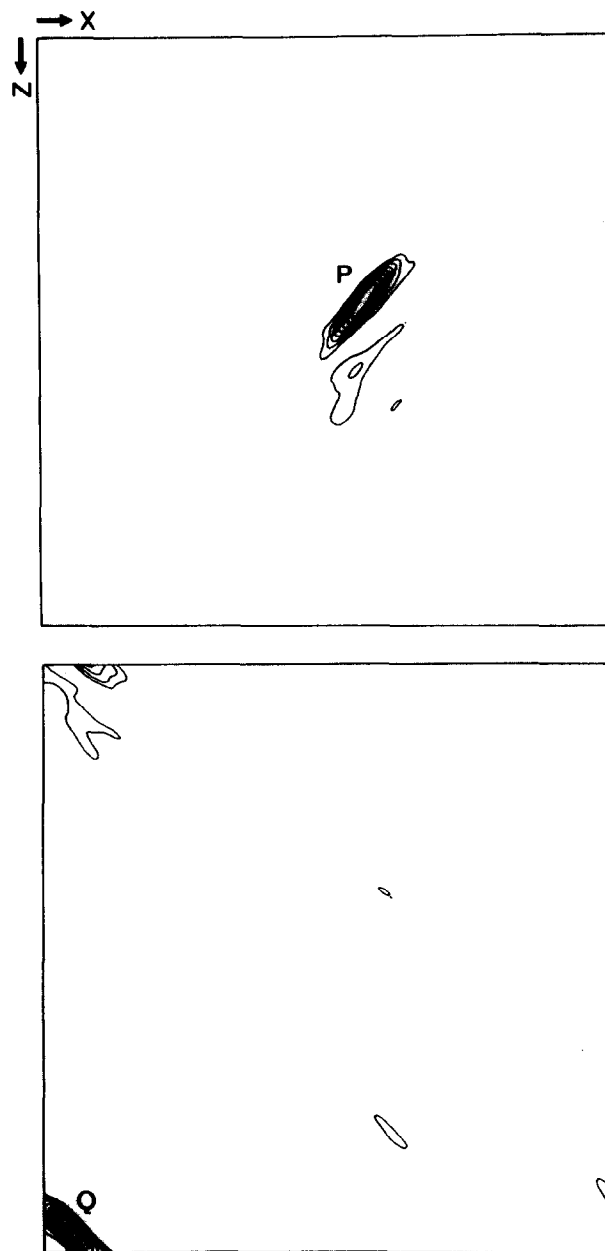


Figure 4. Cross-rotation function using plakalbumin model. Resolution 8 to 4 Å; integration radius 5 to 30 Å; normalized structure factors; contoured at $3 \times$ r.m.s. value of map. The solutions P (section 5) and Q (section 30) are related by the self-rotation matrix.

freedom. The orientation of molecule A was refined, but its position was fixed. Data between 20 Å and 10 Å were used initially and the resolution limit was gradually extended, the R -factor falling to 45.7% for all data to 6 Å.

The overall r.m.s. shift for α -carbon atoms of molecule C was nearly 6 Å, with corresponding shifts of 1 to 2 Å for molecules A, B and D. To improve the model further, each of the 4 molecules was refined as 13 rigid groups, each group broadly corresponding to an element (or group of elements) of secondary structure as indicated from an inspection of the model with FRODO (Jones, 1978). The resolution limit was extended to 5 Å and then to 4 Å, with a final R -factor of 40.0% for all data to 4 Å resolution.

The ovalbumin model after CORELS was refined by simulated annealing with molecular dynamics using

Table 2
Translation functions using plakalbumin as a model

Search	Molecule		Translation vector peaks Fractional co-ordinates			Height ¹	Background ²
	Find	Fix	<i>x</i>	<i>y</i>	<i>z</i>		
1	B	A	0.842	0.070	0.836	7.9	4.6
			0.867	0.572	0.804	6.7	
2	C	A, B	0	0	0	11.0	4.5
			0.958	0.490	0.073	5.7	
3	D	A, B, C	0.773	0.034	0.888	5.7	5.6

¹ Peak height as multiple of r.m.s. of map.

² Biggest noise peak as multiple of r.m.s. of map.

XPLOR (Brunger *et al.*, 1987) implemented on an IRIS (Silicon graphics). The standard procedure described in the manual (Brunger, 1988) was followed, using data between 10 Å and 2.5 Å resolution. In the energy minimization (Jack & Levitt, 1978) prior to molecular dynamics, a soft repulsive potential (the "repel" option), was used to relieve the bad non-bonded contacts at the junctions of CORELS rigid groups. High-temperature molecular dynamics for 2 ps were compared at 3 temperatures (5000, 4000 and 3000 K). A non-crystallographic symmetry restraint (20 kcal/mol Å²) (1 cal = 4.184 J) was applied throughout, restraining the 4 molecules to their average positions after a least-squares superposition (Hendrickson, 1985). After the final energy minimization, the model had poor geometry (r.m.s. deviation from ideal values for bond distances was 0.03 Å and for bond angles was 5°), indicating that the X-ray terms had been weighted too highly in XPLOR. This is consistent with the experience of Weis & Brunger (1989) who find it necessary to reduce the weight during refinement as the model improves. Each XPLOR model was refined to improve its stereochemistry using 10 cycles of the program PROLSQ (Hendrickson & Konnert, 1980).

The XPLOR models resulting from molecular dynamics at the 3 different temperatures had identical *R*-factors (29.5% for all data between 10 Å and 2.5 Å resolution) and stereochemistry. The overall r.m.s. difference for all atoms between equivalent molecules in the 3 models ranged from 0.54 to 0.59 Å. Electron density maps calculated using phases from each model were essentially indistinguishable. The overall r.m.s. atomic shifts during XPLOR, relative to positions in the model after CORELS, were 1.7 Å for all atoms, 1.4 Å for main-chain atoms and 1.9 Å for side-chain atoms (averaged for the 4 molecules). The largest backbone shifts occurred for the sequence preceding helix D (up to 7 Å) and for the region near the site of cleavage in the model structure (up to 11 Å), but these shifts did not correct the model and both regions were manually rebuilt at a later stage of the refinement. The biggest correct shifts occurred mainly during the initial energy minimization before molecular dynamics, including shifts of 4.5 Å for individual α -carbon atoms of helix E.

Subsequent refinement continued with the XPLOR model obtained after molecular dynamics at 4000 K, which was an arbitrary choice. This model, based on the plakalbumin structure, had a missing segment of 6 residues (Ala353 to Ala358) at the site of subtilisin cleavage in each of the 4 molecules. The adjacent residues (Val348 to Glu352 and Ser359) were also deleted from the

model at this stage. The model was improved by 20 rounds of refinement using a modified version of PROLSQ in which structure factors and their derivatives were calculated using a fast Fourier transform based algorithm as implemented in the program DERIV (Jack & Levitt, 1978). The resolution limits for refinement were gradually extended to the limits of the data at 1.9 Å. Individual isotropic atomic temperature factors were refined throughout and a non-crystallographic symmetry restraint was applied.

Each round of refinement with PROLSQ (usually 10 cycles) was followed by manual manipulation of the model using an Evans and Sutherland PS300 interactive graphics display and the program FRODO (Jones, 1978). Model building was based on electron density maps calculated with the Fourier coefficients $(3|F_o| - 2|F_c|)$, α_c and contoured at the r.m.s. value of the electron density in the unit cell. Positive and negative peaks in a difference map, calculated with the Fourier coefficients $(|F_o| - |F_c|)$, α_c and contoured at 3 times the r.m.s. density, were used to locate errors. The largest peaks were identified using the program PEAKMAX and the atoms of the model nearest to each peak were found with ATPEAK. During the initial refinement with 4-fold averaging (see below, section (h)), manual intervention was restricted to molecule A of the model and molecules B, C and D were subsequently generated from A using the symmetry operations determined with SUPERPOSE. The 4 ovalbumin copies were treated independently when unaveraged maps were used. In all, 2 sequences (Phe87 to Val92 and Arg123 to Leu127) in each of the 4 molecules were rebuilt using the "fragment fitting" option of FRODO (Jones & Thirup, 1986). The missing segment of the model around the plakalbumin cleavage site (residues 348 to 359) was modelled in molecules A, B and D at the 11th round of refinement, but in molecule C this sequence is disordered.

Attempts were made to model the post-translational modifications (a single carbohydrate side-chain, 2 phosphoserine groups and an *N*-acetyl group, in each molecule) at a late stage in the refinement. Water molecules in the 1st hydration shell were modelled to account for positive difference map peaks, at least 3 times the r.m.s. value of the map, provided they made at least 1 potential hydrogen bond and were not too close to a carbon atom. They were assigned to 1 of 4 groups (E to H) corresponding to the 4 ovalbumin molecules (A to D). Water molecules with refined temperature factors greater than 40 Å² were deleted from the model at round 16 of the refinement and replaced only if they reappeared as peaks in the difference map after further refinement. A metal ion

was modelled near the end of the refinement to account for an unexplained peak of positive electron density in the difference map (10 times the r.m.s. value).

(h) Molecular averaging

During the initial rounds of restrained least-squares refinement, the electron density maps were improved by molecular averaging. A molecular envelope was determined for each molecule from the atomic model. The symmetry operations relating the 4 molecules were found using the program SUPERPOSE and averaging was performed with SKEWPLANES (Bricogne, 1976). In the early stages of refinement, the averaged maps consistently showed better density than the equivalent unaveraged maps, although a notable exception was the density corresponding to the missing residues 348 to 359, which was better in unaveraged maps. Averaging was not useful beyond the 3rd round of refinement.

3. Ovalbumin Structure

(a) Stereochemistry

The final model is summarized in Table 3. Its stereochemistry is shown in Table 4 and a Ramachandran plot for molecule A in Figure 5. Equivalent plots for molecules B, C and D are almost identical. Nine non-glycine residues have a left-handed α -helical conformation, of which seven residues (Asn47, His93, Asn166, Ser176, Lys236, Ser245a and Asn380) occur at turns in the polypeptide chain and two residues (Arg81 and Asn171) are found in loops. In addition, residue 70 (aspartic acid in ovalbumin) has an unfavourable conformation with a positive ϕ value, as in antitrypsin (Huber & Carrell, 1989) and in α_1 -antichymotrypsin (Baumann *et al.*, 1991).

The variation of main-chain atomic temperature factors along each polypeptide chain is shown in Figure 6. The average temperature factor for all

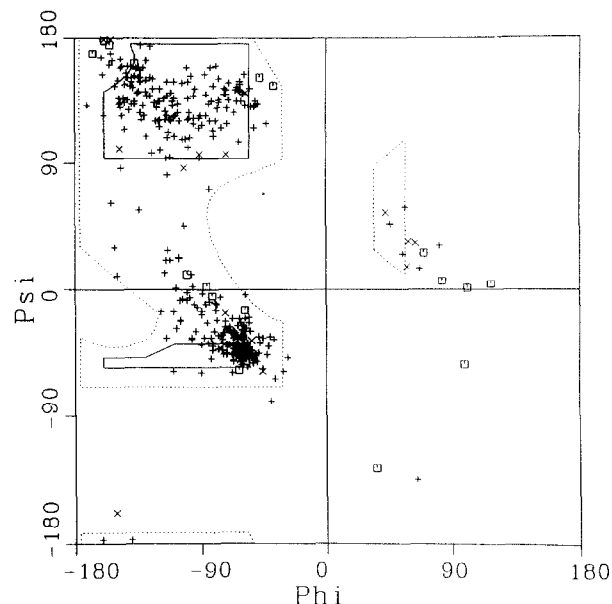


Figure 5. Ramachandran plot of main-chain torsion angles for ovalbumin molecule A. ($\square$) Glycine; ($\times$) asparagine; (+) other residue. The areas enclosed by continuous lines show fully allowed conformations ($C-C-N$ bond angle, τ , 110°) and those enclosed by broken lines indicate acceptable van der Waals' contacts ($C-C-N$ bond angle, τ , 115° ; Ramakrishnan & Ramachandran, 1965).

Table 4

Summary of the restrained least-squares refinement of ovalbumin

	Restraint weight in refinement	Final value
R-factor		17.4%
Resolution range		6 to 1.95 Å
Number of reflections (% of total)		91,865 (94%)
Number of atoms		
protein		11,522
water		678
carbohydrate		56
phosphate		12
metal		1
total		12,269
Number of parameters		49,076
Stereochemistry (r.m.s. deviation from ideal geometry):		
Distances (Å)		
bond	0.02	0.02
angle	0.03	0.04
planar 1 to 4	0.05	0.06
Planes (Å)	0.02	0.01
Chiral volumes (Å ³)	0.15	0.19
Torsion angle (°)		
staggered	15.0	17.3
transverse	20.0	25.5
Van der Waals' contact (Å)		
single torsion	0.20	0.16
multiple torsion	0.20	0.17
Temperature factor (Å ²)		
main-chain bond	2.0	8.3
side-chain bond	3.0	13.2
side-chain angle	4.0	16.3

Table 3
Final ovalbumin model

4 Ovalbumin molecules	
Molecule A	385 amino acid residues
1	N-acetyl group
2	phosphoserine residues
1	N-acetylglucosamine residue
Molecule B	372 amino acid residues (missing 87a to 91, inclusive)
1	N-acetyl group
1	phosphoserine residue
1	N-acetylglucosamine residue
Molecule C	373 amino acid residues (missing 348 to 359, inclusive)
1	N-acetyl group
0	phosphoserine residues
1	N-acetylglucosamine residue
Molecule D	385 amino acid residues
1	N-acetyl group
0	phosphoserine residues
1	N-acetylglucosamine residue
678 Water molecules (groups E to H)	
1 Calcium ion (Ca ²⁺)	

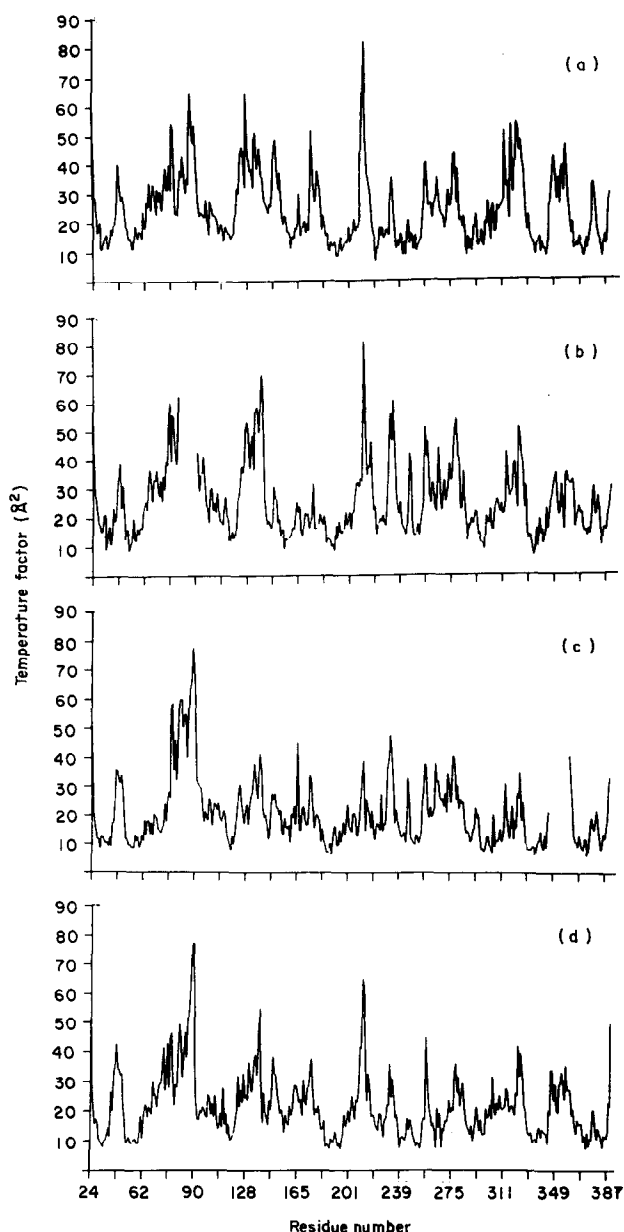


Figure 6. Average main-chain temperature factors for each ovalbumin molecule in the unit cell. (a) Ovalbumin A, (b) ovalbumin B, (c) ovalbumin C and (d) ovalbumin D. Note: Residue numbering is based on sequence alignment with antitrypsin (Huber & Carrell, 1989). Residues 87a to 91 of molecule B and residues 348 to 359 of molecule C are missing in the final model.

main-chain atoms is $23 \text{ \AA}^2$ and for all side-chain atoms is $30 \text{ \AA}^2$. The disulphide bond and the surrounding structure has weak, fragmented electron density with high temperature factors in each molecule and, despite repeated attempts, it could not be modelled in one of the standard configurations described by Richardson (1981).

(b) Description of overall structure

The ovalbumin molecule is approximately ellipsoidal in shape with overall dimensions of $70 \text{ \AA} \times$

Table 5
Ovalbumin secondary structure

A. Helices and β -sheets

Helices		Sheets	
α -Helices	Residues		Residues
A	26 to 44	s1A	142 to 145
B	54 to 65 ¹	s2A	110 to 121
C	70 to 80	s3A	182 to 196
C2	87b to 87e ²	s5A	331 to 340 ⁵
D	94 to 104	s6A	291 to 298
E	128 to 137	s1B	228 to 232
F	150 to 164	s2B	237 to 243
G	260 to 266	s3B	248 to 255
H	269 to 275	s4B	370 to 376
I	299 to 305 ³	s5B	382 to 388
R	350 to 359 ⁴	s6B	50 to 52
		s1C	363 to 365
3_{10} -Helices		s2C	281a to 289
F1	200 to 202	s3C	215 to 227 ⁶
II	310 to 312	s4C	204 to 209 ⁷

B. Turns and β -bulges

Residues	Type	Residues	Type
Reverse turns ($O_i \rightarrow N_{i+3}$ hydrogen bond)			
45 to 48	II	179 to 180	I
82 to 84	I	233 to 236	I
84 to 87	II	244 to 246	II
91 to 94	II	256a to 259	I
121 to 124	III	313 to 316	I
165 to 168	I'	318 to 321	I
174 to 177	II	324 to 327	I

Reverse turns associated with α -helices

65 to 68	I
87e to 87h	I
277 to 280	III

α -Turns ($O_i \rightarrow N_{i+4}$ hydrogen bond)

148 to 150
376 to 380

Classic β -bulge

383, 384, 374 (residues 1, 2 and X respectively)

¹ 54 to 66 in molecule C.

² Disordered in molecule B.

³ 299 to 306 in molecules A and B.

⁴ Disordered in molecule C.

⁵ 329 to 340 in molecule B.

⁶ 214 to 227 in molecule A; 216 to 227 in molecule D.

⁷ 204 to 211 in molecule A.

Ovalbumin secondary structure was assigned by the method of Kabsch & Sander (1983) and labelled according to the analogous structural elements of antitrypsin and plakalbumin. Strand s4A is not present in ovalbumin. The helix in the part of the ovalbumin sequence corresponding to strand s4A of antitrypsin has been labelled helix R.

Turn type was assigned on the basis of $\phi_2 \psi_2$, $\phi_3 \psi_3$ conformation angles (Crawford *et al.*, 1973). The β -bulge is classified according to Richardson (1981).

$45 \text{ \AA} \times 50 \text{ \AA}$ (Fig. 7). Its structure closely resembles that of plakalbumin (Wright *et al.*, 1990) and antitrypsin (Loebermann *et al.*, 1984). Almost all of the polypeptide chain is involved in defined secondary structure (Table 5), with three β -sheets (A to C) and nine α -helices (A to H and helix R). In addition, there are three short helical segments of three to four residues (C2, F1 and II).

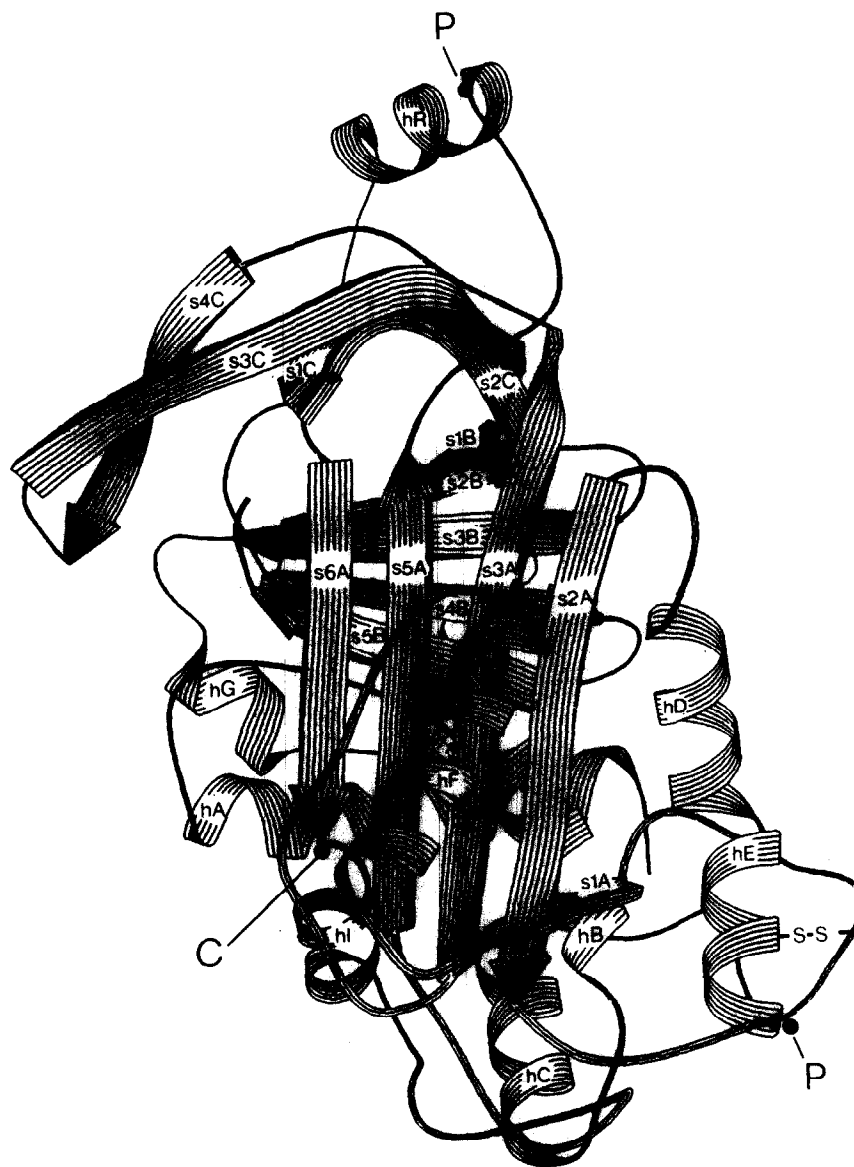


Figure 7. Schematic drawing of the ovalbumin structure produced using the program, RIBBON (Priestle, 1988). Strands of β -sheets are represented by arrows (labelled, s) and α -helices by helical ribbons (labelled, h). Sites of post-translational modification are shown. (C) Carbohydrate side-chain; (P) phosphoserine; (S-S) disulphide bond.

(c) *Disordered residues*

In all, 316 atoms with temperature factors exceeding $80 \text{ \AA}^2$ and with absent or uninterpretable electron density were deleted from the model. Most are side-chain atoms of surface polar residues. Two segments of the polypeptide chain are disordered and have not been modelled. These are Gly87a to Asn91 (13 residues) preceding helix D of molecule B and Val348 to Ser359 (12 residues) at the site of helix R of molecule C. The side-chains of a few buried non-polar residues (Ile87d of molecule A, Leu184 of molecule B, Val134 and Val333 of molecule C) show electron density that suggests two distinct conformations and these were positioned in the conformation corresponding to the stronger density.

(d) *Helix R*

The sequence in ovalbumin analogous to the intact reactive centre loops of the inhibitory serpins, Gly344 to Glu363, has been modelled in molecules A, B and D (Fig. 8). The putative reactive centre (Ala358/Ser359) is found on the final turn of a three-turn α -helix (helix R) formed by Ser350 (phosphorylated) to Ser359. The helix protrudes from the protein core on two peptide stalks. It is joined by the N-terminal stalk (Gly344 to Gly349) to strand s5A and by the C-terminal stalk (Val360 to Glu363) to strand s1C. The N-terminal stalk is anchored to the top of strand s3A by main-chain hydrogen bonds (Arg345 N-O Trp194, Glu346 N-O Glu195 and Glu346 O-N Ala197).

The axis of helix R lies almost perpendicular to

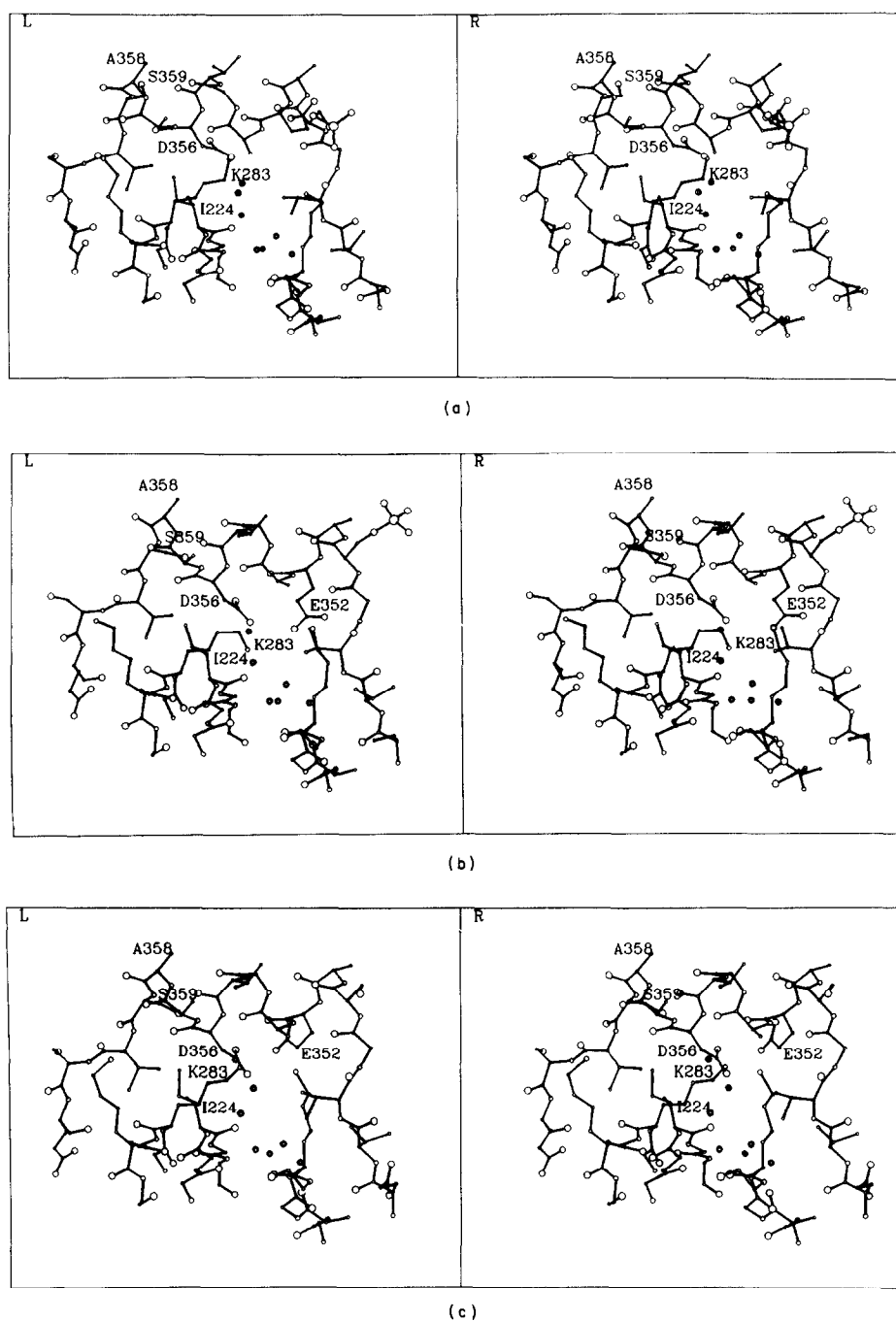


Figure 8. Stereo view of helix R in (a) molecule A, (b) molecule B and (c) molecule D showing its interactions with the protein core and ordered solvent (see Table 6). The reactive centre of ovalbumin by alignment with inhibitory serpins is at Ala358/Ser359.

the strands of the underlying C sheet. The twist in the C sheet separates strand s4C from the N-terminal end of the helix, forming a channel which is occupied by ordered water molecules in the crystal structure. At its C-terminal end, the helix maintains van der Waals' contact with the underlying residues of strands s2C and s3C. The most significant contacts are between Ala353, Asp356 and Ala357 of the helix and Ile224 of strand s3C. Other stabilizing interactions between the helix and the protein core differ in each molecule (Table 6), suggesting that the helix is mobile. In each case, the

helix is packed between two adjacent molecules in the crystal lattice and its position is probably influenced by its environment (Fig. 9). Although the α -helical backbones can be almost exactly superimposed (with an r.m.s. deviation for main-chain atoms of less than 0.25 Å), the position of helix R relative to the protein core differs between molecules (with an r.m.s. deviation for main-chain atoms of 1.7 to 3.3 Å). Consistent with this apparent mobility, the refined temperature factors for residues of this helix are higher than average values for the protein (Fig. 6).

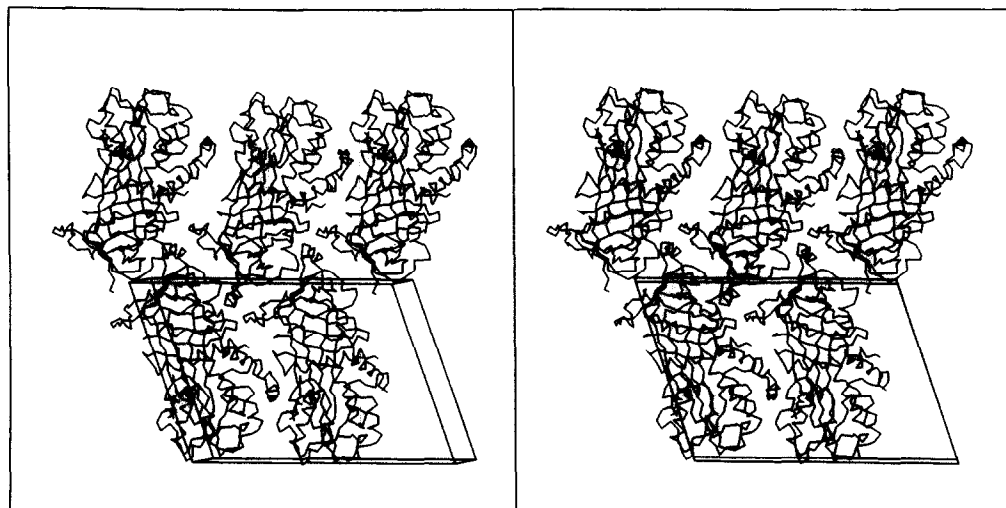


Figure 9. Crystal packing in the triclinic unit cell. The view is down the crystallographic *c* axis. For clarity, only molecules B and D are shown within the outlined unit cell, while molecules A and C are displaced by 1 lattice repeat along the *a* axis. A total of 2 copies of molecule C are shown related by a lattice translation along *b*. The Fig. illustrates the packing of helix R (at the top of molecules B and D and at the bottom of molecule A) against 2 adjacent molecules in the crystal. The glycosylated residue Asn298 and the first *N*-acetylglucosamine residue are also shown. The pseudo *b*/2 repeat within the unit cell, which relates molecules B to D and A to C, is clearly visible.

(e) *Comparison of the four ovalbumin molecules*

The four ovalbumin molecules in the unit cell were superposed using 36 α -carbon atoms of the hydrophobic "core" sequence of helix B and strands s4B, s5B and s6B (Table 7). The r.m.s. deviation in atomic co-ordinates between pairs of molecules is shown in Table 8, and deviations from the average structure as a function of residue number, in Figure 10. The largest shifts involve helix R (1.7 to 3.3 Å). Other marked differences occur in exposed loops, including the sequence preceding helix D (0.6 to 1.0 Å).

(f) *Post-translational modifications*

A single carbohydrate side-chain is covalently linked to the nitrogen amide of Asn298, an external

Table 6
Ion pair interactions stabilizing helix R in ovalbumin

Atoms	Distance (Å)
Molecule A	
Asp356 OD2–NZ Lys283	2.9
Asp356 OD1–O Ile224	via water
Ala357 O–NZ Lys285	2.9
Molecule B	
Glu352 OE1–N Gly349	via water
Glu352 OE2–NZ Lys283	via water
Asp356 OD1–NZ Lys283	via water
Asp356 OD1–O Ile224	via water
Molecule D	
Glu352 OE1–NZ Lys283	2.6
Asp356 OD1–NZ Lys283	2.6
Asp356 OD2–O Ile224	via water
Ala357 O–NZ Lys285	via water

Table 7
Transformations relating the four ovalbumin molecules

B→A			
Rotation =	–0.74109	–0.67004	0.04278
	–0.67054	0.73540	–0.09780
	0.03407	–0.10116	–0.99429
Translation =	35.49	18.48	80.60
C→A			
Rotation =	0.99403	–0.06828	–0.08507
	0.07993	0.98662	0.14209
	0.07423	–0.14804	0.98619
Translation =	23.07	–42.91	9.15
D→A			
Rotation =	–0.81974	–0.57173	0.03395
	–0.57273	0.81800	–0.05340
	0.00276	–0.06322	–0.99800
Translation =	43.03	–26.65	84.39
C→B			
Rotation =	–0.78775	–0.61599	0.00134
	–0.61475	0.78630	0.06169
	–0.03906	0.04777	–0.99809
Translation =	47.94	–29.60	76.52
D→B			
Rotation =	0.99165	–0.12686	–0.02319
	0.12770	0.99103	0.03933
	0.01800	–0.04196	0.99896
Translation =	24.79	–38.65	0.99
D→C			
Rotation =	–0.86039	–0.50766	–0.04482
	–0.50956	0.85541	0.09287
	–0.00881	0.10274	–0.99467
Translation =	26.75	3.54	74.80

B→A means the transformation applied to molecule B to superpose it on molecule A.

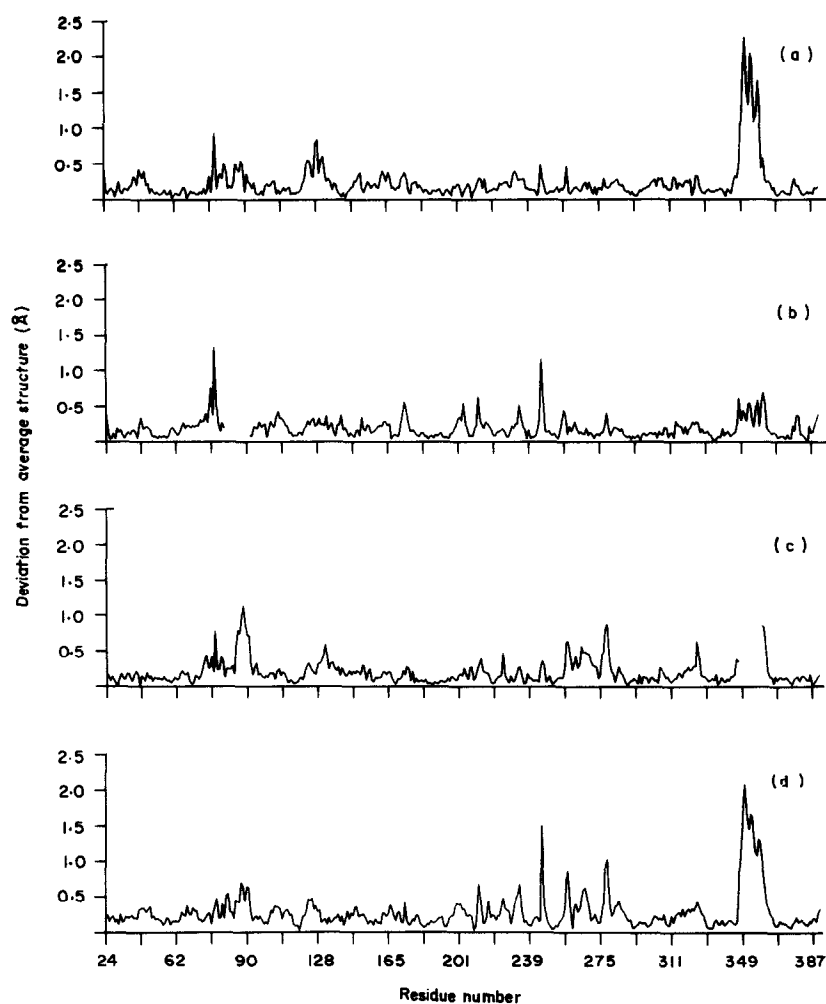


Figure 10. Deviation in α -carbon positions of each ovalbumin molecule in the unit cell from the average structure. (a) Ovalbumin A, (b) ovalbumin B, (c) ovalbumin C and (d) ovalbumin D. Residue numbering is based on sequence alignment with antitrypsin (Huber & Carrell, 1989). Residues 87a to 91 of molecule B and residues 348 to 359 of molecule C are missing in the final model.

site near the end of strand s6A. The ovalbumin molecules in the crystal show electron density only for one or two *N*-acetylglucosamine groups of the core structure. The first of these residues has been modelled in each molecule (Fig. 11).

Electron density for the phosphate groups of the phosphoserine molecules is weak. The phosphate at Ser87c, in the sequence preceding helix D, has been modelled only in molecule A and the phosphate at

Ser350 (on helix R), in molecules A and B. The latter site is a lattice contact, which may explain why diphosphorylated ovalbumin crystallized more successfully than the crude preparation containing a mixture of forms with two, one or zero phosphate groups per ovalbumin molecule.

The acetyl group at the N-terminal glycine has been modelled in each molecule.

(g) Solvent

A total of 678 water molecules have been modelled in the unit cell with multiplicity and temperature factor distribution as shown in Table 9. There is no evidence (from temperature factors or patterns of co-ordination) that any of the solvent sites are occupied by molecules other than water.

There is an extensive interface between pairs of molecules related by the pseudo twofold axis parallel to the crystallographic *b* axis (molecules A/B and C/D) involving residues from strands s1A,

Table 8
r.m.s. deviation of atomic co-ordinates between pairs of ovalbumin molecules in the unit cell

	A	B	C	D
A		0.5	0.4	0.8
B	1.0		0.4	0.6
C	1.1	1.0		0.5
D	0.9	1.0	1.0	

Main-chain, above diagonal; side-chain, below diagonal.

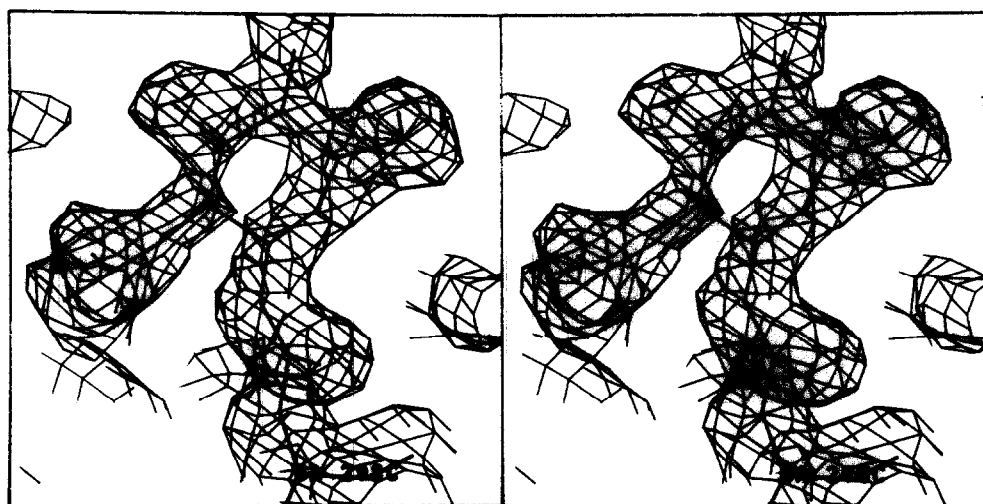


Figure 11. Stereo view of Asn298 and the first *N*-acetylglucosamine residue of the carbohydrate side-chain (molecule C). The electron density map was calculated with Fourier coefficients $(3|F_o| - 2|F_c|)$, α_c and contoured at the r.m.s. electron density in the unit cell.

s2A, helix F and the loop linking helix F to strand s3A (Fig. 12). Within this interface there are numerous hydrogen bonds between the molecules mediated by one or two bridging water molecules. In fact, this interface contains by far the greatest density of ordered water, and also includes the largest networks of hydrogen bonded water molecules (involving up to seven distinct sites).

(h) Metal ion

The model includes a single metal ion co-ordinated by side-chains of two adjacent molecules (Table 10). The metal was probably introduced as a contaminant during crystallization, but its chemical identity is uncertain. The tetrahedral co-ordination geometry would be unusual for calcium or magnesium and the nature of the co-ordinating ligands makes it unlikely to be copper or zinc. The site has been modelled as a calcium ion (Ca^{2+}) with an occupancy of 0.5 and a refined temperature factor of $32 \text{ \AA}^2$.

(i) Comparison of ovalbumin and plakalbumin

The structure of plakalbumin† was superposed in turn on each ovalbumin molecule in the unit cell, using 36 α -carbon atoms of the core sequence of helix B and strands s4B, s5B and s6B (Table 11, Fig. 13). The overall r.m.s. separation for equivalent main-chain atoms ranges from 1.3 to 2.2 $\text{\AA}$. Differences are greatest in the part of the structure surrounding the protease cleavage site. The missing segment of six amino acid residues (353 to 358) in plakalbumin forms part of helix R in ovalbumin. In the ovalbumin structure, main-chain hydrogen bonds are present between residues 345 and 346 (the base of the N-terminal stalk of helix R) and residues 194 to 197 (the top of strand s3A). In plakalbumin, on the other hand, residues 345 to 352 are displaced to form a site of contact between the two molecules in the asymmetric unit. The distortion of the A

† The co-ordinates were supplied by Dr T. Wright and Professor R. Huber for 1 of the 2 plakalbumin molecules in the asymmetric unit.

Table 9
Water molecules

A. <i>Equivalence of sites in relation to protein structure</i>								
Multiplicity ¹	1	2	3	4				
Number	177	136	93	272				
B. <i>Temperature factor distribution</i>								
Temperature factor (Å²)	0 to 10	10 to 20	20 to 30	30 to 40	40 to 50	50 to 60	60 to 70	70 to 80
Number	1	56	149	180	166	96	25	5

¹ Multiplicity n means that the same site (within 1.5 $\text{\AA}$) is observed in n of the 4 ovalbumin molecules in the unit cell.

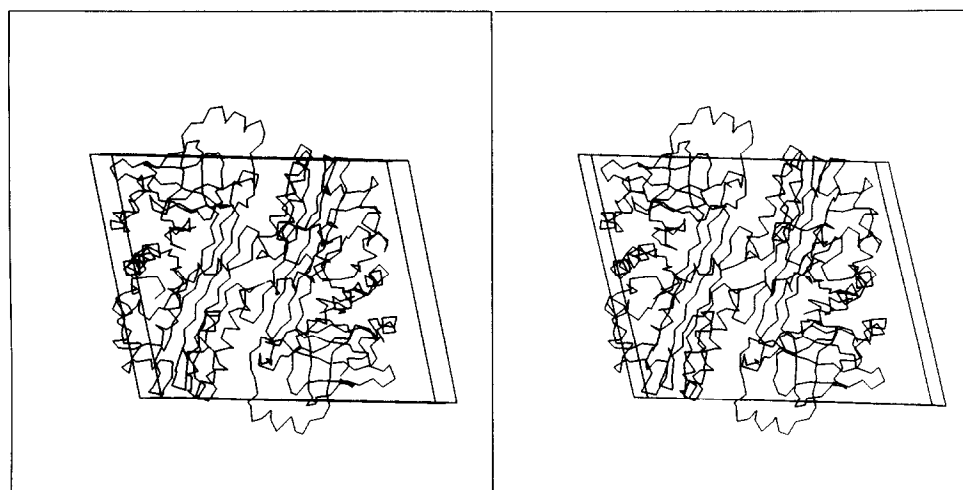


Figure 12. Stereo view of the packing of molecules A and B within the unit cell. The view is down the crystallographic *b* axis, and illustrates the pseudo 2-fold axis that relates molecule A to B. Ordered water molecules play a major role in stabilizing the extensive interface between these 2 molecules. The interface between molecules C and D is very similar to that shown here.

sheet between strands s3A and s5A is similar in both structures.

Other differences between the ovalbumin and plakalbumin crystal structures are unlikely to be a consequence of the cleavage and probably result from disorder, displacements caused by crystal packing and errors in both structures. The plakalbumin structure was determined to a lower resolution (2.8 Å) than ovalbumin (1.95 Å) and refinement was not complete when the co-ordinates were given to us. A 22-residue segment between helices C and D (Phe82 to His93) differs significantly between plakalbumin and ovalbumin, although the electron density for this sequence is weak in both structures.

(j) *Comparison of ovalbumin and antitrypsin*

The structure of cleaved antitrypsin was superposed in turn on each ovalbumin molecule in the unit cell, using 36 α -carbon atoms of the core sequence of helix B and strands s4B, s5B and s6B (Table 12, Fig. 14). The r.m.s. deviation for these atoms between pairs of molecules ranges from 0.65 to 0.70 Å.

The greatest difference between intact ovalbumin and cleaved antitrypsin is in the arrangement of the A sheet. In ovalbumin, the A sheet has only five strands, compared with six strands in antitrypsin. The sequence in ovalbumin corresponding to strand

s4A of antitrypsin (residues 343 to 358) is exposed, forming a new α -helix of three turns (helix R) with short peptide connections to the protein core. In antitrypsin, residues 342 and 343 are thought to act as a "hinge" for the insertion of strand s4A into the A sheet when the reactive centre sequence is cleaved. The A sheet in ovalbumin is distorted near this hinge, with disruption of β -sheet hydrogen bonds between strands s3A and s5A. A hydrogen bond is formed between His337 O–N Phe190, but not between Glu339 N–O Phe190, Glu339 O–N

Table 11

Transformations relating plakalbumin to ovalbumin (molecules A to D)

Plakalbumin→ovalbumin mol A			
Rotation =	0.88758	−0.14129	−0.43844
	0.09857	0.98801	−0.11886
	0.44998	0.06228	−0.89086
Translation =			
	54.91	30.19	20.81
Plakalbumin→ovalbumin mol B			
Rotation =	−0.70843	−0.55574	0.43505
	−0.56765	0.81496	0.11668
	−0.41939	−0.16430	−0.89281
Translation =			
	−24.28	1.64	59.12
Plakalbumin→ovalbumin mol C			
Rotation =	0.92347	−0.05681	−0.37943
	−0.02998	0.97527	−0.21899
	0.38249	0.21361	−0.89893
Translation =			
	38.36	68.22	19.19
Plakalbumin→ovalbumin mol D			
Rotation =	−0.78286	−0.44993	0.42976
	−0.45545	0.88498	0.09687
	−0.42391	−0.11990	−0.89773
Translation =			
	−42.47	43.72	60.84

Plakalbumin→ovalbumin mol A means the transformation applied to plakalbumin to superpose it on molecule A.

Table 10
Co-ordination of metal ion

Ovalbumin molecule	Residue	Atom	Distance (Å)
A	Phosphoserine 350	O6	1.6
D	Glu281b	OE2	1.9
A	Glu201	OE1	1.9
	Water (145 chain F)	O	2.2



Figure 13. A superposition of the α -carbon backbones of ovalbumin (thick bonds) and plakalbumin (thin bonds) achieved by optimal superposition of helix B and strands s4B, s5B and s6B in the 2 structures. The putative reactive centre of ovalbumin is indicated, P1.



Figure 14. A superposition of the α -carbon backbones of ovalbumin (thick bonds) and antitrypsin (thin bonds) achieved by optimal superposition of helix B and strands s4B, s5B and s6B in the 2 structures. The putative reactive centre of ovalbumin is indicated, P1.

Table 12
Transformations relating antitrypsin to ovalbumin
(molecules A to D)

Antitrypsin→ovalbumin mol A			
Rotation =	-0.97968	-0.05451	-0.19304
	-0.14229	0.86717	0.47725
	0.14139	0.49502	-0.85730
Translation =	103.30	-15.21	58.66
Antitrypsin→ovalbumin mol B			
Rotation =	0.82623	-0.52417	-0.20639
	0.53765	0.62436	0.56666
	-0.16817	-0.57915	0.79769
Translation =	-28.39	-67.98	27.97
Antitrypsin→ovalbumin mol C			
Rotation =	-0.97478	0.05180	-0.21709
	-0.09429	0.78605	0.61093
	0.20229	0.61598	-0.76134
Translation =	85.63	14.51	45.97
Antitrypsin→ovalbumin mol D			
Rotation =	0.88499	-0.45039	-0.11804
	0.43506	0.70961	0.55424
	-0.16586	-0.54186	0.82394
Translation =	-55.99	-23.45	27.12

Antitrypsin→ovalbumin mol A means the transformation applied to antitrypsin to superpose it on molecule A.

Gly192 or Asn341 N-O Gly192. Instead, these atoms are bonded to four water molecules lying between the strands. Other differences between ovalbumin and antitrypsin agree with those described by Wright *et al.* (1990) and Engh *et al.* (1990), based on a comparison of the plakalbumin and antitrypsin structures.

4. Discussion

The structure of uncleaved ovalbumin, in agreement with that of plakalbumin (Wright *et al.*, 1990), provides evidence for the dramatic tertiary structure transformation that follows reactive centre cleavage of most serpins, in which the new C terminus generated by the cleavage is inserted in the middle of β -sheet A forming an additional strand as proposed by Loebermann *et al.*, 1984. The mechanism of this tertiary structure transformation and the conformational changes that allow it to take place have been examined by comparing the atomic structure of antitrypsin with the structures of plakalbumin (Engh *et al.*, 1990) and uncleaved ovalbumin (Stein & Chothia, 1991).

The feature of greatest functional interest in the ovalbumin structure is the α -helical conformation of the intact sequence homologous to the reactive centre loops of the inhibitory members of the serpin family (Stein *et al.*, 1990). The overall structural similarity between ovalbumin and antitrypsin and the sequence conservation within the family (Huber & Carrell, 1989) suggest that the ability to adopt this helical form may be common to all serpins, although it would be an unexpected feature of a protease inhibitor.

Crystal structures of the complexes formed between non-serpin inhibitors and their target proteases consistently show that at least eight residues of the reactive centre loop of the inhibitor have an extended conformation (Bode *et al.*, 1986; McPhalen & James, 1987). Comparison of structures of free and complexed inhibitors indicate that this conformation remains relatively fixed, with only small changes to facilitate close packing on complex formation (Hirono *et al.*, 1984). A helical reactive centre could not bind to a protease active site without unfolding, requiring a significant conformational change that is atypical of protein-protein recognition.

Lack of inhibitory activity in ovalbumin might be a consequence of the helix at the putative reactive centre. On the other hand, this part of the structure is able to conform to the active site of a protease, since cleavage sites of pancreatic elastase (Wright, 1984) and subtilisin (Linderstrøm-Lang, 1952) lie within the helix. Such limited proteolysis of a globular protein typically occurs at exposed surface loops and not within regions of defined secondary structure (Fontana, 1989). The susceptibility of helix R to cleavage and evidence of its mobility in the crystal structure (section 3(d)) suggest that this helix must unfold in some environments, although the mechanism for unfolding a helix that is fixed at both ends is unclear.

The sequence in ovalbumin that forms the peptide stalk at the N-terminal side of helix R (residues 344 to 350) differs from the aligned sequences of the inhibitory serpins at the equivalent position (Huber & Carrell, 1989). In the inhibitory serpins, conservation of small hydrophobic residues, particularly Ala and Gly, suggests that this segment is flexible. In contrast, ovalbumin has a large polar Arg at residue 345 and Val at residues 347 and 348 that would restrict movement. Similarly, angiotensinogen, another non-inhibitory serpin, has Arg at 345 and Glu at 349. In inhibitory serpins, natural mutations of the conserved alanines at residues 347 or 349 (12 or 10 residues, respectively, on the N-terminal side of the P_1P_1' reactive centre) result in loss of function, although the reactive centre of the mutant protein can still be cleaved by proteases (Perry *et al.*, 1989; Levy *et al.*, 1990). These mutant serpins resemble ovalbumin in their ability to interact with proteases as substrates but not as inhibitors, implying that the stalk sequence on the N-terminal side of the reactive centre is a crucial determinant of inhibitory function. Small hydrophobic residues may be necessary to allow a conformational change that fixes the reactive centre peptide bond so that it can interact optimally with a target protease. Such a conformational change is likely to involve partial folding of the stalk sequence into sheet A between strands s3A and s5A, resembling an incomplete form of the full tertiary structure transformation that typically takes place following reactive centre cleavage (Loebermann *et al.*, 1984).

The ovalbumin model provides an explanation as

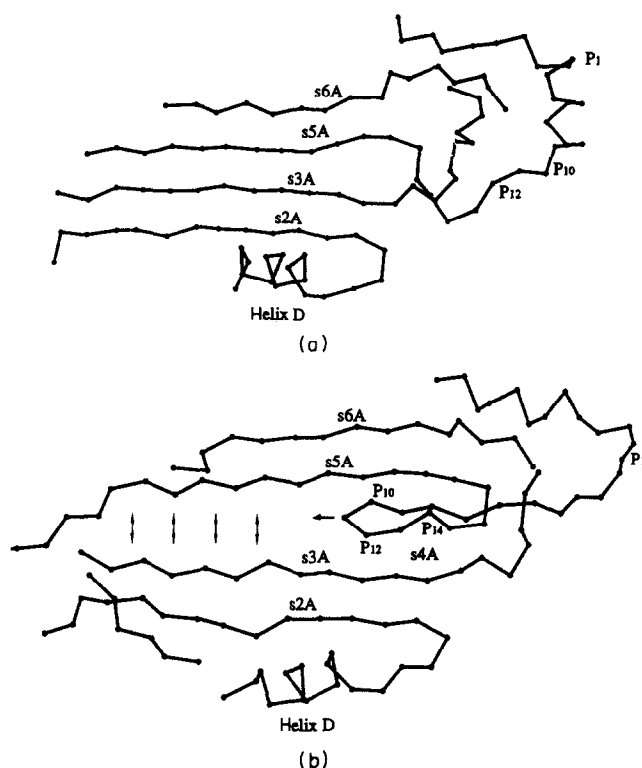


Figure 15. Schematic diagram of the reactive centre loop. (a) A-sheet and helical reactive centre of ovalbumin and (b) the reactive centre of the human serpin antithrombin illustrating the proposed folding of the N-terminal stalk back into the A sheet to give the active inhibitory conformation (Perry *et al.*, 1991).

to why the serpin reactive centre may differ from the fixed extended reactive centre loops of other known protease inhibitors. We have evidence (unpublished results) that the reactive centre in human plasma serpins is mobile and able to adopt various conformations, depending on the environment of the inhibitor. This will enable the reactive centre peptide loop to take up either the relatively protected but inactive helical conformation or to partially fold back into the A sheet to give the exposed active conformation (Fig. 15). A helical reactive centre, as in the ovalbumin crystal structure, could not bind to a protease, but this form would be advantageous in a situation where inhibition was not required, as the reactive centre would be relatively protected from inadvertent proteolytic cleavage with irreversible loss of function. Unfolding of the helix to form an extended reactive centre loop would allow it to interact readily with proteases, either as a substrate or as an inhibitor. For protease inhibition, an associated refolding of the sequence on the N-terminal side of the reactive centre, with partial re-entry into the A sheet would be necessary to fix the optimal conformation of the P_1P_1' peptide bond for tight binding to the target protease. This requirement for partial movement back into the A sheet would explain the lack of inhibitory function in ovalbumin, angiotensinogen and certain mutant serpins; each of these has

changes which will hinder folding in the stalk region and leave the reactive centre in an open conformation that acts as a protease substrate but not as an inhibitor.

We are grateful to Dr T. Wright and Professor R. Huber for the plakalbumin co-ordinates. We also thank Dr I. Tickle for the translation program TFPART; W. Turnell for advice on crystallization; Dr K. Wilson and colleagues at EMBL, Hamburg and Dr P. McLaughlin for help with data collection. This work was supported by the Medical Research Council and the Wellcome Trust. The atomic co-ordinates for the ovalbumin structure have been deposited with the Brookhaven Protein Databank (entry number, 1OVA).

References

- Arndt, U. W. & Gilmore, D. J. (1979). X-ray television area detectors for macromolecular structural studies with synchrotron radiation sources. *J. Appl. Crystallogr.* **12**, 1–9.
- Atkinson, P. H., Grey, A., Carver, J. P., Hakimi, J. & Ceccarini, C. (1981). Demonstration of heterogeneity of chick ovalbumin glycopeptides using 360-MHz proton magnetic resonance spectroscopy. *Biochemistry*, **20**, 3979–3986.
- Baumann, U., Huber, R., Bode, W., Grosse, D., Lesjak, M. & Laurell, C. B. (1991). Crystal structure of cleaved α_1 -antichymotrypsin at 2.7 Å resolution and its comparison with other serpins. *J. Mol. Biol.* **218**, 595–606.
- Bode, W., Papamokos, E., Musil, D., Seemueller, U. & Fritz, H. (1986). Refined 1.2 Å crystal structure of the complex formed between subtilisin Carlsberg and the inhibitor eglin c. Molecular structure of eglin and its detailed interaction with subtilisin. *EMBO J.* **5**, 813–818.
- Bricogne, G. (1976). Methods and programs for direct-space exploitation of geometric redundancies. *Acta Crystallogr. sect. A*, **32**, 832–847.
- Bruch, M., Weiss, V. & Engel, J. (1988). Plasma serine proteinase inhibitors (serpins) exhibit major conformational changes and a large increase in conformational stability on cleavage at their reactive sites. *J. Biol. Chem.* **263**, 16626–16630.
- Brunger, A. T. (1988). Xplor manual. Yale University.
- Brunger, A. T., Kuriyan, J. & Karplus, M. (1987). Crystallographic *R* factor refinement by molecular dynamics. *Science*, **235**, 458–460.
- Cann, J. R. (1949). Electrophoretic analysis of ovalbumin. *J. Amer. Chem. Soc.* **71**, 907–909.
- Carrell, R. W. & Boswell, D. R. (1986). Serpins: the superfamily of plasma serine proteinase inhibitors. In *Protease Inhibitors* (Barrett, A. & Salvesen, G., eds), pp. 403–420, Elsevier, Amsterdam.
- Carrell, R. W. & Owen, M. C. (1985). Plakalbumin, α_1 -antitrypsin, antithrombin and the mechanism of inflammatory thrombosis. *Nature (London)*, **317**, 730–732.
- Carrell, R. W. & Travis, J. (1985). α_1 -Antitrypsin and the serpins: variation and countervariation. *Trends Biochem. Sci.* **10**, 20–24.
- Crawford, J. L., Lipscomb, W. N. & Schellman, C. G. (1973). The reverse turn as a polypeptide conformation in globular proteins. *Proc. Nat. Acad. Sci., U.S.A.* **70**, 538–542.

- Crowther, R. A. (1972). The fast rotation function. In *The Molecular Replacement Method. A Collection of Papers on the Use of Non-crystallographic Symmetry* (Rossmann, M. G., ed.), pp. 173–178, Gordon and Breach, New York.
- Crowther, R. A. & Blow, D. M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallogr.* **23**, 544–548.
- Engh, R. A., Wright, H. T. & Huber, R. (1990). Modeling the intact form of the α_1 -proteinase inhibitor. *Protein Eng.* **3**, 469–477.
- Fontana, A. (1989). Limited proteolysis of globular proteins occurs at exposed and flexible loops. In *Highlights of Modern Biochemistry* (Kotyk, A., et al., eds), vol. 2, pp. 1711–1726, VSP International Science Publishers, Zeist, The Netherlands.
- Gettins, P. (1989). Absence of large scale conformational change upon limited proteolysis of ovalbumin, the prototypic serpin. *J. Biol. Chem.* **264**, 3781–3785.
- Gettins, P. & Harten, B. (1988). Properties of thrombin and elastase modified human antithrombin III. *Biochemistry*, **27**, 3634–3639.
- Goux, W. J. & Venkatasubramanian, P. N. (1986). Metal ion binding properties of hen ovalbumin and S-ovalbumin: characterisation of the metal ion binding site by ^{31}P NMR and water proton relaxation rate enhancements. *Biochemistry*, **25**, 84–94.
- Hendrickson, W. A. (1985). Stereochemically restrained refinement of macromolecular structures. *Methods Enzymol.* **115**, 252–270.
- Hendrickson, W. A. & Konnert, J. H. (1980). Incorporation of stereochemical information into crystallographic refinement. In *Computing in Crystallography* (Diamond, R., Ramesh, S. & Venkatesan, K., eds), pp. 13.01–13.2, Indian Academy of Sciences, Bangalore.
- Hirono, S., Akagawa, H., Mitsui, Y. & Itaka, Y. (1984). Crystal structure at 2.6 Å resolution of a complex of subtilisin BPN with *Streptomyces* subtilisin inhibitor. *J. Mol. Biol.* **178**, 389–413.
- Hopkins, F. G. & Pinkus, S. N. (1898). Observations on the crystallisation of animal proteins. *J. Physiol. Lond.* **23**, 130–136.
- Huber, R. & Carrell, R. W. (1989). Implications of the 3-dimensional structure of α_1 -antitrypsin for the structure and function of serpins. *Biochemistry*, **28**, 8951–8965.
- Hunt, L. T. & Dayhoff, M. O. (1980). A surprising new protein superfamily containing ovalbumin, antithrombin III and α_1 -proteinase inhibitor. *Biochem. Biophys. Res. Commun.* **95**, 864–871.
- Ishihara, H., Takahashi, N., Ito, J., Takeuchi, E. & Teijima, S. (1981). Either high-mannose-type of hybrid-type oligosaccharide is linked to the same asparagine residue in ovalbumin. *Biochim. Biophys. Acta*, **669**, 216–221.
- Jack, A. & Levitt, M. (1978). Refinement of large structures by simultaneous minimisation of energy and R-factor. *Acta Crystallogr. sect. A*, **34**, 931–935.
- Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272.
- Jones, A. T. & Thirup, S. (1986). Using known substructures in protein model building and crystallography. *EMBO J.* **5**, 819–827.
- Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
- Levy, N. J., Ramesh, N., Circardi, M., Harrison, R. A. & Davis, A. E. (1990). Type II hereditary angioneurotic edema that may result from a single nucleotide change in the codon for alanine 436 in the inhibitor gene. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 265–268.
- Linderstrøm-Lang, K. (1952). Formation of plakalbumin from ovalbumin. *Med. Sci.* **6**, 73–92.
- Linderstrøm-Lang, K. & Ottesen, M. (1949). A new protein from ovalbumin. *Compt. rend. trav. Lab. Carlsberg Ser. Chim.* **26**, 403–442.
- Loebermann, H., Tokuko, R., Deisenhofer, J. & Huber, R. (1984). Human α_1 -proteinase inhibitor. Crystal structure analysis of two crystal modifications, molecular model and preliminary analysis of the implications for function. *J. Mol. Biol.* **177**, 531–556.
- Long, W. F. & Williamson, F. B. (1980). Ovalbumin, a protein possessing sequence homologies with antithrombin III and α_1 -antitrypsin, lacks antithrombin and anti-Xa activities. *I.R.C.S. Med. Sci.* **8**, 808.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
- McPhalen, C. A. & James, M. N. G. (1987). Crystal structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry*, **26**, 261–269.
- McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M. & Brownlee, G. G. (1978). Sequence of chicken ovalbumin mRNA. *Nature (London)*, **273**, 723–728.
- Miller, M., Weinstein, J. N. & Wlodawer, A. (1983). Preliminary X-ray analysis of single crystals of ovalbumin and plakalbumin. *J. Biol. Chem.* **258**, 5864–5866.
- Milstein, C. P. (1968). An application of diagonal electrophoresis to the selective purification of serine phosphate peptides. *Biochem. J.* **110**, 127–134.
- Narasimhan, S., Harpaz, N., Longmore, G., Carver, J. P., Grey, A. P. & Schachter, H. (1980). The purification by preparative high voltage paper electrophoresis in borate of glycopeptides containing high mannose and complex oligosaccharide chains linked to asparagine. *J. Biol. Chem.* **255**, 4876–4884.
- Narita, K. & Ishii, J. (1962). N-terminal sequence in ovalbumin. *J. Biochem. (Tokyo)*, **52**, 367–373.
- Nisbet, A. D., Saundry, R. H., Moir, A. J. G., Fothergill, L. A. & Fothergill, J. E. (1981). The complete amino acid sequence of hen ovalbumin. *Eur. J. Biochem.* **115**, 335–345.
- Nyborg, J. & Wonacott, A. J. (1977). Computer programs. In *The Rotation Method in Crystallography* (Arndt, U. W. & Wonacott, A. J., eds), pp. 139–152, North-Holland, Amsterdam.
- Ødum, L. (1987). Trypsin-inhibitory activity of ovalbumin preparations is due to ovomucoid. *Biol. Chem. Hoppe-Seyler*, **368**, 1603–1606.
- Pemberton, P. A., Stein, P. E., Pepys, M. B., Potter, J. M. & Carrell, R. W. (1988). Hormone binding globulins undergo serpin conformational change in inflammation. *Nature (London)*, **336**, 257–258.
- Pemberton, P. A., Harrison, R. A., Lachman, P. J. & Carrell, R. W. (1989). The structural basis for neutrophil inactivation of C1-inhibitor. *Biochem. J.* **258**, 193–198.
- Perlmann, G. E. (1952). Enzymatic dephosphorylation of ovalbumin and plakalbumin. *J. Gen. Physiol.* **35**, 711–726.
- Perry, D. J., Harper, P. L., Fairham, S., Daly, M. & Carrell, R. W. (1989). Antithrombin Cambridge, 384 Ala to Pro: a new variant identified using the polymerase chain reaction. *FEBS Letters*, **254**, 174–176.

- Perry, D. J., Daly, M., Harper, P. L., Tait, R. C., Price, J., Walker, I. D. & Carrell, R. W. (1991). Antithrombin Cambridge II, 384 Ala to Ser: further evidence of the role of the reactive centre loop in the inhibitory function of the serpins. *FEBS Letters*, **285**, 248–250.
- Priestle, J. P. (1988). RIBBON: a stereo cartoon drawing program for proteins. *J. Appl. Crystallogr.* **21**, 572–576.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.
- Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.
- Robinson, A., Meredith, C. & Austen, B. M. (1986). Isolation and properties of the signal region from ovalbumin. *FEBS Letters*, **203**, 243–246.
- Saito, Z. & Martin, W. G. (1966). Ovalbumin and other water-soluble proteins in avian yolk during embryogenesis. *Canad. J. Biochem.* **44**, 293–301.
- Satake, K., Sasakura, S. & Honda, S. (1965). Formation of plakalbumin from ovalbumin under the action of plant proteases. *J. Biochem. (Tokyo)*, **58**, 305–307.
- Shepherd, V. & Montgomery, R. (1978). Interaction of ovalbumin and its asparaginyl-carbohydrate fractions with concanavalin A. *Biochim. Biophys. Acta*, **535**, 356–369.
- Smith, M. B. & Back, J. F. (1962). Modification of ovalbumin in stored eggs detected by heat denaturation. *Nature (London)*, **193**, 878–879.
- Smith, M. B. & Back, J. F. (1965). Studies on ovalbumin. II. The formation and properties of S-ovalbumin, a more stable form of ovalbumin. *Aust. J. Biol. Sci.* **18**, 365–377.
- Stein, P. E. (1990). Structure and function of non-inhibitory serpins. Ph.D. thesis, University of Cambridge.
- Stein, P. E. & Chothia, C. (1991). The Serpin tertiary structure transformation. *J. Mol. Biol.* **221**, 615–621.
- Stein, P. E., Tewksbury, D. A. & Carrell, R. W. (1989). Ovalbumin and angiotensinogen lack serpin S-R conformational change. *Biochem. J.* **262**, 103–107.
- Stein, P. E., Leslie, A. G. W., Finch, J. T., Turnell, W. G., McLaughlin, P. J. & Carrell, R. W. (1990). Crystal structure of ovalbumin as a model for the reactive centre of serpins. *Nature (London)*, **347**, 99–102.
- Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. (1977). A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. *Acta Crystallogr. sect. A*, **33**, 800–804.
- Taborsky, G. (1974). Phosphoproteins. *Advan. Protein Chem.* **28**, 1–210.
- Tai, T., Yamashita, K., Ogata-Arakawa, M., Koide, N., Muramatsu, T., Iwashita, S., Inoue Y. & Kobata, A. (1975). Structural studies of 2 ovalbumin glycopeptides in relation to the endo- β -N-acetyl glucosaminidase specificity. *J. Biol. Chem.* **250**, 8569–8575.
- Tai, T., Yamashita, K., Ito, S. & Kobata, A. (1977a). Structures of the carbohydrate moiety of ovalbumin glycopeptide III and the difference in specificity of endo- β -N-acetylglucosaminidases C11 and H. *J. Biol. Chem.* **252**, 6687–6694.
- Tai, T., Yamashita, K. & Kobata, A. (1977b). The substrate specificities of endo- β -N-acetyl glucosaminidases. *Biochim. Biophys. Res. Commun.* **78**, 434–441.
- Thompson, E. O. P. & Fisher, W. K. (1978). Amino acid sequences containing half-cystine residues in ovalbumin. *Aust. J. Biol. Sci.* **31**, 433–442.
- Warner, R. C. (1954). Egg proteins. In *The Proteins* (Neurath, H. & Bailey, K., eds), pp. 435–485, Academic Press, New York.
- Weis, W. I. & Brunger, A. T. (1989). Crystallographic refinement by simulated annealing. In *Molecular Simulation and Protein Crystallography* (Goodfellow, J., Henrick, K. & Hubbard, R., eds), pp. 16–28, S.E.R.C. Daresbury Laboratory, Daresbury, U.K.
- Wiseman, R. L., Fothergill, J. E. & Fothergill, L. A. (1972). Replacement of asparagine by aspartic acid in hen ovalbumin and a difference in immunochemical reactivity. *Biochem. J.* **127**, 775–780.
- Woo, S. L. C., Beattie, W. G., Catterall, J. F., Dugaiczky, A., Staden, R., Brownlee, G. G. & O'Malley, B. W. (1981). Complete nucleotide sequence of the chicken chromosomal ovalbumin gene and its biological significance. *Biochemistry*, **20**, 6437–6446.
- Wright, H. T. (1984). Ovalbumin is an elastase substrate. *J. Biol. Chem.* **259**, 14335–14336.
- Wright, H. T., Qian, H. X. & Huber, R. (1990). Crystal structure of plakalbumin, a proteolytically nicked form of ovalbumin. Its relationship to the structure of cleaved α -1-proteinase inhibitor. *J. Mol. Biol.* **213**, 513–528.
- Yamashita, K., Tachibana, Y. & Kobata, A. (1978). The structures of the galactose-containing sugar chains of ovalbumin. *J. Biol. Chem.* **253**, 3862–3869.