Structural studies on alpha-lactalbumin and retinol binding protein

Chrysina, Evangelia D.

Award date:
2000

Awarding institution:
University of Bath

STRUCTURAL STUDIES ON

α-LACTALBUMIN

AND

RETINOL BINDING PROTEIN

STRUCTURAL STUDIES ON

α-LACTALBUMIN

AND

RETINOL BINDING PROTEIN

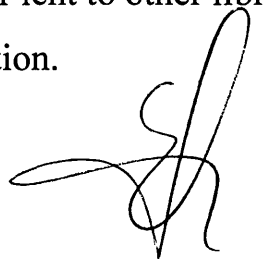
submitted by Evangelia D. Chrysina

for the degree of PhD

of the University of Bath

2000

COPYRIGHT



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To my parents :

Maria Chrysina

Demetrios Chrysinas

ABSTRACT

α-Lactalbumin: Lactose, the major source of carbohydrate in most mammalian milks, is synthesised from glucose and UDP-galactose in a reaction catalysed by the enzyme lactose synthase (LS). The biosynthesis of LS is catalysed by the reversible binding of the regulatory subunit α-Lactalbumin (LA) to Galactosyltransferase (GT, catalytic subunit) promoting the glucose binding (1000-fold decrease in K_m for glucose).

An important feature of LA is that its three dimensional structure is similar to that of c-type lysozyme (LZ) and the two proteins are homologous (about 40% sequence identity) with divergent functions.

LA binds calcium strongly (also other metal ions such as Zn^{2+}, Mn^{2+}, Co^{2+}, Tb^{3+}, Na^{+}, K^{+}) and specifically whereas in the LZs there are two subgroups, representing paralogous gene lines, of which only one binds Ca^{2+} ion. High affinity Ca^{2+} binding to LA stabilises the native structure and is required for the efficient generation of native protein with correct disulphide bonds from the reduced denatured state. At ambient temperature and low ionic strength the apo-protein assumes a molten globule state. Calcium greatly accelerates folding by binding to rate limiting intermediates in the folding process. To investigate the role of calcium in the LA structure with the goal of obtaining high-resolution data relevant to its role in folding and stability, X-ray structures at 2.2 Å resolution have been determined for crystals of the apo- and holo-forms of bovine LA. Although Ca^{2+} removal has little effect on protein structure in the metal binding site, a significant structural change was observed in the LA cleft region at the periphery of the hydrophobic box in the region around Tyr-103 of the helical lobe and Gln-54 of the beta lobe. This change results in a more open cleft structure in the apo-protein and appears to reflect an effect of calcium binding on buried solvent molecules which in turn affects interactions between the lobes. This

provides high-resolution structural information on the mechanism through which Ca^{2+} binding can facilitate the formation of packing interactions in the inter-lobe region at later stages of folding.

Bovine LA had been used as a model system for the design of a series of LA variants at regions proposed to be directly involved in LA action in LS complex in order to follow a 'structure-based' approach. A detailed structural investigation was performed for three LA variants: Ala109-Pro, Tyr103-Pro and Trp118-His while crystallisation conditions for additional four mutants: Phe31-Tyr, His32-Tyr, Lys114-Asn have been established.

Retinol binding protein: Retinol binding protein (RBP) is a specific carrier protein synthesised in the hepatocytes. Its physiological role is the delivery of retinol to the peripheral target tissues. In the plasma it also forms a complex with transthyretin (TTR). Being a carrier protein, RBP molecule appears to be a good model protein for protein engineering studies and for the design of molecules that will enable RBP to transport not only drugs but also other molecules.

Investigations into the relationship between sequence conservation, stability and folding with RBP had been performed and in the context of these studies the high-resolution structures of rRBP and a double variant W67L/W91H was determined at 1.7 and 2.0 Å resolution respectively. The overall structures of the two proteins are very similar and the most ordered region of the molecule was formed by residues surrounding the β-barrel in both cases apart from the flexible 62-68 loop. Glycerol molecules (used during data collection as cryoprotectant) and water molecules were identified bound in the core of the β-barrel at the retinol binding site.

ACKNOWLEDGEMENTS

I would like to thank my supervisor K. Ravi Acharya for giving me the opportunity to be part of his group and work on these two projects. His excitement about research and his dedication to science along with his constant support and encouragement from the first till the last moment I have been in his lab, made possible the completion of this thesis and motivated me to continue in the same field in the future.

I would like to acknowledge Dr. Laurie Irons for his contribution to the Retinol Binding Protein project and his supervision. I would like to thank Professor Keith Brew for providing α-Lactalbumin, Retinol binding protein and the variants used in this Ph.D. project, Dr. Ashley Pike for preparing the α-Lactalbumin variants.

I am indebted to Prof. Nikos G. Oikonomakos who gave me the first insights on protein crystallography and supported my coming to Bath. On the same note I am grateful to Dr. Demetres D. Leonidas who was always willing to share his knowledge and help me whenever it was necessary. I would also like to thank Dr. Tassos C. Papageorgiou along with Dr. Leonidas and Dr. G. Jawahar Swaminathan for their help during data collection and Dr. Dan Holloway for his useful advice on the preparation of calcium-free solutions and sequence alignment. Particularly valuable was the help of Ms. Shalini Iyer during the final compilation of the thesis.

I would also like to thank Dr Vasanta Subramanian and Mrs Nadie Kofina-Leonidas who provided a family-like environment when I arrived in Bath along with my friends Ms. Alexandra Timotheadi, Mr. George Tzimas, Ms. Katerina Ioannou and Ms. Maroula Sidiropoulou who stood by me at good and bad times during my stay here. A special 'thank you'

is saved for my parents for their constant love, support and understanding all these years.

I am grateful to the staff at the Synchrotron Radiation Source (CLRC Laboratory, Daresbury, EMBL-Hamburg and Elettra-Italy) for help with the data collection and finally I would like to thank European Commission for supporting my work through a training fellowship.

February 2000

Abbreviations

Å	:	Ångstrom=10^{-1}nm
ACI	:	Aromatic cluster I
ACII	:	Aromatic cluster II
apo-LA	:	Calcium free α-Lactalbumin
ARP	:	Automated refinement procudure.
B-factor	:	Thermal parameter
bLA	:	Native bovine LA
CaM	:	Calmodulin
CCD	:	Charged coupled detector
CD	:	Circular Dichroism
CG	:	Conjugate Gradient
E-RABP	:	Epididymal retinoic acid binding protein
ER	:	Endoplasmic reticulum
FABP	:	Fatty acid binding proteins
GlcNAc	:	N-acetylglucosaminyl
GT	:	β1,4-galactosyltransferase
HRBP	:	Human-serum retinol binding protein
K_{app}	:	Dissociation constant
LA	:	α-Lactalbumin
LS	:	Lactose synthase
LSQ	:	Least squares
LZ	:	Lysozyme
MLA	:	Recombinant α-Lactalbumin (monoclinic form)
ML	:	Maximum likelihood
MPD	:	2,4 dimethyl-pentanediol
PEG	:	Polyethylene glycol

R	:	Reliability factor
r.m.s.d.	:	Root mean square deviation
rRBP	:	Recombinant retinol binding protein
RBP	:	Retinol binding protein
RABP	:	Retinoic acid binding protein
SRS	:	Synchrotron radiation source
TTR	:	Transthyretin
UDP	:	Uridino-phosphate
UDP-Gal	:	Uridine diphosphogalactose
UV	:	Ultra violet

CONTENTS

CHAPTER I.

CHAPTER II.

-1-

CRYSTALLISATION OF PROTEINS

PRINCIPLES OF X-RAY DIFFRACTION

DATA COLLECTION

STRUCTURE DETERMINATION

REFINEMENT

STRUCTURE ANALYSIS

1.1. CRYSTALLISATION OF PROTEINS

1.1.1. Factors affecting the crystallisation of proteins

Proteins subjected to crystallisation trials are obtained either by means of protein engineering or by isolation from a natural source and should be pure and homogeneous. Screening of various crystallisation conditions are carried out where parameters that affect the formation (or quality in the case of already established conditions) of the crystals are adjusted to achieve the optimum.

A crystal is an orderly three-dimensional array of molecules held together by non-covalent interactions. The crystallisation conditions that are used should fulfil two aims: (i) to reach a point of supersaturating at which crystals are formed and (ii) to enable the crystals to grow large enough for diffraction studies.

A protein is considered to be a large polyvalent ion and its solubility depends on interactions of the molecule with water. Ionic strength, pH, temperature and organic solvents are factors that affect the solubility of a protein in different ways.

Kam *et al.* (1978) have proposed three distinct phases that describe the production of usable crystals judging from their size and their population: (i) *the nucleation*; it is the process during which aggregates begin to grow when they reach a critical size (nucleus), (ii) *the post-nucleation growth*; it is defined by the rate of transport of protein molecules to the crystal and the probability of attachment of a molecule which decreases when the diffusion of protein molecules is slow (the rate of deposition of protein mass at the crystal surface is

given by Fick's law), (iii) *the cessation of growth*; it must be due to a modified property of the crystal surface. The faster the crystals grow, the larger their number and the smaller their terminal size. Re-growth of the terminal sized crystals, cut into pieces to approximately the same terminal size has been observed for the exposed surface and the poisoning of favourable growth sites on the surface by impurities is been considered as a possible cause (Blundell and Johnson, 1976; Feher, 1986; Kam *et al.*, 1978).

1.1.2. Vapour diffusion as a crystallisation technique

In the most common methods of growing protein crystals, purified protein is dissolved in an aqueous buffer containing a precipitant such as ammonium sulphate or polyethylene glycol, at a concentration below that necessary to precipitate the protein (McPherson, 1982). Vapour diffusion is the most widely used method for crystallisation, in which the protein/precipitant solution is allowed to reach equilibrium in a closed container with a larger aqueous reservoir where the precipitant concentration is optimal for producing crystals. A typical example of

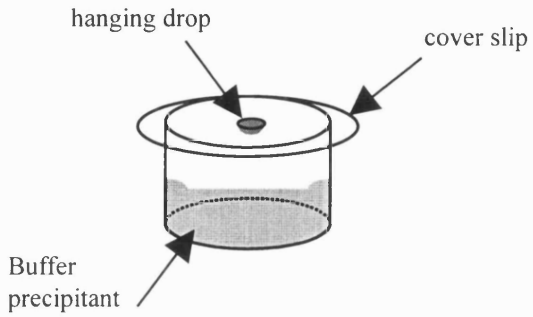


***Figure* 1.1**

this technique is the 'hanging-drop method' according to which equal volumes of the purified protein (usually dissolved in water) and the reservoir solution are mixed, and the new solution is suspended as a droplet underneath a cover slip, sealed onto the top of the reservoir with vacuum grease (*Figure* 1.1).

1.2. PRINCIPLES OF X-RAY DIFFRACTION

X-rays are electromagnetic radiation of wavelengths in the range of 0.1-100Å (1nm=10Å; wavelength of visible light λ=400-700 nm) produced when high-energy electrons collide with and displace an electron from a low-lying orbital in a target metal atom. Consequently, an electron from higher orbital drops into the resulting vacancy, emitting its excess energy as an X-ray photon.

Most widely used X-ray sources are: X-ray tubes, rotating anode tubes, and particle accelerators that produce synchrotron radiation in the X-ray region. The first two are more common as laboratory or conventional sources and accelerate electrons to a high voltage of 40-50 kV at a metal target (usually copper with characteristic wavelength λ=1.5418 Å) filter and focus X-rays by means of a monochromator and curved mirrors. The monochromator is either a piece of nickel or single-crystal that removes any radiation absorbed by the sample and does not contribute to the diffraction pattern. Mirrors deflect focused X-rays at low angle and as a result increase the brilliance and consequently allow resolution of very large unit cells with closely spaced diffraction spots. As brilliance of the beam is designated the fraction of flux over the angle through which radiation is emitted multiplied by the cross sectional area of the source, where flux is the number of photons emitted within a given wavelength range.

However, the use of synchrotron radiation is much more effective in protein crystallography. Synchrotron radiation is electromagnetic radiation produced when electrons or positrons moving at relativistic energies are constrained to follow a circular path defined by dipole

magnets or magnetic insertion devices like undulators (periodic array of permanent magnets) or wiggler magnets (superconducting magnets that force the electron beam to effectively take a higher acceleration and emit radiation at short wavelength). At the Daresbury Synchrotron Radiation Source, electrons are injected into a booster synchrotron by a 12 MeV linear accelerator and they are accelerated to an energy of 600 MeV prior to their extraction from the synchrotron. Injection of the electrons into the 2 GeV storage ring follows until a stored current up to 300 mA has been achieved and finally the electrons obtain an energy of 2000 MeV by increase of the field strengths in the magnets.

The highly monochromatic and tuneable nature of the beam and the great intensity (at least 1000 times that of a conventional source) combined with the small divergence, the high intensity over a continuous wavelength from 0.02 - 0.26 nm make the use of synchrotron radiation essential in protein crystallography. Moreover, the use of short selected wavelengths lead to reduced radiation damage and absorption effects and is beneficiary for the determination of the phases for the reflection (Blundell and Johnson, 1976; McRee, 1993; Rhodes, 1993).

The interaction of the synchrotron light beam with an experimental sample must be recorded in some way so it can be analysed giving information about the innermost structure of the sample. X-ray generators are equipped with an area detector for this purpose. A more advanced detector that has been used lately is the image plate, which is exposed to X-rays as any other detector. The X-ray photon causes a chemical change in the plate coating and as a result it releases a fluorescence scanned with light of the proper wavelength. Most recently charged couple detectors (CCD) have been launched with improved

features compared to image plate such as the dynamic range of the recorded reflections, reduced read out and erase time as well as minimum phi rotation of the crystal along the spindle axis allowing the time that the crystal is in a diffracting position to be almost the same to the oscillation time and as a result reduce the noise.

X-rays can be diffracted even by the smallest molecules (the dimensions of the scattering objects are of the same order of magnitude of the wavelength of X-rays) but they cannot produce a focused image of a molecule because there is no lens capable of focusing X-rays. A computer, which simulates an image-reconstructing lens by measuring the directions and the intensities of the diffracted X-rays gives the answer to this problem (*Figure* 1.2). However, a single molecule is a very weak diffractor of X-rays, so the analysis of crystal diffraction rather than individual molecules is essential since the scattering from any molecule is reinforced by the scattering of all the others.

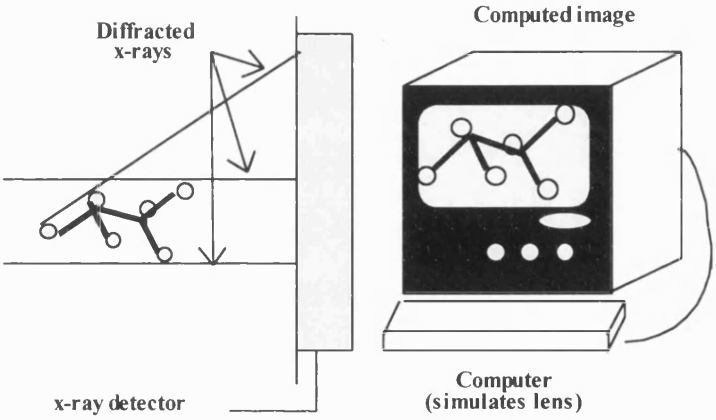


***Figure* 1.2**

The diffraction pattern obtained gives both the beam intensity and direction for each diffracted beam, parameters required from a computer program in order to reconstruct an image of molecules in the unit cell.

A protein crystal is an orderly three-dimensional array of molecules held together by non-covalent interactions and its simplest volume element representative of the whole crystal is the unit cell. The array of points at the corners of the vertices of the unit cell is called lattice. The unit cell can be subdivided into equivalent parts called asymmetric units (*Figure* 1.3). The number of molecules that can be observed per asymmetric unit is indicative of how compact the molecules are packed within the crystal.

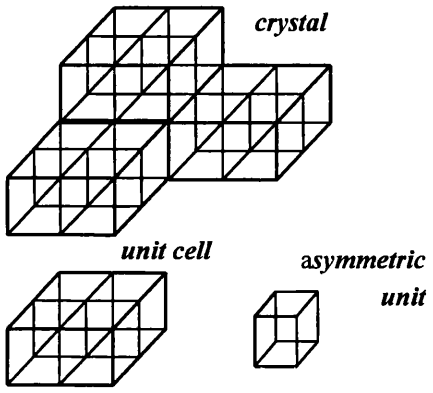
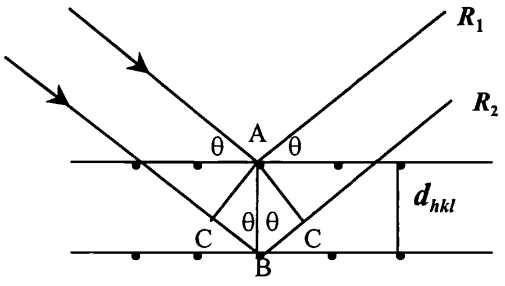


***Figure* 1.3**

Bragg in 1913 was the first who succeeded in visualising the scattering of X-rays by a crystal in terms of reflections from planes of atoms. According to Bragg's law a set of parallel planes with index hkl (Miller indices) and interplanar spacing d_{hkl} produces a diffracted beam (*Figure* 1.4), when X-rays of wavelength λ impinge on the planes at an angle θ and are reflected at the same angle, only if θ meets the condition

$$\mathbf{2}d_{hkl}\cdot\mathbf{sin}\theta = \mathbf{n}\lambda, \text{ n integer}$$



$\mathbf{sin}\theta = \mathbf{BC/AB}$

$\mathbf{BC} = \mathbf{AB}\ \mathbf{sin}\theta = d_{hkl}\cdot\mathbf{sin}\theta$

***Figure* 1.4**

Schematic description of Bragg's law.
(The closer the separation of planes d_{hkl}, the larger the angle of diffraction)

Another approach of X-ray diffraction more general than that of Bragg's was proposed by Ewald in 1921 by the use of geometrical construction (*Figure* 1.5) according to which X-rays will be diffracted in the direction CB if the point B represents a reciprocal lattice point (*h*,*k*,*l*) (the dimensions are inversely proportional to those of the real lattice).

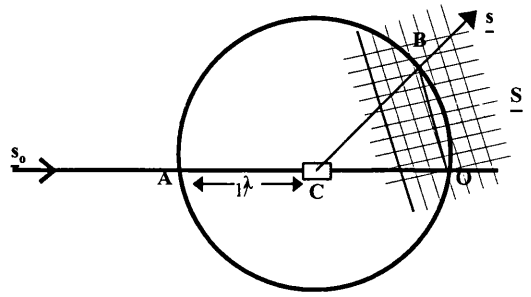


***Figure* 1.5**
The Ewald's geometric construction where $\underline{s}_0$ *is the incident vector,* $\underline{s}$ *the diffracted vector and* $\underline{S}$ *is a reciprocal lattice vector*

The spacing of the reflections in the lattice on the film is designated as reciprocal lattice while real lattice is the spacing of the unit cell in the crystalline lattice. The diffraction pattern that emerges is the product of the diffraction of the molecule (molecular transformation) with the diffraction of the reciprocal lattice.

1.3. DATA COLLECTION

1.3.1. Crystal mounting and alignment

The crystals are usually mounted in thin-walled capillaries in order to minimise the absorption of the scattered X-rays and the background of the glass, provided that the data collection is taking place under room temperature. Nowadays, more and more crystallographic data collections on biological macromolecules are performed at cryogenic temperatures and these data are usually superior to those collected at room temperature in many respects.

The temperatures used in the crystal freezing technique are in the range of 80 to 100°K. The crystal is placed rapidly directly in the stream of cold nitrogen gas, and it is conveyed by a loop made of very fine fibre of nylon or glass wool, held in a thin film of cryoprotectant (*Figure* 1.6). The crystal is held in the loop by surface tension and is frozen immediately. The presence of the cryoprotectant is of great importance but varies considerably because it has to be compatible with the crystal. Substances such as glycerol, polyethylene glycol (PEG), 2,4 dimethyl-pentanodiol (MPD) etc., that aim to form a thin film (amorphous glass) around the crystal which will prevent the formation of ice are widely used cryoprotectants.

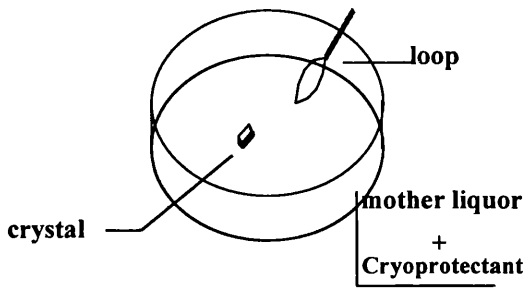


***Figure* 1.6**

Data collection at cryogenic temperatures has tremendous benefits: (a) Reduced crystal radiation damage both primary and secondary that increases significantly the life of the crystal. Primary damage is caused by ionisation effects due to incident photons and depends on the energy the crystal absorbs (dose). Secondary damage is observed only when the data collection takes place at room temperature. The production rate of free radicals increases either by direct hydrolysis of water molecules (macromolecular crystals have usually high solvent content), or by reactions initiated by radiation. Diffusion of radicals to distant regions mediated by crystal water and heating of the sample by intense X-ray follows and damage becomes more severe proportionally to time. At cryogenic temperatures free radicals produced by the incident X-ray photons cannot diffuse away from the site of their production to create damage over the entire crystal. (b) Reduced thermal vibrations that result in lower thermal parameters (*B* factors) and thus render reflections at higher resolution accessible to measurement. (c) Elimination of mechanical stress is of great importance especially for manipulating fragile crystals. (d) It enables studies of the intermediate states of a catalytic process in enzymatic reactions since the rate of the reactions slows down in cold. Molecules are trapped in a particular intermediate state of a catalytic process giving the opportunity to obtain a snapshot of the corresponding step of the chemical reaction (Garman and Schneider, 1997; Rodgers, 1994). (e) Increased signal to noise ratio is observed by avoiding background X-ray scatter from excess of mother liquor and glass/quartz capillary tube. (f) Improvement of the data quality by collecting complete data sets from only one crystal.

As every other technique apart from advantages it also has drawbacks: (i) Non-isomorphism, where changes in the unit cell or in thermal and structural parameters are observed (e.g. between native and derivative data sets). (ii) Alterations in the overall structure or in physical and chemical properties dependent on temperature. (iii) Variations in the atomic displacement parameters and multiple site disorders affected by temperature. (iv) Changes caused by the cryoprotectant binding which might be competitive or synergistic in the case of a possible ligand or just cause conformational changes preventing the aim of the experiment. (v) Ice formation due to not optimal cryoprotectant conditions and (vi) Mosaicity increase.

The crystal's orientation, dimensions and symmetry are some of the factors that should be taken into consideration for the alignment of the crystal. The crystal is rotated around one axis, usually perpendicular to the beam and is aligned at 0-180-90-270°. The alignment of the crystal should be towards the longest axis, since the cell dimension along the spindle axis will never have an effect because it will never lie along the beam (*Figure* 1.7) (Dauter, 1997). Only the axis perpendicular to the rotation angle is limiting.

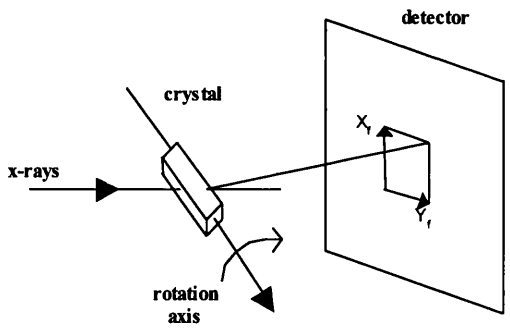


***Figure* 1.7**

Schematic description of the construction used for data collection

The spot size or mosaic spread which corresponds to the width of the spot in the

rotational direction and depends on the crystal quality and the X-ray optics available, as well as the desired resolution should be considered before deciding the rotation angle. The maximum rotation angle in the area of the diffraction pattern perpendicular to the rotation axis and at maximum diffraction angle can be calculated using the following equation:

$$\Delta\text{rotation}_{max} = \tan^{-1}(d_{min}/\text{cell edge}) - \text{spotwidth}$$

Another limitation during data collection is a region of reciprocal space near the rotation axis that cannot be collected because the Lorentz correction is too high. Lorentz correction is dependent on the geometry of the detector and accounts for the amount of time a reflection spends in diffracting conditions while being rotated through the Ewald sphere.

1.3.2. Preliminary characterisation of the crystal.

Preliminary characterisation of the crystal involves the extraction of information that would describe best the features of the exposed crystal and would facilitate X-ray data collection. Visual inspection of the diffraction pattern of a single computer reconstructed image provides a vast amount of information. The quality of the crystal can be assessed by the limit of observable diffraction which is defined as the resolution where at least one third of the possible reflections are still visible above background and the mosaicity of the crystal, a measure of the order within a crystal that can be estimated by the broadness of lunes on the diffraction pattern and the sharpness of the spots. The reciprocal space symmetry can be identified by the lattice representation on the

diffraction pattern, the spacegroup can be determined from the systematic absences and the unit cell dimensions as well as the number of molecules per asymmetric unit can be estimated. The appropriate crystal size can also be decided from the diffraction pattern. Crystals with large unit cells usually give weak diffraction patterns since the amount of photons they diffract is spread out over more reflections.

1.3.3. Data integration and Reduction

The collected data are processed by indexing the reflections. Several algorithms have been developed with the aim to automate the procedure of indexing and the most common for data integration and reduction are the *HKL* package (Otwinowski and Minor, 1997) and *MARXDS* (Kabsch, 1988). In particular the *HKL* package is very widely used and provides several levels of insight into the data at each stage of the measurement and data analysis process: (i) presents data visually, up to a single pixel level (*XdisplayF*), (ii) provides numerical analysis of one oscillation image (*DENZO*) and (iii) provides statistics for the full data set (*SCALEPACK*). Use of the above package is a significant tool especially during data collection since it gives answers to very basic questions like whether to proceed with data collection. One or two images are enough to index, estimate the mosaicity and to work out the data collection parameters (avoiding overlap of reflections).

Autoindexing in *DENZO* involves the determination of a standard lattice. It searches for real-space vectors (3 linearly independent vectors- a basis) that would index all of the observed peaks and reduces the cell by conversion of the basis into a standard cell according the

International Tables for Crystallography, which contain an index for standard space group and symmetry classification. The highest symmetry lattice that fits the data with minimal distortion is selected and the processing of rest of the data will be performed accordingly using the initial estimates as reference.

Once indexed it is necessary to scale the data. Scaling is the operation of setting the sum of one data set to the other. Systematic errors in the data due to absorption and decay mean that different parts of the data require different scale factors. Often scaling is performed in bins based on resolution. The program *SCALEPACK* from the *HKL* package (Otwinowski and Minor, 1997) reduces the bias towards reflections with an integrated intensity below the average. Percentage completeness is the number of unique reflections measured over the number of unique reflections possible and one of the major values obtained after data reduction. The important parameter is the merging *R* (R_m), an internal measure of data precision given by the equation:

$$R_m(I) = \frac{\sum_h \sum_i \left| \left(I_h - I_{hi} \right) \right|}{\sum_h \sum_i I_{hi}} \times 100$$

where I_h is the weighted mean measured intensity of the observations I_{hi} in which the intensities of the symmetry related reflections, which should be the same, are compared and R_m is the estimate of the disagreement (I_h, I_{hi}).

The data quality is also assessed by the ratio of intensity to sigma of intensity I/σ(I). σ of a reflection intensity is both an estimate of the accuracy of an individual measurement as opposed to the accuracy of

the data set as a whole and a combination of counting statistics, background height and variance as well as the number of times the reflection was measured.

1.4. STRUCTURE DETERMINATION

1.4.1. From structure factor to electron density. The phase problem.

The atomic scattering factor is the integration of the individual contributions over the volume of the atom is given by the following expression:

$$\mathbf{F(S)} = \int_{\substack{vol.of \\ atom}} \rho(\boldsymbol{r}) \exp(2\pi i \boldsymbol{r} \cdot \boldsymbol{S}) dv$$

where ρ is the variation in electron density over the entire volume of the atom. The total wave scattered by all the atoms in a molecule or the so-called '*molecular transform*' is given by the equation:

$$\mathbf{G(S)} = \sum_{j=1}^{N} f_j \exp(2\pi i \boldsymbol{r}_j \cdot \boldsymbol{S})$$

where f, r, are the atomic scattering factor and the distance of the atom j from the origin respectively, for a molecule comprised of N atoms. The structure factor for a particular reflection from a crystal can be represented by its amplitude and phase according to the following equation:

$$\mathbf{F(hkl)} = \sum_{j=1}^{N} f_j \exp 2\pi i(\boldsymbol{hx}_j + \boldsymbol{ky}_j + \boldsymbol{lz}_j) = \mathrm{F(hkl)} \exp i\alpha(\boldsymbol{hkl})$$

where F(hkl) is the amplitude and α(hkl) is the phase. Once the structure of the crystal is known the diffraction pattern of a crystal can be calculated from the structure factor equation. Consequently, taking into consideration that the crystal structure is the Fourier transform of the diffraction pattern, if the structure factors F(hkl), are known for all reflections, *hkl*, then the electron density may be calculated for each

point xyz in the unit cell as described in the expression (Blundell and Johnson, 1976; Rhodes, 1993):

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \exp -2\pi i(hx + ky + lz)$$

Both the amplitude F(hkl) and the phase α(hkl) of the structure factor are necessary in order to calculate the electron density but the diffraction pattern can provide information only for the amplitude. Hence the 'phase problem'. This can be solved by using one of the following methods: (i) the 'isomorphous replacement' (IR) in which indirect experimental estimates of the protein phase angles are provided, after the introduction of a heavy atom marker, by observing the interference effects of the intensities on scattered beams (Greene et al., 1954; Ke, 1997), (ii) the 'molecular replacement' (MR) is applicable only in the case of significant amino acid sequence identity between the starting model used to determine the orientation and the position of the crystal structure of a molecule and the molecule itself. Both phase angle estimations and amplitudes give an approximate Fourier synthesis of the unknown structure (Rossmann, 1972) and (iii) anomalous scattering in which all the phase information emerges from the scattering information of an atom whose natural absorption frequency is close to the wavelength of the incident radiation. Resolved anomalous scattering requires intensity measurements at one wavelength, while the multiple wavelength anomalous dispersion method, most commonly used in large molecules requires intensity measurements at several wavelengths (the tuneable SRS is most appropriate for such cases) (Hendrickson, 1991).

1.4.2. Molecular replacement technique using *AMoRe*.

AMoRe is one of the most commonly used Automated Molecular Replacement program package to determine the positions of the molecules in the unit cell (Navaza, 1994). Homologous molecular models to the molecules included in the crystal are placed in tentative positions in the unit cell in such a configuration that give acceptable agreement between the calculated and the observed structure factors (Navaza and Saludjian, 1997).

Some of the factors affecting the success of the method are (i) the quality and the completeness of the collected data, (ii) the percentage of homology between the search molecular models and the actual molecules to be determined, (iii) the relative size of the search model compared to the content of the crystal cell and (iv) the applied criteria used to assess the quality of agreement between the model and the experimental data.

The molecular models are treated in every case as a rigid body and their position in the crystal cell is identified by three rotational and three translational parameters calculated accordingly by the rotation (*RF*) and translation function (*TF*). The combined results of the rotation and translation functions using as criteria not only the peak height from the rotation function (program *ROTING*) but also the correlation coefficient from the translation function (program *TRAING*) to assess the selected positions, are finally submitted to rigid body refinement (program *FITING*).

1.5. STRUCTURE REFINEMENT

1.5.1. General principles of crystallographic refinement.

The molecular models that emerge from molecular replacement need further manipulation since there are several variations to the actual molecules in the crystal in terms of amino acid sequence, side or even main chain conformational alterations, etc. A number of algorithms have been developed over the years with the aim to overcome model discrepancy by applying an automated refinement procedure.

Crystallographic refinement is a technique that optimises the agreement of an atomic model with both observed diffraction data and chemical restraints. The measure of the agreement is a 'cost function' e.g. a function of the observed and calculated values which reduces as the model improves. The target function for crystallographic refinement formulated in such a way to search for global minima is given by the following expression:

$$E = E_{\text{chem}} + w_{\text{xray}} E_{\text{xray}}$$

where E_{chem} is an empirical potential energy function of all atomic positions that contains information about chemical interactions, describing covalent and non-covalent interactions.

$$E_{chem} = \underbrace{\sum_{bonds} k_b (r - r_o)^2 + \sum_{angles} k_\theta (\theta - \theta_o)^2 + \sum_{dihedrals} k_\phi (n\phi - d)2 + \sum_{chiral,planar} k_\omega (\omega - \omega_o)^2 +}_{\text{covalent interactions}}$$

$$+ \underbrace{\sum_{atompairs} (ar^{-12} + br^{-6} + cr^{-1})}_{\text{non-covalentinteractions}}$$

Additional restraints and constraints can be used by fixing the position of various atoms, bond lengths, angles, dihedrals, in order to improve the ratio of the observations to the refined parameters to avoid overparameterisation of the model. In the case of multiple copies of molecules in the asymmetric unit non-crystallographic symmetry or crystallographic symmetry can also be used as a means to increase the signal to noise ratio by using average of all molecules while treating them as equivalent entities.

E_{xray} stands for the difference between observed and calculated diffraction data according the equation:

$$E_{xray} = \sum_{\mathbf{h}} \left[\left| \mathbf{F}_{\text{obs}}(\mathbf{h}) \right| - k \left| \mathbf{F}_{\text{calc}}(\mathbf{h}) \right| \right]^2$$

while w_{xray} is a weight chosen to balance the forces arising from each term. The standard crystallographic residual is enriched by a term of additional phase restraints based on the difference between experimental phases and those calculated from the model so E_{xray} is described by

$$E_{xray} = \sum_{\mathbf{h}} \left[\left| \mathbf{F}_{\text{obs}}(\mathbf{h}) \right| - k \left| \mathbf{F}_{\text{calc}}(\mathbf{h}) \right| \right]^2 + w_{\text{p}} \sum_{\mathbf{h}} f \left[\phi_{\text{obs}}(\mathbf{h}) - \phi_{\text{calc}}(\mathbf{h}) \right]$$

where w_{p} is the weight given to the phase restraint and f is a square-well function with a width equal to the arccosine of the figure of merit $[m(\mathbf{h})]$ for each reflection. Figure of merit is the fraction of the correct structure factors (Adams *et al.*, 1997; Brünger and Rice, 1997).

Optimisation methods such as Conjugate Gradient (CG, belongs to the first derivative methods and identifies the nearest minimum in the target function by increasing the radius of convergence) or Least Squares (LSQ, the weighted sum of squares of the deviations between the observed and the calculated quantities) are the most widely used methods but their major disadvantage is that convergence can be hindered by 'trapping' the initial models in multiple local minima. Hence, the need for maximum likelihood (ML) target functions became apparent to avoid various assumptions made during LSQ refinement. The method of LSQ implies that (i) the current phase is always the correct one, (ii) the atoms should move towards or away from a certain and not an unknown plane, (iii) the effects of coordinate errors lead to a Gaussian error in $|\mathbf{F}_{calc}|$ as an estimate of $|\mathbf{F}_{obs}|$, (iv) the most probable value of $|\mathbf{F}_{obs}|$ is not generally equal to $|\mathbf{F}_{calc}|$ because of the effect of phase errors. In the ML method the probability of a particular $|\mathbf{F}_{obs}|$ amplitude is obtained by integrating the probability of true structure factors with that amplitude over all possible phases.

The above refinement technique's aim is to overcome the local minima that the model might be 'trapped' during minimisation of their target function and the procedure of simulated annealing provides the necessary potential energy to the system to cope with such barriers. Annealing is a physical process wherein a solid is heated until it is converted to a viscous liquid comprising a random arrangement of all particles. The viscous liquid is then cooled slowly allowing particles to arrange in the lowest energy state. In the case of simulated annealing as a molecular dynamics refinement method combined with the

optimisation techniques, the simulation starts at a temperature of 300°K and the system is heated to a temperature of 5000°K before it starts cooling down. A control parameter that determines the likelihood of the target function to overcome barriers is designated as 'temperature and has no physical aspect (Brünger and Rice, 1997).

The most widely used refinement programs for protein crystallography are *X-PLOR* (Brünger, 1992b), *REFMAC* from the *CCP4* package (CCP4, 1994) and *CNS* by Brünger *et al.* (1998). All the programs use a combination of a minimisation technique along with simulated annealing and seem to cope well with the present needs.

CNS in particular has the unique feature of using the combined simulated-annealing/maximum-likelihood model refinement (Adams *et al.*, 1997; Brünger and Rice, 1997). Refinements with the maximum-likelihood target require computation of cross-validated σ_A values. The cross-validated σ_A error estimates and the weight between X-ray diffraction target function and the geometric energy function in the course of refinement are recalculated in the course of refinement. Recalculation of the σ_A error estimates is extremely important since the maximum-likelihood target functions depend on these values, thus it has been incorporated after initial energy minimisation as well as after molecular dynamics simulated annealing (Brünger *et al.*, 1998).

At this stage where the necessity to refine the position (x, y, z) and the thermal parameters (B factor) of atoms for the initial model against the X-ray observations has already been made clear, a new parameter to assess the quality of the structure at each stage of refinement is introduced and is described by the expression:

$$R = \frac{\sum \left| |\mathbf{F}_{obs}| - |\mathbf{F}_{calc}| \right|}{\sum |\mathbf{F}_{obs}|}$$

In this expression each $|\mathbf{F}_{obs}|$ is derived from a measured reflection intensity and each $|\mathbf{F}_{calc}|$ is the amplitude of the corresponding structure factor calculated from the current model. The values of R, the crystallographic reliability index, indicate the agreement of calculated and observed intensities. It is necessary to perform successive cycles of refinement followed by rebuilding of the protein until the value of R is between 0.15 and 0.20 for a well-refined structure. Another reliable and unbiased parameter used to evaluate the information content of a model produced by X-ray crystallography is a statistical quantity (R_{free}) that measures the agreement between observed and computed structure factor amplitudes for a 'test' set of reflections that is omitted in the modelling and refinement process (Brünger, 1992a). The criteria for assessing the quality of a structure apart from the reliability factor is also the stereochemical identity measured by the root mean square deviation (r.m.s.) from covalent bond lengths, which should be in the range of 0.01-0.02 Å.

It is difficult to correct gross errors by any of the refinement procedures mentioned above as it is usually indicated by the reliability factors. Use of either the initial phase information or the current one depending on the stage of the refinement can be used to calculate a variety of maps [usually ($2|\mathbf{F}_{obs}|-|\mathbf{F}_{calc}|$) and ($|\mathbf{F}_{obs}|-|\mathbf{F}_{calc}|$) maps] that would facilitate model building. Some useful protocols such as inclusion of bulk solvent correction that enhances the signal-to-noise ratio of electron density difference maps for missing parts of the model or

simulated annealing omit maps calculated for regions in doubt to remove any model bias are also used.

Model building is performed using the program '*O*' for molecular graphics (Jones *et al.*, 1991). It supplies a set of tools to perform a variety of manipulations such as movement/rotation of small parts of the structure manually, assignment of a side chain rotamer for a specific residue from a database, insertion/removal of residues from a region. The quality of the model can be improved by local real space refinement procedure during model building exercises.

An automated refinement procedure (*ARP*) can also be used based on the iterative use of unrestrained least squares minimisation coupled with constant updating of the model. It is a comparable procedure to the iterative least-squares/Fourier synthesis approach for small molecules and it requires very high-resolution quality data or else it can be used to improve unrestrained parts of the molecule (e.g. solvent or ligand molecules). The philosophy behind this procedure is that the quickest way to move an atom from its current position to another is to remove it and add a new one. The new set of atoms, which does not have to contain the same number of atoms as before, is subjected to least squares optimisation of their parameters against the experimental data. A requirement of the procedure is that the density maps are calculated using model phases, the quality of which defines the convergence limit.

Density at the atomic centre and density shape (and not thermal parameters) are used in order to decide upon rejection of certain atoms. Addition of atoms is performed using grid points with the highest density that satisfy the distance constraints selected for the new atoms.

Real space refinement can be used for the refinement of the atomic positions and for identifying a wrongly placed atom. Geometric restraints are also taken into consideration. The program *ARP* has been incorporated in the CCP4 package (Lamzin and Wilson, 1997).

1.5.2. Twinning.

Twinning is a crystal growth anomaly in which the specimen is composed of separate crystal domains whose orientations differ in a specific way. Multiple crystal growth disorders are common, but twinning refers to special cases where some or all of the lattice directions in the separate domains are parallel. This leads to either partial or complete coincidence between the lattices of the distinct domains (Redinbo and Yeates, 1993).

Twinning can be classified into three categories according to the extent the separate lattices coincide as (i) *nonmerohedral* or *epitaxial*: the overlap occurs in less than three dimensions, (ii) *pseudomerohedral*: the lattices overlap approximately (but not exactly) and (iii) *merohedral twinning*: the lattices of two or more distinct domains coincide exactly in three dimensions. Particularly in *merohedral* twinning, although the diffraction pattern appears normal each observed intensity is a weighted sum of intensities of two reflections that are related by twinning operation but not crystallographic symmetry.

$$I_{obs}(h_1) = (1-\alpha)I(h_1) + \alpha I(h_2)$$

$$I_{obs}(h_2) = \alpha I(h_1) + (1-\alpha)I(h_2)$$

α, is the *twinning fraction* that represents the fractional volume of the crystal occupied by domains in the second orientation. In the case that α equals 0 then no twinning is detected but if the value of α is less than 1/2 then the crystal is considered to be partial twinned. Problems arise when α becomes 1/2, *perfect twinning*, in which reflections related by the twin law contribute equally to each observed intensity (Yeates, 1997).

1.6. STRUCTURE ANALYSIS

The final stage is the analysis of the structure in which the overall architecture of the molecule is studied. Ramachandran plot (φ and ψ angles of the polypeptide backbone are printed) that will assess the geometry of the structure, inter- and intramolecular contacts in the crystal as well as a record of polar and non-polar interactions need to be examined. More extensive analysis involves superposition with other known homologous structures from which useful information can be derived as well as attempts to extrapolate the current finding to other analogous systems.

The overall procedure showing the basic stages in the determination of a protein structure is shown in the following flow diagram (*Figure* 1.8).

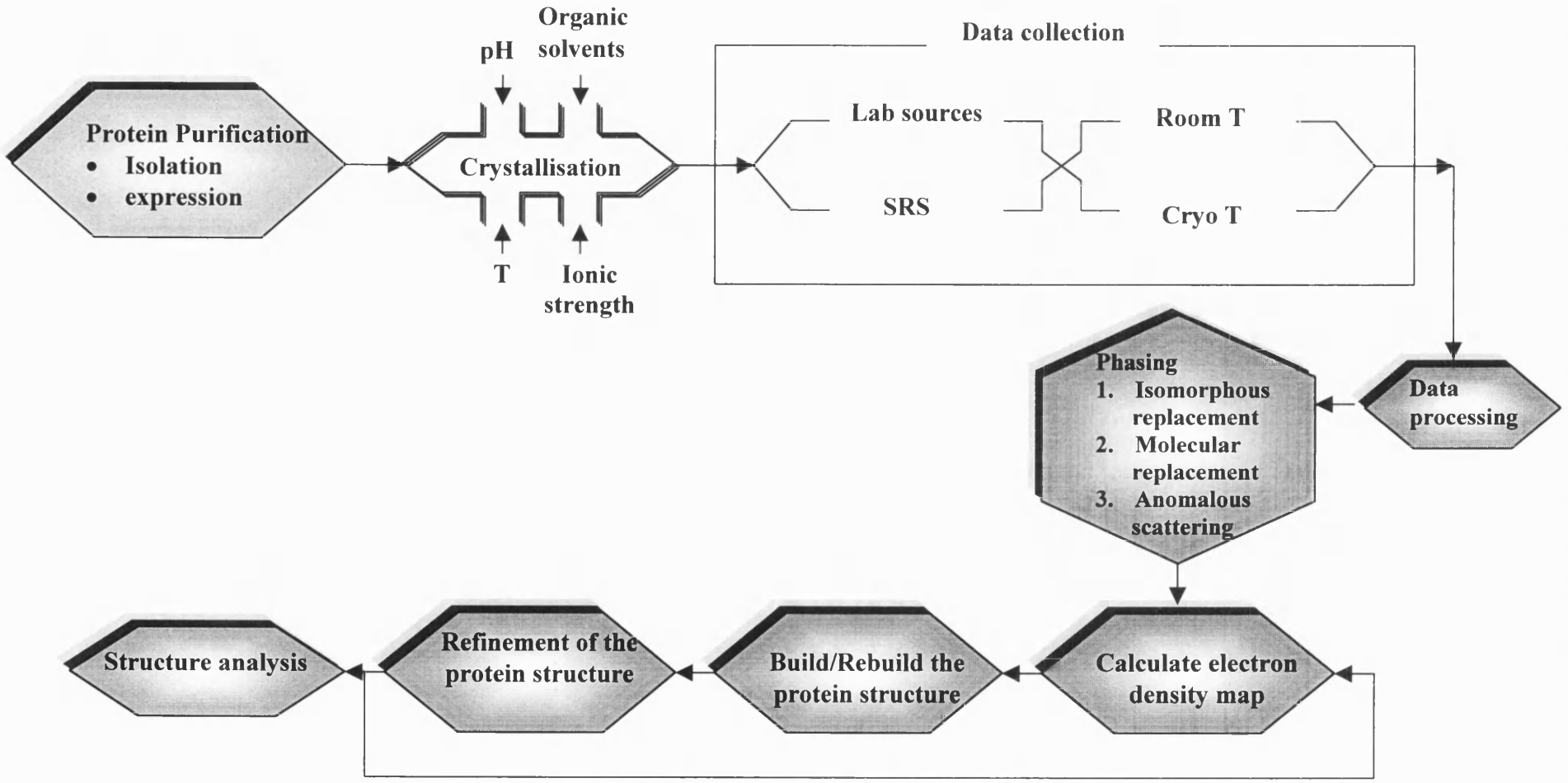


Figure 1.8 :
Flow diagram showing the basic stages in the determination of a protein structure.

-2-

α-LACTALBUMIN

INTRODUCTION

CRYSTAL STRUCTURES OF apo-LA and bLA

α-LACTALBUMIN VARIANTS

INVESTIGATION OF BINDING OF Mn^{2+} TO HUMAN α-LACTALBUMIN

2. α-LACTALBUMIN

2.1. INTRODUCTION

2.1.1. α-Lactalbumin as a modulator of Lactose Synthase complex

Lactose, the major source of carbohydrate in most mammalian milks, is synthesised from glucose and UDP-galactose in a reaction catalysed by the enzyme lactose synthase (LS, EC 2.4.1.22) (Acharya *et al.*, 1989; Smith *et al.*, 1987; Stuart *et al.*, 1986). This enzyme is a complex composed of two distinct protein components: α-Lactalbumin (LA) and UDP-galactose β-*N*-acetylglucosaminide β1-4,galactosyltransferase (EC 2.4.1.38, GT).

GT, is a 55KDa type II membrane glycoprotein of the *trans*-golgi membranes of the mammary gland and other tissues. In most cells, GT participates in the biosynthesis of oligosaccharide chains of secretory and membrane-bound glycoconjugates by catalysing the transfer of galactose from UDP-galactose to non-reducing terminal β-linked *N*-acetylglucosaminyl (GlcNAc) moieties of the oligosaccharide chain according to the following reaction:

UDP-D-galactose + N-acetylglucosamine →

→ UDP + D-galactosyl-N-acetyl-D-glucosamine

However, GT is unable to catalyse the biosynthesis of lactose (transfer of galactose to glucose) at physiological concentrations of glucose because of its poor affinity for this acceptor substrate. During

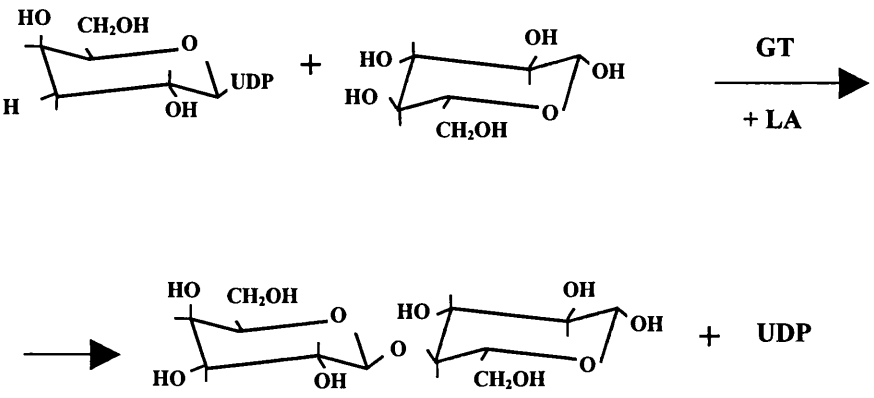


***Figure* 2.1**

The reaction of the biosynthesis of lactose catalysed by GT in the presence of LA

pregnancy GT is present in the mammary gland, as in most tissues, but when lactation begins after parturition, the specificity of GT changes to D-glucose by transferring the D-galactosyl group at a very high rate, thus making lactose according to the reaction:

UDP-D-galactose + D-glucose → UDP + D-lactose

in the presence of a milk protein α-lactalbumin (LA) the synthesis of which is regulated in the mammary gland by hormones promoting lactation. LA is produced in the rough endoplasmic reticulum of the lactating mammary gland together with other milk proteins as a pre-protein with a hydrophobic leader sequence. In lactose synthase, reversible binding of the regulatory subunit α-lactalbumin (LA) to GT promotes glucose binding (reflected in a 1000-fold decrease in the K_m for glucose) and facilitates the biosynthesis of lactose (Wang *et al.*, 1989) (*Figure* 2.1). LA comes into contact with GT in the golgi apparatus

(where GT is located) only during milk secretion after translocation across the ER membrane and proteolytic removal of the leader sequence. Lactose is then produced but its size does not allow it to pass through the golgi membrane to the cytosol and as a result it accumulates generating a flow of cellular water into the golgi to maintain osmotic equilibrium, a process that contributes significantly to the so-called 'aqueous phase' of milk. When the production of lactose is complete, LA is secreted along with lactose by exocytocis in secretory vesicles formed by the golgi membranes (Brew and Grobler, 1992; Linzell and Peaker, 1971).

The mechanism under which the two proteins GT and LA come into contact to form the lactose synthase complex is still unknown. A plausible model has been proposed for the regulatory action of LA by Brew *et al.* (1979) (*Figure* 2.2). According to this model LA molecule should possess proximal areas for interaction with monosaccharides when in complex with the enzyme. Measurements of LA activity that support the assumption concerning the above plausible mechanism have been performed by steady-state kinetic studies of the action of GT in the

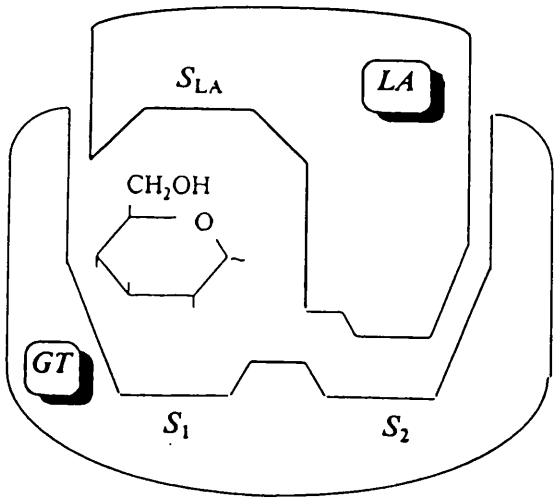


***Figure* 2.2**
Monosaccharide bridge model

absence and presence of LA (Bell *et al.*, 1976; Khatra *et al.*, 1974; Morrison and Ebner, 1971a; Morrison and Ebner, 1971b; Powell and Brew, 1976). The properties also of the covalently cross-linked LS complex and the effect of the presence/absence of various metal ions such as Mn^{2+}, Ca^{2+} or substrates such as monosaccharides or UDP-derivatives have been studied. The emerging results seem to be in favour of the synergistic binding of LA and monosaccharide (glucose) to GT while in the case of oligosaccharides competitive inhibition between LA and the sugar is detected (Bell *et al.*, 1976; Brew *et al.*, 1975).

2.1.2. Description of α-Lactalbumin molecule.

LA is a 14.2 kDa globular 'specifier' protein that consists of a single polypeptide chain of 123 amino acids. An important feature of LA is that its three dimensional structure is similar to that of c-type lysozyme (LZ) (*Figures* 2.3-2.5) and the two proteins are homologous (about 40% sequence identity, *Figure* 2.6) with divergent functions (Acharya *et al.*, 1989; Smith *et al.*, 1987; Stuart *et al.*, 1986).

The overall structure of LA molecule is divided into two domains by a deep cleft, which is the corresponding active site in LZs (*Figure* 2.3). Domain α is comprised of 3 major α-helices and 2 short 3_{10} helices (residues 1-34, 84-123) while domain β consists of a small 3-stranded antiparallel β-pleated sheet and a short 3_{10} helix (residues 35-85). Four disulphide bridges occur between residues 6 and 120, 28 and 111, 61 and 77, 73 and 91, and stabilise the overall polypeptide fold of LA molecule (Acharya *et al.*, 1989).

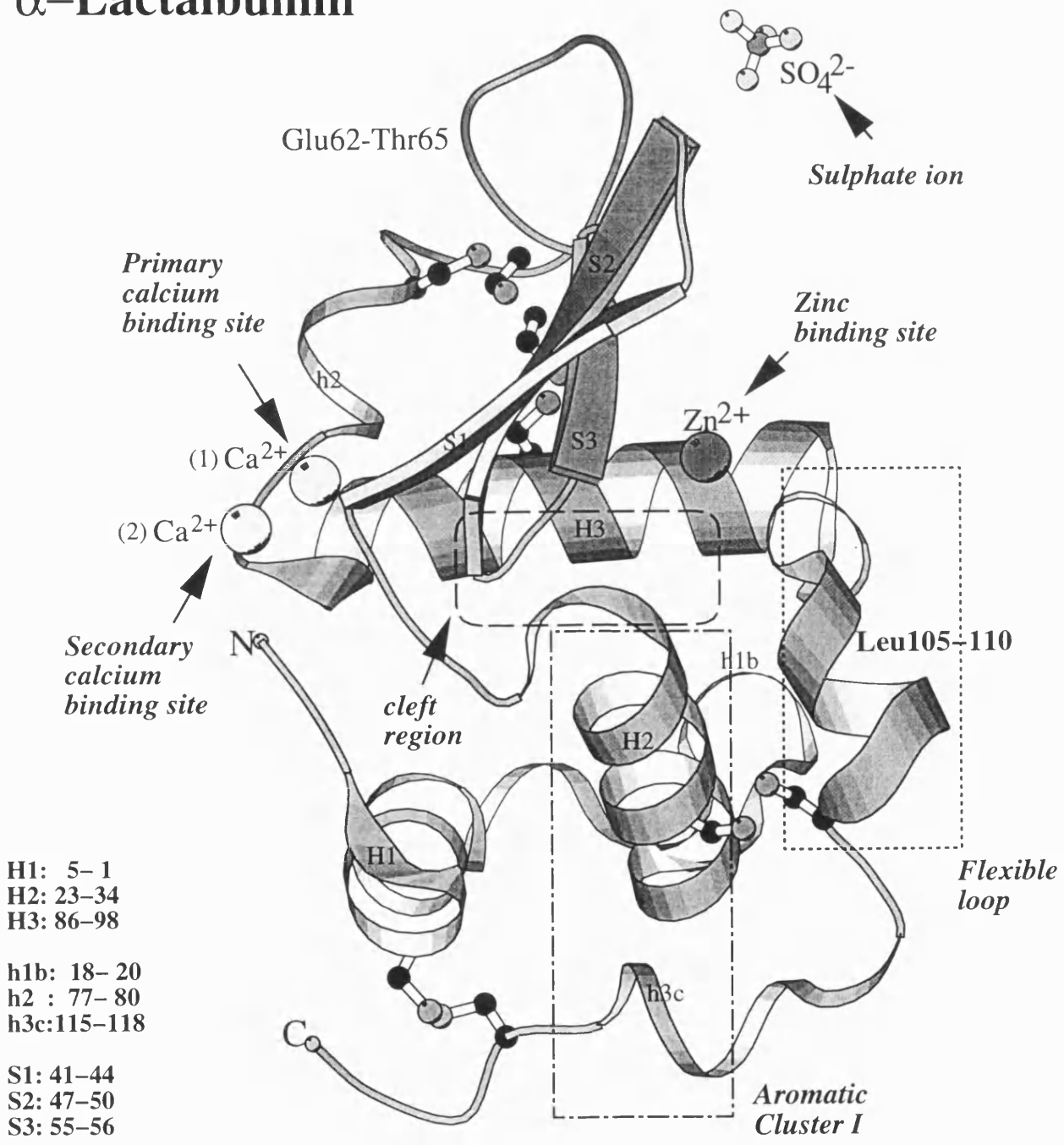


***Figure* 2.3**

Structure of LA (Acharya et al., 1991). The boxed areas represent the functional regions of the molecule. The figure was generated using the program BOBSCRIPT (Esnouf, 1997)

Lysozyme

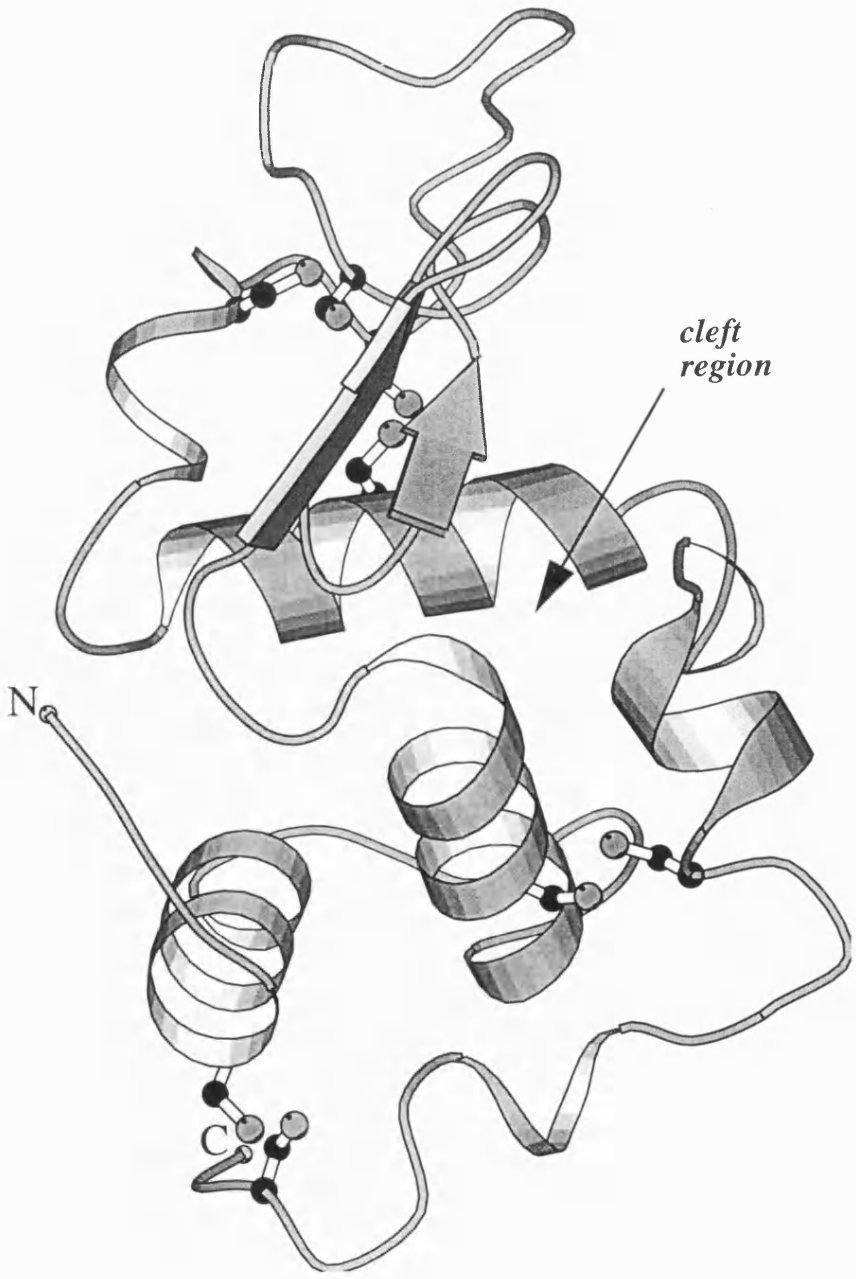


***Figure* 2.4**

Structure of ***hen egg white*** *LZ (Steinrauf, 1998). The figure was generated using the program BOBSCRIPT (Esnouf, 1997)*

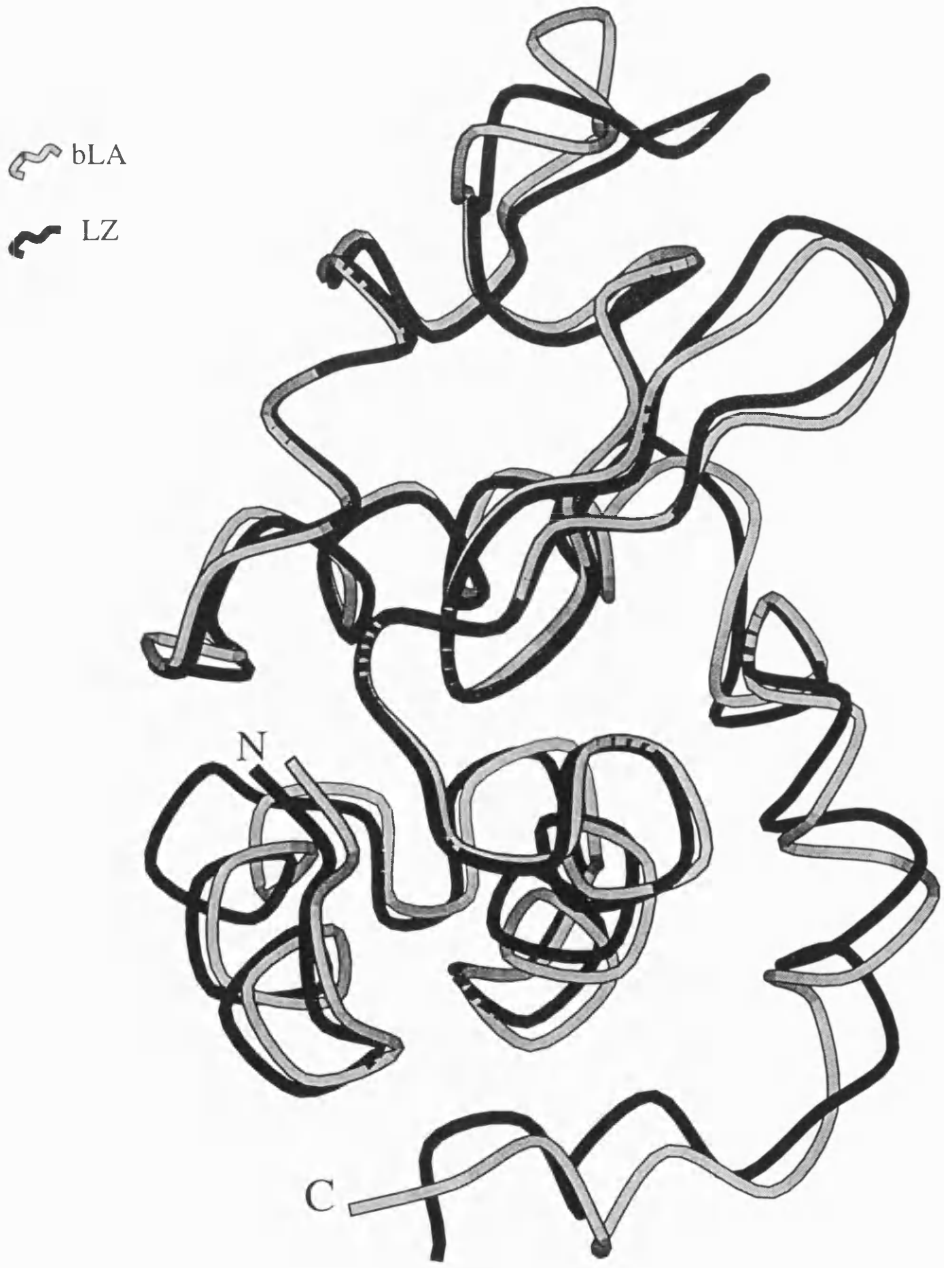


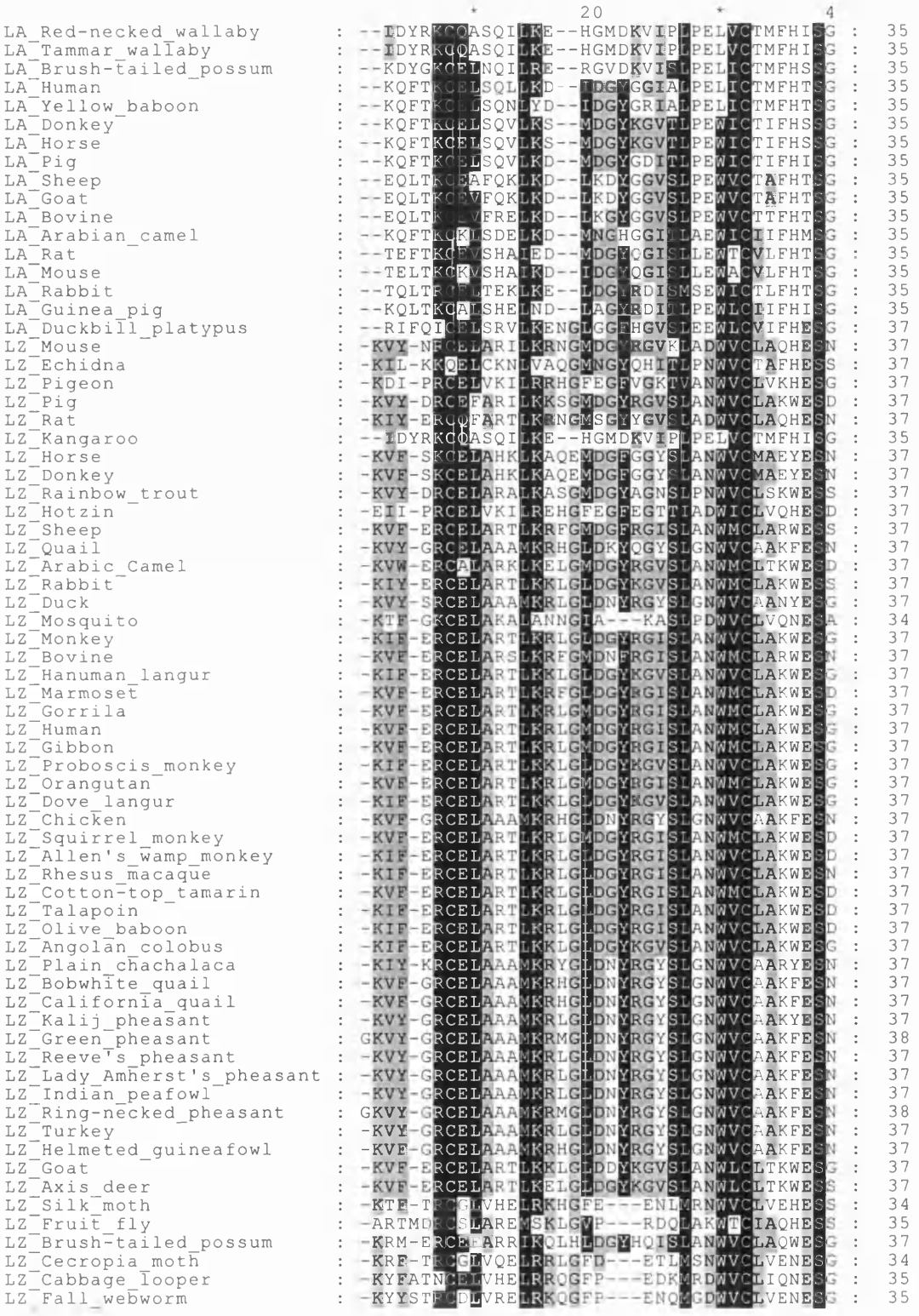
***Figure* 2.5**

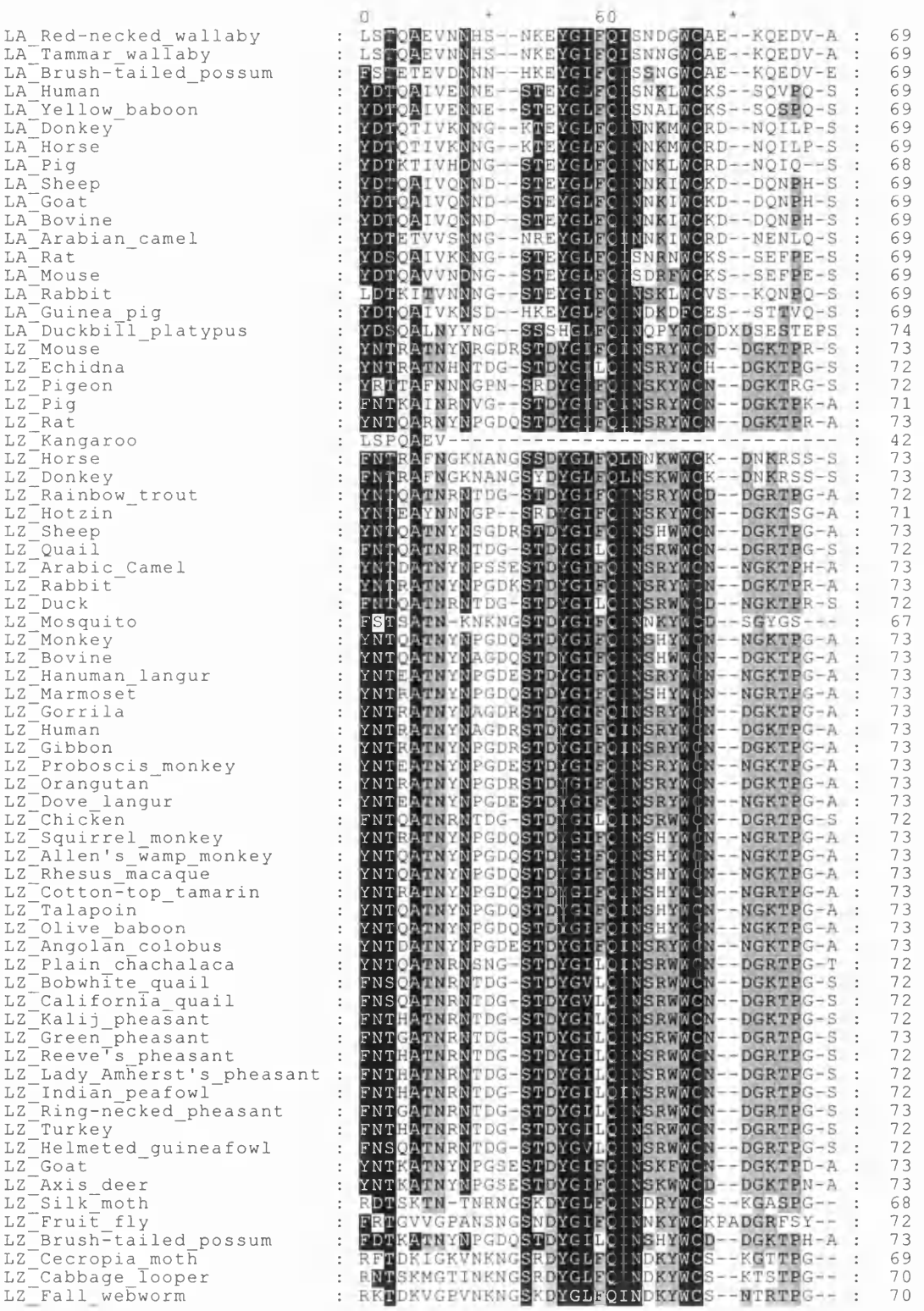
Superposition of LA and LZ structure indicating the similarity of their 3D structure

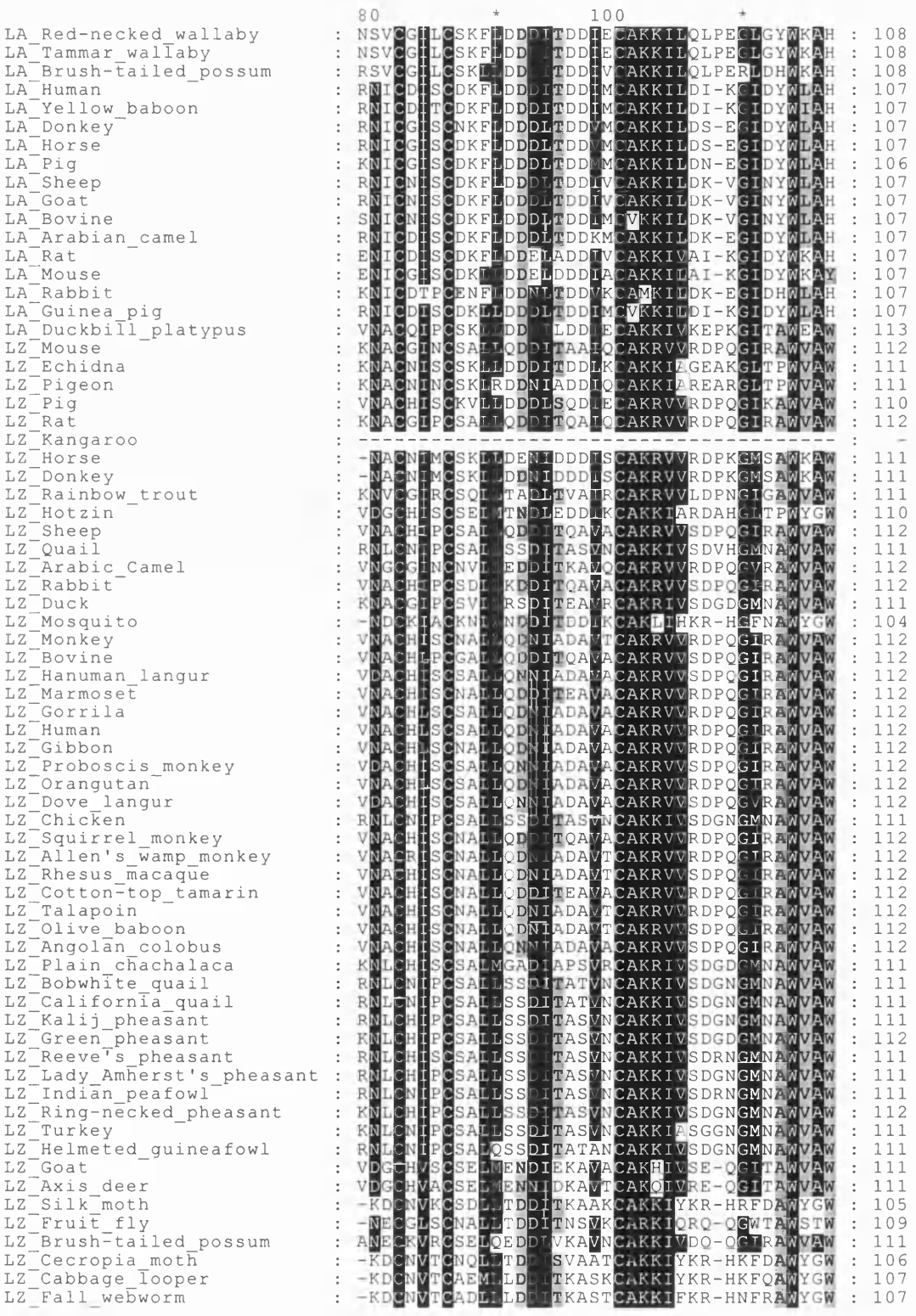
LA binds calcium strongly (also other metal ions such as Zn^{2+}, Mn^{2+}, Co^{2+}, Tb^{3+}, Na^{+}, K^{+}) and specifically whereas in the LZs there are two subgroups, representing paralogous gene lines, of which one binds Ca^{2+} ion and the other (the 'conventional' LZ group) does not (*Figure* 2.6).

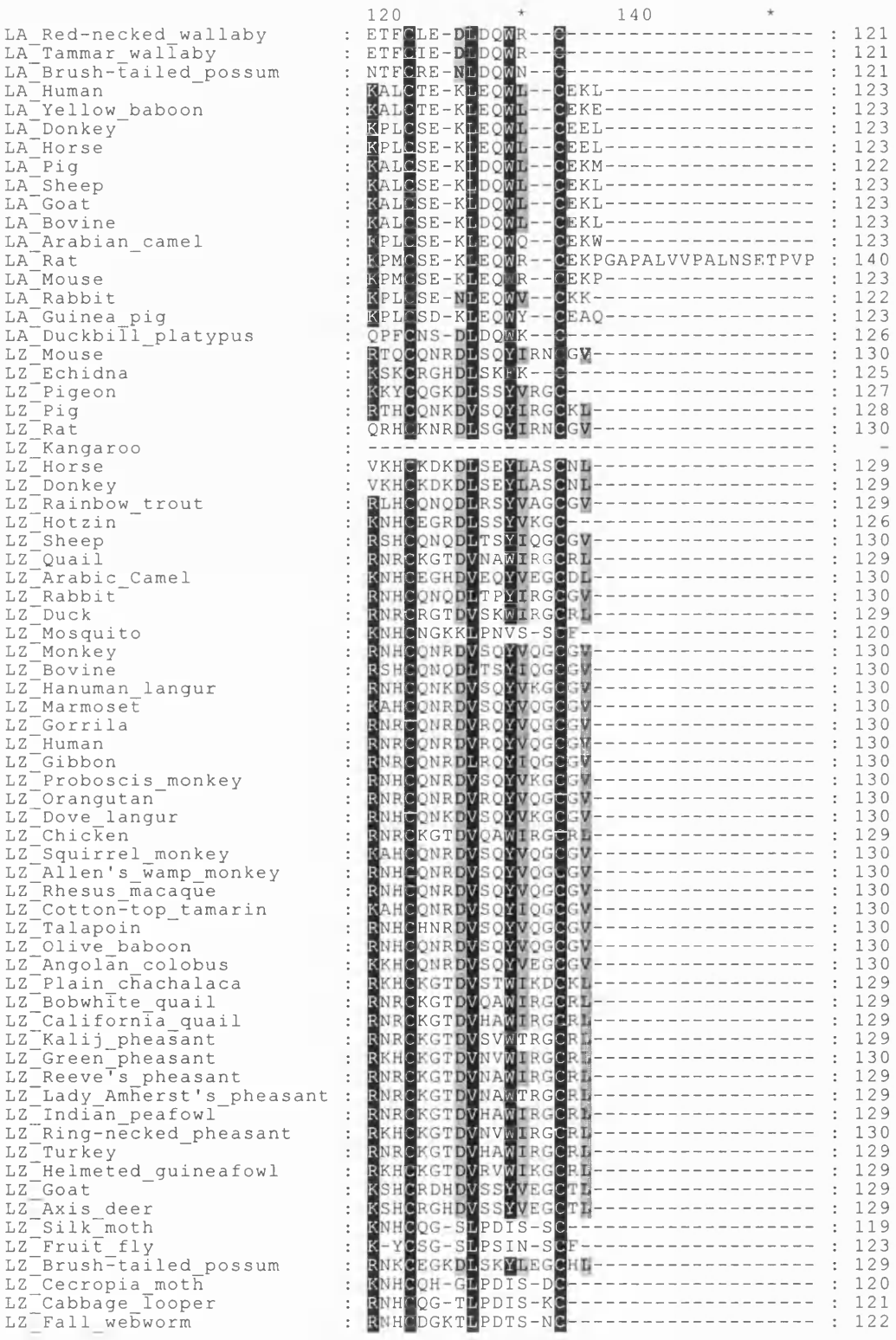
***Figure* 2.6 (overleaf).**

Sequence comparison for LA and LZ. The alignment shows a selection of LA and LZ (Ca^{2+}-binding and non Ca^{2+}-binding) sequences. The shaded areas correspond to conservative residues. The sequence retrieval was performed using the program WU-Blust (version 2.0) and the alignment using the program Clustalx.


LA_Red-necked_wallaby : --IDYRKCQASQILKE--HGMDKVIPLPELVCTMFHISG : 35
LA_Tammar_wallaby : --IDYRKCQASQILKE--HGMDKVIPLPELVCTMFHISG : 35
LA_Brush-tailed_possum : --KDYGKCELNQILRE--RGVDKVISLPELICTMFHSSG : 35
LA_Human : --KQFTKCELSQLLKD--TDGYGGIALPELICTMFHTSG : 35
LA_Yellow_baboon : --KQFTKCELSQNLYD--IDGYGRIALPELICTMFHTSG : 35
LA_Donkey : --KQFTKCELSQVLKS--MDGYKGVTLPEWICTIFHSSG : 35
LA_Horse : --KQFTKCELSQVLKS--MDGYKGVTLPEWICTIFHSSG : 35
LA_Pig : --KQFTKCELSQVLKD--MDGYGDITLPEWICTIFHISG : 35
LA_Sheep : --EQLTKCEAFQKLKD--LKDYGGVSLPEWVCTAFHTSG : 35
LA_Goat : --EQLTKCEVFQKLKD--LKDYGGVSLPEWVCTAFHTSG : 35
LA_Bovine : --EQLTK_EVFRELKD--LKGYGGVSLPEWVCTTFHTSG : 35
LA_Arabian_camel : --KQFTKCKLSDELKD--MNGHGGITLAEWICIIFHMSG : 35
LA_Rat : --TEFTKCEVSHAIED--MDGYQGISLLEWTCVLFHTSG : 35
LA_Mouse : --TELTKCKVSHAIKD--IDGYQGISLLEWACVLFHTSG : 35
LA_Rabbit : --TQLTKCELTEKLKE--LDGYRDISMSEWICTLFHTSG : 35
LA_Guinea_pig : --KQLTKCALSHELND--LAGYRDITLPEWLCTIFHISG : 35
LA_Duckbill_platypus : --RIFQICELSRVLKENGLGGFHGVSLEEWLCVIFHESG : 37
LZ_Mouse : -KVY-NFCELARILKRNGMDGYRGVKLADWVCLAQHESN : 37
LZ_Echidna : -KIL-KKQELCKNLVAQGMNGYQHITLPNWVCTAFHESS : 37
LZ_Pigeon : -KDI-PRCELVKILRRHGFEGFVGKTVANWVCLVKHESG : 37
LZ_Pig : -KVY-DRCEFARILKKSGMDGYRGVSLANWVCLAKWESD : 37
LZ_Rat : -KIY-ERCQFARTLKENGMSGYYGVSLADWVCLAQHESN : 37
LZ_Kangaroo : --IDYRKCQASQILKE--HGMDKVIPLPELVCTMFHISG : 35
LZ_Horse : -KVF-SKCELAHKLKAQEMDGFGGYSLANWVCMAEYESN : 37
LZ_Donkey : -KVF-SKCELAHKLKAQEMDGFGGYSLANWVCMAEYESN : 37
LZ_Rainbow_trout : -KVY-DRCELARALKASGMDGYAGNSLPNWVCLSKWESS : 37
LZ_Hotzin : -EII-PRCELVKILREHGFEGFEGTTIADWICLVQHESD : 37
LZ_Sheep : -KVF-ERCELARTLKRFGMDGFRGISLANWMCLARWESS : 37
LZ_Quail : -KVY-GRCELAAAMKRHGLDKYQGYSLGNWVCAAKFESN : 37
LZ_Arabic_Camel : -KVW-ERCALARKLKELGMDGYRGVSLANWMCLTKWESD : 37
LZ_Rabbit : -KIY-ERCELARTLKKLGLDGYKGVSLANWMCLAKWESS : 37
LZ_Duck : -KVY-SRCELAAAMKRLGLDNYRGYSLGNWVCAANYESG : 37
LZ_Mosquito : -KTF-GKCELAKALANNGIA---KASLPDWVCLVQNESA : 34
LZ_Monkey : -KIF-ERCELARTLKRLGLDGYRGISLANWVCLAKWESG : 37
LZ_Bovine : -KVF-ERCELARSLKRFGMDNFRGISLANWMCLARWESN : 37
LZ_Hanuman_langur : -KIF-ERCELARTLKKLGLDGYKGVSLANWVCLAKWESG : 37
LZ_Marmoset : -KVF-ERCELARTLKRFGLDGYRGISLANWMCLAKWESD : 37
LZ_Gorrila : -KVF-ERCELARTLKRLGMDGYRGISLANWMCLAKWESG : 37
LZ_Human : -KVF-ERCELARTLKRLGMDGYRGISLANWMCLAKWESG : 37
LZ_Gibbon : -KVF-ERCELARTLKRLGMDGYRGISLANWMCLAKWESG : 37
LZ_Proboscis_monkey : -KIF-ERCELARTLKKLGLDGYKGVSLANWVCLAKWESG : 37
LZ_Orangutan : -KVF-ERCELARTLKRLGMDGYRGISLANWMCLAKWESG : 37
LZ_Dove_langur : -KIF-ERCELARTLKKLGLDGYKGVSLANWVCLAKWESG : 37
LZ_Chicken : -KVF-GRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Squirrel_monkey : -KVF-ERCELARTLKRLGMDGYRGISLANWMCLAKWESD : 37
LZ_Allen's_wamp_monkey : -KIF-ERCELARTLKRLGLDGYRGISLANWVCLAKWESD : 37
LZ_Rhesus_macaque : -KIF-ERCELARTLKRLGLDGYRGISLANWVCLAKWESN : 37
LZ_Cotton-top_tamarin : -KVF-ERCELARTLKRLGLDGYRGISLANWMCLAKWESD : 37
LZ_Talapoin : -KIF-ERCELARTLKRLGLDGYRGISLANWVCLAKWESD : 37
LZ_Olive_baboon : -KIF-ERCELARTLKRLGLDGYRGISLANWVCLAKWESD : 37
LZ_Angolan_colobus : -KIF-ERCELARTLKKLGLDGYKGVSLANWVCLAKWESG : 37
LZ_Plain_chachalaca : -KIY-KRCELAAAMKRYGLDNYRGYSLGNWVCAARYESN : 37
LZ_Bobwhite_quail : -KVF-GRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESN : 37
LZ_California_quail : -KVF-GRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Kalij_pheasant : -KVY-GRCELAAAMKRLGLDNYRGYSLGNWVCAAKYESN : 37
LZ_Green_pheasant : GKVY-GRCELAAAMKRMGLDNYRGYSLGNWVCAAKFESN : 38
LZ_Reeve's_pheasant : -KVY-GRCELAAAMKRLGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Lady_Amherst's_pheasant : -KVY-GRCELAAAMKRLGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Indian_peafowl : -KVY-GRCELAAAMKRLGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Ring-necked_pheasant : GKVY-GRCELAAAMKRMGLDNYRGYSLGNWVCAAKFESN : 38
LZ_Turkey : -KVY-GRCELAAAMKRLGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Helmeted_guineafowl : -KVF-GRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESN : 37
LZ_Goat : -KVF-ERCELARTLKKLGLDDYKGVSLANWLCLTKWESG : 37
LZ_Axis_deer : -KVF-ERCELARTLKELGLDGYKGVSLANWLCLTKWESS : 37
LZ_Silk_moth : -KTF-TKCGLVHELRKHGFE---ENLMRNWVCLVEHESS : 34
LZ_Fruit_fly : -ARTMDKCSLARELSKLGVP---RDQLAKWTCIAQHESS : 35
LZ_Brush-tailed_possum : -KRM-ERCEFARRIKQLHLDGYHQISLANWVCLAQWESG : 37
LZ_Cecropia_moth : -KRF-TKCGLVQELRRLGFD---ETLMSNWVCLVENESG : 34
LZ_Cabbage_looper : -KYFATNCELVHELRRQGFP---EDKMRDWVCLIQNESG : 35
LZ_Fall_webworm : -KYYSTFCDLVRELRKQGFP---ENQMGDWVCLVENESG : 35


0 + 60 +
LA_Red-necked_wallaby : LSTQAEVNNHS--NKEYGIFQISNDGWCAE--KQEDV-A : 69
LA_Tammar_wallaby : LSTQAEVNNHS--NKEYGIFQISNNGWCAE--KQEDV-A : 69
LA_Brush-tailed_possum : FSTETEVDNNN--HKEYGIFQISSNGWCAE--KQEDV-E : 69
LA_Human : YDTQAIVENNE--STEYGLFQISNKLWCKS--SQVPQ-S : 69
LA_Yellow_baboon : YDTQAIVENNE--STEYGLFQISNALWCKS--SQSPQ-S : 69
LA_Donkey : YDTQTIVKNNG--KTEYGLFQINNKMWCRD--NQILP-S : 69
LA_Horse : YDTQTIVKNNG--KTEYGLFQINNKMWCRD--NQILP-S : 69
LA_Pig : YDTKTIVHDNG--STEYGLFQINNKLWCRD--NQIQ--S : 68
LA_Sheep : YDTQAIVQNND--STEYGLFQINNKIWCKD--DQNPH-S : 69
LA_Goat : YDTQAIVQNND--STEYGLFQINNKIWCKD--DQNPH-S : 69
LA_Bovine : YDTQAIVQNND--STEYGLFQINNKIWCKD--DQNPH-S : 69
LA_Arabian_camel : YDTETVVSNNG--NREYGLFQINNKIWCRD--NENLQ-S : 69
LA_Rat : YDSQAIVKNNG--STEYGLFQISNRNWCKS--SEFPE-S : 69
LA_Mouse : YDTQAVVNDNG--STEYGLFQISDRFWCKS--SEFPE-S : 69
LA_Rabbit : LDTKITVNNNG--STEYGIFQINSKLWCVS--KQNPQ-S : 69
LA_Guinea_pig : YDTQAIVKNSD--HKEYGLFQINDKDFCES--STTVQ-S : 69
LA_Duckbill_platypus : YDSQALNYYNG--SSSHGLFQINQPYWCDDXDSESTEPS : 74
LZ_Mouse : YNTRATNYNRGDRSTDYGIFQINSRYWCN--DGKTPR-S : 73
LZ_Echidna : YNTRATNHNTDG-STDYGILQINSRYWCH--DGKTPG-S : 72
LZ_Pigeon : YRTTAFNNNGPN-SRDYGIFQINSKYWCN--DGKTRG-S : 72
LZ_Pig : FNTKAINRNVG--STDYGIFQINSRYWCN--DGKTPK-A : 71
LZ_Rat : YNTQARNYNPGDQSTDYGIFQINSRYWCN--DGKTPR-A : 73
LZ_Kangaroo : LSPQAEV---------- : 42
LZ_Horse : FNTRAFNGKNANGSSDYGLFQLNNKWWCK--DNKRSS-S : 73
LZ_Donkey : FNTRAFNGKNANGSYDYGLFQLNSKWWCK--DNKRSS-S : 73
LZ_Rainbow_trout : YNTQATNRNTDG-STDYGIFQINSRYWCD--DGRTPG-A : 72
LZ_Hotzin : YNTEAYNNNGP--SRDYGIFQINSKYWCN--DGKTSG-A : 71
LZ_Sheep : YNTQATNYNSGDRSTDYGIFQINSHWWCN--DGKTPG-A : 73
LZ_Quail : FNTQATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 72
LZ_Arabic_Camel : YNTDATNYNPSSESTDYGIFQINSRYWCN--NGKTPH-A : 73
LZ_Rabbit : YNTRATNYNPGDKSTDYGIFQINSRYWCN--DGKTPR-A : 73
LZ_Duck : FNTQATNRNTDG-STDYGILQINSRWWCD--NGKTPR-S : 72
LZ_Mosquito : FSTSATN-KNKNGSTDYGIFQINNKYWCD--SGYGS--- : 67
LZ_Monkey : YNTQATNYNPGDQSTDYGIFQINSHYWCN--NGKTPG-A : 73
LZ_Bovine : YNTQATNYNAGDQSTDYGIFQINSHWWCN--DGKTPG-A : 73
LZ_Hanuman_langur : YNTEATNYNPGDESTDYGIFQINSRYWCN--NGKTPG-A : 73
LZ_Marmoset : YNTRATNYNPGDQSTDYGIFQINSHYWCN--NGRTPG-A : 73
LZ_Gorrila : YNTRATNYNAGDRSTDYGIFQINSRYWCN--DGKTPG-A : 73
LZ_Human : YNTRATNYNAGDRSTDYGIFQINSRYWCN--DGKTPG-A : 73
LZ_Gibbon : YNTRATNYNPGDRSTDYGIFQINSRYWCN--DGKTPG-A : 73
LZ_Proboscis_monkey : YNTEATNYNPGDESTDYGIFQINSRYWCN--NGKTPG-A : 73
LZ_Orangutan : YNTRATNYNPGDRSTDYGIFQINSRYWCN--DGKTPG-A : 73
LZ_Dove_langur : YNTEATNYNPGDESTDYGIFQINSRYWCN--NGKTPG-A : 73
LZ_Chicken : FNTQATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 72
LZ_Squirrel_monkey : YNTRATNYNPGDQSTDYGIFQINSHYWCN--NGRTPG-A : 73
LZ_Allen's_wamp_monkey : YNTQATNYNPGDQSTDYGIFQINSHYWCN--NGKTPG-A : 73
LZ_Rhesus_macaque : YNTQATNYNPGDQSTDYGIFQINSHYWCN--NGKTPG-A : 73
LZ_Cotton-top_tamarin : YNTRATNYNPGDQSTDYGIFQINSHYWCN--NGRTPG-A : 73
LZ_Talapoin : YNTQATNYNPGDQSTDYGIFQINSHYWCN--NGKTPG-A : 73
LZ_Olive_baboon : YNTQATNYNPGDQSTDYGIFQINSHYWCN--NGKTPG-A : 73
LZ_Angolan_colobus : YNTDATNYNPGDESTDYGIFQINSRYWCN--NGKTPG-A : 73
LZ_Plain_chachalaca : YNTQATNRNSNG-STDYGILQINSRWWCN--DGRTPG-T : 72
LZ_Bobwhite_quail : FNSQATNRNTDG-STDYGVLQINSRWWCN--DGKTPG-S : 72
LZ_California_quail : FNSQATNRNTDG-STDYGVLQINSRWWCN--DGRTPG-S : 72
LZ_Kalij_pheasant : FNTHATNRNTDG-STDYGILQINSRWWCN--DGKTPG-S : 72
LZ_Green_pheasant : FNTGATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 73
LZ_Reeve's_pheasant : FNTHATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 72
LZ_Lady_Amherst's_pheasant : FNTHATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 72
LZ_Indian_peafowl : FNTHATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 72
LZ_Ring-necked_pheasant : FNTGATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 73
LZ_Turkey : FNTHATNRNTDG-STDYGILQINSRWWCN--DGRTPG-S : 72
LZ_Helmeted_guineafowl : FNSQATNRNTDG-STDYGVLQINSRWWCN--DGRTPG-S : 72
LZ_Goat : YNTKATNYNPGSESTDYGIFQINSKFWCN--DGKTPD-A : 73
LZ_Axis_deer : YNTKATNYNPGSESTDYGIFQINSKWWCD--DGKTPN-A : 73
LZ_Silk_moth : RDTSKTN-TNRNGSKDYGLFQINDRYWCS--KGASPG-- : 68
LZ_Fruit_fly : FRTGVVGPANSNGSNDYGIFQINNKYWCKPADGRFSY-- : 72
LZ_Brush-tailed_possum : FDTKATNYNPGDQSTDYGILQINSHYWCD--DGKTPH-A : 73
LZ_Cecropia_moth : RFTDKIGKVNKNGSRDYGLFQINDKYWCS--KGTTPG-- : 69
LZ_Cabbage_looper : RNTSKMGTINKNGSRDYGLFQINDKYWCS--KTSTPG-- : 70
LZ_Fall_webworm : RKTDKVGPVNKNGSKDYGLFQINDKYWCS--NTRTPG-- : 70


80 * 100 *
LA_Red-necked_wallaby : NSVCGILCSKFLDDDITDDIECAKKILQLPECLGYWKAH : 108
LA_Tammar_wallaby : NSVCGILCSKFLDDDITDDIECAKKILQLPEGLGYWKAH : 108
LA_Brush-tailed_possum : RSVCGILCSKLLDDDITDDIVCAKKILQLPERLDHWKAH : 108
LA_Human : RNICDISCDKFLDDDITDDIMCAKKILDI-KGIDYWLAH : 107
LA_Yellow_baboon : RNICDITCDKFLDDDITDDIMCAKKILDI-KGIDYWIAH : 107
LA_Donkey : RNICGISCNKFLDDDLTDDVMCAKKILDS-EGIDYWLAH : 107
LA_Horse : RNICGISCDKFLDDDLTDDVMCAKKILDS-EGIDYWLAH : 107
LA_Pig : KNICGISCDKFLDDDLTDDVMCAKKILDN-EGIDYWLAH : 106
LA_Sheep : RNICNISCDKFLDDDLTDDIVCAKKILDK-VGINYWLAH : 107
LA_Goat : RNICNISCDKFLDDDLTDDIVCAKKILDK-VGINYWLAH : 107
LA_Bovine : SNICNISCDKFLDDDLTDDIMCVKKILDK-VGINYWLAH : 107
LA_Arabian_camel : RNICDISCDKFLDDDLTDDKMCAKKILDK-EGIDYWLAH : 107
LA_Rat : ENICDISCDKFLDDELADDIVCAKKIVAI-KGIDYWKAH : 107
LA_Mouse : ENICGISCDKLLDDELDDDIACAKKILAI-KGIDYWKAY : 107
LA_Rabbit : KNICDTPCENFLDDNLTDDVKCAMKILDK-EGIDHWLAH : 107
LA_Guinea_pig : RNICDISCDKLLDDDLTDDIMCVKKILDI-KGIDYWLAH : 107
LA_Duckbill_platypus : VNACQIPCSKLLDDILDDIECAKKIVKEPKGITAWEAW : 113
LZ_Mouse : KNACGINCSALLQDDITAAIQCAKRVVRDPQGIRAWVAW : 112
LZ_Echidna : KNACNISCSKLLDDDITDDLKCAKKIAGEAKGLTPWVAW : 111
LZ_Pigeon : KNACNINCSKLRDDNIADDIQCAKKIAREARGLTPWVAW : 111
LZ_Pig : VNACHISCKVLLDDDLSQDLECAKRVVRDPQGIKAWVAW : 110
LZ_Rat : KNACGIPCSALLQDDITQAIQCAKRVVRDPQGIRAWVAW : 112
LZ_Kangaroo : ---------- : -
LZ_Horse : -NACNIMCSKLLDENIDDDISCAKRVVRDPKGMSAWKAW : 111
LZ_Donkey : -NACNIMCSKLLDDNIDDDISCAKRVVRDPKGMSAWKAW : 111
LZ_Rainbow_trout : KNVCGIRCSQLLTALLTVAIRCAKRVVLDPNGIGAWVAW : 111
LZ_Hotzin : VDGCHISCSELLTNDLEDDLKCAKKIARDAHGLTPWYGW : 110
LZ_Sheep : VNACHIPCSALLQDDITQAVACAKRVVSDPQGIRAWVAW : 112
LZ_Quail : RNLCNIPCSALLSSDITASVNCAKKIVSDVHGMNAWVAW : 111
LZ_Arabic_Camel : VNGCGINCNVLLEDDITKAVQCAKRVVRDPQGVRAWVAW : 112
LZ_Rabbit : VNACHIPCSDLLKDDITQAVACAKRVVSDPQGIRAWVAW : 112
LZ_Duck : KNACGIPCSVLLRSDITEAVRCAKRIVSDGDGMNAWVAW : 111
LZ_Mosquito : -NDCKIACKNLLNDDITDDIKCAKLIHKR-HGFNAWYGW : 104
LZ_Monkey : VNACHISCNALLQDNIADAVTCAKRVVRDPQGIRAWVAW : 112
LZ_Bovine : VNACHLPCGALLQDDITQAVACAKRVVSDPQGIRAWVAW : 112
LZ_Hanuman_langur : VDACHISCSALLQNNIADAVACAKRVVSDPQGIRAWVAW : 112
LZ_Marmoset : VNACHISCNALLQDDITEAVACAKRVVRDPQGIRAWVAW : 112
LZ_Gorrila : VNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAW : 112
LZ_Human : VNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAW : 112
LZ_Gibbon : VNACHLSCNALLQDNIADAVACAKRVVRDPQGIRAWVAW : 112
LZ_Proboscis_monkey : VDACHISCSALLQNNIADAVACAKRVVSDPQGIRAWVAW : 112
LZ_Orangutan : VNACHLSCSALLQDNIADAVACAKRVVRDPQGTRAWVAW : 112
LZ_Dove_langur : VDACHISCSALLQNNIADAVACAKRVVSDPQGVRAWVAW : 112
LZ_Chicken : RNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAW : 111
LZ_Squirrel_monkey : VNACHISCNALLQDDITQAVACAKRVVRDPQGIRAWVAW : 112
LZ_Allen's_wamp_monkey : VNACRISCNALLQDNIADAVTCAKRVVRDPQGIRAWVAW : 112
LZ_Rhesus_macaque : VNACHISCNALLQDNIADAVTCAKRVVSDPQGIRAWVAW : 112
LZ_Cotton-top_tamarin : VNACHISCNALLQDDITEAVACAKRVVRDPQGIRAWVAW : 112
LZ_Talapoin : VNACHISCNALLQDNIADAVTCAKRVVRDPQGIRAWVAW : 112
LZ_Olive_baboon : VNACHISCNALLQDNIADAVTCAKRVVSDPQGIRAWVAW : 112
LZ_Angolan_colobus : VNACHISCNALLQNNIADAVACAKRVVSDPQGIRAWVAW : 112
LZ_Plain_chachalaca : KNLCHISCSALMGADIAPSVRCAKRIVSDGDGMNAWVAW : 111
LZ_Bobwhite_quail : RNLCNIPCSALLSSDITATVNCAKKIVSDGNGMNAWVAW : 111
LZ_California_quail : RNLCNIPCSALLSSDITATVNCAKKIVSDGNGMNAWVAW : 111
LZ_Kalij_pheasant : RNLCHIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAW : 111
LZ_Green_pheasant : KNLCHIPCSALLSSDITASVNCAKKIVSDGDGMNAWVAW : 112
LZ_Reeve's_pheasant : RNLCHIPCSALLSSDITASVNCAKKIVSDRNGMNAWVAW : 111
LZ_Lady_Amherst's_pheasant : RNLCHIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAW : 111
LZ_Indian_peafowl : RNLCNIPCSALLSSDITASVNCAKKIVSDRNGMNAWVAW : 111
LZ_Ring-necked_pheasant : KNLCHIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAW : 112
LZ_Turkey : KNLCNIPCSALLSSDITASVNCAKKIASGGNGMNAWVAW : 111
LZ_Helmeted_guineafowl : RNLCNIPCSALQSSDITATANCAKKIVSDGNGMNAWVAW : 111
LZ_Goat : VDGCHVSCSELMENDIEKAVACAKHIVSE-QGITAWVAW : 111
LZ_Axis_deer : VDGCHVACSELMENNIDKAVTCAKQIVRE-QGITAWVAW : 111
LZ_Silk_moth : -KDCNVKCSDLLTDDITKAAKCAKKIYKR-HRFDAWYGW : 105
LZ_Fruit_fly : -NECGLSCNALLTDDITNSVKCARKIQRQ-QGWTAWSTW : 109
LZ_Brush-tailed_possum : ANECKVRCSELQEDDIVKAVNCAKKIVDQ-QGIRAWVAW : 111
LZ_Cecropia_moth : -KDCNVTCNQLLTDDISVAATCAKKIYKR-HKFDAWYGW : 106
LZ_Cabbage_looper : -KDCNVTCAEMLLDDITKASKCAKKIYKR-HKFQAWYGW : 107
LZ_Fall_webworm : -KDCNVTCADLLLDDITKASTCAKKIFKR-HNFRAWYGW : 107


120 * 140 *
LA_Red-necked_wallaby : ETFCLE-DLDQWR--C---------- : 121
LA_Tammar_wallaby : ETFCIE-DLDQWR--C---------- : 121
LA_Brush-tailed_possum : NTFCRE-NLDQWN--C---------- : 121
LA_Human : KALCTE-KLEQWL--CEKL---------- : 123
LA_Yellow_baboon : KALCTE-KLEQWL--CEKE---------- : 123
LA_Donkey : KPLCSE-KLEQWL--CEEL---------- : 123
LA_Horse : KPLCSE-KLEQWL--CEEL---------- : 123
LA_Pig : KALCSE-KLDQWL--CEKM---------- : 122
LA_Sheep : KALCSE-KLDQWL--CEKL---------- : 123
LA_Goat : KALCSE-KLDQWL--CEKL---------- : 123
LA_Bovine : KALCSE-KLDQWL--CEKL---------- : 123
LA_Arabian_camel : KPLCSE-KLEQWQ--CEKW---------- : 123
LA_Rat : KPMCSE-KLEQWR--CEKPGAPALVVPALNSETPVP : 140
LA_Mouse : KPMCSE-KLEQWR--CEKP---------- : 123
LA_Rabbit : KPLCSE-NLEQWV--CKK---------- : 122
LA_Guinea_pig : KPLCSD-KLEQWY--CEAQ---------- : 123
LA_Duckbill_platypus : QPFCNS-DLDQWK--C---------- : 126
LZ_Mouse : RTQCQNRDLSQYIRNCGV---------- : 130
LZ_Echidna : KSKCRGHDLSKFK--C---------- : 125
LZ_Pigeon : KKYCQGKDLSSYVRGC---------- : 127
LZ_Pig : RTHCQNKDVSQYIRGCKL---------- : 128
LZ_Rat : QRHCKNRDLSGYIRNCGV---------- : 130
LZ_Kangaroo : ---------- : -
LZ_Horse : VKHCKDKDLSEYLASCNL---------- : 129
LZ_Donkey : VKHCKDKDLSEYLASCNL---------- : 129
LZ_Rainbow_trout : RLHCQNQDLRSYVAGCGV---------- : 129
LZ_Hotzin : KNHCEGRDLSSYVKGC---------- : 126
LZ_Sheep : RSHCQNQDLTSYIQGCGV---------- : 130
LZ_Quail : RNRCKGTDVNAWIRGCRL---------- : 129
LZ_Arabic_Camel : KNHCEGHDVEQYVEGCDL---------- : 130
LZ_Rabbit : RNHCQNQDLTPYIRGCGV---------- : 130
LZ_Duck : RNRCRGTDVSKWIRGCRL---------- : 129
LZ_Mosquito : KNHCNGKKLPNVS-SCF---------- : 120
LZ_Monkey : RNHCQNRDVSQYVQGCGV---------- : 130
LZ_Bovine : RSHCQNQDLTSYIQGCGV---------- : 130
LZ_Hanuman_langur : RNHCQNKDVSQYVKGCGV---------- : 130
LZ_Marmoset : KAHCQNRDVSQYVQGCGV---------- : 130
LZ_Gorrila : RNRCQNRDVRQYVQGCGV---------- : 130
LZ_Human : RNRCQNRDVRQYVQGCGV---------- : 130
LZ_Gibbon : RNRCQNRDLRQYIQGCGV---------- : 130
LZ_Proboscis_monkey : RNHCQNRDVSQYVKGCGV---------- : 130
LZ_Orangutan : RNRCQNRDVRQYVQGCGV---------- : 130
LZ_Dove_langur : RNHCQNKDVSQYVKGCGV---------- : 130
LZ_Chicken : RNRCKGTDVQAWIRGCRL---------- : 129
LZ_Squirrel_monkey : KAHCQNRDVSQYVQGCGV---------- : 130
LZ_Allen's_wamp_monkey : RNHCQNRDVSQYVQGCGV---------- : 130
LZ_Rhesus_macaque : RNHCQNRDVSQYVQGCGV---------- : 130
LZ_Cotton-top_tamarin : KAHCQNRDVSQYIQGCGV---------- : 130
LZ_Talapoin : RNHCHNRDVSQYVQGCGV---------- : 130
LZ_Olive_baboon : RNHCQNRDVSQYVQGCGV---------- : 130
LZ_Angolan_colobus : KKHCQNRDVSQYVEGCGV---------- : 130
LZ_Plain_chachalaca : RKHCKGTDVSTWIKDCKL---------- : 129
LZ_Bobwhite_quail : RNRCKGTDVQAWIRGCRL---------- : 129
LZ_California_quail : RNRCKGTDVHAWIRGCRL---------- : 129
LZ_Kalij_pheasant : RNRCKGTDVSVWTRGCRL---------- : 129
LZ_Green_pheasant : RKHCKGTDVNVWIRGCRL---------- : 130
LZ_Reeve's_pheasant : RNRCKGTDVNAWIRGCRL---------- : 129
LZ_Lady_Amherst's_pheasant : RNRCKGTDVNAWTRGCRL---------- : 129
LZ_Indian_peafowl : RNRCKGTDVHAWIRGCRL---------- : 129
LZ_Ring-necked_pheasant : RKHCKGTDVNVWIRGCRL---------- : 130
LZ_Turkey : RNRCKGTDVHAWIRGCRL---------- : 129
LZ_Helmeted_guineafowl : RKHCKGTDVRVWIKGCRL---------- : 129
LZ_Goat : KSHCRDHDVSSYVEGCTL---------- : 129
LZ_Axis_deer : KSHCRGHDVSSYVEGCTL---------- : 129
LZ_Silk_moth : KNHCQG-SLPDIS-SC---------- : 119
LZ_Fruit_fly : K-YCSG-SLPSIN-SCF---------- : 123
LZ_Brush-tailed_possum : RNKCEGKDLSKYLEGCHL---------- : 129
LZ_Cecropia_moth : KNHCQH-GLPDIS-DC---------- : 120
LZ_Cabbage_looper : RNHCQG-TLPDIS-KC---------- : 121
LZ_Fall_webworm : RNHCDGKTLPDTS-NC---------- : 122

2.1.3. Metal ion binding properties and functional regions of LA.

The K_{app} for Ca^{2+} binding to apo-LA under physiological conditions is of the order of 10^{-7} M (Kronman *et al.*, 1981; Permyakov *et al.*, 1981; Schaer *et al.*, 1985). It is known that bound Ca^{2+} in LA is not required for its activity and can be deduced to be not needed for enhanced stability against denaturation (Ewbank and Creighton, 1991; Rao and Brew, 1989) although it plays an essential structural role in the protein. This may reflect the need by LA for additional free energy to stabilise the native state because of features that are not present in LZ or a regulatory action of calcium. LA and LZ also differ significantly in their molecular stabilities (Greene *et al.*, 1999; Rao and Brew, 1989) and metal ion binding properties. LA can bind a number of different metal ions in solution (Musci and Berliner, 1985a; Musci and Berliner, 1985b; Musci and Berliner, 1986; Permyakov *et al.*, 1991; Permyakov *et al.*, 1988) and exhibits a variety of different conformational states depending on its ligation state (Kronman, 1989; Kronman *et al.*, 1981). It undergoes a more global transconformation to a molten globule state (neither fully folded nor fully unfolded) in response to mildly disruptive treatments such as low pH and exposure to high concentrations of certain metal ions or intermediate concentrations of denaturants (Kronman, 1989; Kuwajima *et al.*, 1989). Exposure of native LA (N state) to pHs below 4.0 promotes formation of a conformer (A state) whose properties are markedly different from those of the native protein. The side chains have increased freedom of movement with no spatial correlation while the molecule is in an apparent folded state, although less compact compared to the N state. Such differences include a markedly increased propensity for self-association and characteristic changes in spectroscopic

properties. The enthalpy of the molten globule state is very nearly the same as that of the fully unfolded state and substantially different from that of the folded state, hence LA molecule is seen as folded although somewhat less compactly than that in the N state, and of more 'fluid-like' character than that of the native protein. This enhanced molecular flexibility of LA, as compared with LZs, may have evolved to suit the functional demands of its role in lactose synthase (Greene *et al.*, 1999).

Crystal structures of different species variants of LA in different crystal forms and metal complexes (Acharya *et al.*, 1991; Acharya *et al.*, 1989; Calderone *et al.*, 1996; Chandra *et al.*, 1998; Harata and Muraki, 1992; Pike *et al.*, 1996; Ren *et al.*, 1993; Stuart *et al.*, 1986) have shown that a high affinity Ca^{2+} binding site is located at the junction of these sub domains and is composed of a continuous section of polypeptide chain (residues 79-88) that contains highly conserved Ca^{2+} liganding aspartates (*Figure* 2.3). Although the presence of this site in LA was originally thought to represent a new evolutionary development (Stuart *et al.*, 1986) recently, several LZs with Ca^{2+} binding sites have been identified and molecular structure analyses indicate that the site is an ancient feature of the superfamily (Acharya *et al.*, 1994; Grobler *et al.*, 1994a).

Recently a secondary Ca^{2+} binding site (different from Zn^{2+} or SO_4^{2-} site), 7.9Å away from the primary binding site was reported (Chandra *et al.*, 1998) *Figure* 2.3. This site had common features with the Mn^{2+} binding site as described by Gerken (1984). Based on the proximity of the Mn^{2+} and Ca^{2+} binding region and the location of the functional site on one side of the charged surface of LA molecule it was proposed that these binding sites might have a definitive role in the

formation of the lactose synthase complex (Chandra *et al.*, 1998). It is worth noting that all of the metal ions that activate GT also bind to LA [for a review see Kronman (1989)].

The functional site of LA encompasses a region corresponding to a subsite in LZ and parts of the two adjacent substructures (Malinovskii *et al.*, 1996). The available structural information on LA has demonstrated that part of the interaction site for GT (involving residues 105–110) adopts two distinct conformations and that it is probably highly fluctuating in solution. The environments of both residues in LA that have been implicated in stabilising glucose binding by lactose synthase (Phe-31 and His-32) are affected by the conformation of the 'flexible-loop' region (*Figure* 2.3) and may be critical for LA's modulatory properties. Certainly, the unusual degree of conformational adaptability exhibited by LA in solution, depending on its interaction with metal ions, and the general observation that the interaction site is localised in one of the most flexible parts of the molecule, suggests that LA's ability to adjust its conformational state may be important in the regulatory mechanisms mediated by this unusual protein (Malinovskii *et al.*, 1996; Pike *et al.*, 1996).

2.1.4. Site-directed mutagenesis as a tool to understand the functional regions of LA molecule.

In the absence of a 3D structure for the LS complex, only tentative conclusions could be made about the role of specific residues from kinetic results from the variants designed using site directed mutagenesis. However, it is possible to obtain significant amount of knowledge on the altered functional property of a variant by following a

structure-based approach in combination with 'Protein engineering experiments'.

Bovine LA was used as a model system. A series of LA variants at regions proposed to be directly involved in LA action in LS complex were produced in our collaborator's laboratory (Prof. Keith Brew, University of Miami School of Medicine, USA) according to the protocol described by Grobler *et al.* (1994b). The sequence changes that have been introduced to the recombinant bovine protein (mLA) are summarised in Table 2.1.

Briefly, a facile bacterial expression system for a variant of bovine LA, which could yield fully active recombinant LA (mLA) was generated in high yields. This system succeeds an earlier system in which bovine LA was expressed as a fusion protein from which a native, active protein was generated by limited proteolysis after extraction and treatment to generate native folding and disulphide bond formation (Wang *et al.*, 1989).

Single site substitutions targeted selectively to residues of the aromatic clusters I and II, the functional regions of the protein, lying adjacent to the cleft, critical for both LA activity and glucose binding in LA complex (Grobler *et al.*, 1994b; Malinovskii *et al.*, 1996; Pike *et al.*, 1996).

A detailed structural investigation was performed for three LA variants (*Figure* 2.7): Ala109-Pro, Tyr103-Pro and Trp118-His, which seem to possess unusual kinetic properties, while crystallisation conditions for additional three mutants: His32-Tyr, Phe31-Tyr, Lys114-Asn have been established.

Table 2.1 Amino acid substitutions in the functional sites of LA molecule

Functional Regions	*Residues involved*	*Sequence changes*
aromatic cluster I	Phe 31	Phe31-Tyr
	His 32	His32-Tyr
	Gln 117	His32-Ala
	Trp 118	Trp118-His
	Tyr 36	
aromatic cluster II	Trp 26	
	Phe 53	
	Trp 60	
	Tyr 103	Tyr103-Pro
	Trp 104	
flexible-loop	Residues 105-110	His107-Ala
		His107-Trp
		His107-Tyr
		Ala109-Pro
		Leu110-Glu
		Leu110-His
		Leu110-Arg
other mutants	Cys 6	Cys6-Ser
	Ile 55	Ile55-Val
	Lys 114	Lys114-Asn

The functional role of aromatic clusters I and II (*Figure* 2.3) in the LA molecule has been studied by Grobler *et al.* (1994b). Aromatic cluster I comprises the side chains of the invariant amino acids Phe 31, His 32, Gln 117 and Trp 118 together with Tyr 36, while aromatic cluster II designated also as 'hydrophobic box' includes Trp 26, Phe 53,

Trp 60, Tyr 103 and Trp 104 (Table 2.1).

The rationale for the sequence substitutions was based either on conservation criteria with other LA primary structures, direct comparison with LZ sequence, stereochemical properties of particular residues that give them certain freedom of movements or the extent of perturbations caused to the activity of LA when particular residues were subjected to chemical modification. For example, His 32 was changed to an Asn and Tyr (as in several LZ), a Glu (to investigate the effect of introducing a charged group into the cluster) and an Ala (to remove the side chain contribution).

AROMATIC CLUSTER I.

Histidine 32 to Tyrosine or Alanine : Histidine 32 participates in the formation of the surface that lies adjacent to the cleft region in LA (*Figure* 2.7, see page 50). This region is partly conserved and may provide part of the binding site for the sugar acceptor substrates of the GT. It has to be noted that the specific changes in these residues, which are quite close together on the surface of the LA molecule, and lack of effects on other residues, indicates the involvement of a region of LA that includes a number of conserved residues, including His32, in the interaction with GT (Brew and Grobler, 1992). It has been shown that modification of the imidazole group of histidine 32 in LA by ethoxyformylation leads to a major reduction in activity (Brew *et al.*, 1979; Schindler *et al.*, 1976). His 32 of human and bovine LA may play an important role in the interaction of LA and GT that leads to formation of LS. It has been suggested that the N-1 atom of the imidazole ring of His 32 cannot be substituted without inactivation of the LA while

substitution of the N-3 atom does not lead to complete inactivation suggesting a high degree of specificity in the interaction between LA and GT (Preels *et al.*, 1979).

Phenylalanine 31 to Tyrosine: As in the case of histidine 32, Phe 31 also lies adjacent to the cleft region and its ability to promote glucose binding (reflected by a 7-fold increase in K_m for glucose) affecting only slightly affinity for GT when compared to mLA (Grobler *et al.*, 1994b) makes it rather interesting for further investigation at a structural level.

Tryptophan 118 to Histidine: Tryptophan 118 side chain, as well as the side chains of the conserved residues Phe 31, His 32 and Gln 17 lie in a region adjacent to the cleft (*Figure* 2.3). These are flanked by lysines 5 and 114, whose reactivities have been shown in previous studies to be perturbed upon binding of LA to GT (Brew and Grobler, 1992; Brew *et al.*, 1979). Sulphenylation of Trp 118 causes probably not complete inactivation of LA (Brew *et al.*, 1979; Hill and Brew, 1975). Of the four tryptophans of bovine LA (26, 60, 104 and 118), residues 60 and 118 are the most susceptible to modification and reaction at either site reduces its affinity for galactosyltransferase or causes complete inactivation depending on the nature of the modification (Brew and Grobler, 1992). According to measurements performed by Grobler *et al.* (1994b), 40% of the free energy of the LA-GT interaction is lost on replacing Trp 118 with either histidine or tyrosine.

AROMATIC CLUSTER II.

Tyrosine 103 to Proline: Comparison of LA with LZ structures

suggests that the two molecules have evolved from a common ancestor. The striking difference between these two proteins is the loss of activity for LA. The cleft region in LA structure is the corresponding binding site in LZ. Although the residues that line the cleft and act in substrate binding in LZ are conserved in all known LAs, one particular substitution of a tyrosine at position 103 (LA) for alanine or proline in LZ (*Figure* 2.7), has the effect of blocking the upper region of the cleft containing the A and B monosaccharide binding sites in LZ (Brew and Grobler, 1992; Malinovskii *et al.*, 1996). Tyrosine 103 does not have a direct functional role by analogy with its function in the LZs, the possibility that part of the cleft in LA may act to facilitate glucose binding in the LS complex is raised (Brew and Grobler, 1992; Malinovskii *et al.*, 1996; Shewale *et al.*, 1984). Tyrosines 103 and 18 of human LA are susceptible to nitration with tetranitromethane (Tyr 103 > Tyr 18) with essentially no effect on its activity (Brew and Grobler, 1992).

FLEXIBLE LOOP LEU105 - LEU110.

Alanine 109 to Proline: The conformation that the flexible loop (residues 105-110) of LA molecule adopts is critical for its functional properties. X-ray crystal structure analysis has revealed that changes in the crystallisation conditions result in either a helical or a loop conformation of the residues comprising this highly mobile part of the protein (Harata and Muraki, 1992; Pike *et al.*, 1996). The substitution of alanine 109 to proline (*Figure* 2.7) has been performed with the aim to restrict the flexibility of the polypeptide chain, although the impact on LA activity from this substitution is very little (Malinovskii *et al.*, 1996;

Pike *et al.*, 1996).

Histidine 107 to Alanine: Histidine 107 is part of the flexible loop and is in close proximity to Lys 108 which as it has been suggested by Brew *et al.* (1975) it is the major site of cross linking in the LS complex. The adaptability to two different conformations of this loop region as exhibited in the 3D structure of LA depending on the crystallisation environment was considered. His 107 to Ala variant was designed to favour the helical conformation observed in human LA crystals at physiological conditions, in which the side chain of residue 107 is buried (Pike, 1995).

Leucine 110 to Arginine: As in the case of His 107 and Ala 109 variants from the flexible region, leucine 110 was mutated to arginine based on sequence comparisons with human and chicken LZs, in order to investigate the effects of various side chain types on the loop conformation. More variants of leucine substituted by Glu or His were also designed (Pike, 1995).

OTHER VARIANTS.

Lysine 114 to Asparagine: Lysine 114 is located at the beginning of the C-terminal tail of LA molecule, which is known to exist in a number of energetically degenerate conformations, and belongs to the group of residues, which are proximal to aromatic cluster, I, a substructure previously identified to be crucial to LA action (Grobler *et al.*, 1994b). Substitution of Asn for Lys 114 appears to produce a more

than 30-fold reduction in activity that implicates a decrease in affinity for GT and increased thermal stability indicating a possible effect on the structure and/or dynamics of the flexible C-terminus (Greene *et al.*, 1999; Malinovskii *et al.*, 1996).

Cysteine 6 to Serine: Cysteine 6 is involved in disulphide bond formation with cysteine 120 located at the C-terminus of the molecule (*Figure* 2.7). Out of four disulphide bridges that stabilise the overall structure of LA, chemical studies have shown that only Cys6-Cys120 is the most solvent-exposed disulphide bond, it gets immediately reduced in the absence of denaturants and is not essential for the structure or activity of LA (Shechter *et al.*, 1973). It is known that the C-terminus of LA molecule is highly mobile and to a certain extent this mobility is anchored by this 'reactive' disulphide bridge. It appears that residues beyond Cys120 are not necessary for either the structure or activity of LAs since they are variable or even absent in certain LAs or LZs (Grobler *et al.*, 1994a). The high mobility of these residues is reflected in to the electron density map of the most LA structures that have been determined so far, where they appear to be disordered after position 120 of the molecule. Moreover, residues from the N-terminus are involved in interactions with residues lining the calcium-binding site (Chaudhuri *et al.*, 1999). Thus, with the aim to investigate the involvement of the tail of LA molecule as well as the minimal requirements of that region for LA activity Cys 120 was mutated to Ser to abolish the disulphide bond. Progressive truncation of the polypeptide chain at positions 119, 118 and 116 showed that residues from Trp 118 to the C-terminus are not

essential for LA activity (Malinovskii *et al.*, 1996) and the structure determination of this variant is of particular interest.

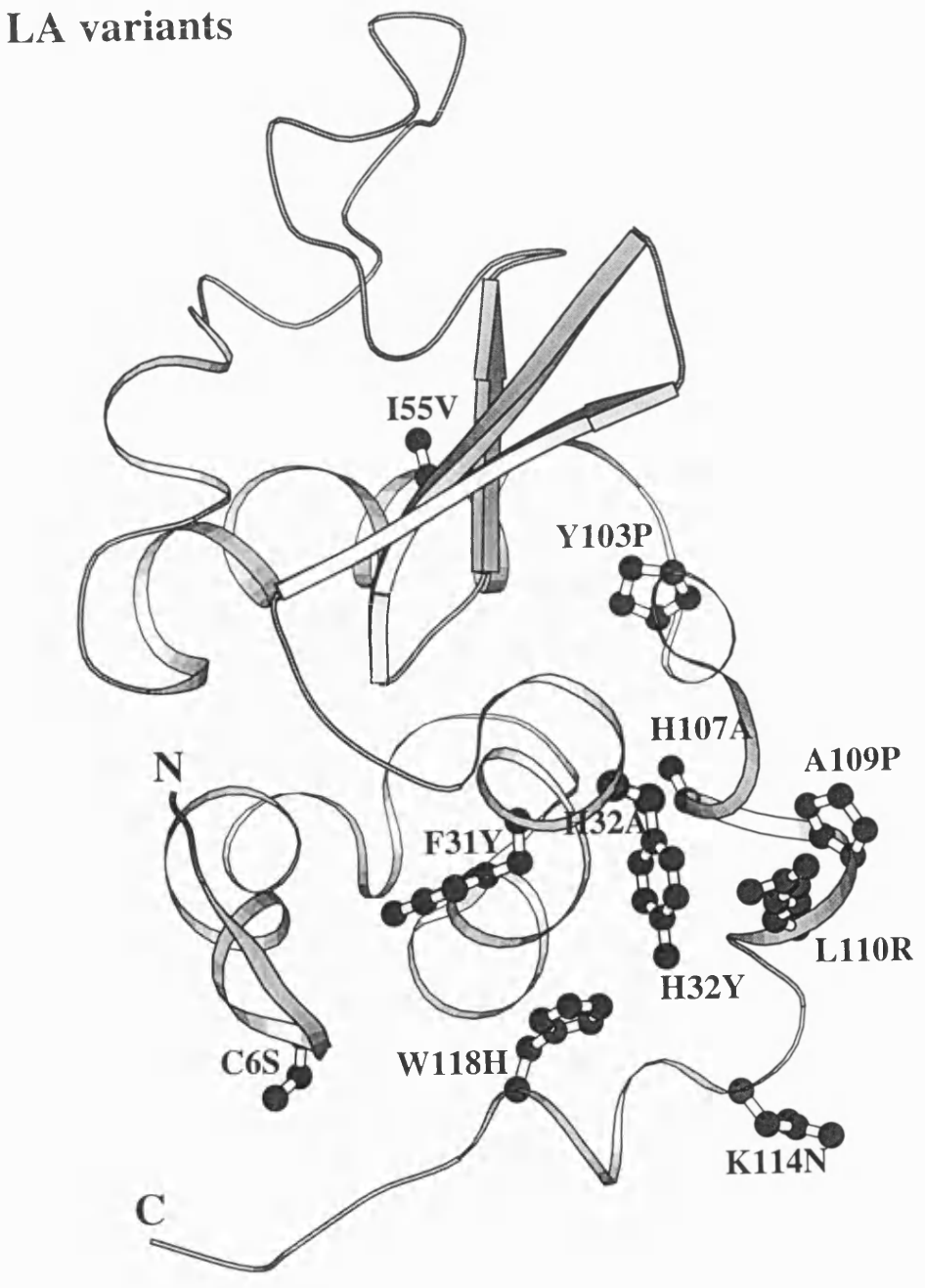


***Figure* 2.7.**

Sequence substitutions in the mLA molecule. The figure was generated using the program BOBSCRIPT (Esnouf, 1997).

Isoleucine 55 to Valine: Isoleucine 55 is buried (~2% exposed to the solvent) and a component of a region of sequence that is conserved in even the most distant relatives of LA such as bacterial transglycosylases and goose LZ. The isoleucine to valine mutation is highly conservative but the LA variant appears to have a perturbed structure or at least a near UV CD spectrum with a similar shape but reduced intensity indicating a less fixed tertiary structure. The secondary structure content seems to be unchanged based on far UV CD but the activity is greatly reduced (Prof. K. Brew, personal communication).

2.1.5. The 3D structure of GT molecule.

Most recently the crystal structure of the catalytic domain of GT in complex with uridine diphosphogalactose (UDP-Gal) was determined by (Gastinel *et al.*, 1999) at 2.4Å resolution (*Figure* 2.8). The catalytic domain of GT comprises 20% β-strands, 23% α-helices and 2% 3_{10} helices and involves residues Leu 131 to Ser 402.

The core of the structure consists of an eight stranded β-sheet surrounded by two α-helices on the one side and by four α-helices on the other. Although the structure of GT is of considerable importance in understanding the functional aspects of glycosyltransferases, very little information can be extracted towards understanding the interaction of GT with LA without the complex structure. However, it is now possible to perform modelling studies taking into consideration previous results obtained by chemical modification and site-directed mutagenesis studies. It also stresses the need for further direct structural studies on LA-GT complex.

β1,4–Galactosyltranferase

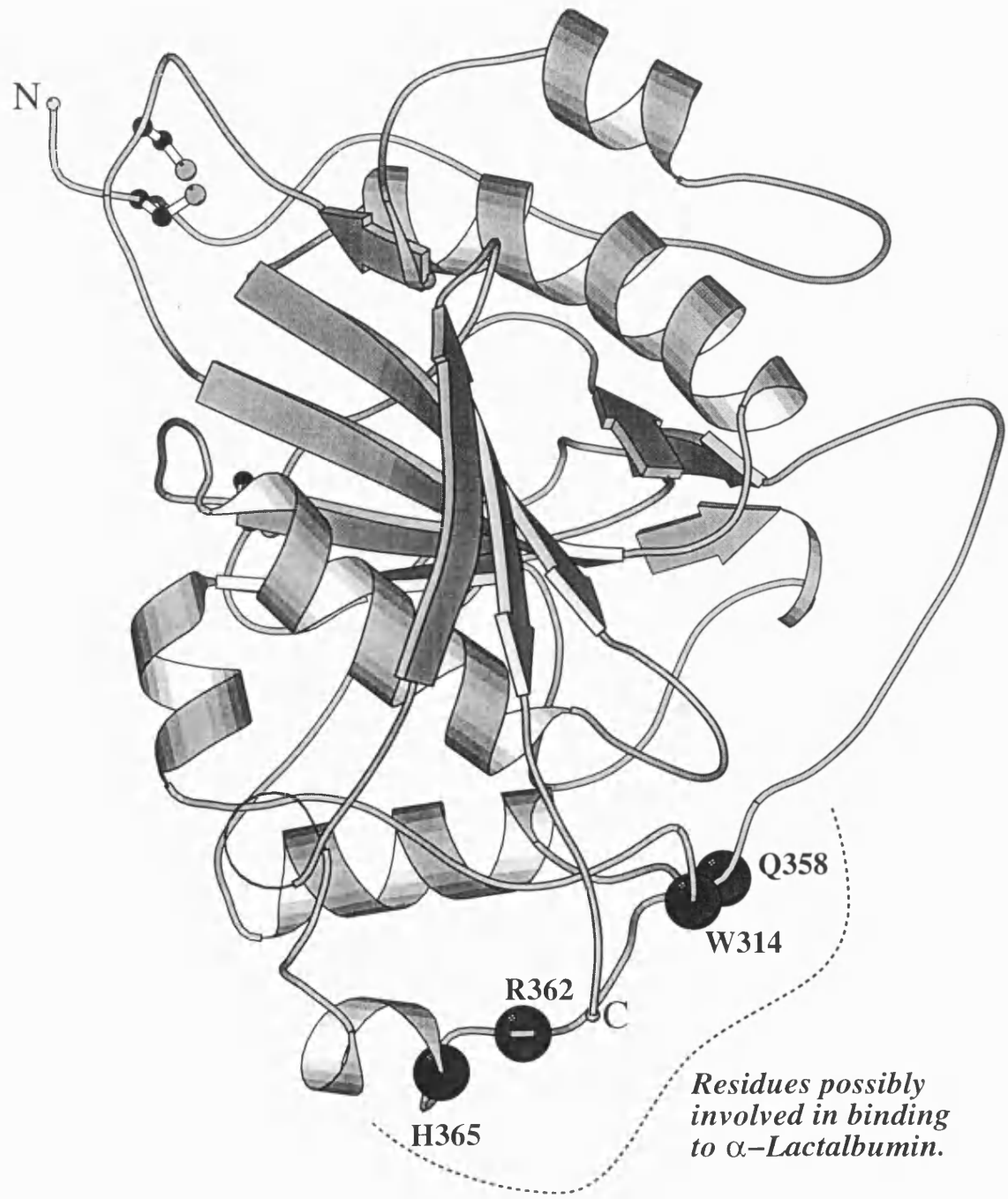


***Figure* 2.8**

Structure of galactosyltransferase.

2.2. CRYSTAL STRUCTURES OF apo-LA and bLA (holo-LA)

2.2.1. MATERIALS AND METHODS

2.2.1.1. Preparation of proteins

Pure apo-bovine LA used for crystallographic studies (described in this chapter) was purified by our collaborator Prof. K. Brew using the method developed by Martin Kronman (Kronman *et al.*, 1981) based on a modified protocol by Aschaffenburg and Drewry (1957) (Aschaffenburg and Drewry, 1957). Apo-LA was isolated by gel filtration of the protein at pH 2.0 ('Suprapure' HCl), with subsequent removal of the acid by lyophilisation and reconstitution in very dilute pH 5 to 6 metal-free Tris buffer. Solutions of apo-LA are stored in the frozen state at concentrations of 3 to 5mM. All chemicals used were certified to contain < 2.0 ppm of Ca^{2+} or Mg^{2+} ions and solutions had been passed through a column of Chelex 100. Water used for dilutions and other purposes was deionised and then passed through columns of Chelex 100. All glassware and optical cells were treated with concentrated HNO_3; concentrated H_2SO_4 while all plastic vials were soaked in EDTA and then rinsed with metal free water. Native bovine LA (bLA) was obtained from Sigma chemical company and used without further purification.

2.2.1.2.Crystallisation

Apo-LA: Crystals were grown with the method of vapour diffusion. Equal volumes (2.0 μl) of protein (~20mg/ml) and reservoir

solution (2.0 M ammonium sulphate, Tris/HCl 0.1 M pH 6.0) were mixed on siliconised coverslips and left to equilibrate at 16°C against the reservoir solution. Crystals appeared in ~20 days as tetrahedra (*Figure* 2.9). Care was taken to avoid contamination by adventitious metal ions during crystallisation experiments. All glassware was acid washed using nitric acid and rinsed with chelex treated water according to the protocol of Veillon and Vallee (1978).

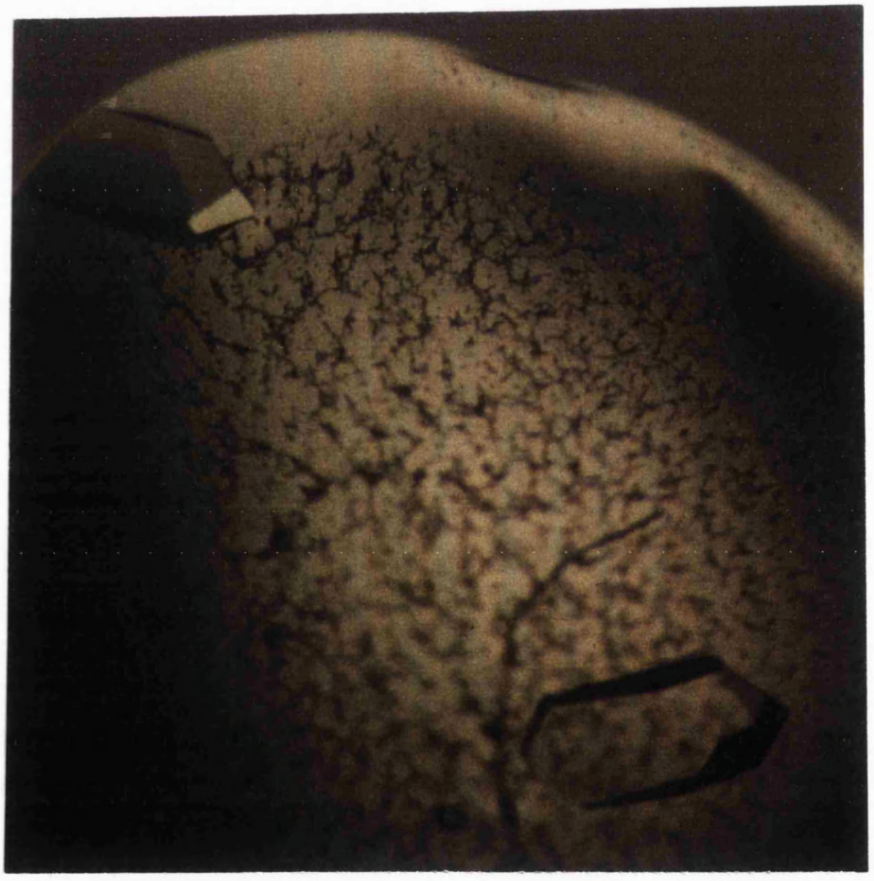


***Figure* 2.9**

Crystals of apo-LA

bLA: Native bovine LA was crystallised by the vapour diffusion method at 16°C against a reservoir solution containing 15% PEG 8K and 0.05 M potassium dihydrogen orthophosphate. The protein concentration was 20.0 mg/ml prior to crystallisation. Equal volumes (2.0 µl) of the protein and reservoir solution were used. Parallelepiped crystals appeared overnight and continued to grow for about five days (*Figure* 2.10).

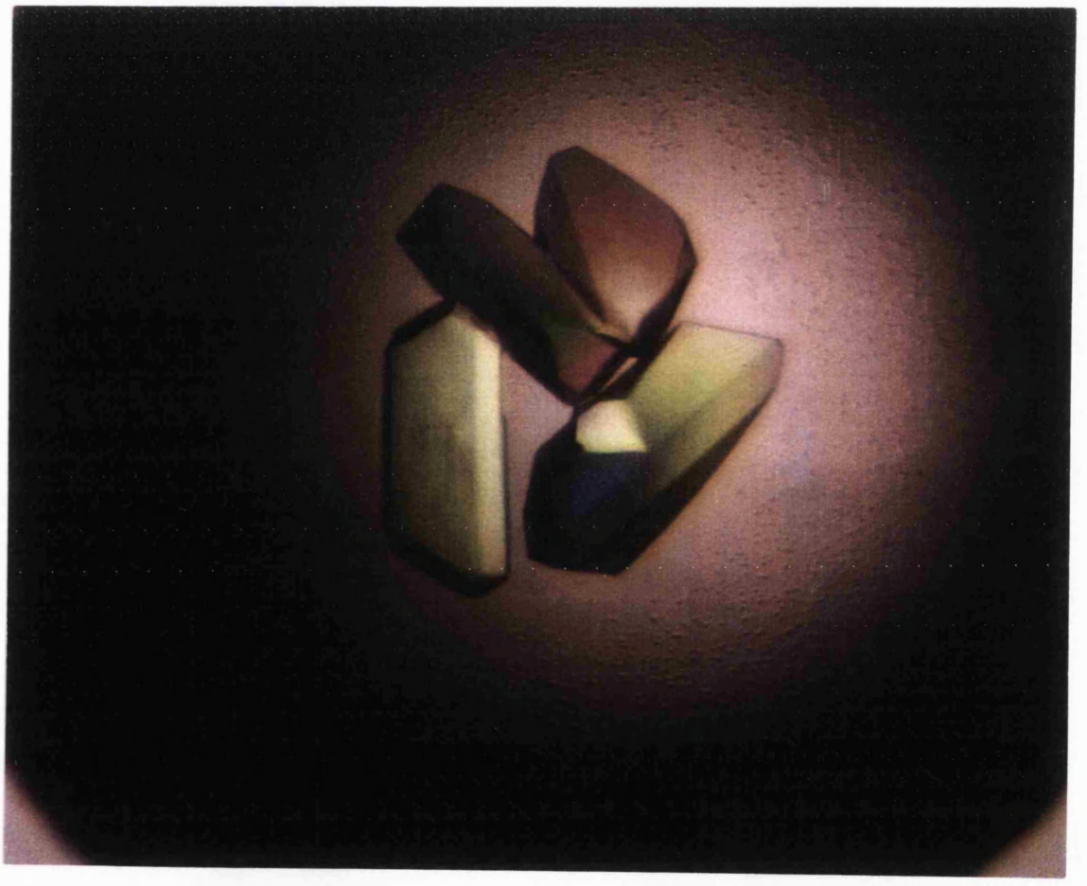


***Figure* 2.10**

Crystals of bLA

2.2.1.3. Diffraction data collection

Apo-LA: Data were collected from one crystal at the Synchrotron Radiation Source, Daresbury on station PX9.5 (oscillation range 1.0°, exposure time ~150 sec/image, λ=0.90 Å) at room temperature using a 30cm MAR Research image plate. Absence of (0 0 l) reflections hindered the immediate spacegroup determination for apo-LA. The data set was processed in tetragonal lattice which indicated the possibility of 6 or 8 molecules in the asymmetric unit with a solvent content of ~60 % (Matthews coefficient of ~3.2 $Å^3$/D). Data integration and reduction were performed with the programs *DENZO* and *SCALEPACK* (Otwinowski and Minor, 1997). Diffraction spots up to 2.0 Å were observed (*Figure* 2.11) but completeness dropped beyond 2.2 Å resolution (Tables 2.2, 2.3). The data collection statistics are presented in Table 2.4.

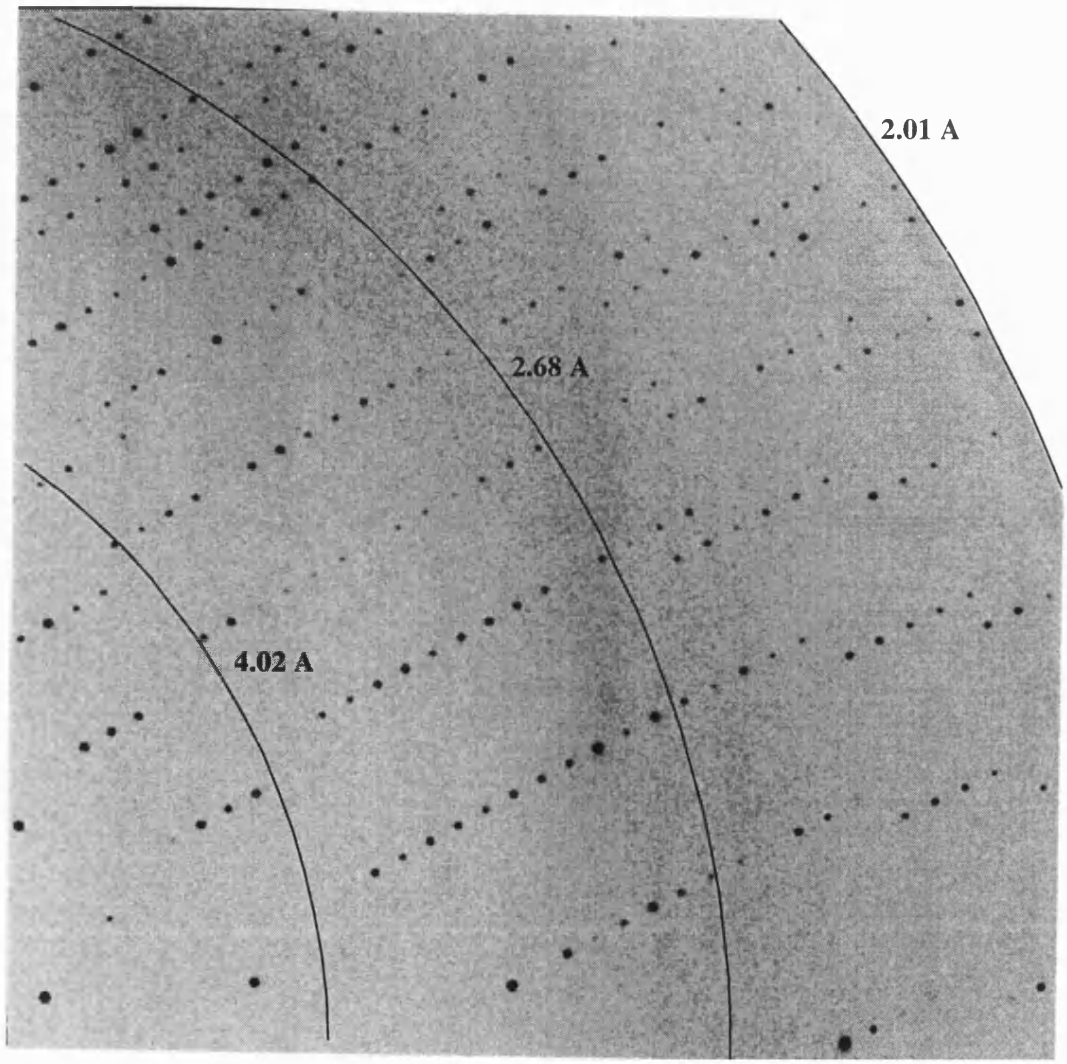


***Figure* 2.11**

apo-LA diffraction pattern

Table 2.2 Distribution of reflections for the apo-LA data

Shell		*I / Sigma in resolution shells*								
Lower	*Upper*	*No. of reflections with I / Sigma less than*								
limit	*limit*	*0*	*1*	*2*	*3*	*5*	*10*	*20*	*>20*	*total*
∝	4.74	53	110	166	218	315	590	1627	4388	6015
4.74	3.76	48	97	166	228	359	729	2202	3584	5786
3.76	3.29	72	156	275	401	639	1381	5210	566	5776
3.29	2.99	142	302	501	749	1249	3177	5716	0	5716
2.99	2.77	279	567	993	1487	2642	5563	5677	0	5677
2.77	2.61	407	841	1407	2122	3567	5672	5672	0	5672
2.61	2.48	462	970	1716	2514	4145	5678	5678	0	5678
2.48	2.37	582	1184	2005	2959	4754	5655	5655	0	5655
2.37	2.28	702	1434	2382	3380	5069	5645	5645	0	5645
2.28	2.2	786	1623	2671	3801	5343	5653	5653	0	5653
	All hkl	***3533***	***7284***	***12282***	***17859***	***28082***	***39743***	***48735***	***8538***	***57273***

Table 2.3 Completeness of the apo-LA data

Shell		*I / Sigma in resolution shells*								
Lower	*Upper*	*% of reflections with I / Sigma less than*								
limit	*limit*	*0*	*1*	*2*	*3*	*5*	*10*	*20*	*>20*	*total*
∝	4.74	0.9	1.8	2.7	3.6	5.1	9.6	26.5	71.5	98
4.74	3.76	0.8	1.7	2.8	3.9	6.2	12.5	37.8	61.4	99.2
3.76	3.29	1.2	2.7	4.7	6.9	11	23.8	89.7	9.7	99.5
3.29	2.99	2.5	5.3	8.7	13	21.8	55.3	99.5	0	99.5
2.99	2.77	4.9	9.9	17.4	26	46.2	97.3	99.3	0	99.3
2.77	2.61	7.1	14.8	24.7	37.3	62.6	99.6	99.6	0	99.6
2.61	2.48	8.1	17	30.1	44.2	72.8	99.7	99.7	0	99.7
2.48	2.37	10.3	20.9	35.3	52.1	83.8	99.6	99.6	0	99.6
2.37	2.28	12.4	25.3	42	59.6	89.4	99.6	99.6	0	99.6
2.28	2.2	13.8	28.6	47	66.9	94.1	99.5	99.5	0	99.5
	All hkl	***6.1***	***12.6***	***21.3***	***31***	***48.7***	***68.9***	***84.5***	***14.8***	***99.3***

Table 2.4 Data collection statistics for apo-LA and bLA

Spacegroup details	**apo-LA**	**BLA**
Cell dimensions (a, b, c) (Å)	a=b=119.57, c=152.74	a=72.05, b=104.65, c=117.41
α, β, γ (°)	90, 90, 90	90, 90, 90
Space group	$P4_12_12$	$P2_12_12$
No of molecules per asymmetric unit	6	6
Data collection and processing statistics		
Station (Synchrotron)	Daresbury PX95	EMBL, BW7B
Image plate	Mar30	Mar30
Wavelength (Å)	0.9	0.847
No of observations	561254	371734#
No of unique reflections	57273	45363#
Max. resolution (Å)	2.2	2.2#
*Completeness (outermost shell) (%)	99.3 (99.5)	99.2 (99.2)#
$^{\diamond}R_{sym}$ (I) (outermost shell) (%)	8.6 (63.3)	8.1 (51.7)#
$< I / \sigma (I) >$ (outermost shell)	8.8 (2.5)	11.5 (2.1)#
Outermost shell (Å)	2.28 - 2.20	2.32 - 2.20#

*Completeness in the range ∞-resmax, where resmax is the maximum resolution to which data were collected. #for bLA includes reflections after merging with the in-house data collected on a Siemens detector mounted on a Siemens X-ray generator (λ=1.5418Å, 0.25° per frame.

bLA: Data from one bLA crystal were collected to 2.2 Å on station BW7B at EMBL-Hamburg Synchrotron Radiation Source, using a MAR300 image plate (Table 2.4). The oscillation range per image was 1.0° (λ=0.847 Å) and the exposure time was 40-50 sec/image. Both *DENZO* and *SCALEPACK* programs were used for data processing (Otwinowski and Minor, 1997). To improve the completeness at low

resolution shells, a medium resolution data set which was collected previously, using crystals exposed to graphite monochromated CuK_α X-rays (λ=1.5418 Å) produced by a Siemens rotating anode generator operating at 50kV and 80mA at room temperature was merged using *SCALA* and subsequently converted to structure factor amplitudes with the program *TRUNCATE* (Table 2.5) as implemented by the CCP4 suite of programs (CCP4, 1994). The systematic absences were found to agree with orthorhombic $P2_12_12$ space group with either 6 or 8 molecules in the asymmetric unit with a solvent content of ~50 % (Matthews coefficient of ~2.6 $Å^3$/D). The statistics of the data collection and reduction are presented in Table 2.4.

Table 2.5 Completeness, multiplicity, R_{meas} vs. resolution for bLA

Completeness and multiplicity, including reflections measured only once
Note: completeness figures are approximate, as they are calculated by volume, (so there are grid-sampling errors).
%poss. is completeness in the shell.

N	Dmin	Nmeas	Nref	Ncent	%poss	Mlplcty	(Rsym)
1	6.73	6586	1755	470	99.8	3.8	0.046
2	4.76	12223	3009	481	99.7	4.1	0.046
3	3.89	15702	3852	490	100.3	4.1	0.052
4	3.37	18204	4486	482	99.7	4.1	0.085
5	3.01	20205	5079	483	100.2	4.0	0.157
6	2.75	19988	5550	486	99.4	3.6	0.249
7	2.55	15944	5974	460	99.0	2.7	0.236
8	2.38	15869	6423	444	99.5	2.5	0.346
9	2.24	14492	6674	398	97.6	2.2	0.479
10	2.13	8373	5071	215	70.8	1.7	0.537
Overall		**147586**	**47873**	**4409**	**95.1**	**3.1**	**0.081**

2.2.1.4. Structure determination

Apo-LA: The structure of recombinant bLA (mLA, monoclinic form) determined at 2.3 Å (Pike *et al.*, 1996) was used for molecular replacement with the program *AMoRe* (Navaza, 1994). Due to the difficulty in assigning the correct spacegroup amongst possible spacegroups in the tetragonal lattice ($P42_12$, $P4_12_12$, $P4_22_12$, $P4_32_12$), the molecular replacement technique was applied for all four of them. First the self-rotation function was calculated with *POLARRFN* (CCP4, 1994). The stereographic projection of section κ=180° (in polar convention) indicated that the molecules are related by three local 2-fold symmetry axes to each other (*Figure* 2.12).

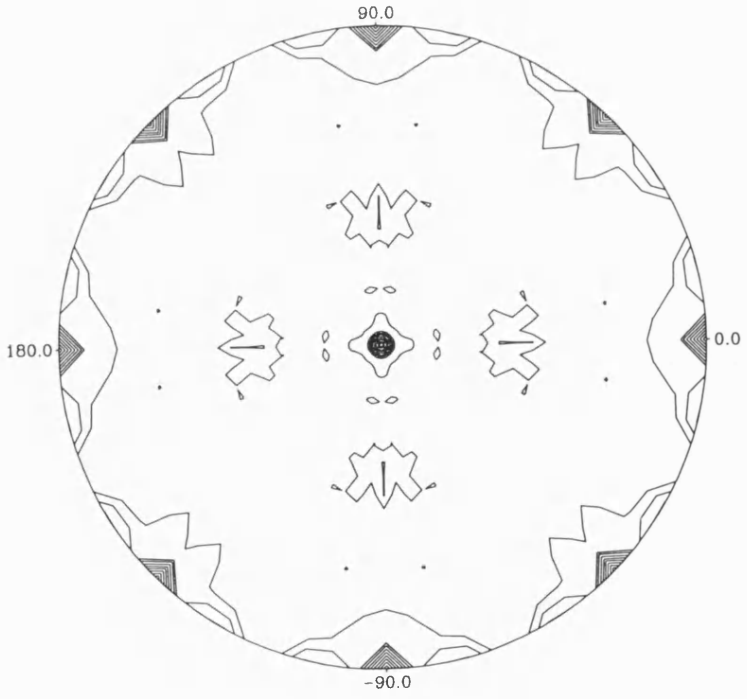


***Figure* 2.12**
Rotation function from molecular replacement

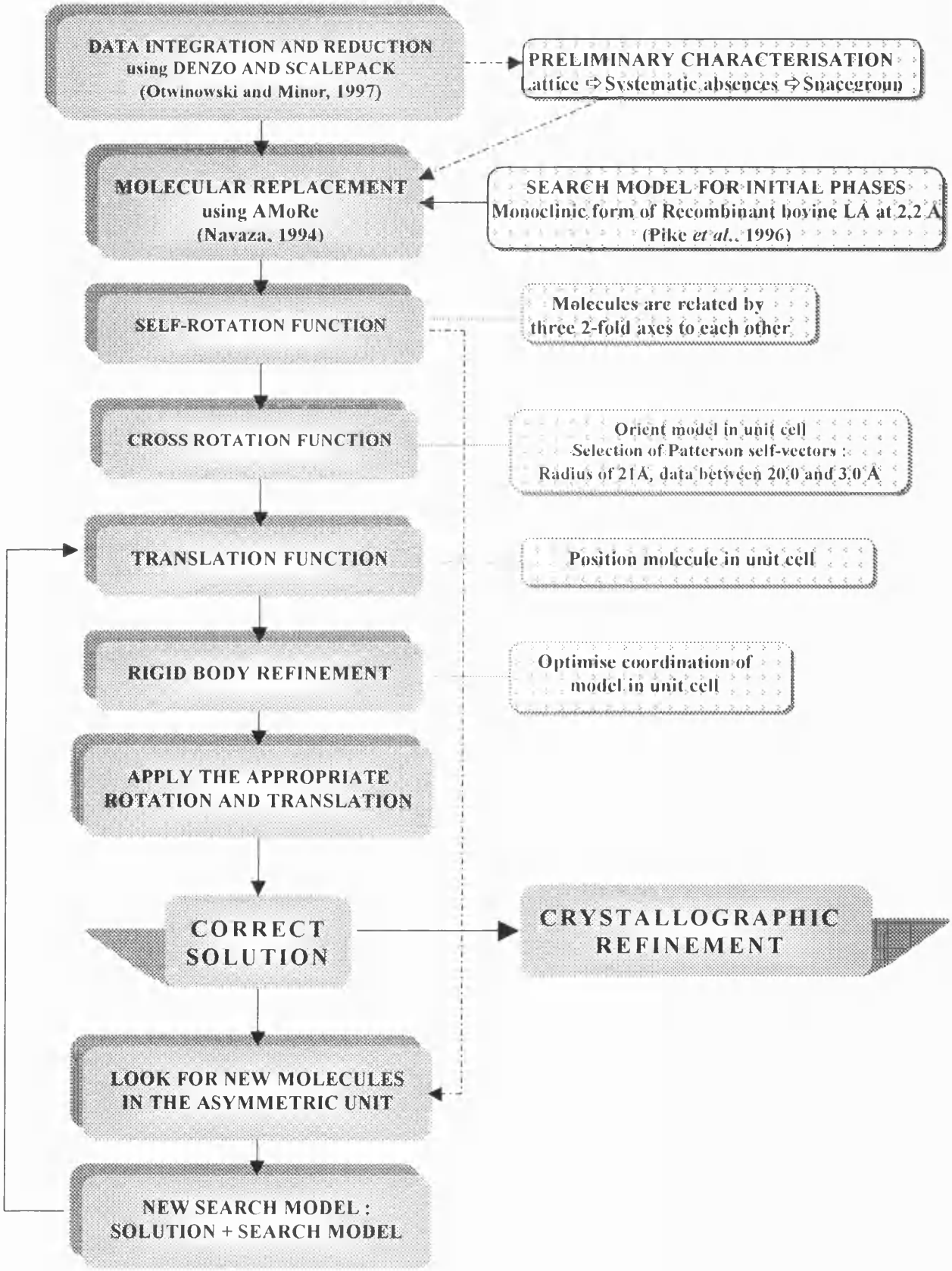


Figure 2.13

Flow diagram describing the procedure followed for the structure determination of the apo- and holo- forms of LA using the molecular replacement technique.

All data between 20.0 and 3.0 Å were used to calculate the cross rotation and translation functions. Rigid body refinement against the data in the above range was used to optimise the position of the initial model. The appropriate rotation and translation parameters were applied to the starting model and the position of one molecule was determined. Combination of molecular replacement, rigid body refinement using X-PLOR (Brünger, 1992b) and visual inspection of the $(2F_o\text{-}F_c)$ electron density maps were helpful in identifying positions for five apo-LA molecules in the asymmetric unit in spacegroup $P4_12_12$. An outline of the procedure adopted for the structure determination is summarised in *Figure* 2.13.

Further attempts to determine the sixth molecule (the search model consisted of five molecules) using molecular replacement were not successful. However, visual inspection of the $(F_o\text{-}F_c)$ electron density map calculated using the refined model with five molecules, identified the position of the sixth molecule as a large piece of contiguous density sufficient to accommodate one more molecule. Careful examination of the packing features for five molecules and the symmetry information derived from the self-rotation function enabled accurate positioning of the sixth molecule in the unaccounted electron density region. These results confirmed the presence of only six molecules in the asymmetric unit (*Figure* 2.14). The list of the best translation function solutions and their progression in terms of correlation coefficient and *R* factor are summarised in Tables 2.6-2.10. The final solutions after the rigid body refinement with the highest correlation coefficient and the best *R* factor are shown in Table 2.11.

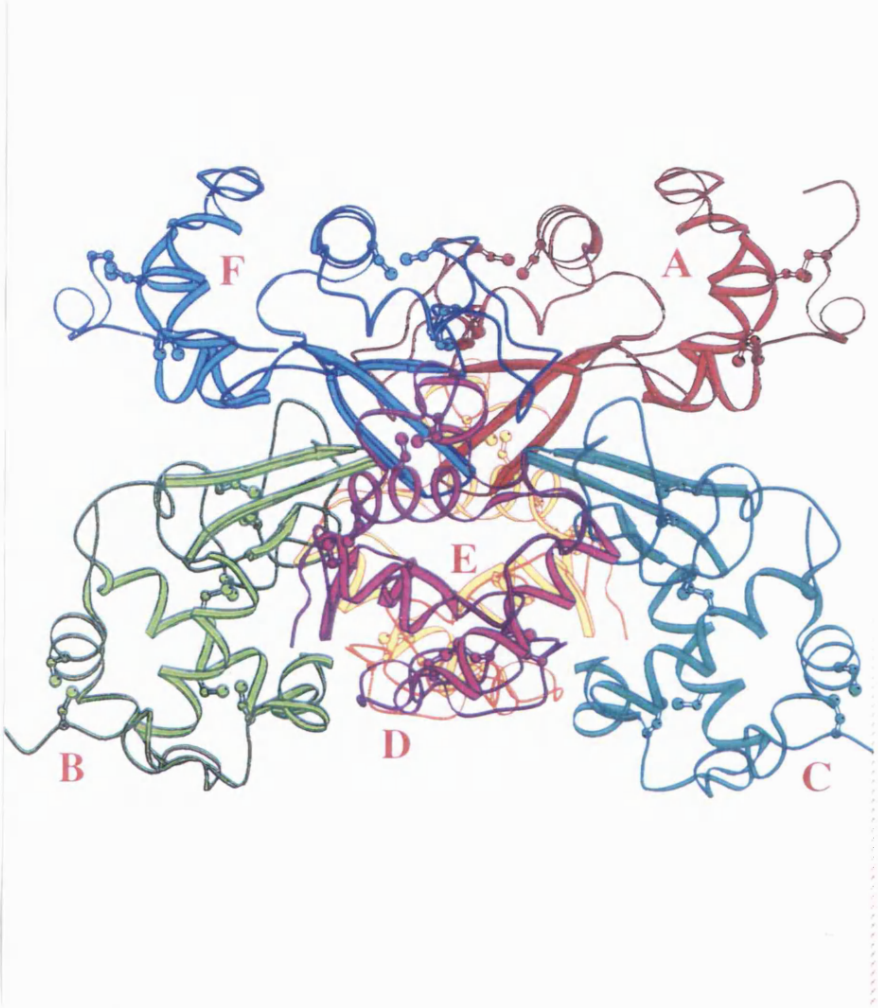


***Figure* 2.14**

Packing of apo-LA (6 molecules/a.u)

MOLECULAR REPLACEMENT RESULTS FOR apo-LA

Tables 2.6-2.10: *The columns α, β, and γ correspond to the orientation Euler angles; T_x, T_y and T_z represent the positional parameters (fractions of the unit cell), cc the correlation coefficient, R_f the R-factor and no- to the sorting number of the peak when the translation function was calculated for the first time. The solutions are shown in shaded areas and the peak number is in bold face.*

Table 2.11: *See Tables 2.6-2.10. The highest peak numbers that correspond to five out of six molecules in the asymmetric unit after rigid body refinement are shown in* bold face.

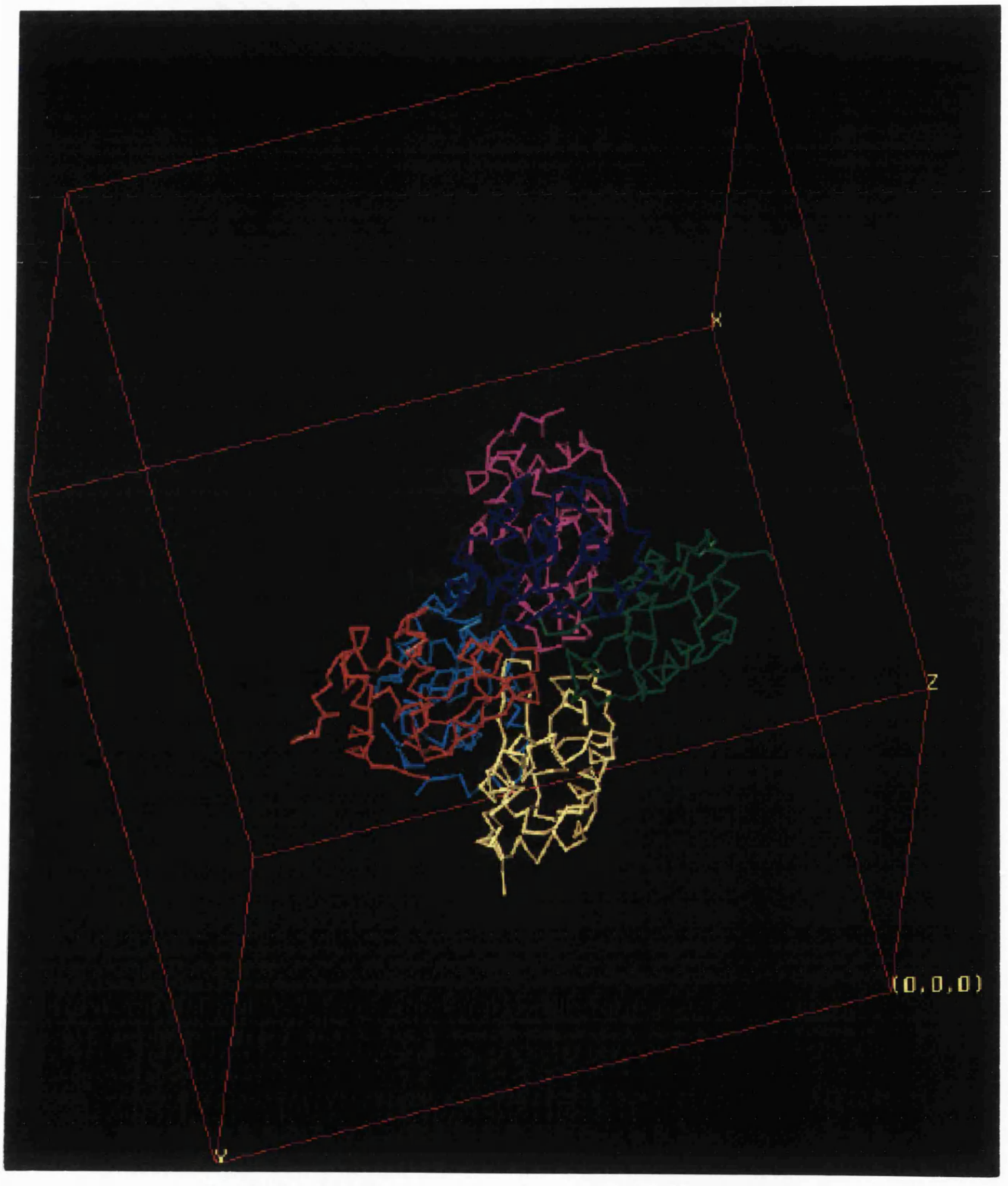


***Figure* 2.14a**

Packing of apo-LA (6 molecules/a.u)

Table 2.6 Molecule I (apo-LA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	28.11	97.92	252.18	0.8243	0.805	0.015	23.7	56.7	*1*
SOLUTIONTF1	61.89	82.08	72.18	0.3037	0.3256	0.4846	23.5	56.6	*2*
	....	....	....	....	....	....	....	....	......
SOLUTIONTF1	72.99	67.73	249.33	0.6266	0.3861	0.2491	22.4	56.6	*32*
SOLUTIONTF1	76.06	31.98	192.53	0.8851	0.8538	0.0212	22.8	56.2	*33*
SOLUTIONTF1	13.94	148.02	12.53	0.5932	0.4859	0.2759	22.9	56.2	*34*
	....	....	....	....	....	....	....	....	......
SOLUTIONTF1	19.01	143.28	4.87	0.2307	0.7207	0.1949	22.5	56.6	*55*
SOLUTIONTF1	71	36.72	184.87	0.7472	0.7294	0.0104	22.3	56.5	*56*
* *									
SOLUTIONTF1	**76.5**	**22.26**	**183.64**	**0.1285**	**0.7264**	**0.1105**	**25.4**	**54.7**	***57***
* *									
SOLUTIONTF1	13.5	157.74	3.64	0.7224	0.3442	0.2938	23.1	56	*58*
SOLUTIONTF1	1.81	136.65	108.9	0.3516	0.6333	0.2719	22.6	56.3	*59*
SOLUTIONTF1	88.19	43.35	288.9	0.881	0.2206	0.232	22.5	56.5	60
SOLUTIONTF1	69.43	62.85	64.6	0.0706	0.082	0.4976	22.5	56.9	61
SOLUTIONTF1	20.57	117.15	244.6	0.7955	0.2125	0.3098	23.1	56.2	62
SOLUTIONTF1	38.5	92.74	252.04	0.8445	0.7832	0.4549	22.3	56.7	63
	....	....	....	....	....	....	....	....	......
SOLUTIONTF1	56	123.37	113.83	0.9098	0.9024	0.4976	22.5	57	*73*
SOLUTIONTF1	34	56.63	293.83	0.4029	0.4221	0.0148	22.5	56.5	*74*
SOLUTIONTF1	5.86	119.95	194	0.6512	0.638	0.031	22.6	56.3	*75*
SOLUTIONTF1	84.14	60.05	14	0.219	0.2483	0.0453	22.8	56.5	*76*
SOLUTIONTF1	29.92	64.89	55.33	0.9266	0.8007	0.0134	23.5	55.9	*77*
SOLUTIONTF1	60.08	115.11	235.34	0.3026	0.4254	0.4871	23.2	55.9	*78*

Table 2.7 Molecule II (apo-LA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
* *									
SOLUTIONTF1	**76.5**	**22.26**	**183.64**	**0.1285**	**0.7264**	**0.1105**	**25.4**	**54.7**	
SOLUTIONTF2	**28.11**	**97.92**	**252.18**	**0.8399**	**0.634**	**0.5938**	**32**	**53.1**	***1***
* *									
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	61.89	82.08	72.18	0.634	0.8399	0.4062	32	53.1	*2*
	....	....	....	....	....	....	....	....	......

Table 2.8 Molecule III (apo-LA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	69.43	62.85	64.6	0.6702	0.7794	0.5124	30.3	54	*63*
	.....	.....	.....	.....	.....	.....	.....	.....	.....
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	34	56.63	293.83	0.8035	0.138	0.6646	30.5	53.6	*72*
**********									
SOLUTIONTF1	**76.5**	**22.26**	**183.64**	**0.1285**	**0.7264**	**0.1105**	**25.4**	**54.7**	
SOLUTIONTF2	**28.11**	**97.92**	**252.18**	**0.8399**	**0.634**	**0.5938**	**32**	**53.1**	
SOLUTIONTF3	**5.86**	**119.95**	**194**	**0.1876**	**0.6528**	**0.9514**	**35.2**	**51.9**	*75*
**********									
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	84.14	60.05	14	0.6528	0.1876	0.0486	35.2	51.9	*78*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.9 Molecule IV (apo-LA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	5.86	119.95	194	0.1876	0.6528	0.9514	35.2	51.9	
SOLUTIONTF4	44	146	1.5	0.8085	0.304	0.6054	33.2	53	*21*
	.....	.....	.....	.....	.....	.....	.....	.....	.....
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	5.86	119.95	194	0.1876	0.6528	0.9514	35.2	51.9	
SOLUTIONTF4	72.99	67.73	249.33	0.4314	0.2074	0.0326	32.7	53.1	*29*
**********									
SOLUTIONTF1	**76.5**	**22.26**	**183.64**	**0.1285**	**0.7264**	**0.1105**	**25.4**	**54.7**	
SOLUTIONTF2	**28.11**	**97.92**	**252.18**	**0.8399**	**0.634**	**0.5938**	**32**	**53.1**	
SOLUTIONTF3	**5.86**	**119.95**	**194**	**0.1876**	**0.6528**	**0.9514**	**35.2**	**51.9**	
SOLUTIONTF4	**76.06**	**31.98**	**192.53**	**0.4768**	**0.3495**	**0.9451**	**38.9**	**50.7**	*33*
**********									
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	5.86	119.95	194	0.1876	0.6528	0.9514	35.2	51.9	
SOLUTIONTF4	13.94	148.02	12.53	0.3495	0.4768	0.0549	38.9	50.7	*37*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.10 Molecule V (apo-LA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	5.86	119.95	194	0.1876	0.6528	0.9514	35.2	51.9	
SOLUTIONTF4	76.06	31.98	192.53	0.4768	0.3495	0.9451	38.9	50.7	
SOLUTIONTF5	69.43	62.85	64.6	0.8166	0.195	0.8556	36.4	51.8	*53*
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	5.86	119.95	194	0.1876	0.6528	0.9514	35.2	51.9	
SOLUTIONTF4	76.06	31.98	192.53	0.4768	0.3495	0.9451	38.9	50.7	
SOLUTIONTF5	20.57	117.15	244.6	0.195	0.8166	0.1444	36.4	51.8	*58*
**********									
SOLUTIONTF1	**76.5**	**22.26**	**183.64**	**0.1285**	**0.7264**	**0.1105**	**25.4**	**54.7**	
SOLUTIONTF2	**28.11**	**97.92**	**252.18**	**0.8399**	**0.634**	**0.5938**	**32**	**53.1**	
SOLUTIONTF3	**5.86**	**119.95**	**194**	**0.1876**	**0.6528**	**0.9514**	**35.2**	**51.9**	
SOLUTIONTF4	**76.06**	**31.98**	**192.53**	**0.4768**	**0.3495**	**0.9451**	**38.9**	**50.7**	
SOLUTIONTF5	**38.5**	**92.74**	**252.04**	**0.4243**	**0.9944**	**0.4737**	**39.7**	**50.5**	***63***
**********									
SOLUTIONTF1	76.5	22.26	183.64	0.1285	0.7264	0.1105	25.4	54.7	
SOLUTIONTF2	28.11	97.92	252.18	0.8399	0.634	0.5938	32	53.1	
SOLUTIONTF3	5.86	119.95	194	0.1876	0.6528	0.9514	35.2	51.9	
SOLUTIONTF4	76.06	31.98	192.53	0.4768	0.3495	0.9451	38.9	50.7	
SOLUTIONTF5	51.5	87.26	72.04	0.9944	0.4243	0.5263	39.7	50.5	*68*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.11 Final solutions for molecules I-V after *FITING* (apo-LA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONF	**78.75**	**20.1**	**180.88**	**0.128**	**0.7261**	**0.1105**	**54.5**	**45.2**	**57**
SOLUTIONF	**27.51**	**97.82**	**253.22**	**0.8388**	**0.6345**	**0.5938**	**54.5**	**45.2**	**1**
SOLUTIONF	**5.5**	**122.83**	**189.5**	**0.1878**	**0.6536**	**0.9514**	**54.5**	**45.2**	**75**
SOLUTIONF	**69.77**	**34.16**	**201.95**	**0.4754**	**0.349**	**0.9451**	**54.5**	**45.2**	**33**
SOLUTIONF	**39.34**	**83.49**	**253.22**	**0.4258**	**0.995**	**0.4737**	**54.5**	**45.2**	**63**

bLA: The structure of bLA was also determined by molecular replacement with the program *AMoRe* (Navaza, 1994) using the structure of recombinant bovine LA (Pike *et al.*, 1996) using a similar protocol as shown in *Figure* 2.13. As in the case of apo-LA, packing

considerations and visual inspection of electron density maps revealed the positions for all six molecules in the asymmetric unit (*Figure* 2.15). The solutions that exhibited greatest improvement in terms of correlation coefficient and *R* factor are summarised in Tables 2.12-2.17 while in Table 2.18 the results after rigid body refinement are shown.

MOLECULAR REPLACEMENT RESULTS FOR bLA

Tables 2.12-2.17: *The columns α, β, and γ correspond to the orientation Euler angles; T_X, T_Y and T_Z represent the positional parameters (fractions of the unit cell), cc the correlation coefficient, R_f the R-factor and no- to the sorting number of the peak when the translation function was calculated for the first time. The solutions are shown in shaded areas and the peak number is in bold face.*
Table 2.18: *See Tables 2.12-2.17. The highest peak numbers that correspond to the six molecules in the asymmetric unit after rigid body refinement are shown in* bold face.

Table 2.12 Molecule I (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	30.82	88.44	265.36	0.4458	0.0068	0.0931	27.1	56	*1*
SOLUTIONTF1	149.18	91.56	85.36	0.0543	0.0064	0.4101	27	56.2	*2*
SOLUTIONTF1	114	153.66	131.5	0.3644	0.0137	0.1528	23.8	57.1	*3*
	.....	.....	.....	.....	.....	.....	.....	.....	.....
SOLUTIONTF1	69.35	126.02	15.06	0.4774	0.4644	0.2966	24.6	57.4	*14*
SOLUTIONTF1	70.33	75.07	327.35	0.1167	0.1218	0.0349	24.3	56.7	*15*
SOLUTIONTF1	109.67	104.93	147.35	0.4745	0.0155	0.0324	24.4	57	*16*
**********									
SOLUTIONTF1	**86.78**	**92.41**	**81.76**	**0.4378**	**0.1093**	**0.0815**	**29.7**	**55**	***17***
**********									
SOLUTIONTF1	93.22	87.59	261.76	0.0623	0.1085	0.4194	29.7	54.9	*18*
SOLUTIONTF1	93.52	26.27	66.5	0.4617	0.0027	0.3299	24	57.4	*19*
SOLUTIONTF1	86.48	153.73	246.5	0.3018	0.4252	0.4209	24.2	56.7	*20*
SOLUTIONTF1	93.22	87.59	261.76	0.0623	0.1085	0.4194	29.7	54.9	*21*
SOLUTIONTF1	93.52	26.27	66.5	0.4617	0.0027	0.3299	24	57.4	*22*
SOLUTIONTF1	86.48	153.73	246.5	0.3018	0.4252	0.4209	24.2	56.7	*23*
SOLUTIONTF1	10.81	91.33	80.03	0.4876	0.412	0.295	26.9	56	*24*
SOLUTIONTF1	169.19	88.67	260.03	0.0123	0.412	0.2037	26.9	56	*25*
SOLUTIONTF1	25.06	86.88	84.38	0.0116	0.493	0.2443	24	57.1	*26*
SOLUTIONTF1	154.94	93.12	264.38	0.4705	0.4774	0.1036	24.3	57.4	*27*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.13 Molecule II (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	**86.78**	**92.41**	**81.76**	**0.4378**	**0.1093**	**0.0815**	**29.7**	**55**	
SOLUTIONTF2	**30.82**	**88.44**	**265.36**	**0.6791**	**0.5074**	**0.5913**	**36.1**	**52.8**	*1*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	149.18	91.56	85.36	0.8209	0.0074	0.4087	36.1	52.8	*2*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	114	153.66	131.5	0.1504	0.4082	0.9582	28.8	55.6	*4*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.14 Molecule III (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	**86.78**	**92.41**	**81.76**	**0.4378**	**0.1093**	**0.0815**	**29.7**	**55**	
SOLUTIONTF2	**30.82**	**88.44**	**265.36**	**0.6791**	**0.5074**	**0.5913**	**36.1**	**52.8**	
SOLUTIONTF3	**30.82**	**88.44**	**265.36**	**0.68**	**0.5087**	**0.0932**	**40.4**	**55**	*1*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	149.18	91.56	85.36	0.82	0.0087	0.9068	40.4	55	*2*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	114	153.66	131.5	0.0534	0.4755	0.9506	33.8	54	*3*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.15 Molecule IV (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	**86.78**	**92.41**	**81.76**	**0.4378**	**0.1093**	**0.0815**	**29.7**	**55**	
SOLUTIONTF2	**30.82**	**88.44**	**265.36**	**0.6791**	**0.5074**	**0.5913**	**36.1**	**52.8**	
SOLUTIONTF3	**30.82**	**88.44**	**265.36**	**0.68**	**0.5087**	**0.0932**	**40.4**	**55**	
SOLUTIONTF4	**10.81**	**91.33**	**80.03**	**0.7834**	**0.7753**	**0.7525**	**46**	**51**	*21*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	30.82	88.44	265.36	0.68	0.5087	0.0932	40.4	55	
SOLUTIONTF4	169.19	88.67	260.03	0.7166	0.2753	0.2475	46	51	*25*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	30.82	88.44	265.36	0.68	0.5087	0.0932	40.4	55	
SOLUTIONTF4	25.06	86.88	84.38	0.8456	0.7815	0.2431	44.1	51.8	*29*
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Table 2.16 Molecule V (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	30.82	88.44	265.36	0.68	0.5087	0.0932	40.4	55	
SOLUTIONTF4	10.81	91.33	80.03	0.7834	0.7753	0.7525	46	51	
SOLUTIONTF5	169.19	88.67	260.03	0.7177	0.275	0.2472	43.4	50.8	*18*
**									
SOLUTIONTF1	**86.78**	**92.41**	**81.76**	**0.4378**	**0.1093**	**0.0815**	**29.7**	**55**	
SOLUTIONTF2	**30.82**	**88.44**	**265.36**	**0.6791**	**0.5074**	**0.5913**	**36.1**	**52.8**	
SOLUTIONTF3	**30.82**	**88.44**	**265.36**	**0.68**	**0.5087**	**0.0932**	**40.4**	**55**	
SOLUTIONTF4	**10.81**	**91.33**	**80.03**	**0.7834**	**0.7753**	**0.7525**	**46**	**51**	
SOLUTIONTF5	**25.06**	**86.88**	**84.38**	**0.8437**	**0.7822**	**0.243**	**50**	**48.8**	*23*
**									
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	30.82	88.44	265.36	0.68	0.5087	0.0932	40.4	55	
SOLUTIONTF4	10.81	91.33	80.03	0.7834	0.7753	0.7525	46	51	
SOLUTIONTF5	154.94	93.12	264.38	0.6563	0.2822	0.757	50	48.8	*28*
	....	....	....	....	....	....	....	....	......

Table 2.17 Molecule VI (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
	....	....	....	....	....	....	....	....	.....
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	30.82	88.44	265.36	0.68	0.5087	0.0932	40.4	55	
SOLUTIONTF4	10.81	91.33	80.03	0.7834	0.7753	0.7525	46	51	
SOLUTIONTF5	25.06	86.88	84.38	0.8437	0.7822	0.243	50	48.8	
SOLUTIONTF6	167.73	90.16	260.11	0.6704	0.2778	0.7517	55.9	48.1	*2*
**									
SOLUTIONTF1	**86.78**	**92.41**	**81.76**	**0.4378**	**0.1093**	**0.0815**	**29.7**	**55**	
SOLUTIONTF2	**30.82**	**88.44**	**265.36**	**0.6791**	**0.5074**	**0.5913**	**36.1**	**52.8**	
SOLUTIONTF3	**30.82**	**88.44**	**265.36**	**0.68**	**0.5087**	**0.0932**	**40.4**	**55**	
SOLUTIONTF4	**10.81**	**91.33**	**80.03**	**0.7834**	**0.7753**	**0.7525**	**46**	**51**	
SOLUTIONTF5	**25.06**	**86.88**	**84.38**	**0.8437**	**0.7822**	**0.243**	**50**	**48.8**	
SOLUTIONTF6	**102.74**	**90.74**	**268.24**	**0.0333**	**0.6129**	**0.4288**	**62.9**	**44.5**	*3*
**									
SOLUTIONTF1	86.78	92.41	81.76	0.4378	0.1093	0.0815	29.7	55	
SOLUTIONTF2	30.82	88.44	265.36	0.6791	0.5074	0.5913	36.1	52.8	
SOLUTIONTF3	30.82	88.44	265.36	0.68	0.5087	0.0932	40.4	55	
SOLUTIONTF4	10.81	91.33	80.03	0.7834	0.7753	0.7525	46	51	
SOLUTIONTF5	25.06	86.88	84.38	0.8437	0.7822	0.243	50	48.8	
SOLUTIONTF6	77.26	89.26	88.24	0.4667	0.1129	0.5712	62.9	44.5	*4*
......	......	......	......	......	......	......	......	......	...

Table 2.18 Final solutions for molecules I-VI after *FITING* (bLA)

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONF	85.07	93.75	81.94	0.4386	0.1093	0.0815	70.3	38.4	*17*
SOLUTIONF	31.26	88.14	265.07	0.6786	0.5059	0.5913	70.3	38.4	*1*
SOLUTIONF	34.7	89.29	268.52	0.682	0.5113	0.0932	70.3	38.4	*1*
SOLUTIONF	12.55	93.32	78.43	0.7825	0.775	0.7525	70.3	38.4	*21*
SOLUTIONF	22.16	87.72	82.65	0.8423	0.7827	0.243	70.3	38.4	*23*
SOLUTIONF	101.21	87.88	270.05	0.032	0.6085	0.4288	70.3	38.4	*3*

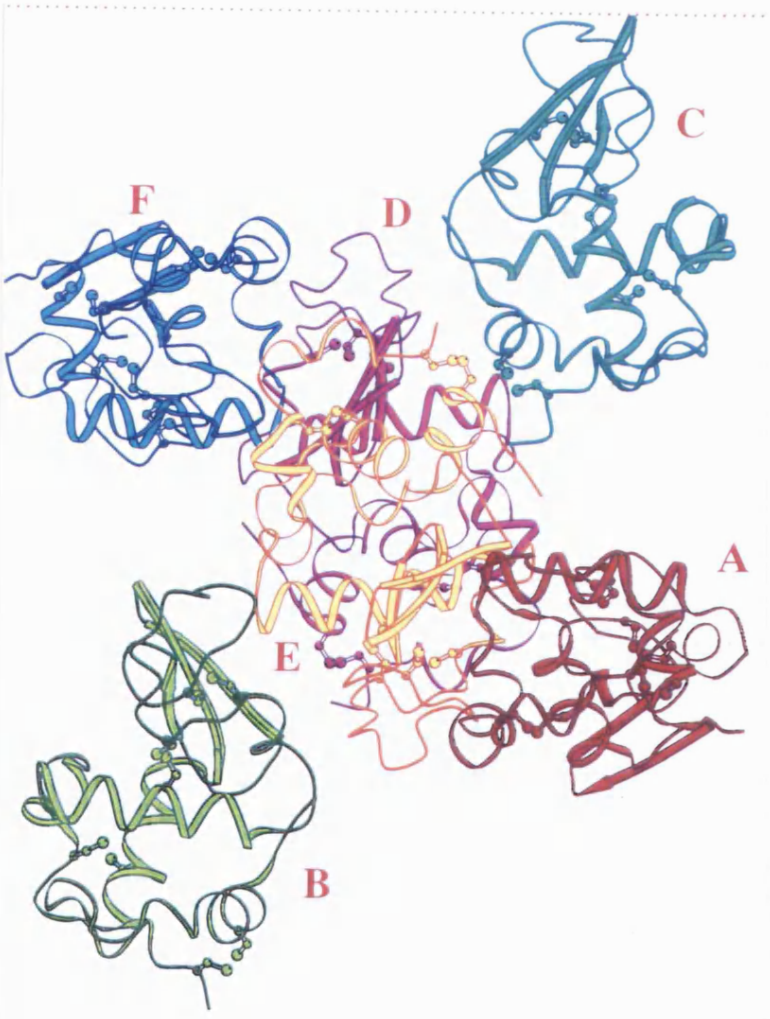


***Figure* 2.15**

Packing of bLA (6 molecules/a.u)

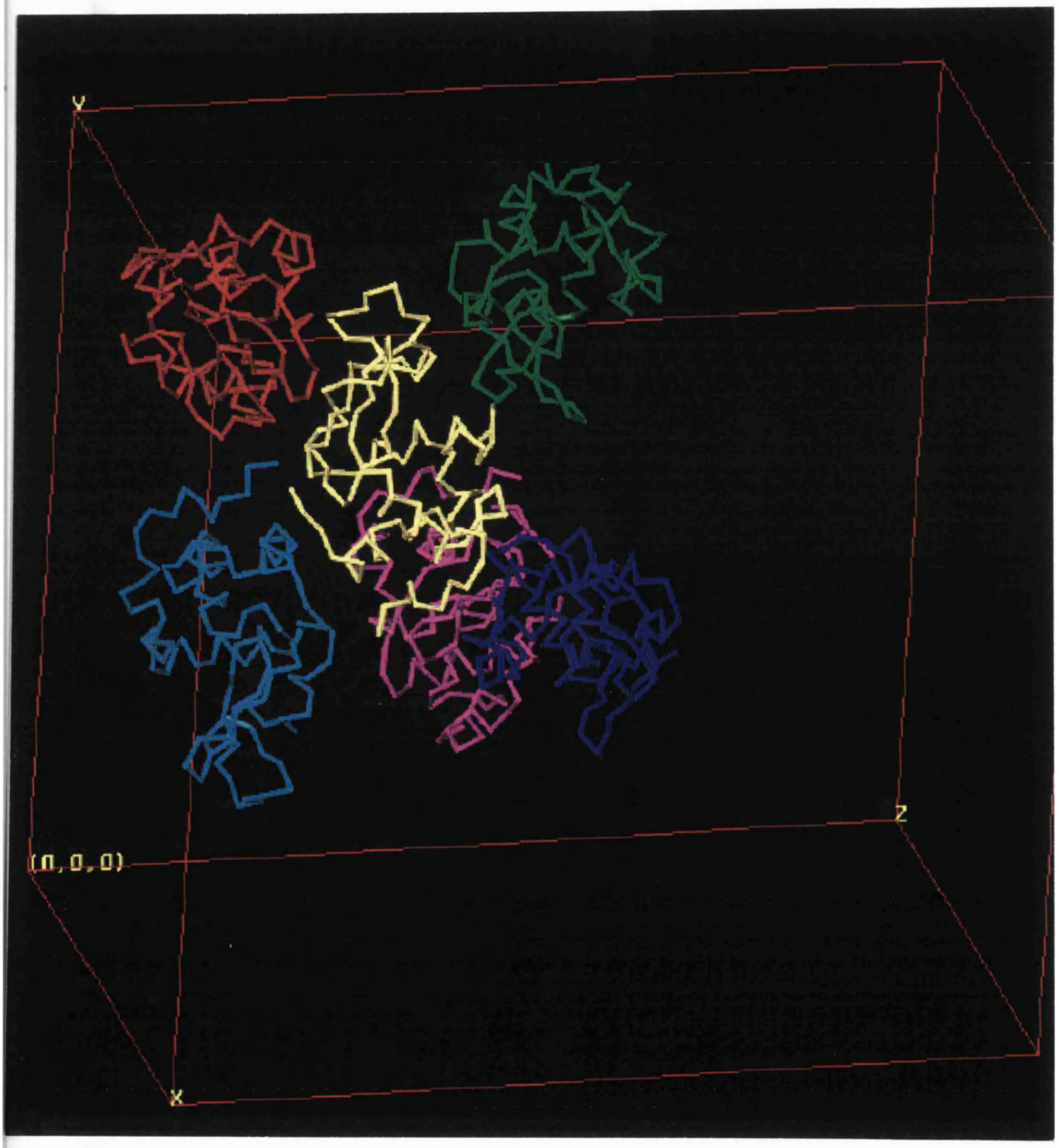


***Figure* 2.15a**

Packing of bLA (6 molecules/a.u)

2.2.1.5. Refinement

The resultant models of apo-LA and bLA structures (six molecules) were first subjected to rigid body refinement. Cycles of refinement were performed using the slowcool protocol at moderate temperatures with the program *X-PLOR* (Brünger, 1992b) using non-crystallographic restraints for apo-LA and with *CNS* (Brünger *et al.*, 1998) for bLA. The progress of refinement was monitored through both free and conventional R-factors (Brünger, 1992a). Alternative cycles of manual rebuilding with the graphics program 'O' (Jones *et al.*, 1991), intertwined with refinement cycles using the standard protocol improved the quality of the model.

Table 2.19 Refinement statistics for apo-LA and bLA

		apo-LA	BLA
Resolution range (Å)		40.0 - 2.2	30.0 - 2.2
Reflections		56422	45348
No of protein atoms		5844	5856
No of solvent molecules		167	89
Ca^{2+} ions		–	6
$^{\star}R_{free}$ (%)		24.8	25.3
$^{\text{¤}}R_{conv}$ (%)		19.1	21.6
Rms deviations in	*bond lengths (Å)	0.011	0.006
	*bond angles (°)	1.582	1.27
Temperature factors (Å^2):	overall	36.1	49.7
	average main-chain	34.1	48.9
	average side-chain	37.8	50.7
	Ca^{2+} ions	–	42.3

$^{\star}R_{free}=\Sigma hkl \subset T||Fobs-k|Fcalc||/\Sigma hkl \subset T|Fobs| \times 100$, where $hkl \subset T$ represents the test set (5% of the diffraction data). $^{\text{¤}}R_{conv}=\Sigma hkl||Fobs-k|Fcalc||/\Sigma hkl|Fobs| \times 100$. *Deviations from ideal values (Engh and Huber, 1991).

In the case of bLA structure, simulated annealing omit maps, calculated using *CNS*, indicated density (remained as a persistent feature of the structure at high sigma level) at the calcium site sufficient to accommodate a calcium ion in all six molecules. Six calcium ions were incorporated in to the model (one per molecule) at final stages of the refinement. In each case, water molecules were inserted in the model according to the ($2F_o$-F_c) and (F_o-F_c) electron density maps. During the final cycles of refinement the non-crystallographic restraints were released and individual molecules were examined on the graphics with the aid of electron density maps. Analysis of the Ramachandran plot [calculated with the program *PROCHECK* (CCP4, 1994)] for both structures showed that all residues lie in the allowed regions. The details of refinement statistics and model accuracy are listed in Table 2.19. Structural superpositions were performed with the program '*O*' (Jones *et al.*, 1991), *SHP* (Stuart *et al.*, 1979) and *MAPS* (Lu, 1998).

2.2.2. RESULTS

2.2.2.1. Overall structures

Even though the crystals showed diffraction to 2.0 Å, due to poor quality of data at high-resolution shells, the structures of both apo-LA and bLA were refined to 2.2 Å resolution. The two structures are closely similar to that previously described for recombinant mLA (Pike *et al.*, 1996) (*Figure* 2.16) and other known LA species variants (Table 2.20). Some differences were observed in the flexible, solvent exposed loop regions between residues 43-47, 62-65 and the C-terminal tail (which is highly disordered) between LA molecules. The functionally important

Table 2.20 Comparison of different LA structures

	apo-LA	bLA	mLA (bovine)	Baboon	Guinea Pig	Goat	Buffalo	Human	Human in complex with Zn^{2+}	Human with two Ca^{2+} ions
apo-LA	—	**0.68**	**0.80**	**0.93**	**1.01**	**0.75**	**0.80**	**0.99**	**0.97**	**0.94**
bLA	0.39	—	**0.51**	**0.89**	**0.82**	**0.61**	**0.60**	**0.87**	**0.87**	**0.89**
mLA (bovine)	0.41	0.28	—	**0.94**	**0.84**	**0.77**	**0.70**	**0.92**	**0.90**	**0.93**
Baboon	0.65	0.65	0.65	—	**1.06**	**1.06**	**1.04**	**0.41**	**0.36**	**0.47**
Guinea Pig	0.57	0.44	0.44	0.66	—	**1.11**	**1.08**	**1.01**	**1.03**	**1.04**
Goat	0.60	0.45	0.54	0.80	0.70	—	**0.52**	**1.05**	**1.01**	**1.02**
Buffalo	0.53	0.95	0.48	0.76	0.63	0.27	—	**1.01**	**1.01**	**1.02**
Human	0.66	1.00	0.65	0.27	0.63	0.81	0.78	—	**0.35**	**0.35**
Human in complex with Zn^{2+} ion	0.62	0.94	0.58	0.21	0.61	0.71	0.69	0.26	—	**0.35**
Human with two Ca^{2+} ions	0.63	1.37	0.60	0.23	0.59	0.74	0.71	0.23	0.16	—

The rms deviations are given after superposition using the least square fitting option in 'O' applied to C_α atoms of residues 1-120 (values in boldface) and to the core of the molecule (values in normal typeface). LA-core: residues 5-11, 23-40, 50-61 and 71-104.mLA: recombinant bovine LA at 2.3 Å [(Pike *et al.*, 1996); PDB code: 1HFZ]. Baboon LA at 1.7 Å [(Acharya *et al.*, 1989); PDB code: 1ALC]. Guinea Pig LA at 1.9 Å [(Pike *et al.*, 1996); PDB code: 1HFX]. Goat LA at 2.3 Å [(Pike *et al.*, 1996); PDB code: 1HFY]. Buffalo LA at 2.3 Å (Calderone *et al.*, 1996). Human LA at 1.7 Å (Acharya *et al.*, 1991). Human LA in complex with Zn^{2+} at 1.7 Å [(Ren *et al.*, 1993); PDB code: 1HML]. Human LA with two Ca^{2+} ions at 1.8 Å [(Chandra *et al.*, 1998); PDB code: 1A4V].

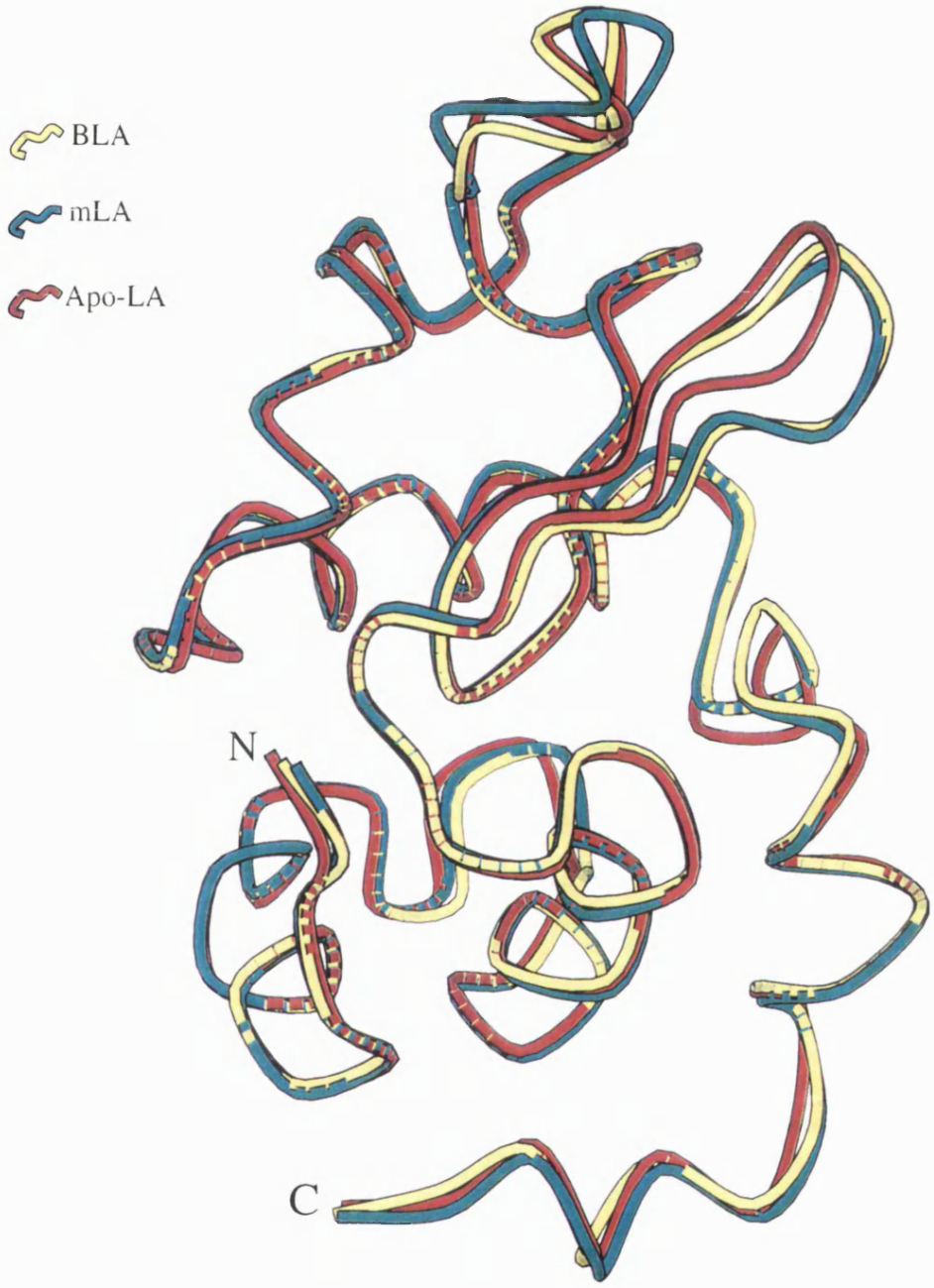


***Figure* 2.16**

Superposition of apo-LA, bLA and mLA

flexible loop region (involving residues 105-111) located adjacent to the lower end of the cleft adopts 'helical' conformation in both apo-LA and

bLA structures (similar 'helical' conformation was observed in the previously reported mLA structure crystallised at pH 8.0, (Pike *et al.*, 1996). It is known that at high pH (6.5-8.0) this loop adopts the 'helical' conformation and at low pH (4.6), the 'looped-out' conformation. The helical conformation is predominant at pH above 6.0 in different LA structures reported so far (Harata and Muraki, 1992; Pike *et al.*, 1996).

The monomers in the two structures have nearly identical structures (Table 2.21).

Table 2.21 Comparison of different LA molecules

	apo-LA					
	mol1	**Mol2**	**mol3**	**mol4**	**Mol5**	**mol6**
mol1	—	**0.39**	**0.40**	**0.39**	**0.28**	**0.32**
mol2	0.21	—	**0.47**	**0.45**	**0.33**	**0.38**
mol3	0.18	0.22	—	**0.49**	**0.46**	**0.37**
mol4	0.26	0.30	0.30	—	**0.36**	**0.42**
mol5	0.18	0.18	0.17	0.24	—	**0.32**
mol6	0.23	0.27	0.19	0.28	0.20	—
	BLA					
	mol1	**Mol2**	**mol3**	**mol4**	**Mol5**	**mol6**
mol1	—	**0.41**	**0.57**	**0.53**	**0.58**	**0.62**
mol2	0.19	—	**0.44**	**0.40**	**0.45**	**0.42**
mol3	0.23	0.18	—	**0.53**	**0.67**	**0.58**
mol4	0.24	0.20	0.23	—	**0.51**	**0.58**
mol5	0.30	0.25	0.28	0.21	—	**0.48**
mol6	0.21	0.19	0.17	0.22	0.28	—

The rms deviations are given after superposition using the least square fitting option in 'O' (Jones *et al.*, 1991) applied to C_α atoms for residues 1-120 (values in boldface) and to the core of the molecule (values in normal typeface). LA-core: residues 5-11, 23-40, 50-61 and 71-104.

Each monomer in apo-LA and bLA has ~6,512 $Å^2$ and ~6,820 $Å^2$ accessible surface area respectively. Upon formation of dimers, a loss of 463.14 $Å^2$ (~3.6%) in apo-LA and 535.68 $Å^2$ (~3.9%) in bLA of the accessible surface area is observed, consistent with the results from other protein-protein interfaces.

The interactions between the monomers are mainly mediated through charged residues (Table 2.22). The packing arrangements in the two structures are shown in *Figures* 2.14 and 2.15.

A number of water molecules (167 in the apo-LA and 89 in the bLA structures respectively, Table 2.19) were identified with average temperature factors < 45$Å^2$. Seven water molecules are conserved in bLA and other known LA structures, and only two of these are conserved in apo-LA.

2.2.2.2. Calcium binding site

The Ca^{2+} binding site in bLA has maintained identical conformation (distorted pentagonal bipyramid co-ordination), known to be present in all LA structures (*Figure* 2.17, Table 2.23). As in the mLA structure, the average temperature factor for the Ca^{2+} ion is slightly higher. However, the presence of Ca^{2+} ion and the two water molecules that form the co-ordination sphere were clearly visible (except in mol5, where only one water molecule was identified) in the electron density map.

Examination of the apo-LA structure revealed that the absence of Ca^{2+} ion did not cause any structural change in this region (r.m.s. devia-

Table 2.22. Crystal packing contacts between monomers of bLA and apo-LA structures

	bLA		apo-LA	
	protein-protein	water mediated	protein-protein	water mediated
Mol1	S70/L81-*E*, S70/Q39-*E*, K98/Q43-*E*, V99/Q43-*E*$^{(2)}$		Q43/L05-*E*, Q43/Y103-*E*, Q43/E49-*E*, N44/N44-*C*, D46/K58-*B*, E49/Q43-*C*, K58/D46-*B*, K58/N45-*C*, D64/K108-*E*, N102/H68-*F*, Y103/Q43-*C*, L105/Q43-*C*, K108/D64-*C*	Q43/A106-*E*/O12-*W*, H68/Y103-*F*/O12-*W*, H68/Q65-*B*/O135-*W*, H68/D64-*B*/O296-*W*, S70/K98-*F*/O311-*W*, A106/Q43-*C*/O317-*W*
Mol2	Q39/S70-*E*, Q39/K98-*E*, Q43/V99-*E*$^{(3)}$, Q43/G100-*E*	Q43/K98-*E*/D97-*E*/O105-*W*, D83/H68-*E*/O33-*W*	Q43/L105-*F*, Q43/Y103-*F*, Q44/N44-*D*, N45/K58-*F*, E49/Q43-*D*, N56/N45-*D*, D64/K108-*F*, H58/N102-*E*, N102/H68-*E*, Y103/Q43-*D*, L105/Q43-*D*, K108/D64-*D*	I41/T33-*F*/O107-*W*, I41:2/H32-*F*/O177-*W*, Q43/A106-*F*/O44-*W*, N44/N44-*D*/O284-*W*, N56/N45-*D*/O156-*W*, A106/Q43-*D*/O115-*W*
Mol3			N44/N44-*E*, D46/K58-*F*, E49/Q43-*E*$^{(2)}$, N56/N45-*E*, K58/D46-*F*, H68/N102-*D*, N102/H68-*D*, Y103/Q43-*E*, L105/Q43-*E*, K108/D64-*E*	Q65/H68-*F*/O148-*W*, H68/Q65-*F*/O148-*W*, K98/N71-*D*/O268-*W*, A106/Q43-*E*/O75-*W*
Mol4	N45/A121-*E*, Q39/K98-*F*	D78/*K*58-*F*/O58-*W*, D82/H68-*F*/O59-*W*	N44/N44-*F*, D46/K58-*E*$^{(2)}$, E49/N44-*F*, N56/N45-*F*, Y103/Q43-*F*, L105/Q43-*F*, K108/D64-*F*	H32/Q39-*F*/O300-*W*, H68/Q65-*E*/O136-*W*, A106/Q43-*F*/O287-*W*
Mol5	K114/E25-*F*, L115/G20-*F*	L115/E25-*F*/O71-*W*, D116/E25-*F*/O119-*W*, D116/H107-*F*/N102-*F*/O130-*W*		

Note : The crystal packing contacts listed above were calculated using CCP4 (CCP4, 1994) and include only H-bond interactions between the monomers with a cutoff distance 3.4 Å. The letters in italics represent the molecule identity while numbers in superscript represent the number of contacts.

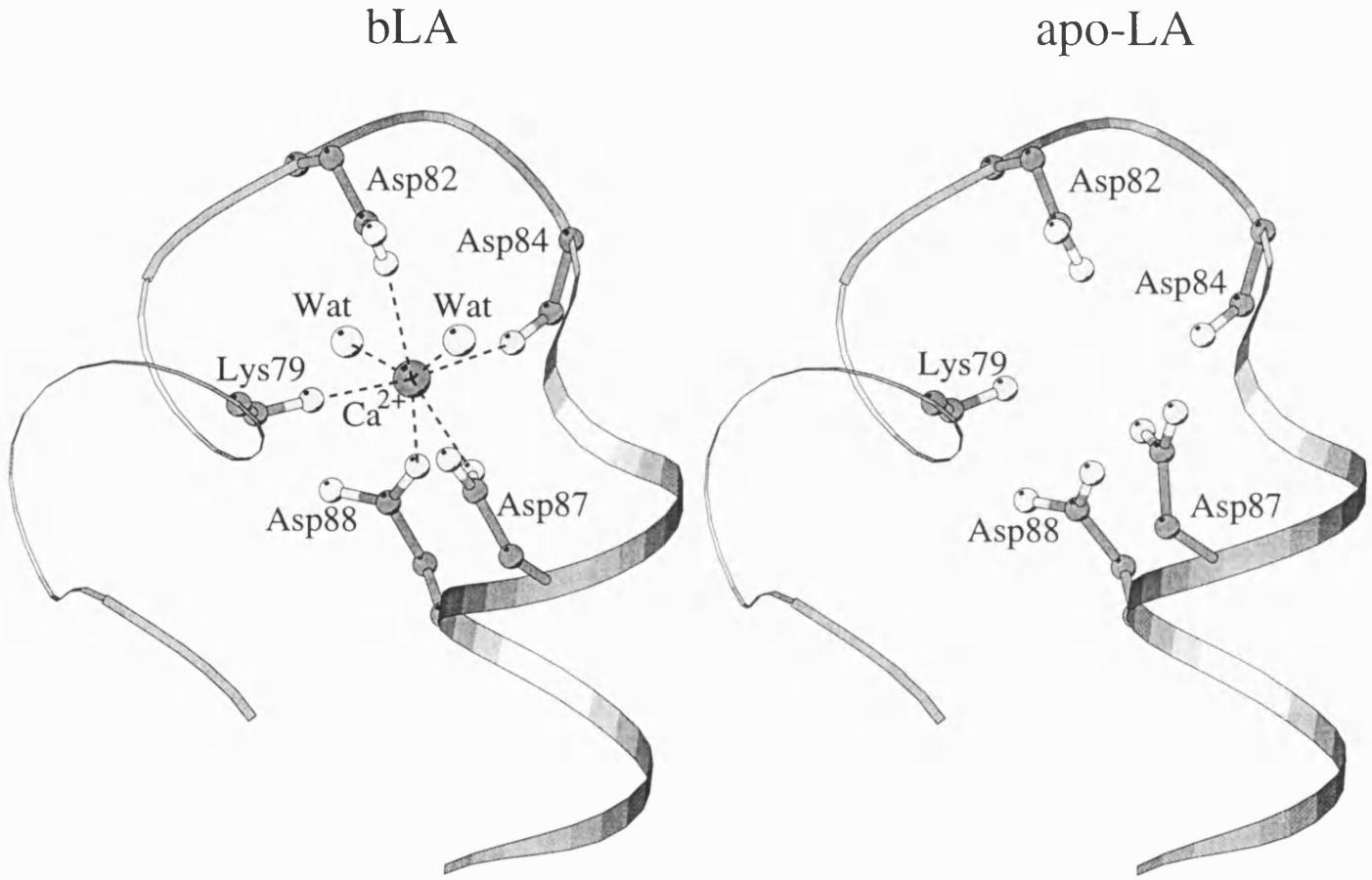


***Figure* 2.17**

Comparison of apo-LA and bLA structures at the calcium binding site. The figure was produced using BOBSCRIPT (Esnouf, 1997).

tion for Cα atoms between residues 78-90 for apo-LA and bLA is 0.26 Å) to compensate for the large net negative charge in this region except slight change in orientation of Asp-87 side-chain and some perturbation of Lys 79 side chain beyond Cβ atom (the Nζ atoms of apo- and holo-forms are separated by 1.7 Å). Comparison of the thermal parameter distribution for apo- and holo-LA confirms also the above observations as it has been indicated in *Figures* 2.18-2.20.

Table 2.23 Thermal parameters at the Calcium binding site (Å²)

	bLA						
	Lys79 O	**Asp82 OD1**	**Asp84 O**	**Asp87 OD1**	**Asp88 OD1**	**Wat (1)**	**Wat (2)**
Mol1	37.68	35.58	38.16	38.23	32.34	35.71	31.82
Mol2	38.45	52.59	50.39	40.58	33.80	46.88	42.57
Mol3	43.24	49.74	54.50	44.35	43.04	40.12	40.61
Mol4	40.96	43.17	46.41	34.57	36.63	32.68	39.32
Mol5	48.31	62.25	54.45	53.05	51.10	40.97	—
Mol6	39.11	41.37	38.25	39.84	30.87	33.86	35.17

	Equivalent region in apo-LA				
	Lys79 O	**Asp82 OD1**	**Asp84 O**	**Asp87 OD1**	**Asp88 OD1**
Mol1	29.76	55.50	35.78	56.53	28.38
Mol2	30.99	64.83	35.44	42.54	34.26
Mol3	36.30	62.98	39.62	42.83	34.20
mol4	37.52	57.51	38.07	49.20	37.74
mol5	35.04	64.89	36.57	51.08	33.74
mol6	39.87	69.30	43.10	50.59	37.72

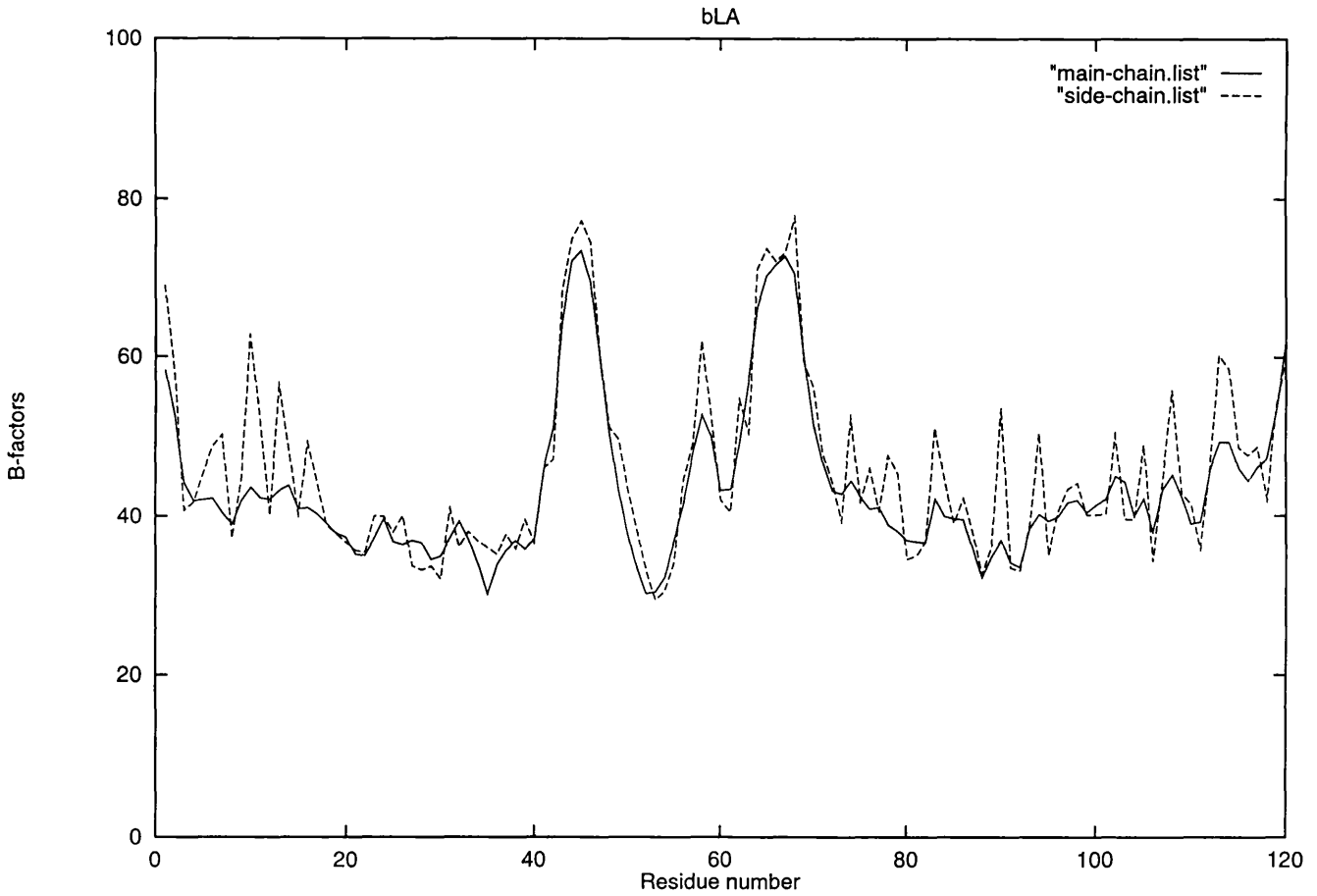

bLA
"main-chain.list"
"side-chain.list"
B-factors
Residue number
0
20
40
60
80
100
120

Figure 2.18

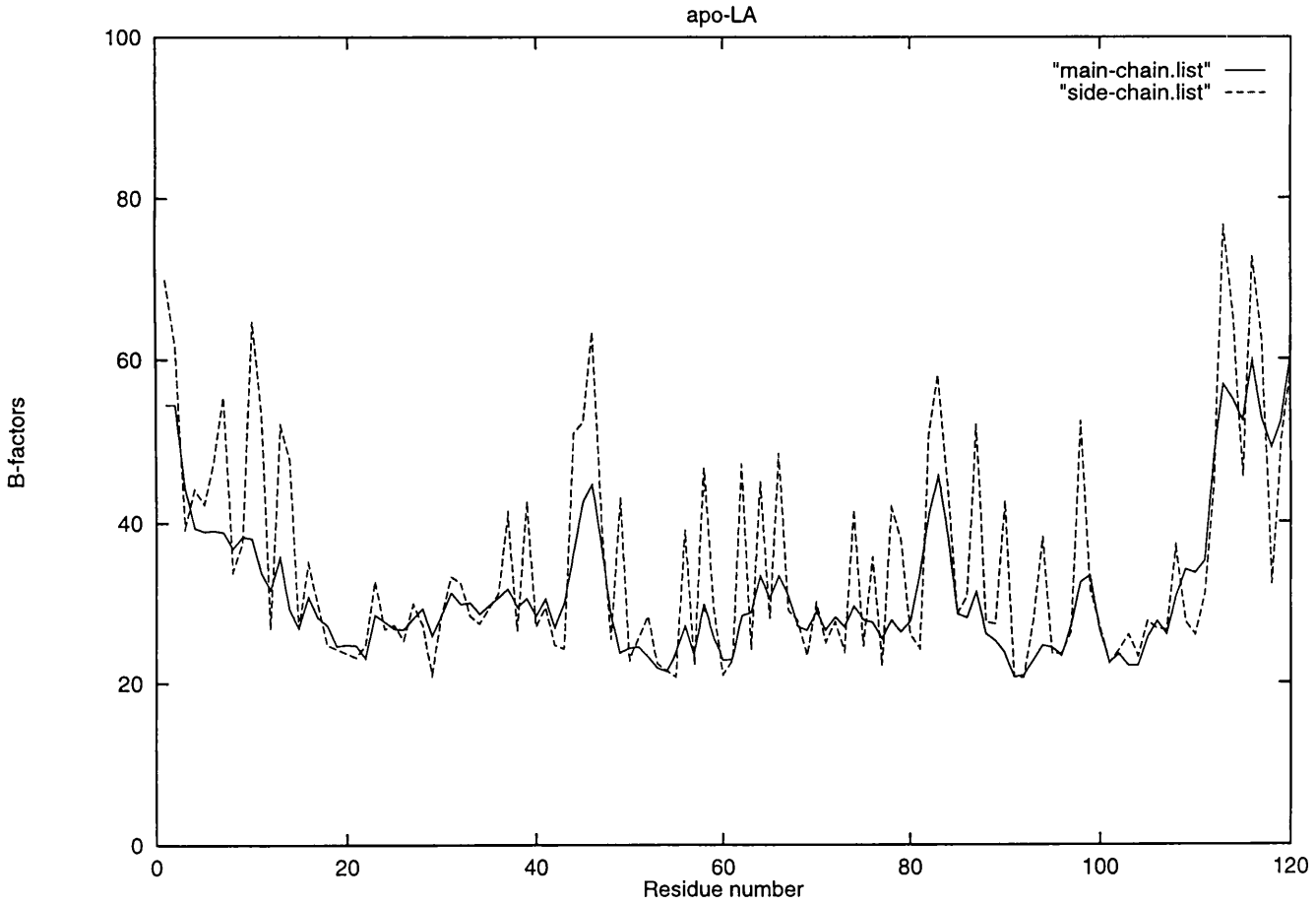

apo-LA
"main-chain.list"
"side-chain.list"
B-factors
Residue number
0
20
40
60
80
100
120

Figure 2.19.

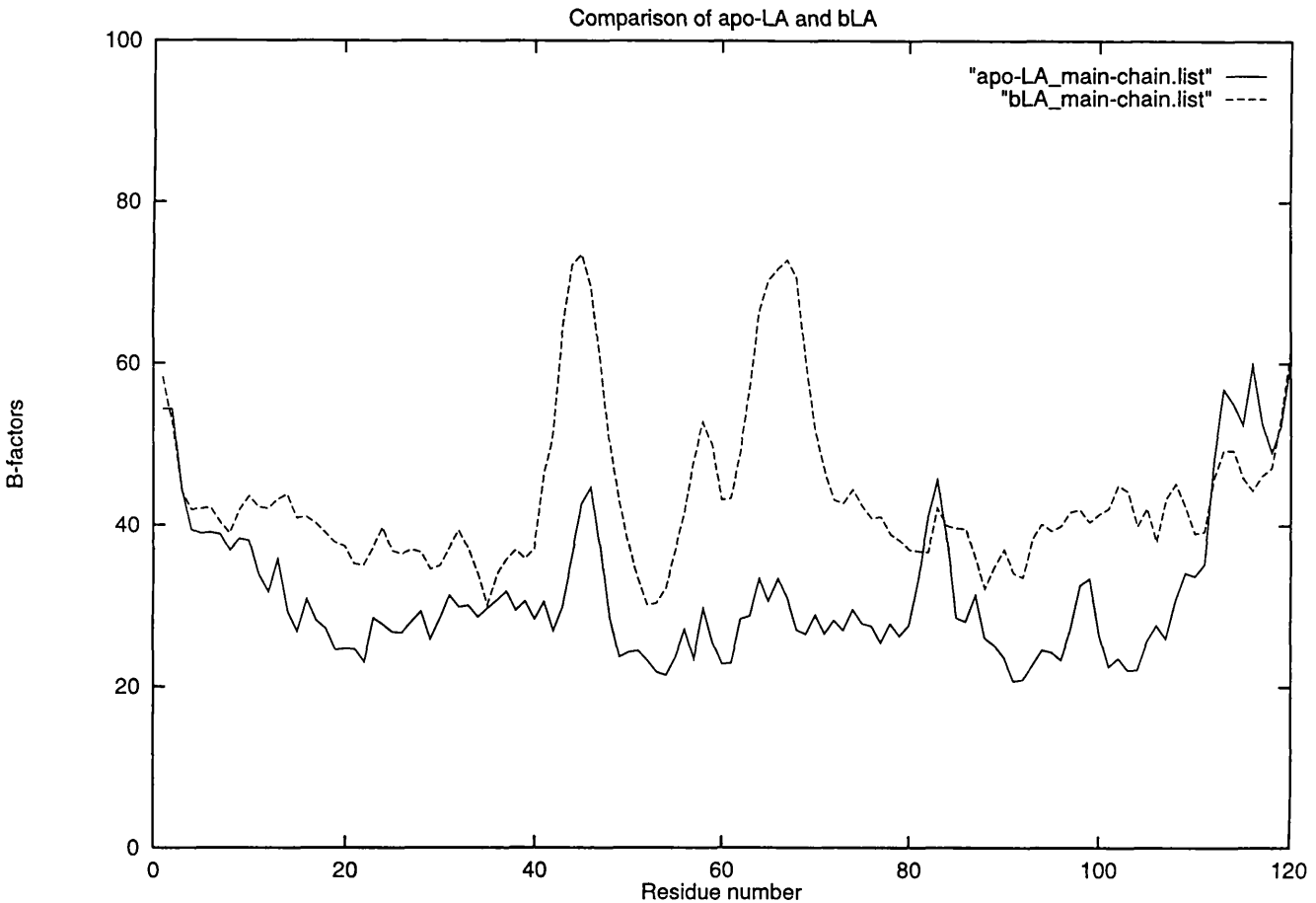

Comparison of apo-LA and bLA
"apo-LA_main-chain.list"
"bLA_main-chain.list"
B-factors
Residue number
0
20
40
60
80
100
120

Figure 2.20

The two conserved water molecules (part of the Ca^{2+} site), which were observed in all LA structures were absent in the apo-LA structure.

This was confirmed independently by increasing the resolution limit to 2.0 Å. One round of refinement was performed at this resolution (even though the quality of data between resolution shells 2.2 –2.0 Å were weak) and the electron density map did not indicate the presence of these two conserved water molecules.

Similarly independent checks (refinement followed by analysis of thermal parameters and careful examination of electron density maps) were performed to rule out the presence of any residual Ca^{2+} ion binding/sodium ion/replacement of Ca^{2+} ion by a water molecule.

In none of the above cases, the results were convincing to support the existence of any ligand at the calcium-binding site.

Comparison of the calcium-binding loop of bLA with the corresponding one in apo-LA in terms of accessibility shows that this region in apo-LA has slightly higher solvent accessibility. In the high-resolution structure of baboon LA (1.7 Å resolution), an internal channel starting at Ile 27 and Asp 88 (from the calcium binding site), partly occupied by internal water molecules has been described (Acharya *et al.*, 1989). The corresponding channel is present in both bLA and apo-LA structures. However, less number of water molecules are trapped in this channel (which may be partly due to the lower resolution of these two structures, at 2.2 Å resolution) and three of these water molecules are conserved in all three structures (baboon LA, bLA and apo-LA).

2.2.2.3. Cleft region

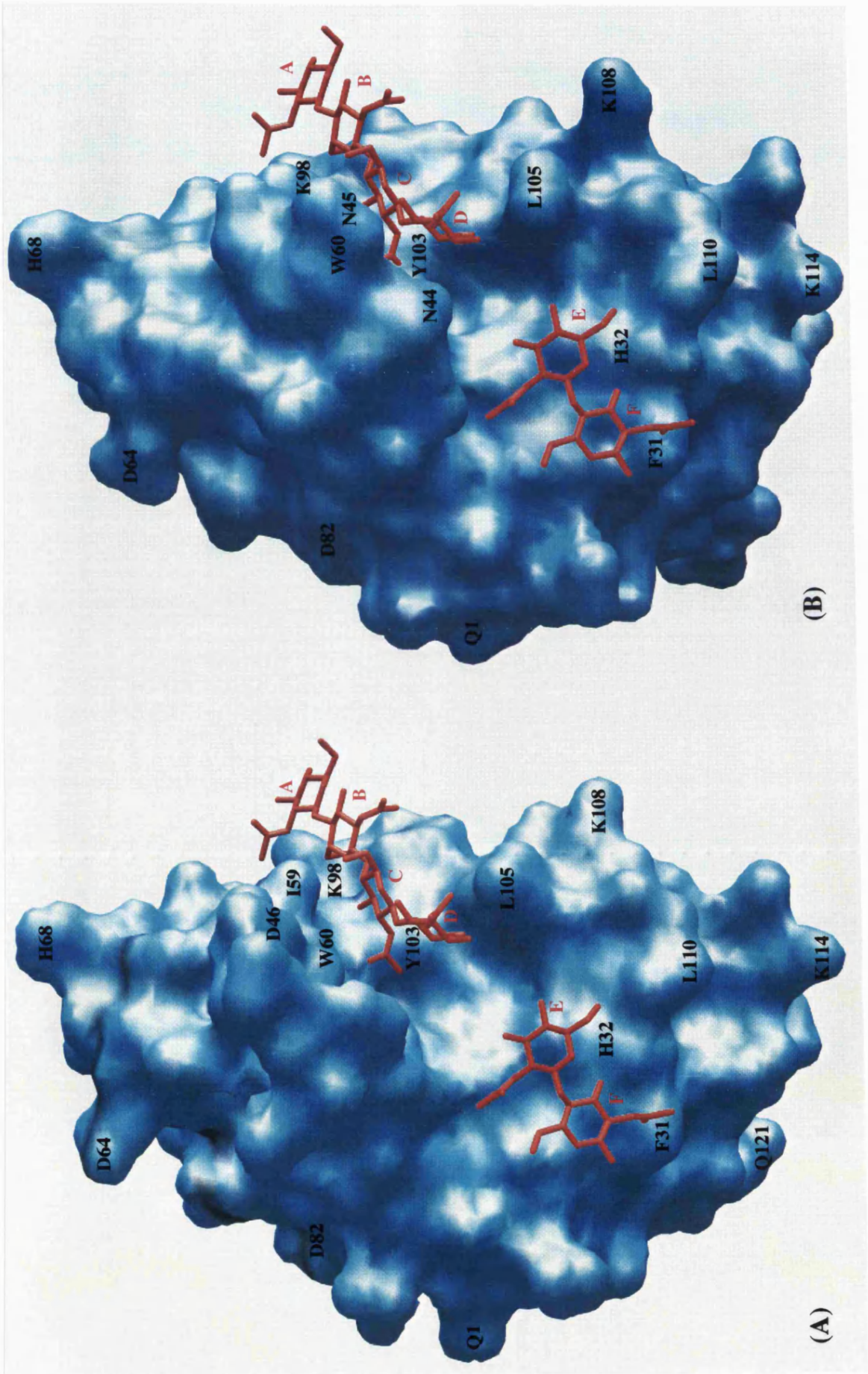
It has been speculated that LA utilises the residues in the cleft region to stabilise monosaccharide binding in lactose synthase (Grobler *et al.*, 1994b; Malinovskii *et al.*, 1996). A similar cleft (the active site region) has been described for the homologous protein, LZ and consists of six binding subsites (A-F) to which the hexasaccharide binds. In the case of LA, subsites A and B are blocked by Tyr-103 side-chain (Acharya *et al.*, 1989).

However, available data based on mutagenesis studies and crystallographic analyses of LA structures suggest that the region corresponding to subsite F of LZ and an adjacent surface participate either directly or indirectly in monosaccharide binding by the complex (Grobler *et al.*, 1994b; Pike *et al.*, 1996).

In the bLA structure, there is a hydrogen bond between the phenolic OH group of Tyr 103 and the carbonyl oxygen of Gln-54 (2.9 Å); this is part of the interaction interface between the two lobes Tyr 103 being in the helical lobe and Gln 54 in the beta-lobe. The most

***Figure* 2.21** (overleaf).

The molecular surfaces of apo-LA and bLA are shown in (A) and (B) respectively. Minor perturbation of Tyr-103 in apo-LA resulted in slight opening of the cleft region. The carbohydrate of hexa-N-acetylchitohexaose [from the human LZ structure, Song et al., 1994)] was modelled [followed by molecular dynamics using the program CNS (Brünger et al., 1998) in apo-LA (A). The predicted position of the hexasaccharide based on modelling studies in apo-LA structure has been retained in bLA (B) for comparative purpose. The figure was prepared using the program GRASP (Nicholls et al., 1991).


(A)
Q1
D82
D64
H68
D46
I59
W60
K98
Y103
L105
K108
L110
K114
Q121
H32
F31
A
B
C
D
E
F
(B)
Q1
D82
D64
H68
W60
K98
N45
N44
Y103
L105
K108
L110
K114
H32
F31
A
B
C
D
E
F

significant structural change observed in apo-LA is found in this region; Tyr 103 is shifted towards the interior of the cleft and a water-mediated interaction with Gln 54 replaces the direct hydrogen bond. This perturbation of Tyr 103 is found in all 6 molecules of apo-LA and results in a more open cleft that is reflected in the altered position of the beta-lobe (*Figure* 2.21).

Superposition of the two lobes of LA molecule for both apo-LA and bLA suggests a slight outward movement of the two lobes (α-domain and β-domain) with respect to each other. More specifically the angle between the 3_{10} helix h_2 (residues 77-80) and helix C (residues 86-98) is 68.50° for apo-LA and 69.55° for bLA. Moreover superposition of the backbone of apo-LA and bLA reveals that apo-LA deviates more in the β-sheet region than in other parts of the molecule. The superimposed structures as well as a more detailed view of the two aromatic clusters are shown in *Figure* 2.22.

In order to investigate further the reorientation of Tyr-103 side chain, an attempt to fit an oligosaccharide in the cleft was made using modelling approaches. The carbohydrate of hexa-N-acetyl-chitohexaose was selected from the human LZ structure determined by Song *et al.* (1994) and a simple molecular dynamics exercise was performed using the program CNS (Brünger *et al.*, 1998). The results indicated that in both apo-LA and bLA, subsites D and F had maintained their conformation, while the more open cleft of apo-LA, unlike the closed cleft in bLA could potentially accommodate carbohydrate residues in sites A-E. This may be partly due to the flexibility of the loop region between residues 44-46 and the fact that Tyr-103 and Gln-46 seem to

adopt a different conformation in apo-LA (*Figure* 2.21). The functional implications of these observations are currently being investigated.

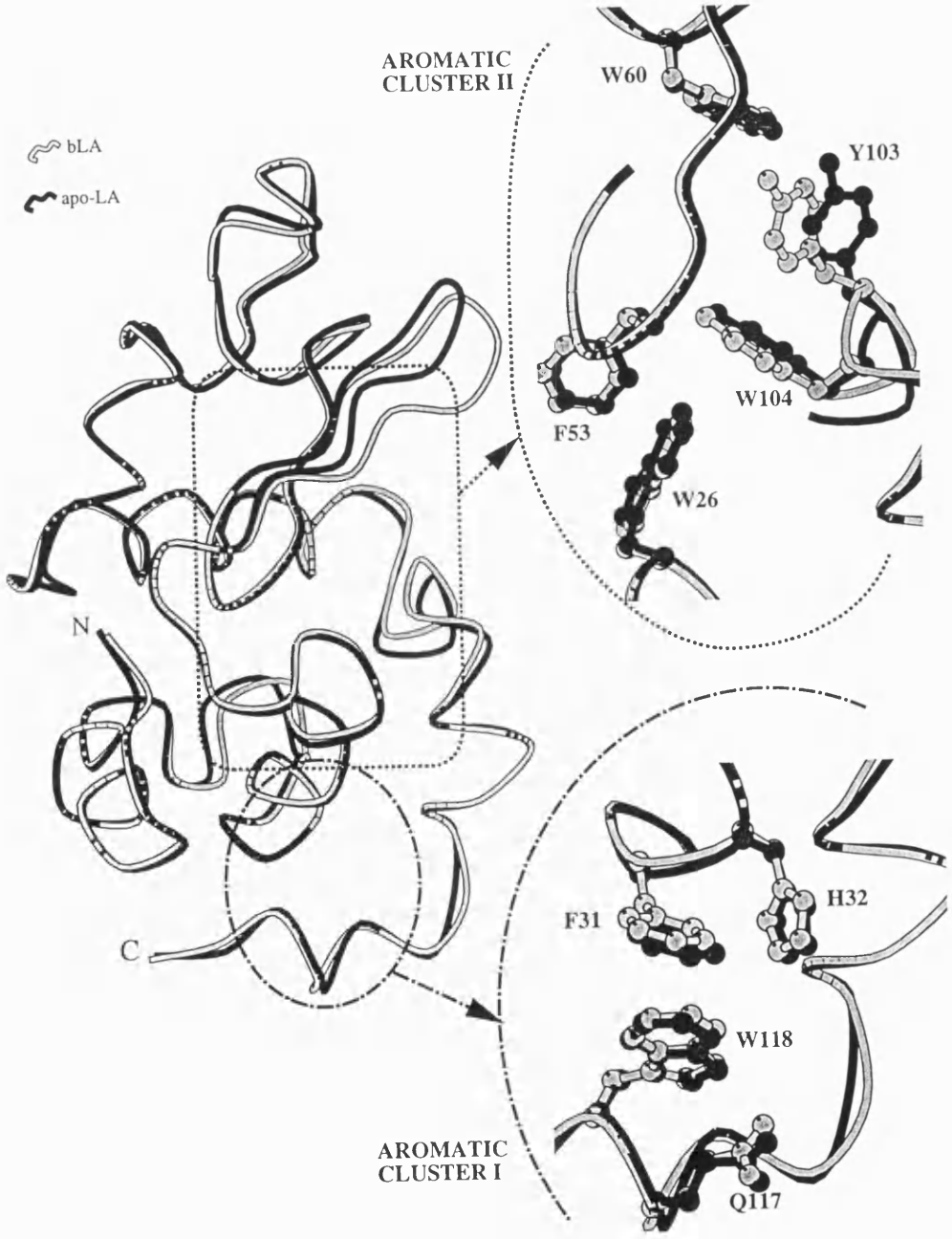


***Figure* 2.22**

Superposition of apo-LA and bLA structures giving a more detailed view of the two aromatic clusters

2.2.3. DISCUSSION

Different conformational states of LA have varying affinities for Ca^{2+}. The native state binds Ca^{2+} with submicromolar affinity; although the molten globule state has a relatively weak millimolar affinity for calcium (Kuwajima, 1989), Ca^{2+} does also stabilise this compact but disordered state (Forge *et al.*, 1999; Vanderheeren and Hanssens, 1994). Calcium accelerates the rate of refolding in denatured bLA (disulphide bonds intact) by more than two orders of magnitude (Forge *et al.*, 1999; Kuwajima, 1989) reflecting its binding to high energy intermediates in the folding process which have pre-formed Ca^{2+} binding sites. The requirement for Ca^{2+} for the formation of native protein with correct disulphide bond arrangements in the reduced denatured protein (Ewbank and Creighton, 1991; Rao and Brew, 1989) suggests that it also acts by selective binding to similar rate-limiting intermediates in the oxidative folding process. The 'molten globule' state and the 'transition state' for folding are arrays of conformations as opposed to discreet species and Ca^{2+} binding appears to be a common property of a range of partially folded states of LA in which affinity for the ion increases as their Ca^{2+} binding sites become more native.

It has been established that the binding of calcium is not required for activity in LA (Berliner *et al.*, 1983; Kronman *et al.*, 1981; Kuwajima *et al.*, 1986) in keeping with the structural studies described here that show that overall fold of LA is unchanged and the Ca^{2+} binding site has similar structures in the holo- and apo-forms of bLA. A cation from the crystallisation medium does not replace the calcium in the apo-form (for example sodium ion). This confirms that stabilisation of the

native state of the apo-protein by higher concentrations of monovalent cations is related to ionic strength effects as opposed to specific binding. The calcium-chelating water molecules are also absent. Although changes in the immediate vicinity of the site appear to be trivial, a larger structural change is found at the inter-lobe interface (which connects the calcium binding site and the cleft involving Tyr-103 through a channel with trapped water molecules) located on the opposite face of the apo-LA structure.

The architecture of the LA Ca^{2+}-binding site is superficially similar to the E-F hand of calmodulin (CaM) and related Ca^{2+} binding proteins. Although the affinity of LA for Ca^{2+} is similar to that of CaM (Kronman, 1989; Kronman *et al.*, 1981), in other ways the interaction of Ca^{2+} with LA is unique (the dissociation constant of Ca^{2+} to apo-LA is of the order of 10^{-7}M). The association of Ca^{2+} to CaM and other intracellular regulatory proteins, as well as organic chelators, involves a positive entropy change. This change is thought to reflect the fact that the increase in entropy resulting from the loss of hydration from the Ca^{2+} and coordinating groups in the binding site, exceeds the decrease in entropy resulting from enhanced order in the Ca^{2+} complex (Schaer *et al.*, 1985). In LA, Ca^{2+} binding is accompanied by a large negative entropy change, indicating that the Ca^{2+} ion has a major role in the protein. The binding of Ca^{2+} to CaM is essential for its ability to activate enzymes whereas both apo and holo-LAs are active in lactose synthase (Berliner *et al.*, 1983; Kronman *et al.*, 1981). Moreover recent studies by Anderson *et al.* (1997) have shown that the disruption of the Ca^{2+} binding site in bLA did not prevent the protein from stimulating lactose

synthase activity even though the binding ability of the Ca^{2+} ion was lost. The precise role of Ca^{2+} has remained unclear. Although it is not uncommon to find Ca^{2+} or another metal ligand acting in stabilising a secreted protein (e.g. trypsinogen) and the Ca^{2+} in LA does increase stability, this seems unlikely to be biologically significant since the site of action of LA is in an intracellular compartment (the lumen of the golgi) and involves the short time span in which it is in contact with GT in the golgi.

However, we can rule out a 'functional role' for the Ca^{2+} ion in LA by drawing analogy with LZ molecule. It is known that LA arose through gene duplication for a Ca^{2+} binding LZ (Brew and Grobler, 1992; Grobler *et al.*, 1994a). Interestingly, the crystal structure analyses of the apo- and holo mutant human LZs with an introduction of a Ca^{2+} binding site did not seem to have had any effect on the overall structure or change in molecular rigidity of the proteins (Inaka *et al.*, 1991). Only subtle, local perturbations were noticed to accommodate the Ca^{2+} ion in the LZ structures. Based on calorimetric studies, these small changes were attributed to an increase in entropy due to the release of water molecules bound to the Ca^{2+} in solution (Kuroki *et al.*, 1992). Similarly comparison of the Ca^{2+} loop in equine LZ (Tsuge *et al.*, 1992) and the corresponding loop in hen-egg-white LZ (a prototype non-Ca^{2+} binding LZ (Steinrauf, 1998), did not show any significant difference in conformation (r.m.s. deviation of 0.59 Å for C^{α} atoms between residues 81-93). These results are complementary to those observed in apo-LA and bLA structures as presented here.

Calcium appears to affect folding by binding to the 'transition

state' or array of transition states, facilitating the locking in of tertiary interactions between the two lobes of the molecule. Iyer and Qasba (1999) have performed a molecular dynamics simulation of LA and calcium binding c-type LZ and the results of their investigation suggest the possibility of a general mechanism by which the protein dynamics and function are modulated by metal binding. It may be possible that the function of metal ion binding site may be not only to maintain the geometry of the cleft, but also to increase the flexibility of the protein at regions both away from the binding site and those involved in monosaccharide binding during the formation of LS complex. Hence, it is plausible that the conformational alteration observed in the cleft region of apo-LA and bLA structures could be attributed to the Ca^{2+}.

It has been observed that at temperatures above 20°C, on removal of Ca^{2+} from native LA, it adopts a molten globule structure, a compact conformer with a large proportion of native secondary structure but little fixed tertiary structure. This state also predominates at low or high pH and at higher temperatures (Kuwajima, 1989). There is much interest in the molten globule state of LA (Alexandrescu *et al.*, 1992; Alexandrescu *et al.*, 1993; Baum *et al.*, 1989; Ewbank and Creighton, 1991) since it is closely similar to an early intermediate in the folding pathway. However, we may not be able to determine the molten globule state structure of LA using X-ray crystallography. On the other hand, the comparison of the structures of apo-LA and bLA presented here clearly points to a 'structural role' for the Ca^{2+} ion that seems to be important for the stability of the protein with Ca^{2+} binding region being the most rigid part of LA molecule.

2.2.4. FURTHER CRYSTALLOGRAPHIC STUDIES WITH apo-LA

The crystal structure of apo-LA could be set as a template for further binding studies of metal ions in LA molecules. Several studies over the years have shown that possible binding site/sites for ions such as Tb^{3+}, Mn^{2+}, Co^{2+} etc. exist in LA molecule and since most of these ions can bind also to GT it is very important to establish the structural details of various metal ion interactions with LA molecule.

With this aim, co-crystallisation experiments of apo-LA with excess of Ca^{2+}, Co^{2+} and Tb^{3+} were performed and low quality data have been collected for the complex of apo-LA with calcium ion. The crystals co-crystallised either with Co^{2+} or Tb^{3+} showed very poor diffraction and turned out to be disordered after exposure to X-rays. Recent studies concerning the effect of pH on apo-LA showed that apo-LA crystallises in the tetragonal lattice in pH 8.0 (buffer: Tris/HCl) also in the presence of 2.2 M ammonium sulphate.

2.3. α-LACTALBUMIN VARIANTS

2.3.1. PREVIOUS WORK ON α-LACTALBUMIN VARIANTS

2.3.1.1. Preparation of mLA variants

The gene for bovine LA has been cloned and expressed at high levels in *Escherichia coli* and recombinant bovine LA (mLA) was obtained by limited proteolysis of the fusion protein, consisting of the mature bovine LA sequence connected to the NH2-terminal 50 residues of human cathepsin D by a linker sequence containing protease cleavage sites (Wang *et al.*, 1989). This fusion protein was expressed in an insoluble form and accumulated to about 50% of the total bacterial protein within 3h after induction of T7 RNA polymerase synthesis and was converted to an active form, the maximum specific activity of which was only 25% of that of native LA. However, after cleavage with trypsin a product indistinguishable in structure and activity from native LA was obtained. The non-specific cleavage of the fusion protein could be destructive in the study of LA variants, hence a methionine residue was introduced prior to the coding sequence using M13 site-directed mutagenesis. The coding region of LA was also rendered resistant to CNBr cleavage by local substitution of methionine 90 with a valine, a residue that appears to be conserved in several LA primary structures. This new construct (pC-LA), provided a recombinant form of bovine LA (M90V-LA) by specific cleavage of the expressed fusion protein with CNBr, that was essentially identical to the wild type protein in terms of kinetic parameters and CD spectra while the final yields were improved to 10mg folded protein per litre of bacterial culture.

Although the fusion expression system was used successfully the

quantities of the produced proteins were not sufficient for crystallisation studies, thus the construction of a new non-fusion expression system was performed (Grobler *et al.*, 1994). Bovine LA was expressed in *Escherichia coli* strain BL21(DE3) by using a vector (pMLA) generated by cloning the coding sequence for LA into pET3a vector at a site adjacent to the codon for the initiator methionine. The coding region for mature LA (in which the codon for methionine 90 was modified by M13 site directed mutagenesis to valine as described above) was amplified by PCR using the primers designated NF-N and NF-C. These primers were designed to introduce a *Nde*I site at the 5' end and a *Bam*HI site at the 3' end, together with an initiator methionine codon immediately preceding the bovine LA sequence. The product of the amplification was purified by agarose gel electrophoresis, digested with *Nde*I and *Bam*HI, and cloned into a preparation of the expression vector (pET3a) that had been previously digested with the same restriction enzymes. The ligated product was transformed into competent DH5α subcloning efficiency cells (Gibco BRL, Gaithersburg, MD) and the tranformants were screened for correctly inserted DNA using *Nde*I/*Bam*HI and *Bam*HI/*Sal*I double digestions. Twelve of the transformant colonies screened the correct insert. The new expression vector (p-MLA) was transformed intro competent BL21(DE3) expression cells, a lysogenic *E. coli* strain carrying the bacteriophage T7 polymerase gene, and ampicillin resistant transformants were selected. p-MLA was sequenced using an automatic DNA sequencing system to confirm that the LA coding sequence was correctly inserted and that no undesired mutations had been introduced during the PCR amplification (Pike, 1995; Wang *et al.*, 1989).

2.3.1.2. Generation and purification of mLA variants.

The single site substitutions were introduced into p-MLA (template in all amplifications) using the PCR Megaprimer method (Sarkar and Sommer, 1990). Megaprimer method involves two separate PCRs and makes use of appropriate oligonucleotide primers depending on the site of the mutation: T7 promoter or T7 terminator primer. The final PCR product (it has undergone two successive amplifications) was isolated, purified, digested and cloned into cut pET3a as in the case of mLA (section 2.3.1.1.). In particular changes at positions 117 and 118 had to be constructed using a modified mutagenesis protocol.

Replacement of aromatic clusters I and II with non-aromatic residues reduced the final yields and such an effect was attributed either to destabilisation of the native conformation or to alterations in the *in vitro* folding kinetics of the mutant (Pike, 1995). However, sufficient quantities of pure variant proteins (reported in this thesis) were produced by Dr. A.C.W. Pike for crystallographic study.

2.3.1.3. Kinetic studies on mLA variants

A plausible model has been suggested in order to explain the mechanism under which LA comes into contact with GT to form the LS complex and three schemes based on this model have been proposed. A random but highly synergistic binding of LA with glucose to GT is a common point of view for all three mechanisms, where LA associates with GT prior to the release of products and dissociates from LS complex before the completion of each catalytic cycle. The proposed schemes variations appear in defining the moment that LA gets involved

in the complex formation as well as in the level of randomness of the reactions. According to (Morrison and Ebner, 1971) GT substrates are Mn^{2+}, UDP-galactose and finally LA while (Khatra *et al.*, 1974) and (Powell and Brew, 1976) consider a partially ordered mechanism as more appropriate to interpret LA's action in which UDP-galactose binds to GT·Mn^{2+} complex prior to the random equilibrium binding of LA and glucose. Finally the third scheme supports a completely random addition of substrates (UDP-galactose, LA and acceptors) to GT·Mn^{2+} complex (Bell *et al.*, 1976).

2.3.1.4. Results from other studies on mLA variants

The LA variants have been designed with the aim to investigate regions of LA molecule to be directly involved in the formation of LS complex such as aromatic clusters I and II, as well as the flexible loop (see chapter-2 introduction for more details). Most of the substitutions involving residues from aromatic cluster II and more specifically Tyr 103 and Trp 104 affect the folding kinetics of the protein, in agreement with the NMR results that this region is of great importance for the stability of molten globule state of LA (Alexandrescu *et al.*, 1993; Pike, 1995; Smith *et al.*, 1994). However, residues like proline seem to replace tyrosine at position 103 without affecting either the yields or the folding of the protein. Mutations at both aromatic cluster II and flexible loop appear to lead in functionally deficient variants. Point mutations at positions 31 and 32 exhibit reduced affinity to GT and increased glucose binding in the LS complex. In the case of tyrosine substitution for phenylalanine at position 31 only a lower ability in glucose binding is observed and minimal effects on GT binding supporting the idea that Phe

31 is participating in the monosaccharide binding. On the contrary residues 117 and 118 affect the strength of LA binding to GT.

The above observations confirm that only a small surface area in the tail of LA molecule (residues 105-118) is implicated in the interaction with GT a result that is strengthened also by the weak affinity of LA for GT (k_d 10^{-5}M) (Pike, 1995).

2.3.2. MATERIALS AND METHODS

2.3.2.1. Crystallisation and diffraction data collection.

Crystals of LA variants had been produced using the vapour diffusion method with hanging drops. Equal volumes (2.0 μl) of protein (~20mg/ml) and reservoir solution were mixed on siliconised coverslips and left to equilibrate against the reservoir solution.

Diffraction data were collected using the Synchrotron Radiation Source (SRS) at Daresbury, UK or at EMBL Hamburg outstation either at room temperature or under cryogenic conditions. Data were recorded on a 18cm or 30cm image plate (MAR Research, Hamburg, Germany) and integrated and reduced with the programs *DENZO* and *SCALEPACK* (Otwinowski and Minor, 1997). In the case of His 32 to Tyr data integration and reduction was performed with the program *MARXDS* and *MARSCALE* (Kabsch, 1988).

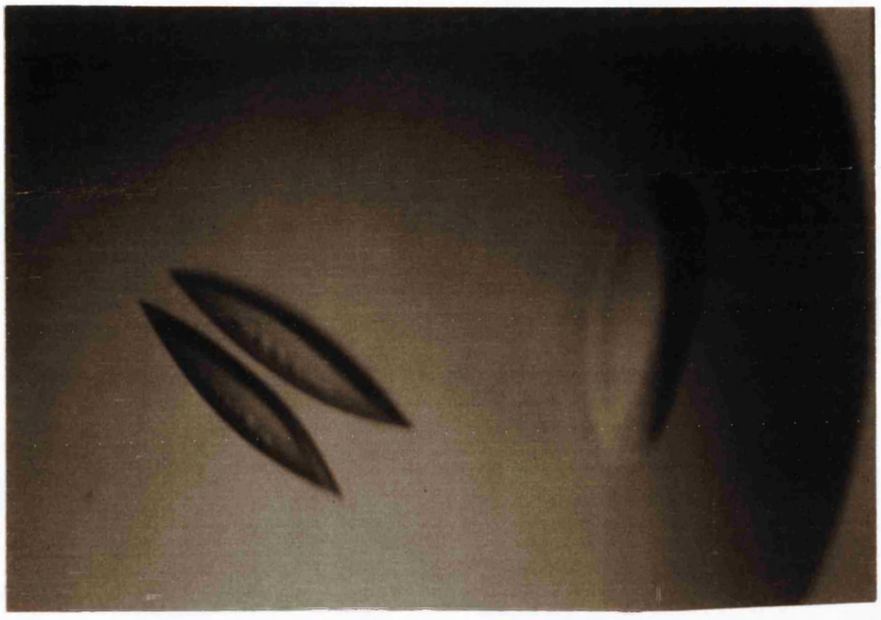
Ala109-Pro, Tyr103-Pro, Trp118-His variants: Crystals appear at 37°C in the hexagonal form (Table 2.23, *Figure* 2.23a) using ammonium sulphate as the precipitant at pH 6.5. Although no surface anomaly was detected under the microscope, all three variants exhibit '*merohedral twinning*'-an inherent problem of most LA crystals grown in the hexagonal form- which can be identified by careful examination of the diffraction pattern or by the use of algorithms that have recently been developed.

Phe31-Tyr: Crystallisation trials using conditions analogous to those optimised for the above three variants were initially performed by

increasing the concentration of ammonium sulphate concentration from 2.1 – 2.4 M. The crystals seem to belong to the hexagonal form, judging from their shape and when they were exposed to X-rays their diffraction pattern was disordered. Matrices of various concentrations of precipitating agent and pH were used and a new crystal form in tetragonal spacegroup, $P4_12_12$ (somewhat similar to apo-LA described earlier) was obtained with six molecules per asymmetric unit (Table 2.23, *Figure* 2.23b).

His32-Tyr: Crystals of His 32 to Tyr variant appeared fairly quickly as very thin, almost two-dimensional plates forming a cluster of crystals and it was rather difficult to isolate a single crystal at a size suitable for collecting data. Optimisation of the crystallisation conditions was attempted by means of detergents but did not improve the quality of the crystals. An incomplete data set was collected to 2.8 Å under cryogenic conditions using glycerol as cryoprotectant and preliminary characterisation of the crystal indicated that it belonged to orthorhombic lattice (Table 2.23).

(a)



(b)

***Figure* 2.23**

Crystals of LA in different forms: (a) Hexagonal, (b) Tetragonal

Table 2.23 Crystallisation, Data collection and processing statistics of LA variants

	A109P	Y103P	W118H	F31Y
Crystallisation conditions	Ammonium Sulph. 1.8M $CaCl_2$ 10mM PIPES 0.1M pH 6.5	Ammonium Sulph. 1.8M $CaCl_2$ 10mM PIPES 0.1M pH 6.5	Ammonium Sulph. 1.8M $CaCl_2$ 10mM PIPES 0.1M pH 6.5	Ammonium Sulph. 1.6M $CaCl_2$ 10mM Tris 0.1M pH 6.0
Temperature	37°C	37°C	37°C	16°C
Cell dimensions (a, b, c) (Å)	a=b=92.94, c=66.87	a=b=93.36, c=67.33	a=b=93.30, c=67.17	a=b=119.96 c=152.56
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 90
Space group	*P*3	*P*3	*P*3	$P4_12_12$
No of mol. per asymmetric unit	2	2	2	6
Station (Synchrotron)	Daresbury PX96 (room temp. data)	Daresbury PX95 (room temp. data)	Daresbury PX95 (room temp. data)	Daresbury PX72 (room temp. data)
Image plate	Mar18	Mar30	Mar30	Mar30
Wavelength (Å)	0.87	1.0	1.0	1.488
Distance (mm)	195	270	270	241
Oscillation range (degrees)	1.0	1.0	1.0	1.0
No of images (degrees)	104 (104)	58 (58)	62 (62)	42 (42)
No of observations	257443	164024	159565	104602
No of unique reflections	36113	41053	42824	18019
Max. resolution (Å)	2.03	2.0	2.0	3.2
*Completeness (outermost shell) (%)	87.4 (68.3)	90.6 (89.7)	95.1 (95.2)	94.5 (96.4)
*R_{sym} (I) (outermost shell) (%)	10.4 (48.8)	4.2 (21.5)	4.2 (29.7)	16.6 (48.1)
< I / σ (I) > (outermost shell)	8.0 (2.6)	14.2 (3.7)	8.9 (3.9)	4.9 (2.9)
Outermost shell (Å)	2.10 - 2.03	2.07 - 2.00	2.07 - 1.99	3.3 – 3.2

Table 2.23 Crystallisation, Data collection and processing statistics of LA variants (continued...)

	H32Y	K114N
Crystallisation conditions	PEG 4K 25%, $CaCl_2$ 10mM Tris 0.1M pH 6.0 0.05% β-o-glucoside	Ammonium sulph. 2.0M Tris 0.1M pH 8.5
Temperature	16°C	16°C
Cell dimensions (a, b, c) (Å)	a=b=101.34, c=71.98	a=b=98.7, c=274.2
α, β, γ (°)	90, 90, 90	90, 90, 120
Space group	Primitive orthorhombic	Hexagonal
No of mol. per asymmetric unit	1	8
Station (Synchrotron)	EMBL-Hamburg X11 (cryo data)	In-house (room temp. data)
Image plate	Mar18	Mar18
Wavelength (Å)	0.9057	1.542
Distance (mm)	215	160
Oscillation range (degrees)	1.0	1.0
No of images (degrees)	31 (31)	38 (38)
No of observations	2278	10891
No of unique reflections	980	
Max. resolution (Å)	2.8	3.0
*Completeness (outermost shell) (%)	32.1 (32.1)	
$^{\diamond}R_{sym}$ (I) (outermost shell) (%)	8.7 (16.3)	MARXDS statistics
< I / σ (I) > (outermost shell)	17.7 (7.7)	
Outermost shell (Å)	3.0 - 2.8	

*Completeness in the range ∞-resmax, where resmax is the maximum resolution to which data were collected. $^{\diamond}R_{sym}(I)=\Sigma_{hkl}\Sigma_i| I_{hkl,i}-I_{average,hkl}| / |\Sigma_{hkl}\Sigma_i| I_{hkl,i}|$ x100.

Lys114-Asn: Single crystals were obtained from crystallisation conditions suggested by sparse matrix crystallisation screening kit, Crystal Screen I from Hampton Research (Jancarik and Kim, 1991) that were diffracting to 3.0 Å resolution. Preliminary characterisation indicated that Lys114-Asn variant belongs to the hexagonal spacegroup with 8 molecules per asymmetric unit (Table 2.23, Mathews coefficient, V_m= 3.9 $Å^3$/D). Attempts to vary the crystallisation conditions by changing the precipitant from ammonium sulphate to lithium sulphate were made with an aim to obtain another form with reduced number of molecules but they were not successful.

His32-Ala: Preliminary indications showed that this LA variant might crystallise in the presence of PEG 4K 20%, $CaCl_2$ 10mM in Tris/HCl buffer pH 7.5. In addition a single crystal was grown using condition 5 of Hampton Research Crystal screen I (the ingredients of which were 30% MPD, 0.1M Sodium Hepes pH 7.5 and 0.2 M sodium citrate) at 16°C. This crystal was exposed to X-rays at SRS but further optimisation of the condition seem to be necessary.

Crystallisation trials have also been performed for His107-Ala as well as Phe31-Ala, Leu110-Arg, Cys6-Ser and I66-Val variants, starting with the Crystal Screen kit or with matrices of various concentrations of PEG and ammonium sulphate against different pH ranges. Some indications of crystals have been observed which can be used to screen further conditions.

Due to lack of time structural determination of these variants could not be performed.

2.4. INVESTIGATION OF BINDING OF Mn^{2+} to human α-LACTALBUMIN

In the context of the previous studies on metal ion binding in LA, before the structure determination of apo-LA, crystals of commercially available human LA (Sigma, chemical company) co-crystallised with 5mM $MnCl_2$ in the orthorhombic form and a high resolution data set was collected (Tables 2.24, 2.25), yet difference Fourier analysis did not show direct binding of Mn^{2+} to human LA protein.

Table 2.24. Crystallisation conditions

- Native Human α-Lactalbumin + Mn^{2+} (cocrystallisation)

Reservoir solution	:	Ammonium Sulphate	2.0 M	Temp.: 37°C
		$CaCl_2$	20 mM	
		PIPES	0.2 M	pH 7.5
		$MnCl_2$	10 mM	
		Glycerol	1%	
Drop	:	HLA	30 mg/ml	
		Ammonium Sulphate	1.6 M	
		$CaCl_2$	5 mM	
		PIPES	0.1 M	pH 7.0
		$MnCl_2$	5 mM	

Table 2.25 Data Processing Statistics for HLA + Mn^{2+} complex

Cell dimensions (a, b, c) (Å)	a=33.64 b=69.8 c=47.12
α, β, γ (°)	90 90 90
Space group	$P2_12_12$
No of mol. per asymmetric unit	1.0
Station	In-house data
Image plate	Mar 30
Wavelength (Å)	1.542
Distance (mm)	109.0
Oscillation range (degrees)	1.0°
No of images (degrees)	62 (62°)
No of observations	65167
No of unique reflections	11063
Max. resolution (Å)	1.7
¤Completeness (outermost shell) (%)	86.7 (85.9)
✧R_{sym} (I) (outermost shell) (%)	7.9 (60.2)
$< I / \sigma (I) >$ (outermost shell)	5.37 (1.81)
Outermost shell (Å)	1.76 - 1.70

-3-

HUMAN SERUM RETINOL BINDING PROTEIN

INTRODUCTION

CRYSTAL STRUTURES OF NATIVE RECOMBINANT RBP (rRBP) & A DOUBLE VARIANT OF rRBP

3. HUMAN SERUM RETINOL BINDING PROTEIN

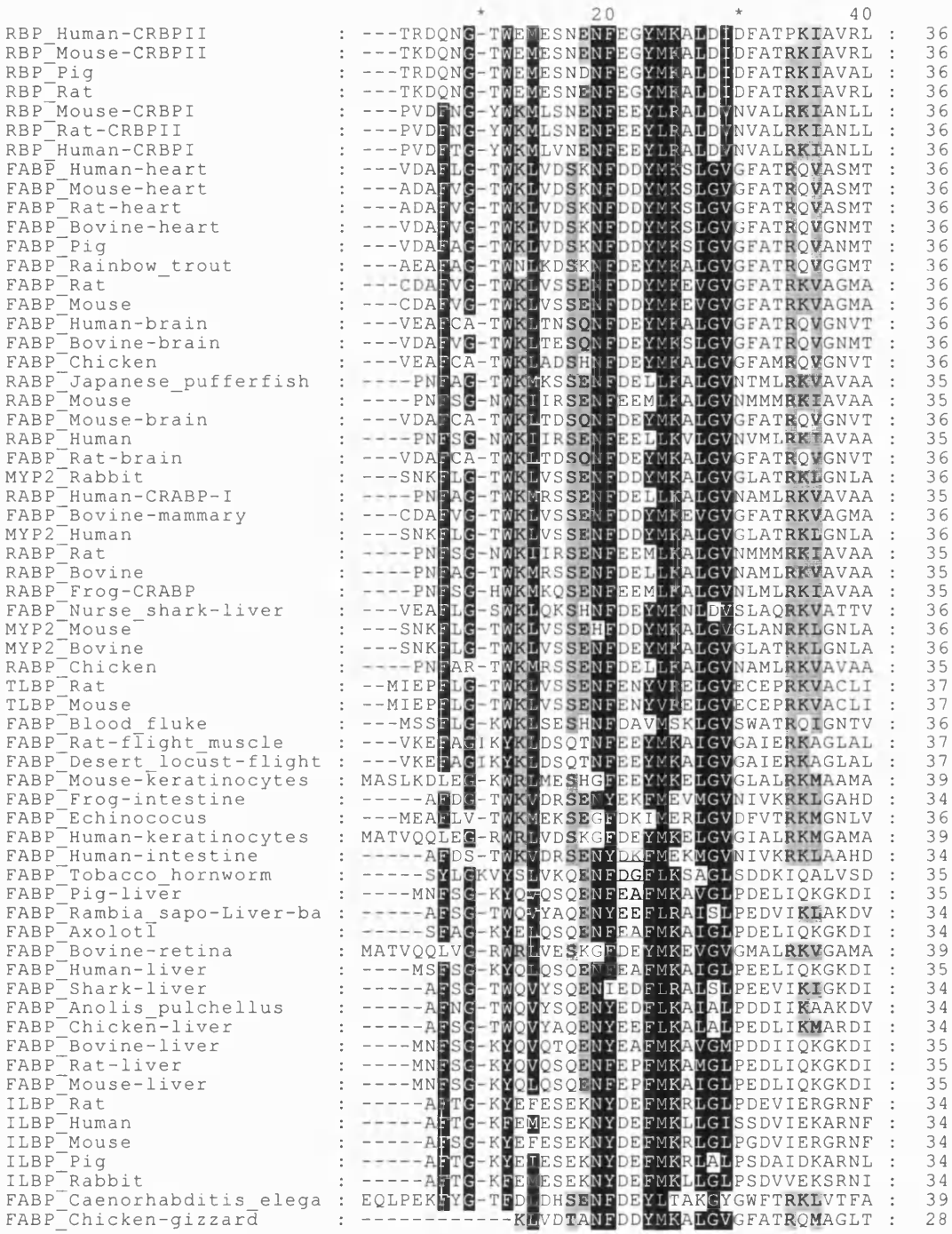
3.1. INTRODUCTION

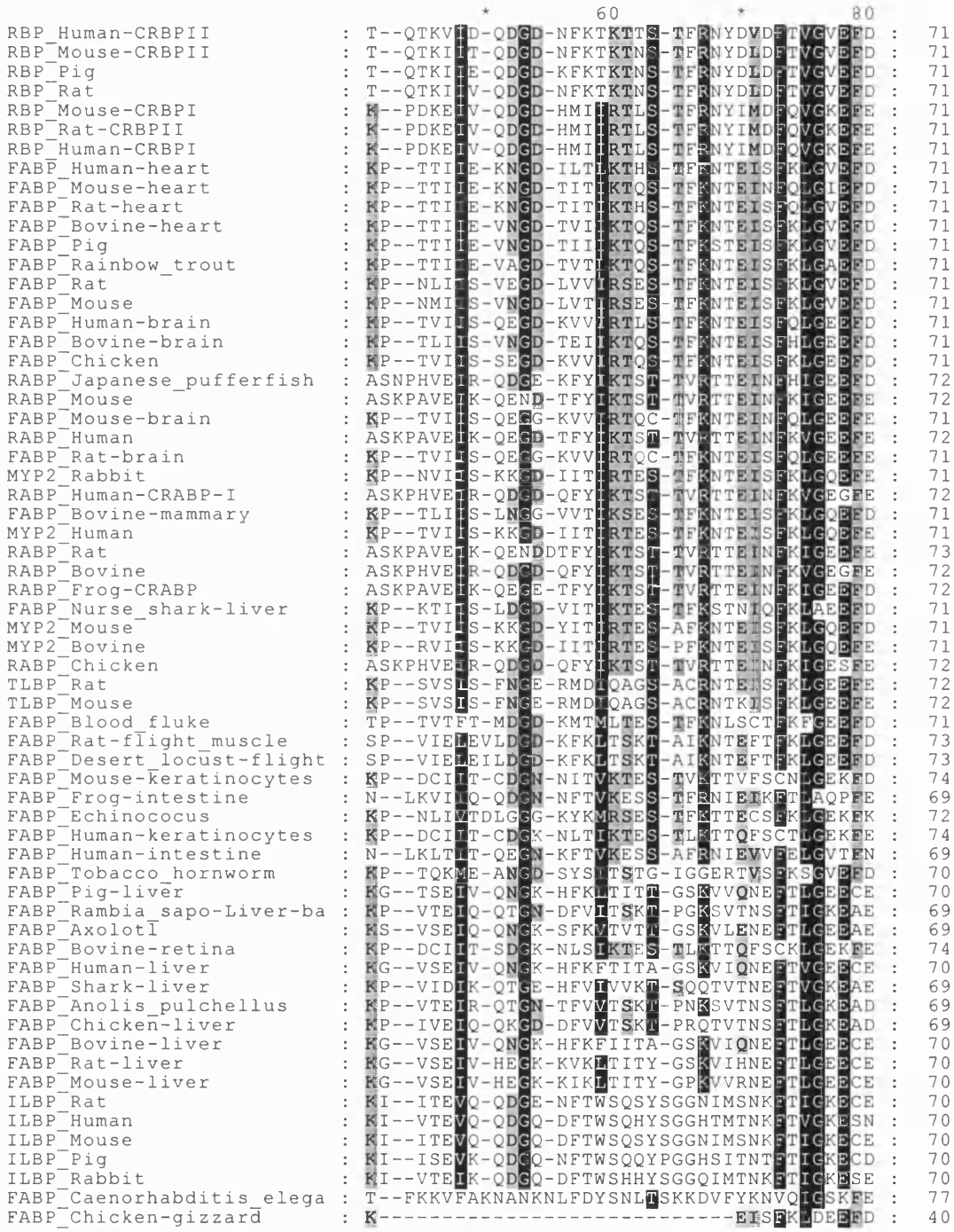
3.1.1. Human Serum Retinol Binding Protein as a member of the family of lipocalins.

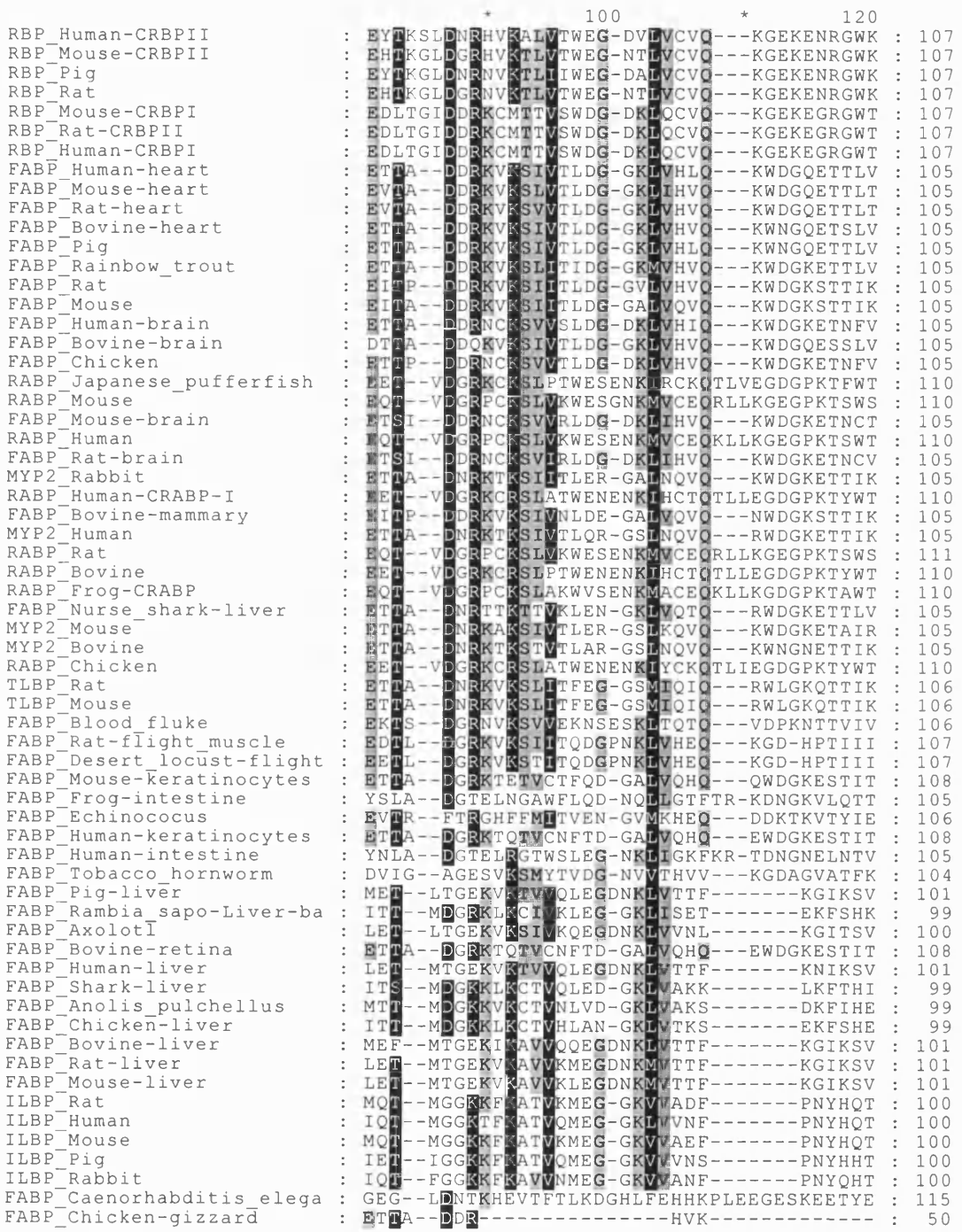
The lipocalins are a structurally and functionally diverse family of proteins and they are only found in eukaryotic organisms. Members of the lipocalin protein family are typically low molecular weight secreted proteins, characterised by a range of different molecular-recognition properties such as ability to bind small, principally hydrophobic molecules, binding to specific cell-surface receptors and formation of macromolecular complexes. Sequence identity within the lipocalin family is very often below 25%. However, there are conserved sequence motifs characteristic of the family, and the family members also share a basic feature present in all lipocalins: the structural framework of an 8 stranded 'up and down' β-barrel (*Figure* 3.1) (Huber *et al.*, 1987) (Flower, 1996) (Brownlow *et al.*, 1997).

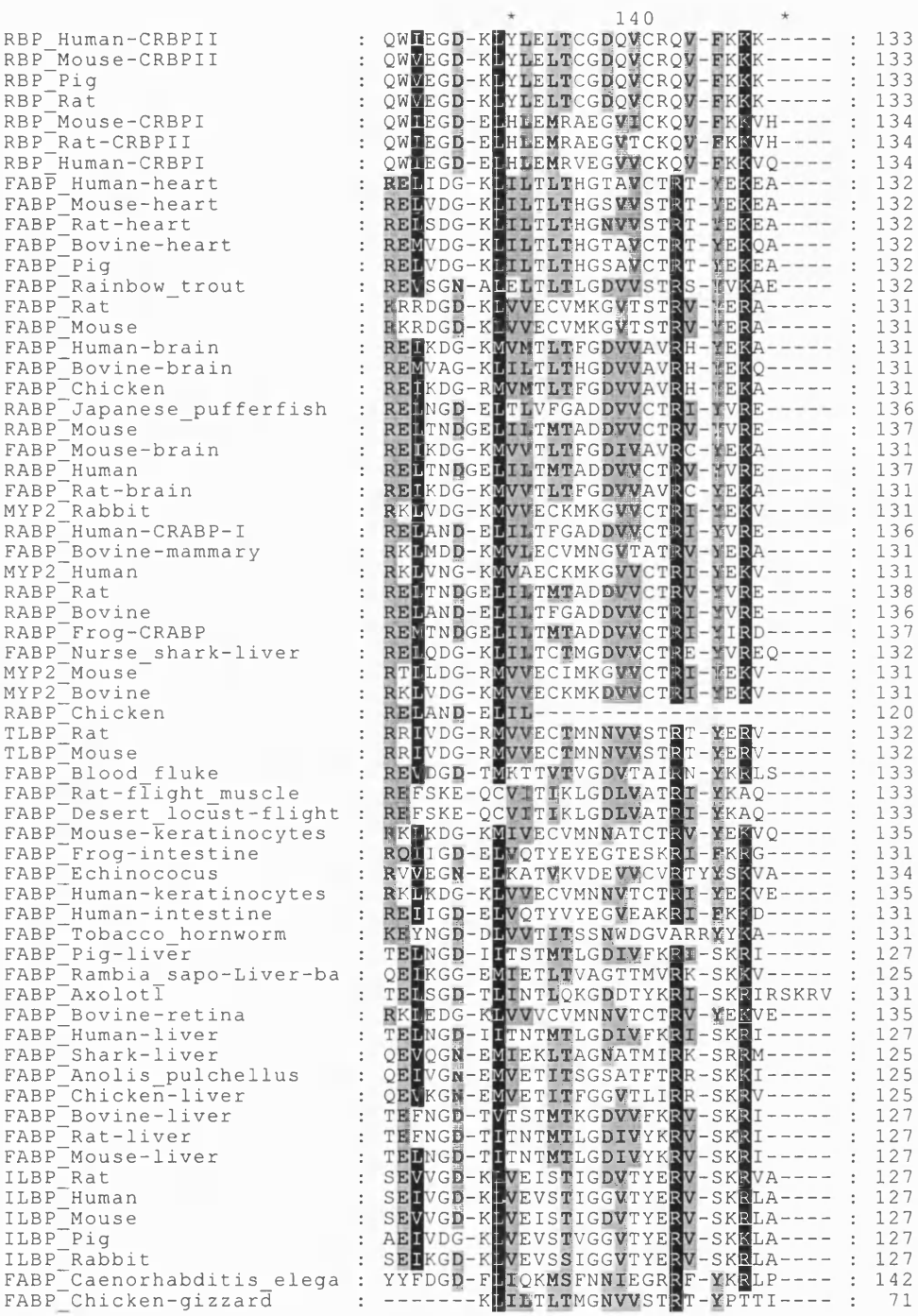
***Figure* 3.1** (overleaf)

Sequence comparison for RBP. The alignment shows a selection of Retinol binding proteins (RBP), fatty acid binding proteins (FABP) and retinoic acid RBP (RABP). The shaded areas correspond to conservative residues. The sequence retrieval was performed using the program WU-Blust (version 2.0) and the alignment using the program Clustalx.


* 20 * 40
RBP_Human-CRBPII : ---TRDQNG-TWEMESNENFEGYMKALDIDFATPKIAVRL : 36
RBP_Mouse-CRBPII : ---TKDQNG-TWEMESNENFEGYMKALDIDFATRKIAVRL : 36
RBP_Pig : ---TRDQNG-TWEMESNDNFEGYMKALDIDFATRKIAVAL : 36
RBP_Rat : ---TKDQNG-TWEMESNENFEGYMKALDIDFATRKIAVRL : 36
RBP_Mouse-CRBPI : ---PVDFNG-YWKMLSNENFEEYLRALDVNVALRKIANLL : 36
RBP_Rat-CRBPII : ---PVDFNG-YWKMLSNENFEEYLRALDVNVALRKIANLL : 36
RBP_Human-CRBPI : ---PVDFTG-YWKMLVNENFEEYLRALDVNVALRKIANLL : 36
FABP_Human-heart : ---VDAFLG-TWKLVDSKNFDDYMKSLGVGFATRQVASMT : 36
FABP_Mouse-heart : ---ADAFVG-TWKLVDSKNFDDYMKSLGVGFATRQVASMT : 36
FABP_Rat-heart : ---ADAFVG-TWKLVDSKNFDDYMKSLGVGFATRQVASMT : 36
FABP_Bovine-heart : ---VDAFVG-TWKLVDSKNFDDYMKSLGVGFATRQVGNMT : 36
FABP_Pig : ---VDAFAG-TWKLVDSKNFDDYMKSIGVGFATRQVANMT : 36
FABP_Rainbow_trout : ---AEAFAG-TWNLKDSKNFDEYMKALGVGFATRQVGGMT : 36
FABP_Rat : ---CDAFVG-TWKLVSSENFDDYMKEVGVGFATRKVAGMA : 36
FABP_Mouse : ---CDAFVG-TWKLVSSENFDDYMKEVGVGFATRKVAGMA : 36
FABP_Human-brain : ---VEAFCA-TWKLTNSQNFDEYMKALGVGFATRQVGNVT : 36
FABP_Bovine-brain : ---VDAFVG-TWKLTESQNFDEYMKSLGVGFATRQVGNMT : 36
FABP_Chicken : ---VEAFCA-TWKLADSHNFDEYMKALGVGFAMRQVGNVT : 36
RABP_Japanese_pufferfish : ----PNFAG-TWKMKSSENFDELLKALGVNTMLRKVAVAA : 35
RABP_Mouse : ----PNFSG-NWKIIRSENFEEMLKALGVNMMMRKIAVAA : 35
FABP_Mouse-brain : ---VDAFCA-TWKLTDSQNFDEYMKALGVGFATRQVGNVT : 36
RABP_Human : ----PNFSG-NWKIIRSENFEELLKVLGVNVMLRKIAVAA : 35
FABP_Rat-brain : ---VDAFCA-TWKLTDSQNFDEYMKALGVGFATRQVGNVT : 36
MYP2_Rabbit : ---SNKFLG-TWKLVSSENFDDYMKALGVGLATRKLGNLA : 36
RABP_Human-CRABP-I : ----PNFAG-TWKMRSSENFDELLKALGVNAMLRKVAVAA : 35
FABP_Bovine-mammary : ---CDAFVG-TWKLVSSENFDDYMKEVGVGFATRKVAGMA : 36
MYP2_Human : ---SNKFLG-TWKLVSSENFDDYMKALGVGLATRKLGNLA : 36
RABP_Rat : ----PNFSG-NWKIIRSENFEEMLKALGVNMMMRKIAVAA : 35
RABP_Bovine : ----PNFAG-TWKMRSSENFDELLKALGVNAMLRKVAVAA : 35
RABP_Frog-CRABP : ----PNFSG-HWKMKQSENFEEMLKALGVNLMLRKIAVAA : 35
FABP_Nurse_shark-liver : ---VEAFLG-SWKLQKSHNFDEYMKNLDVSLAQRKVATTV : 36
MYP2_Mouse : ---SNKFLG-TWKLVSSEHFDDYMKALGVGLANRKLGNLA : 36
MYP2_Bovine : ---SNKFLG-TWKLVSSENFDEYMKALGVGLATRKLGNLA : 36
RABP_Chicken : ----PNFAR-TWKMRSSENFDELLKALGVNAMLRKVAVAA : 35
TLBP_Rat : --MIEPFLG-TWKLVSSENFENYVRELGVECEPRKVACLI : 37
TLBP_Mouse : --MIEPFLG-TWKLVSSENFENYVRELGVECEPRKVACLI : 37
FABP_Blood_fluke : ---MSSFLG-KWKLSESHNFDAVMSKLGVSWATRQIGNTV : 36
FABP_Rat-flight_muscle : ---VKEFAGIKYKLDSQTNFEEYMKAIGVGAIERKAGLAL : 37
FABP_Desert_locust-flight : ---VKEFAGIKYKLDSQTNFEEYMKAIGVGAIERKAGLAL : 37
FABP_Mouse-keratinocytes : MASLKDLEG-KWRLMESHGFEEYMKELGVGLALRKMAAMA : 39
FABP_Frog-intestine : -----AFDG-TWKVDRSENYEKFMEVMGVNIVKRKLGAHD : 34
FABP_Echinococus : ---MEAFLV-TWKMEKSEGFDKIMERLGVDFVTRKMGNLV : 36
FABP_Human-keratinocytes : MATVQQLEG-RWRLVDSKGFDEYMKELGVGIALRKMGAMA : 39
FABP_Human-intestine : -----AFDS-TWKVDRSENYDKFMEKMGVNIVKRKLAAHD : 34
FABP_Tobacco_hornworm : -----SYLGKVYSLVKQENFDGFLKSAGLSDDKIQALVSD : 35
FABP_Pig-liver : ----MNFSG-KYQLQSQENFEAFMKAVGLPDELIQKGKDI : 35
FABP_Rambia_sapo-Liver-ba : -----AFSG-TWQMYAQENYEEFLRAISLPEDVIKLAKDV : 34
FABP_Axolotl : -----SFAG-KYELQSQENFEAFMKAIGLPDELIQKGKDI : 34
FABP_Bovine-retina : MATVQQLVG-RWRLVESKGFDEYMKEVGVGMALRKVGAMA : 39
FABP_Human-liver : ----MSFSG-KYQLQSQENFEAFMKAIGLPEELIQKGKDI : 35
FABP_Shark-liver : -----AFSG-TWQVYSQENIEDFLRALSLPEEVIKIGKDI : 34
FABP_Anolis_pulchellus : -----AFNG-TWQVYSQENYEDFLKAIALPDDIIKAAKDV : 34
FABP_Chicken-liver : -----AFSG-TWQVYAQENYEEFLKALALPEDLIKMARDI : 34
FABP_Bovine-liver : ----MNFSG-KYQVQTQENYEAFMKAVGMPDDIIQKGKDI : 35
FABP_Rat-liver : ----MNFSG-KYQVQSQENFEPFMKAMGLPEDLIQKGKDI : 35
FABP_Mouse-liver : ----MNFSG-KYQLQSQENFEPFMKAIGLPEDLIQKGKDI : 35
ILBP_Rat : -----AFTG-KYEFESEKNYDEFMKRLGLPDEVIERGRNF : 34
ILBP_Human : -----AFTG-KFEMESEKNYDEFMKLLGISSDVIEKARNF : 34
ILBP_Mouse : -----AFSG-KYEFESEKNYDEFMKRLGLPGDVIERGRNF : 34
ILBP_Pig : -----AFTG-KYELESEKNYDEFMKRLALPSDAIDKARNL : 34
ILBP_Rabbit : -----AFTG-KFEMESEKNYDEFMKLLGLPSDVVEKSRNI : 34
FABP_Caenorhabditis_elega : EQLPEKFYG-TFDLDHSENFDEYLTAKGYGWFTRKLVTFA : 39
FABP_Chicken-gizzard : ----------KLVDTANFDDYMKALGVGFATRQMAGLT : 28


* 60 * 80
RBP_Human-CRBPII : T--QTKVID-QDGD-NFKTKTTS-TFRNYDVDFTVGVEFD : 71
RBP_Mouse-CRBPII : T--QTKIIT-QDGD-NFKTKTNS-TFRNYDLDFTVGVEFD : 71
RBP_Pig : T--QTKIIE-QDGD-KFKTKTNS-TFRNYDLDFTVGVEFD : 71
RBP_Rat : T--QTKIIV-QDGD-NFKTKTNS-TFRNYDLDFTVGVEFD : 71
RBP_Mouse-CRBPI : K--PDKEIV-QDGD-HMIIRTLS-TFRNYIMDFQVGKEFE : 71
RBP_Rat-CRBPII : K--PDKEIV-QDGD-HMIIRTLS-TFRNYIMDFQVGKEFE : 71
RBP_Human-CRBPI : K--PDKEIV-QDGD-HMIIRTLS-TFRNYIMDFQVGKEFE : 71
FABP_Human-heart : KP--TTIIE-KNGD-ILTLKTHS-TFKNTEISFKLGVEFD : 71
FABP_Mouse-heart : KP--TTIIE-KNGD-TITIKTQS-TFKNTEINFQLGIEFD : 71
FABP_Rat-heart : KP--TTIIE-KNGD-TITIKTHS-TFKNTEISFQLGVEFD : 71
FABP_Bovine-heart : KP--TTIIE-VNGD-TVIIKTQS-TFKNTEISFKLGVEFD : 71
FABP_Pig : KP--TTIIE-VNGD-TIIIKTQS-TFKSTEISFKLGVEFD : 71
FABP_Rainbow_trout : KP--TTIIE-VAGD-TVTIKTQS-TFKNTEISFKLGAEFD : 71
FABP_Rat : KP--NLITS-VEGD-LVVIRSES-TFKNTEISFKLGVEFD : 71
FABP_Mouse : KP--NMITS-VNGD-LVTIRSES-TFKNTEISFKLGVEFD : 71
FABP_Human-brain : KP--TVIIS-QEGD-KVVIRTLS-TFKNTEISFQLGEEFD : 71
FABP_Bovine-brain : KP--TLIIS-VNGD-TEIIKTQS-TFKNTEISFHLGEEFD : 71
FABP_Chicken : KP--TVIIS-SEGD-KVVIRTQS-TFKNTEISFKLGEEFD : 71
RABP_Japanese_pufferfish : ASNPHVEIR-QDGE-KFYIKTST-TVRTTEINFHIGEEFD : 72
RABP_Mouse : ASKPAVEIK-QENDTFYIKTST-TVRTTEINFKIGEEFE : 72
FABP_Mouse-brain : KP--TVIIS-QEGG-KVVIRTQC-TFKNTEINFQLGEEFE : 71
RABP_Human : ASKPAVEIK-QEGDTFYIKTST-TVRTTEINFKVGEEFE : 72
FABP_Rat-brain : KP--TVIIS-QEGG-KVVIRTQC-TFKNTEISFQLGEEFE : 71
MYP2_Rabbit : KP--NVITS-KKGD-IITIRTES-TFKNTEISFKLGQEFE : 71
RABP_Human-CRABP-I : ASKPHVEIR-QDGD-QFYIKTST-TVRTTEINFKVGEGFE : 72
FABP_Bovine-mammary : KP--TLIIS-LNGG-VVTIKSES-TFKNTEISFKLGQEFD : 71
MYP2_Human : KP--TVIIS-KKGD-IITIRTES-TFKNTEISFKLGQEFE : 71
RABP_Rat : ASKPAVEIK-QENDDTFYIKTST-TVRTTEINFKIGEEFE : 73
RABP_Bovine : ASKPHVEIR-QDGD-QFYIKTST-TVRTTEINFKVGEGFE : 72
RABP_Frog-CRABP : ASKPAVEIK-QEGE-TFYIKTST-TVRTTEINFKIGEEFD : 72
FABP_Nurse_shark-liver : KP--KTIIS-LDGD-VITIKTES-TFKSTNIQFKLAEEFD : 71
MYP2_Mouse : KP--TVIIS-KKGD-YITIRTES-AFKNTEISFKLGQEFD : 71
MYP2_Bovine : KP--RVITS-KKGD-IITIRTES-PFKNTEISFKLGQEFE : 71
RABP_Chicken : ASKPHVEIR-QDGD-QFYIKTST-TVRTTEINFKIGESFE : 72
TLBP_Rat : KP--SVSIS-FNGE-RMDIQAGS-ACRNTEISFKLGEEFE : 72
TLBP_Mouse : KP--SVSIS-FNGE-RMDIQAGS-ACRNTKISFKLGEEFE : 72
FABP_Blood_fluke : TP--TVTFT-MDGD-KMTMLTES-TFKNLSCTFKFGEEFD : 71
FABP_Rat-flight_muscle : SP--VIELEVLDGD-KFKLTSKT-AIKNTEFTFKLGEEFD : 73
FABP_Desert_locust-flight : SP--VIELEILDGD-KFKLTSKT-AIKNTEFTFKLGEEFD : 73
FABP_Mouse-keratinocytes : KP--DCIIT-CDGN-NITVKTES-TVKTTVFSCNLGEKFD : 74
FABP_Frog-intestine : N--LKVIIQ-QDGN-NFTVKESS-TFRNIEIKFTLAQEFE : 69
FABP_Echinococcus : KP--NLIVTDLGGG-KYKMRSES-TFKTTECSFKLGEKFK : 72
FABP_Human-keratinocytes : KP--DCIIT-CDGK-NLTIKTES-TLKTTQFSCTLGEKFE : 74
FABP_Human-intestine : N--LKLTIT-QEGN-KFTVKESS-AFRNIEVVFELGVTFN : 69
FABP_Tobacco_hornworm : KP--TQKME-ANGD-SYSITSTG-IGGERTVSFKSGVEFD : 70
FABP_Pig-liver : KG--TSEIV-QNGK-HFKLTITT-GSKVVQNEFTLGEECE : 70
FABP_Rambia_sapo-Liver-ba : KP--VTEIQ-QTGN-DFVITSKT-PGKSVTNSFTIGKEAE : 69
FABP_Axolotl : KS--VSEIQ-QNGK-SFKVTVTT-GSKVLENEFTLGEEAE : 69
FABP_Bovine-retina : KP--DCIIT-SDGK-NLSIKTES-TLKTTQFSCKLGEKFE : 74
FABP_Human-liver : KG--VSEIV-QNGK-HFKFTITA-GSKVIQNEFTVGEECE : 70
FABP_Shark-liver : KP--VIDIK-QTGE-HFVIVVKT-SQQTVTNEFTVGKEAE : 69
FABP_Anolis_pulchellus : KP--VTEIR-QTGN-TFVVTSKT-PNKSVTNSFTLGKEAD : 69
FABP_Chicken-liver : KP--IVEIQ-QKGD-DFVVTSKT-PRQTVTNSFTLGKEAD : 69
FABP_Bovine-liver : KG--VSEIV-QNGK-HFKFIITA-GSKVIQNEFTLGEECE : 70
FABP_Rat-liver : KG--VSEIV-HEGK-KVKLTITY-GSKVIHNEFTLGEECE : 70
FABP_Mouse-liver : KG--VSEIV-HEGK-KIKLTITY-GPKVVRNEFTLGEECE : 70
ILBP_Rat : KI--ITEVQ-QDGE-NFTWSQSYSGGNIMSNKFTIGKECE : 70
ILBP_Human : KI--VTEVQ-QDGQ-DFTWSQHYSGGHTMTNKFTVGKESN : 70
ILBP_Mouse : KI--ITEVQ-QDGQ-DFTWSQSYSGGNIMSNKFTIGKECE : 70
ILBP_Pig : KI--ISEVK-QDGQ-NFTWSQQYPGGHSITNTFTIGKECD : 70
ILBP_Rabbit : KI--VTEIK-QDGQ-DFTWSHHYSGGQIMTNKFTIGKESE : 70
FABP_Caenorhabditis_elega : T--FKKVFAKNANKNLFDYSNLTSKKDVFYKNVQIGSKFE : 77
FABP_Chicken-gizzard : K----------EISFKLDEEFD : 40


* 100 * 120
RBP_Human-CRBPII : EYTKSLDNRHVKALVTWEG-DVLCVQ---KGEKENRGWK : 107
RBP_Mouse-CRBPII : EHTKGLDGRHVKTLVTWEG-NTLVCVQ---KGEKENRGWK : 107
RBP_Pig : EYTKGLDNRNVKTLIIWEG-DALVCVQ---KGEKENRGWK : 107
RBP_Rat : EHTKGLDGRNVKTLVTWEG-NTLVCVQ---KGEKENRGWK : 107
RBP_Mouse-CRBPI : EDLTGIDDRKCMTTVSWDG-DKLQCVQ---KGEKEGRGWT : 107
RBP_Rat-CRBPII : EDLTGIDDRKCMTTVSWDG-DKLQCVQ---KGEKEGRGWT : 107
RBP_Human-CRBPI : EDLTGIDDRKCMTTVSWDG-DKLQCVQ---KGEKEGRGWT : 107
FABP_Human-heart : ETTA--DDRKVKSIVTLDG-GKLVHLQ---KWDGQETTLV : 105
FABP_Mouse-heart : EVTA--DDRKVKSLVTLDG-GKLIHVQ---KWDGQETTLT : 105
FABP_Rat-heart : EVTA--DDRKVKSVVTLDG-GKLVHVQ---KWDGQETTLT : 105
FABP_Bovine-heart : ETTA--DDRKVKSIVTLDG-GKLVHVQ---KWNGQETSLV : 105
FABP_Pig : ETTA--DDRKVKSIVTLDG-GKLVHLQ---KWNGQETTLV : 105
FABP_Rainbow_trout : ETTA--DDRKVKSLITIDG-GKMVHVQ---KWDGKETTLV : 105
FABP_Rat : EITP--DDRKVKSIITLDG-GVLVHVQ---KWDGKSTTIK : 105
FABP_Mouse : EITA--DDRKVKSIITLDG-GALVQVQ---KWDGKSTTIK : 105
FABP_Human-brain : ETTA--DDRNCKSVVSLDG-DKLVHIQ---KWDGKETNFV : 105
FABP_Bovine-brain : DTTA--DDQKVKSIVTLDG-GKLVHVQ---KWDGQESSLV : 105
FABP_Chicken : ETTP--DDRNCKSVVTLDG-DKLVHVQ---KWDGKETNFV : 105
RABP_Japanese_pufferfish : EET--VDGRKCKSLPTWESENKLRCKQTLVEGDGPKTFWT : 110
RABP_Mouse : EQT--VDGRPCTSLVKWESGNKMVCEQRLLKGEGPKTSWS : 110
FABP_Mouse-brain : ETSI--DDRNCKSVVRLDG-DKLIHVQ---KWDGKETNCT : 105
RABP_Human : EQT--VDGRPCKSLVKWESENKMVCEQKLLKGEGPKTSWT : 110
FABP_Rat-brain : ETSI--DDRNCKSVIRLDG-DKLIHVQ---KWDGKETNCV : 105
MYP2_Rabbit : ETTA--DNRKTKSIITLER-GALNQVQ---KWDGKETTIK : 105
RABP_Human-CRABP-I : EET--VDGRKCRSLATWENENKIHCTQTLLEGDGPKTYWT : 110
FABP_Bovine-mammary : EITP--DDRKVKSIVNLDE-GALVQVQ---NWDGKSTTIK : 105
MYP2_Human : ETTA--DNRKTKSIVTLQR-GSLNQVQ---RWDGKETTIK : 105
RABP_Rat : EQT--VDGRPCKSLVKWESENKMVCEQRLLKGEGPKTSWS : 111
RABP_Bovine : EET--VDGRKCRSLPTWENENKIHCTQTLLEGDGPKTYWT : 110
RABP_Frog-CRABP : EQT--VDGRPCKSLAKWVSENKMACEQKLLKGDGPKTAWT : 110
FABP_Nurse_shark-liver : ETTA--DNRTTKTTVKLEN-GKLVQTQ---RWDGKETTLV : 105
MYP2_Mouse : ETTA--DNRKAKSIVTLER-GSLKQVQ---KWDGKETAIR : 105
MYP2_Bovine : ETTA--DNRKTKSTVTLAR-GSLNQVQ---KWNGNETTIK : 105
RABP_Chicken : EET--VDGRKCRSLATWENENKIYCKQTLIEGDGPKTYWT : 110
TLBP_Rat : ETTA--DNRKVKSLITFEG-GSMTQIQ---RWLGKQTTIK : 106
TLBP_Mouse : ETTA--DNRKVKSLITFEG-GSMIQIQ---RWLGKQTTIK : 106
FABP_Blood_fluke : EKTS--DGRNVKSVVEKNSESKLTQTQ---VDPKNTTVIV : 106
FABP_Rat-flight_muscle : EDTL--DGRKVKSIITQDGPNKLVHEQ---KGD-HPTIII : 107
FABP_Desert_locust-flight : EETL--DGRKVKSTITQDGPNKLVHEQ---KGD-HPTIII : 107
FABP_Mouse-keratinocytes : ETTA--DGRKTETVCTFQD-GALVQHQ---QWDGKESTIT : 108
FABP_Frog-intestine : YSLA--DGTELNGAWFLQD-NQLLGTFTR-KDNGKVLQTT : 105
FABP_Echinococus : EVTR--FTRGHFFMITVEN-GVMKHEQ---DDKTKVTYIE : 106
FABP_Human-keratinocytes : ETTA--DGRKTQTVCNFTD-GALVQHQ---EWDGKESTIT : 108
FABP_Human-intestine : YNLA--DGTELRGTWSLEG-NKLIGKFKR-TDNGNELNTV : 105
FABP_Tobacco_hornworm : DVIG--AGESVKSMYTVDG-NVVTHVV---KGDAGVATFK : 104
FABP_Pig-liver : MET--LTGEKVKTVVQLEGDNKLVTTF-------KGIKSV : 101
FABP_Rambia_sapo-Liver-ba : ITT--MDGRKLKCIVKLEG-GKLISET-------EKFSHK : 99
FABP_Axolotl : LET--LTGEKVKSIVKQEGDNKLVVNL-------KGITSV : 100
FABP_Bovine-retina : ETTA--DGRKTQTVCNFTD-GALVQHQ---EWDGKESTIT : 108
FABP_Human-liver : LET--MTGEKVKTVVQLEGDNKLVTTF-------KNIKSV : 101
FABP_Shark-liver : ITS--MDGKKLKCTVQLED-GKLVAKK-------LKFTHI : 99
FABP_Anolis_pulchellus : MTT--MDGKKVKCTVNLVD-GKLVAKS-------DKFIHE : 99
FABP_Chicken-liver : ITT--MDGKKLKCTVHLAN-GKLVTKS-------EKFSHE : 99
FABP_Bovine-liver : MEF--MTGEKIKAVVQQEGDNKLVTTF-------KGIKSV : 101
FABP_Rat-liver : LET--MTGEKVKAVVKMEGDNKMVTTF-------KGIKSV : 101
FABP_Mouse-liver : LET--MTGEKVKAVVKLEGDNKMVTTF-------KGIKSV : 101
ILBP_Rat : MQT--MGGKKFKATVKMEG-GKVVADF-------PNYHQT : 100
ILBP_Human : IQT--MGGKTFKATVQMEG-GKLVVNF-------PNYHQT : 100
ILBP_Mouse : MQT--MGGKKFKATVKMEG-GKVVAEF-------PNYHQT : 100
ILBP_Pig : IET--IGGKKFKATVQMEG-GKVVVNS-------PNYHHT : 100
ILBP_Rabbit : IQT--FGGKKFKAVVNMEG-GKVVANF-------PNYQHT : 100
FABP_Caenorhabditis_elega : GEG--LDNTKHEVTFTLKDGHLFEHHKPLEEGESKEETYE : 115
FABP_Chicken-gizzard : ETTA--DDR----------HVK---------- : 50


* 140 *
RBP_Human-CRBPII : QWIEGD-KLYLELTCGDQVCRQV-FKKK----- : 133
RBP_Mouse-CRBPII : QWVEGD-KLYLELTCGDQVCRQV-FKKK----- : 133
RBP_Pig : QWVEGD-KLYLELTCGDQVCRQV-FKKK----- : 133
RBP_Rat : QWVEGD-KLYLELTCGDQVCRQV-FKKK----- : 133
RBP_Mouse-CRBPI : QWIEGD-ELHLEMRAEGVICKQV-FKKVH---- : 134
RBP_Rat-CRBPII : QWIEGD-ELHLEMRAEGVTCKQV-FKKVH---- : 134
RBP_Human-CRBPI : QWIEGD-ELHLEMRVEGVVCKQV-FKKVQ---- : 134
FABP_Human-heart : RELIDG-KLILTLTHGTAVCTRT-YEKEA---- : 132
FABP_Mouse-heart : RELVDG-KLILTLTHGSVVSTRT-YEKEA---- : 132
FABP_Rat-heart : RELSDG-KLILTLTHGNVVSTRT-YEKEA---- : 132
FABP_Bovine-heart : REMVDG-KLILTLTHGTAVCTRT-YEKQA---- : 132
FABP_Pig : RELVDG-KLILTLTHGSAVCTRT-YEKEA---- : 132
FABP_Rainbow_trout : REVSGN-ALELTLTLGDVVSTRS-TVKAE---- : 132
FABP_Rat : KRRDGD-KLVVECVMKGVTSTRV-YERA----- : 131
FABP_Mouse : RKRDGD-KLVVECVMKGVTSTRV-YERA----- : 131
FABP_Human-brain : REIKDG-KMVMTLTFGDVVAVRH-YEKA----- : 131
FABP_Bovine-brain : REMVAG-KLILTLTHGDVVAVRH-YEKQ----- : 131
FABP_Chicken : REIKDG-RMVMTLTFGDVVAVRH-YEKA----- : 131
RABP_Japanese_pufferfish : RELNGD-ELTLVFGADDVVCTRI-YVRE----- : 136
RABP_Mouse : RELTNDGELILTMTADDVVCTRV-YVRE----- : 137
FABP_Mouse-brain : REIKDG-KMVVTLTFGDIVAVRC-YEKA----- : 131
RABP_Human : RELTNDGELILTMTADDVVCTRV-YVRE----- : 137
FABP_Rat-brain : REIKDG-KMVVTLTFGDVVAVRC-YEKA----- : 131
MYP2_Rabbit : RKLVDG-KMVVECKMKGVVCTRI-YEKV----- : 131
RABP_Human-CRABP-I : RELANDGELILTFGADDVVCTRI-YVRE----- : 136
FABP_Bovine-mammary : RKLMDD-KMVLECVMNGVTATRV-YERA----- : 131
MYP2_Human : RKLVNG-KMVAECKMKGVVCTRI-YEKV----- : 131
RABP_Rat : RELTNDGELILTMTADDVVCTRV-YVRE----- : 138
RABP_Bovine : RELANDGELILTFGADDVVCTRI-YVRE----- : 136
RABP_Frog-CRABP : REMTNDGELILTMTADDVVCTRI-YIRD----- : 137
FABP_Nurse_shark-liver : RELQDG-KLILTCTMGDVVCTRE-YVREQ---- : 132
MYP2_Mouse : RTLLDG-RMVVECIMKGVVCTRI-YEKV----- : 131
MYP2_Bovine : RKLVDG-KMVVECKMKDVVCTRI-YEKV----- : 131
RABP_Chicken : RELANDGELIL---------- : 120
TLBP_Rat : RRIVDG-RMVVECTMNNVVSTRT-YERV----- : 132
TLBP_Mouse : RRIVDG-RMVVECTMNNVVSTRT-YERV----- : 132
FABP_Blood_fluke : REVDGD-TMKTTVTVGDVTAIRN-YKRLS---- : 133
FABP_Rat-flight_muscle : REFSKE-QCVITIKLGDLVATRI-YKAQ----- : 133
FABP_Desert_locust-flight : REFSKE-QCVITIKLGDLVATRI-YKAQ----- : 133
FABP_Mouse-keratinocytes : RKLKDG-KMIVECVMNNATCTRV-YEKVQ---- : 135
FABP_Frog-intestine : RQIIGD-ELVQTYEYEGTESKRI-FKRG----- : 131
FABP_Echinococus : RVVEGN-ELKATVKVDEVVCVRTYYSKVA---- : 134
FABP_Human-keratinocytes : RKLKDG-KLVVECVMNNVTCTRI-YEKVE---- : 135
FABP_Human-intestine : REIIGD-ELVQTYVYEGVEAKRI-FKKD----- : 131
FABP_Tobacco_hornworm : KEYNGD-DLVVTITSSNWDGVARRYYKA----- : 131
FABP_Pig-liver : TELNGD-IITSTMTLGDIVFKRI-SKRI----- : 127
FABP_Rambia_sapo-Liver-ba : QEIKGG-EMIETLTVAGTTMVRK-SKKV----- : 125
FABP_Axolotl : TELSGD-TLINTLQKGDDTYKRI-SKRIRSKRV : 131
FABP_Bovine-retina : RKLEDG-KLVVVCVMNNVTCTRV-YEKVE---- : 135
FABP_Human-liver : TELNGD-IITNTMTLGDIVFKRI-SKPI----- : 127
FABP_Shark-liver : QEVQGN-EMIEKLTAGNATMIRK-SRRM----- : 125
FABP_Anolis_pulchellus : QEIVGN-EMVETITSGSATFTRR-SKKI----- : 125
FABP_Chicken-liver : QEVKGN-EMVETITFGGVTLIRR-SKRV----- : 125
FABP_Bovine-liver : TEFNGD-TVTSTMTKGDVVFKRV-SKRI----- : 127
FABP_Rat-liver : TEFNGD-TITNTMTLGDIVYKRV-SKRI----- : 127
FABP_Mouse-liver : TELNGD-TITNTMTLGDIVYKRV-SKRI----- : 127
ILBP_Rat : SEVVGD-KLVEISTIGDVTYERV-SKRVA---- : 127
ILBP_Human : SEIVGD-KLVEVSTIGGVTYERV-SKRLA---- : 127
ILBP_Mouse : SEVVGD-KLVEISTIGDVTYERV-SKRLA---- : 127
ILBP_Pig : AEIVDG-KLVEVSTVGGVTYERV-SKKLA---- : 127
ILBP_Rabbit : SEIKGD-KLVEVSSIGGVTYERV-SKRLA---- : 127
FABP_Caenorhabditis_elega : YYFDGD-FLIQKMSFNNIEGRRF-YKRLP---- : 142
FABP_Chicken-gizzard : -------KLILTLTMGNVVSTRT-YPTTI---- : 71

The lipocalins have been classified as extracellular transport proteins and a good representative of such a function is the retinol binding protein (RBP). RBP is the sole specific carrier protein for retinol (vitamin A, $C_{20}H_{30}O$) and its single known physiological function is to protect from oxidation and deliver bound retinol from the storage sites in the liver to the peripheral target tissues. Within tissues, retinol is activated to retinoic acid (local mediator of cellular proliferation and differentiation), which in turn binds to nuclear receptors to regulate transcription of more than 300 diverse target genes.

Dietary retinoids are the only retinoids present in the body. These are packaged by the intestine as retinyl ester, along with other dietary lipids, in chylomicrons. The majority of dietary retinoid is cleared by and stored within the liver. To meet tissue retinoid needs, the liver secretes into the circulation retinol bound to a single polypeptide chain of molecular weight 21kDa synthesised in the hepatocytes, RBP (Quadro *et al.*, 1999; Soprano and Blaner, 1994). Only retinol triggers secretion of RBP, which in turn forms a complex with transthyretin (TTR, formerly called thyroxin-binding prealbumin) when circulating in the plasma. Although *in vitro* it has been shown that the complex is formed by one tetramer of TTR (55kDa) and two RBP molecules, the RBP concentration in plasma is very limiting and the complex isolated from the serum consists of one molecule of each component. The formation of such a complex seems to be essential in order (i) to prevent extensive loss of the low molecular weight RBP during its filtration and degradation through the kidney glomeruli into the intercellular space outside the vascular system and (ii) to facilitate the newly synthesised

retinol-RBP secretion (Naylor and Newcomer, 1999; Ronne *et al.*, 1983; Soprano and Blaner, 1994). It has been shown that cellular RBP expression in hepatic stellate cells (the major site of metabolism and storage: more than 80% of the liver retinoids), relevant for both uptake and esterification of retinol, responds to the extracellular retinol status, and is correlated to the retinol binding capacity of the cytosol (Vicente *et al.*, 1998). The transfer of retinol from serum RBP to cellular RBP is shown to be mediated by a receptor (Sundaram *et al.*, 1998). Such a receptor-mediated delivery of retinol to specific sites on the plasma membrane that influence overall retinol metabolism is not in agreement with studies on human keratinocytes. According to these studies the rate and extent of retinol esterification by keratinocytes and the types of retinyl esters synthesised are the same whether the retinol is delivered to the cells free or bound to RBP (Hodam and Creek, 1998). Coordinate function of co-localised cellular RBP, retinol and retinal dehydrogenase with the RBP-receptor in tissues dependent upon vitamin A for normal development, is responsible for the metabolic pathway from vitamin A to retinoic acid (Bavik *et al.*, 1997).

Attempts to understand the role of RBP in retinoid physiology have been made by means of gene targeting. Experiments performed on RBP knockout mice suggest that the physiological role of RBP is to ensure that retinol is available for maintaining normal cellular functions in times of inadequate vitamin A intake (Quadro *et al.*, 1999).

RBP has also been validated as a sensitive and accurate marker in assays concerning the production of monoclonal antibodies in dairy cows (Lindberg *et al.*, 1999), the effect of 'renal-dose' dopamine on

renal tubular function following cardiac surgery (Bhutta *et al.*, 1999) and renal tubular dysfunction (Dillon *et al.*, 1998), the detection of serum retinol in children with shigellosis (Mitra *et al.*, 1998), the deficiency of vitamin A during inflammation by measuring the molar ratio of RBP to TTR (Rosales *et al.*, 1999), the diabetic nephropathy diagnosis (Hong and Chia, 1998).

Apparent dissociation constants for human RBP vary from 1.5 to 1.9 x 10^{-7} M depending on the method of preparation of the apo-form. RBP appears to accommodate not only retinol but also retinol analogs such as retinoid acid with a dissociation constant comparable to that of retinol ($2.1x10^{-7}$ M vs $1.9x10^{-7}$ M) (Cogan *et al.*, 1976). The relative affinities of vitamin A derivatives for RBP can be placed in the order retinoid acid, retinol, retinal, retinyl ester, retinoyl ester. From binding studies of terpenoids to RBP it can be concluded that the more hydrophobic the ligand the higher the affinity it exhibits explaining the enhanced binding when β-ionone ring is present along with the hydrophobic isoprene chain with the five conjugated double bonds. The hydrophobic moiety alone though, may be insufficient to allow binding, since esters of retinol do not bind at all to RBP while the hydroxyl group of retinol might do the difference to the affinity it exhibits (Hase *et al.*, 1976).

About 5% of the total amount of RBP in the plasma does not participate in complex formation with TTR, 80-90% is apo- and the rest is holo RBP. The total pool of RBP in the body is probably in the order of 410mg, slightly more than half of which (220mg) is present in the plasma compartment (Fex *et al.*, 1979).

3.1.2. Description of RBP molecule.

According to the complete human RBP amino acid sequence determined by Rask *et al.* (1979), the primary structure of RBP consists of 182 amino acids. Crystallisation in the rhombohedral spacegroup and preliminary X-ray data collection of human plasma RBP was first reported by (Ottonello *et al.*, 1983) followed by (Newcomer *et al.*, 1984) who crystallised human and rabbit plasma RBP in the orthorhombic spacegroup. Previous x-ray crystallographic work has shown that RBP is composed of a single globular domain (~40Å diameter) that comprises an N terminal coil, 8 antiparallel β-strands (β-barrel), a short helix and a C terminal strand (*Figure* 3.2). The loops connecting the strands are all rather short. The β-strands correspond to regions of relatively low mobility whereas loop residues between the strands display larger positional fluctuations. The secondary structure of RBP is stabilised by three disulphide bridges formed between residues 4-160, 70-174 and 120-129 (Cowan *et al.*, 1990) (*Figure* 3.2).

The overall structure encompasses both transthyretin and retinol binding sites. The TTR binding site is a hydrophobic area on the protein surface formed by some of the loops that surround the opening of the β-barrel (residues 35, 63, 64, 67). The structures of the complex of human RBP with its carrier protein transthyretin were reported recently by Naylor and Newcomer (1999) and chicken RBP in complex with TTR by Monaco *et al.* (1995) it is evident from these structures that RBP binds at a 2-fold axis of symmetry in the TTR tetramer thereby creating a 2-fold symmetry at the interaction site. Four amino acids from the TTR molecule (Arg 21, Val 20, Leu 82 and Ile-84) are contributed by

the two monomers, while amino acids from RBP that appear at the recognition site (Trp 67, Phe 96, Leu 63, Leu 97) are flanked by the

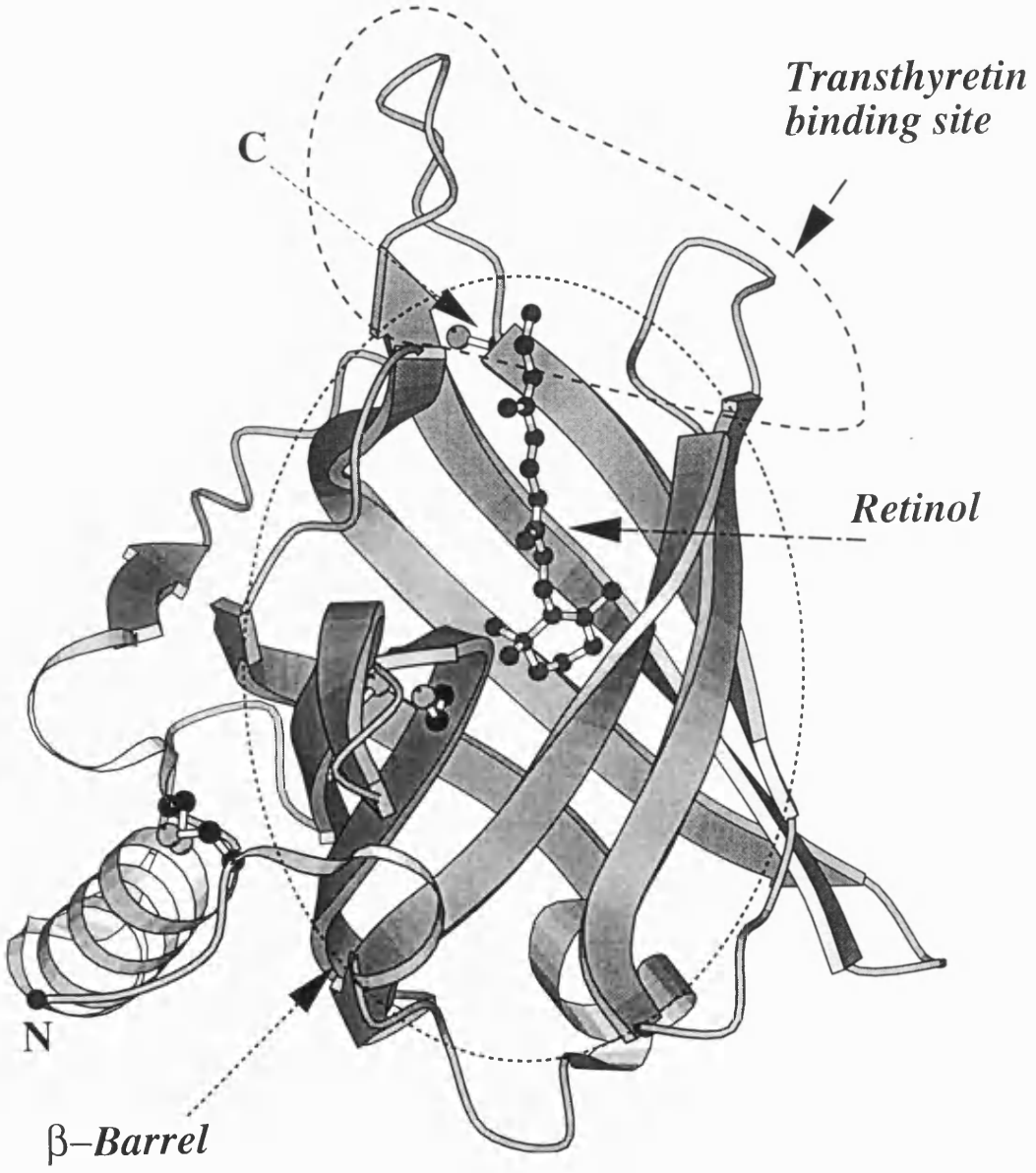


***Figure* 3.2.**

Structure of Retinol Binding Protein. The figure was generated using the program BOBSCRIPT (Esnouf, 1997)

TTR symmetry related side chains. An interaction of the carboxy terminus of RBP with TTR has also been revealed and it has been suggested that such an interaction might be of physiological significance.

Residues involved in the formation of the retinol binding site are mainly those that contribute to the formation of the β-barrel which completely surrounds the retinol molecule with the β-ionone ring in the centre and the isoprene chromophore stretching along the barrel axis almost to the solvent surface (residues 32-37, 47-53, 64-68, 92-98). The retinol is buried in the barrel with no covalent attachment to the protein. Although the β-barrel has two entrances, there is access to the core only through one due to blockage of the second entrance by a cluster of phenylalanine rings (Cowan *et al.*, 1990). The β-barrel also forms a great number of hydrogen bonds (they are responsible for the formation of two orthogonal β-sheets) and its rigidity can be attributed either to hydrophobic or to interchain interactions (Zanotti *et al.*, 1993).

The structure of human plasma RBP, in both the holo and apo forms, has been determined at 2.5 Å resolution (Zanotti *et al.*, 1993). The flexibility of some regions of the molecule leads to different conformations and as a result the holo and apo forms of the protein crystallise in orthorhombic and trigonal space groups respectively (Newcomer *et al.*, 1984; Zanotti *et al.*, 1993).

3.1.3. Human serum RBP as a model lipocalin.

Human serum RBP is a well characterised protein from both functional and structural points of view, thus it was selected as a model lipocalin (Greene, 1998). The presence of four tryptophan residues located in different regions of the structure can serve as probes of folding and unfolding processes. Being a carrier protein, RBP molecule appears to be a good model protein for protein engineering studies and for the design of molecules that will enable RBP to transport not only drugs but also other molecules.

Investigations into the relationship between sequence conservation, stability and folding with RBP were performed in our collaborator's laboratory (Prof. Keith Brew, University of Miami School of Medicine, Miami) with the specific aims: (i) to elucidate the nature of the conserved sequence information in the lipocalin superfamily and evaluate its structural significance; (ii) to develop a high yield expression system, purification and *in vitro* refolding procedures for RBP; (iii) to generate and express mutants with tryptophan replacements used to experimentally investigate the roles of conserved residues versus non conserved tryptophans, in native and molten globule state stability in human serum RBP and (iv) to characterise the folding and unfolding behaviour of human serum RBP (Greene, 1998).

In the context of above studies the cDNA encoding RBP was cloned into the prokaryotic expression vector pET3a. This vector and the bacterial host BL21(DE3) were selected for the expression system and provided high yields of recombinant protein in the form of inclusion bodies.

Site directed mutagenesis by PCR using the 'mega-primer method' was applied to construct RBP variants. The sequence

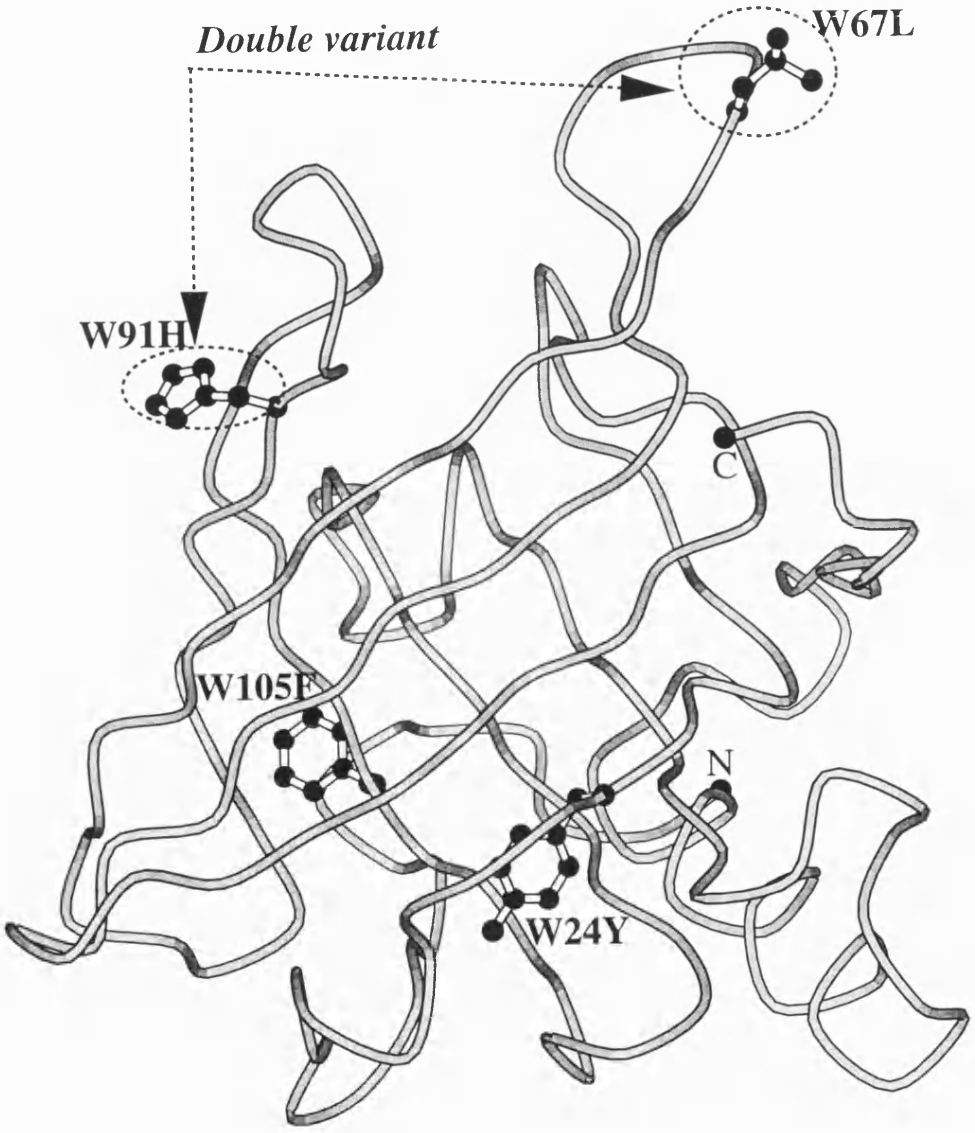


***Figure* 3.3.**

Worm representation of rRBP variants showing the sequence substitutions. The figure was generated using the program BOBSCRIPT (Esnouf, 1997).

substitutions in rRBP targeted to tryptophan residues located at the entrance of the β-barrel close to the retinol hydroxyl group. Selection of substitutions was determined by a homology-based approach using an alignment of RBPs from different species and closest relatives or in consideration of similar side chain chemistry as to minimise structural disruption and effects on the folding pathway. Experimental investigations were conducted into the role of the different tryptophans in structure, as well as native and molten globule state stability. Apart from the homology consideration tryptophans were subjected to mutagenesis for stability and folding studies based on criteria such as : (i) they are ideal intrinsic fluorophors; (ii) their excitation and emission can be selectively monitored at λ_{ex}=295 and λ_{em} =350nm; (iii) they contribute significantly to CD spectrum; (iv) the range of locations make them ideal probes of different structural regions; (v) differences in conservation enable comparison between conserved and non-conserved regions.

The sequence changes that have been introduced to the recombinant parent protein (rRBP) by mutagenesis are (a) a double substitution at positions 67 and 91 from tryptophan to leucine and histidine respectively, (b) a single point substitution at position 24 from tryptophan to tyrosine and (c) a single point substitution at position 105 from tryptophan to phenylalanine as presented in *Figure* 3.3. According to the near- and far-UV circular dichroism (CD) measurements, the substitutions were found not to disrupt the overall native secondary or tertiary rRBP structure. Substitutions to the most conserved tryptophans had different effects on the CD spectrum in comparison to the other

tryptophans. Thermal and chemical denaturation studies indicated that mutations had varying effects on the stability of the native state and there was no correlation between sequence conservation and molten globule stability suggesting that the molten globule state of RBP was not stabilised by native interactions (Greene, 1998).

rRBP : The high resolution structure of rRBP (1.7 Å resolution) was determined using crystals of rRBP grown in the trigonal spacegroup. Although the apo-form of RBP had been previously determined (Zanotti *et al.*, 1993) at 2.5 Å resolution, the high resolution structure presented here was aimed at describing the fine details of the β-barrel architecture and the structure of the water molecules in the interior of the barrel in the absence of retinol molecule and set a template for structural analysis of RBP variants.

W67L/W91H rRBP double variant : Kinetic studies indicated that rRBP folds through a biphasic process (Greene, 1998). The properties of an intermediate suggest that some or all tryptophan residues are present in a less polar environment in the intermediate than in the native structure. In order to investigate further how the non-conserved tryptophans could affect the second rate of folding, sequence substitutions were performed considering the functional role of particular residues in RBP. The basis of the double sequence substitution of Trp 67 and Trp 91 was the primary structure of frog RBP where Leu and His are found at positions 67 and 91 respectively.

Tryptophans 67 and 91 of RBP are also positioned at the surface of the molecule in loop regions that connect the β-strands forming the β-barrel. The crystal structure of the TTR-RBP complex showed that these two tryptophans are located at the interface between the two proteins (Monaco *et al.*, 1995; Naylor and Newcomer, 1999). Hence, in an effort to identify structural perturbations that could directly affect the TTR-RBP complex, the structure of the double variant W67L/W91H was determined at high resolution (2.0 Å).

Tryprophan 24 to tyrosine rRBP variant : There appears to be a large difference between the contributions of Trp24 and other tryptophans in stability by thermal and solvent denaturation. The change in stability to thermal unfolding exhibits significant difference (four fold) in substitutions to conserved compared to non-conserved tryptophans (Greene, 1998). The equilibrium unfolding studies revealed that substitutions for the conserved Trp24 destabilises the native structure by about 2 kcal/mol to chemical denaturation. Molecular model building analysis with the hRBP 3-D structure, indicates that the Trp24-Tyr substitution did not disrupt the conserved tertiary H-bond. The decrease in stability may be attributed to the combined effect of a more polar side-chain with a lower volume, decrease in non-polar van der Waals packing interactions, and the reduced electrostatic interaction with Arg139. Preliminary crystallisation trials for this variant have been performed (see Table 3.11 for more details). It is hoped that the 3-D structure of this rRBP variant might reveal the functional role of Trp 24 in the RBP structure.

Tryptophan 105 to phenylalanine rRBP variant: Trp 105 is conserved in all known RBPs and appears to facilitate the folding process of the molecule by increasing the folding rates. Trp 105 also flanks a basic residue, Lys85 (Cowan *et al.*, 1990). The folding kinetics of guanidine HCl-denatured rRBP appears to be biphasic (Greene, 1998). The first phase is studied via the formation of Trp 24 used as a probe while the second phase reflects changes in the environment of one or more of the non conserved tryptophans 67, 91 or 105 in the superfamily of the lipocalins (only in RBP family) which appear to be more buried during the initial folding phase and become more exposed during the subsequent structural rearrangement to the native state. Attempts to crystallise this single rRBP variant are in progress.

3.2.STRUCTURAL STUDIES ON NATIVE RECOMBINANT RETINOL BINDING PROTEIN (rRBP) AND VARIANTS.

3.2.1. MATERIALS AND METHODS

3.2.1.1. Preparation of proteins

rRBP: The cDNA encoding human serum RBP was amplified by PCR using two primers designed specifically for directional cloning into the pET-3a expression vector. The cloning vector comprised the restriction enzyme *Nde*I cloning site, the initiator methionine codon Met-1, the *Bam*HI cloning site and stop codon. Twelve extra bases were included outside each restriction site to allow for efficient cutting with restriction enzymes and the cDNA of interest was ligated to the vector. The product of the amplification was purified by agarose gel electrophoresis, digested with *Nde*I and *Bam*HI, and cloned into a preparation of the expression vector (pET3a) that had been previously digested with the same restriction enzymes. The cells after being transformed they were plated on LB/ampicillin plates and colonies were screened for the presence of the recombinant vector by restriction endonuclease mapping of miniprep purified plasmids. Amplification of recombinant plasmids (pRBP) for sequencing and for use as templates in PCR mutagenesis experiments were carried out using *E. coli* strain DH5α as a host (Greene, 1998).

RBP mutants: Mutations were introduced using the 'megaprimer' method (Sarkar and Sommer, 1990). During amplification the synthetic T7 promoter primer or T7 terminator primer together with the appropriate mutagenic primer were used in order to generate the

'megaprimer'. Purification of the 'megaprimer' followed by gel electrophoresis that was used in the second amplification with the same template and the cognate T7 primer. The final amplification product was purified and cloned into pET3α as described for RBP. The recombinant vectors containing RBP mutants were sequenced by automated DNA sequencing using both T7 promoter and T7 terminator primers (Greene, 1998).

3.2.1.2. Protein expression and purification of rRBP. *In vitro* folding and purification of folded rRBP or variants.

Cultures of *E. coli* strain BL21(DE3) transformed with the recombinant vector (pRBP) or variant were grown, and inclusion bodies containing recombinant RBP (rRBP) were obtained by cell lysis and isolated by centrifugation. After their solubilisation by extraction with urea, they were passed through a Macroprep column equilibrated with tris-HCl, and SDS gel electrophoresis indicated that rRBP was in a relatively pure form.

Fractions containing rRBP or variants were collected and the protein concentration was estimated from the absorbance at 280 nm and the final concentration was between 50 μg/ml and 100 μg/ml. Properly folded rRBP along with some peaks standing for misfolded protein was recovered after further concentration and purification by anion exchange chromatography.

Functional characterisation of the produced rRBP followed along with determination of near and far UV CD spectra of rRBP and mutants.

Equilibrium thermal unfolding studies were also conducted and guanidine hydrochloride equilibrium unfolding was determined. Urea unfolding of molten globule state studies and stopped flow fluorescence unfolding studies were finally performed. Expression and purification of rRBP and variants of RBP were performed by our collaborator, Prof. Keith Brew, at Miami and the required quantities of proteins for crystallographic studies at Bath were provided by his group.

3.2.1.3.Crystallographic studies.

rRBP : Large single needle shaped crystals of native recombinant human RBP (rRBP) were grown successfully using the vapour diffusion method with hanging drops (*Figure* 3.4). 2.0 μl of protein (~12mg/ml) dissolved in 10mM sodium cacodylate and 3.9 M sodium chloride pH 6.8 are placed on siliconised coverslips and left to equilibrate against the reservoir solution that consists of 20 mM sodium cacodylate and 4.5 M NaCl, pH 6.8. The crystallisation conditions are recorded in Table 3.1.

Diffraction data from one rRBP crystal were collected on station 9.5 (λ=0.9 Å) at the Synchrotron Radiation Source, CLRC Laboratory, Daresbury, UK under cryogenic conditions using glycerol as cryoprotectant at 1.7 Å resolution. Prior to data collection at 100°K, the crystals were transferred to buffer solution (same as the reservoir solution) containing 25% v/v glycerol for about 30 sec. Data were recorded on a 30 cm image plate (MAR Research, Hamburg, Germany) with the crystal oscillating through 2.0° steps. Data integration and reduction were performed with the programs *DENZO* and *SCALEPACK* (Otwinowski and Minor, 1997). Data collection details are summarised

in Table 3.1 and Tables 3.2, 3.3 and the diffraction pattern of rRBP data is shown in *Figure* 3.5.

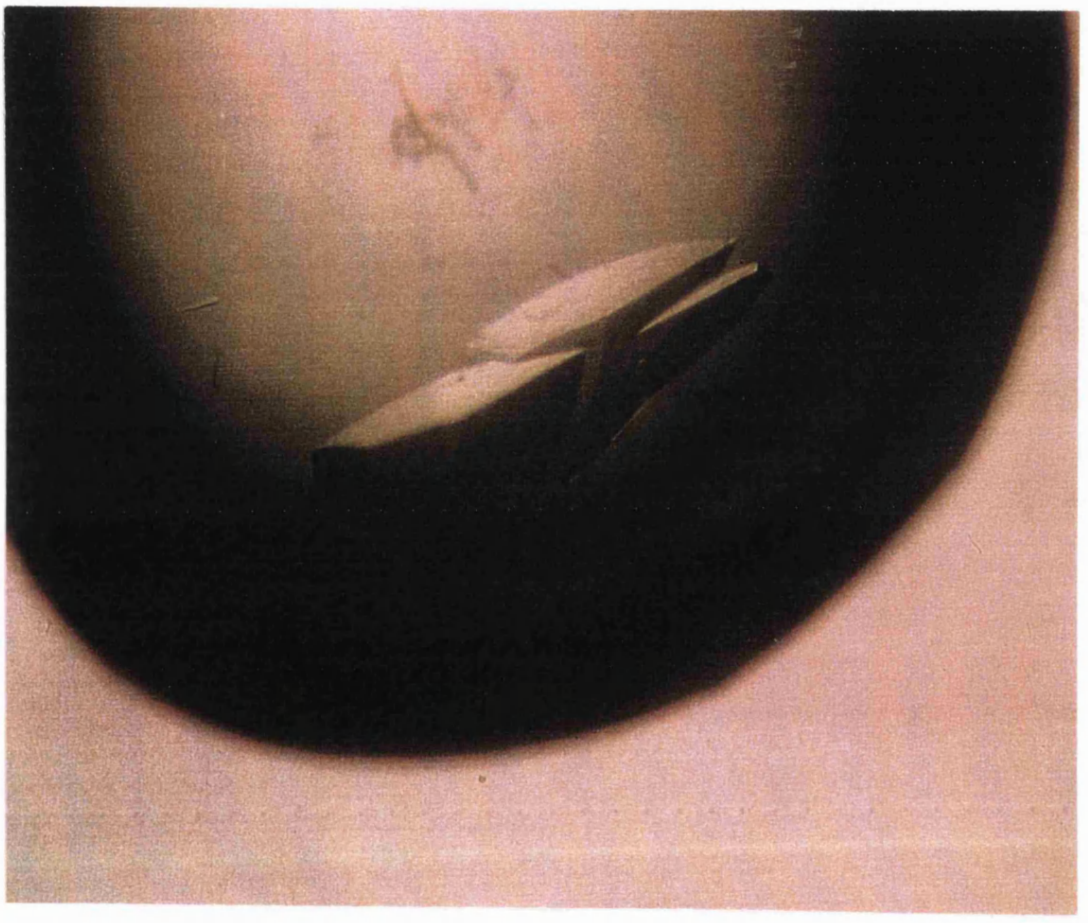


***Figure* 3.4**

Crystal of recombinant Retinol Binding Protein

Table 3.1 Crystallisation and Data collection statistics for rRBP

Crystallisation	
Reservoir solution	20 mM Sodium Cacodylate 4.5 M Sodium Chloride pH 6.8
Drop	12 mg/ml protein concentration 10 mM Sodium Cacodylate 3.9 M NaCl pH 6.8
Temperature	16°C
Cell dimensions (a, b, c) (Å)	a=b=101.34 c=71.98
α, β, γ (°)	90, 90, 120
Space group	*R*3
Molecules per asymmetric unit	1
Data collection & processing	
Temperature	100°K
Station (Synchrotron)	Daresbury PX95
Image plate	Mar30
Wavelength (Å)	0.90
Distance (mm)	250
Oscillation range (degrees)	2.0
No of images (degrees)	25 (50)
No of observations	67359
No of unique reflections	28189
Maximum resolution (Å)	1.7
[#]Completeness (outermost shell) (%)	90.5 (91.0)
[✧]R_{sym} (I) (outermost shell) (%)	4.7 (40.5)
$< I / \sigma (I) >$ (outermost shell)	7.68 (2.59)
Outermost shell (Å)	1.72 - 1.70

[#]Completeness in the range ∞-resmax, where resmax is the maximum resolution to which data were collected. [✧]$R_{sym}(I)=\Sigma_{hkl}\Sigma_i| I_{hkl,i}-I_{average,hkl}| / | \Sigma_{hkl}\Sigma_i| I_{hkl,i}| x100$.

Table 3.2 Distribution of reflections for rRBP data

Shell		*I / Sigma in resolution shells*								
Lower	*Upper*	*No. of reflections with I / Sigma less than*								
limit	*limit*	*0*	*1*	*2*	*3*	*5*	*10*	*20*	*>20*	*total*
∝	4.97	1	1	2	4	15	379	928	0	928
4.97	3.94	0	0	0	1	7	383	867	0	867
3.94	3.45	0	1	3	4	9	477	1007	0	1007
3.45	3.13	3	5	13	18	34	525	1062	0	1062
3.13	2.91	2	8	21	32	57	593	1103	0	1103
2.91	2.74	8	15	31	43	81	648	1126	0	1126
2.74	2.6	7	20	45	68	119	704	1149	0	1149
2.6	2.49	8	21	54	87	169	760	1157	0	1157
2.49	2.39	16	53	86	127	216	817	1146	0	1146
2.39	2.31	11	29	63	122	242	845	1160	0	1160
2.31	2.24	19	50	94	139	237	473	919	249	1168
2.24	2.17	19	57	113	169	276	553	980	188	1168
2.17	2.11	27	68	136	201	324	610	1009	159	1168
2.11	2.06	22	79	159	229	378	719	1040	113	1153
2.06	2.02	37	105	186	272	478	775	1101	95	1196
2.02	1.97	36	127	222	325	509	843	1108	49	1157
1.97	1.93	43	134	236	335	542	884	1133	31	1164
1.93	1.9	59	156	288	408	660	959	1156	16	1172
1.9	1.86	75	205	379	540	764	1018	1148	10	1158
1.86	1.83	77	247	429	591	813	1054	1156	6	1162
1.83	1.8	86	248	466	642	883	1107	1189	0	1189
1.8	1.77	101	306	499	683	888	1083	1126	0	1126
1.77	1.75	143	354	570	761	959	1145	1191	1	1192
1.75	1.72	159	380	612	788	975	1128	1148	0	1148
1.72	1.7	179	423	663	845	1041	1155	1163	0	1163
	All hkl	*1138*	*3092*	*5370*	*7434*	*10676*	*19637*	*27272*	*917*	*28189*

Table 3.3 Completeness of the rRBP data

Shell		*I / Sigma in resolution shells*								
Lower	*Upper*	*% of reflections with I / Sigma less than*								
limit	*limit*	*0*	*1*	*2*	*3*	*5*	*10*	*20*	*>20*	*total*
∝	4.97	0.1	0.1	0.2	0.3	1.2	30.5	74.7	0	74.7
4.97	3.94	0	0	0	0.1	0.6	30.8	69.6	0	69.6
3.94	3.45	0	0.1	0.2	0.3	0.7	38.2	80.6	0	80.6
3.45	3.13	0.2	0.4	1	1.4	2.7	42	84.9	0	84.9
3.13	2.91	0.2	0.6	1.7	2.6	4.6	48.2	89.6	0	89.6
2.91	2.74	0.6	1.2	2.5	3.4	6.5	51.6	89.7	0	89.7
2.74	2.6	0.6	1.6	3.6	5.5	9.5	56.5	92.1	0	92.1
2.6	2.49	0.6	1.7	4.3	7	13.5	60.8	92.6	0	92.6
2.49	2.39	1.3	4.3	7	10.3	17.5	66.2	92.8	0	92.8
2.39	2.31	0.9	2.3	5	9.8	19.4	67.7	92.9	0	92.9
2.31	2.24	1.5	4	7.5	11.1	19	37.9	73.6	19.9	93.5
2.24	2.17	1.5	4.6	9.1	13.6	22.3	44.6	79.1	15.2	94.3
2.17	2.11	2.2	5.4	10.9	16.1	26	48.9	80.8	12.7	93.6
2.11	2.06	1.8	6.5	13.1	18.9	31.2	59.3	85.8	9.3	95.1
2.06	2.02	2.9	8.3	14.6	21.4	37.6	60.9	86.6	7.5	94
2.02	1.97	2.9	10.2	17.9	26.2	41	67.9	89.3	3.9	93.2
1.97	1.93	3.5	10.9	19.2	27.2	44	71.8	92	2.5	94.6
1.93	1.9	4.7	12.5	23	32.6	52.7	76.5	92.3	1.3	93.5
1.9	1.86	6	16.5	30.4	43.4	61.4	81.8	92.2	0.8	93
1.86	1.83	6.2	19.8	34.4	47.4	65.2	84.6	92.8	0.5	93.3
1.83	1.8	6.8	19.5	36.7	50.6	69.6	87.2	93.7	0	93.7
1.8	1.77	8.4	25.6	41.7	57.1	74.2	90.5	94.1	0	94.1
1.77	1.75	11.2	27.7	44.6	59.6	75.1	89.7	93.3	0.1	93.3
1.75	1.72	13	31.2	50.2	64.6	80	92.5	94.2	0	94.2
1.72	1.7	14	33.1	51.9	66.1	81.5	90.4	91	0	91
	All hkl	***3.7***	***9.9***	***17.2***	***23.9***	***34.3***	***63.1***	***87.6***	***2.9***	***90.5***

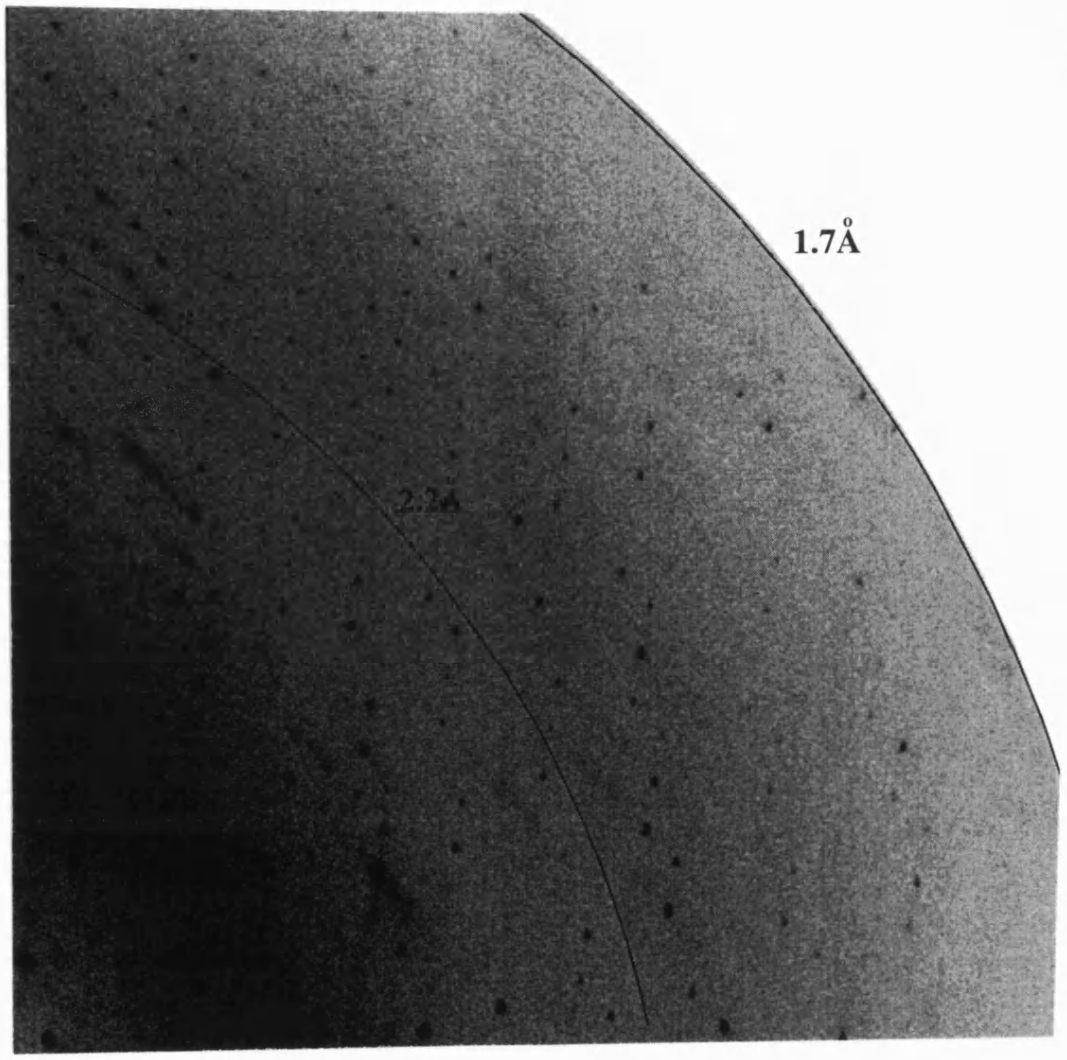


***Figure* 3.5**

rRBP diffraction pattern

rRBP double mutant W67L/W91H: rRBP variant Trp67-Leu and Trp91-His, was crystallised by the vapour diffusion method at 16°C (*Figure* 3.6). Single crystals were obtained by applying the seeding technique from crystals of the native recombinant protein and their size was optimised by addition of 0.05% β-octyl-glucoside. The crystallisation conditions are listed in Table 3.4. The data collection was

performed on station X31 (λ=1.09 Å) at EMBL Hamburg outstation, Hamburg, Germany using glycerol as cryoprotectant. Prior to data collection at 100°K the crystals were transferred to buffer solution (same as reservoir solution) containing 20% v/v glycerol for about 30 sec. The crystals diffracted to 2.0 Å resolution and data were recorded on a 18 cm image plate (MAR Research, Hamburg, Germany) with the crystal oscillating through 1.2° steps. Data integration and data reduction were performed with the programs *DENZO* and *SCALEPACK* (Otwinowski and Minor, 1997). Data collection details are summarised in Tables 3.4-3.6.

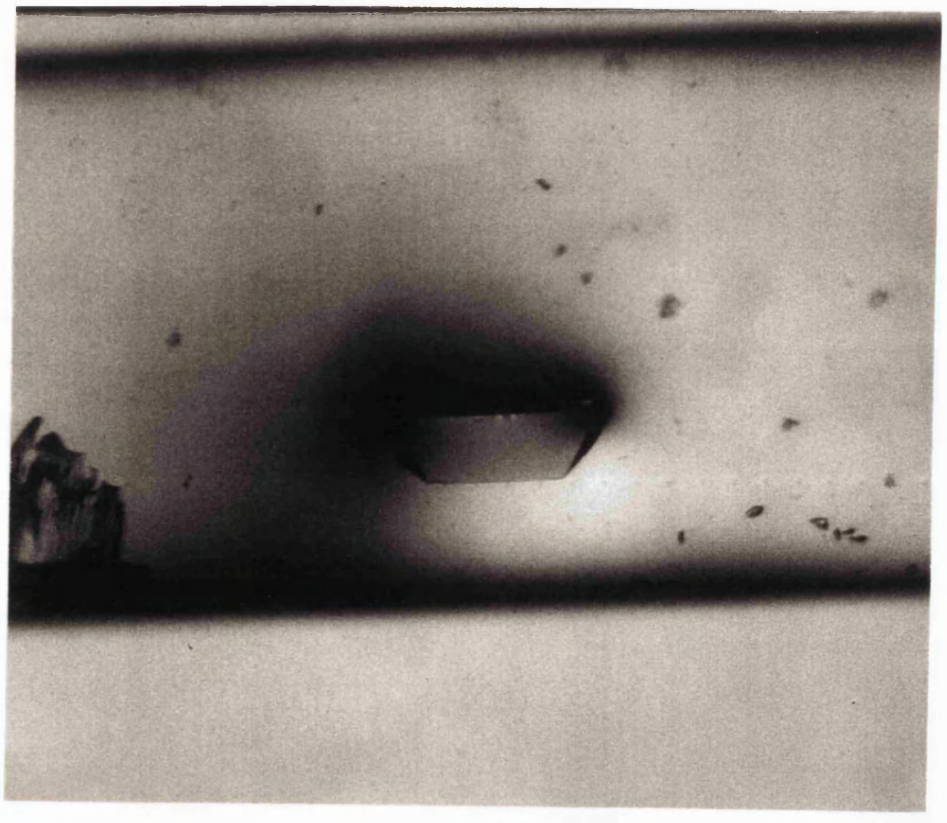


***Figure* 3.6**

Crystal of the double variant W67L/W91H of rRBP.

Table 3.4 Crystallisation and Data collection statistics for W67L/W91H rRBP variant

Crystallisation	
Reservoir solution	Sodium Chloride 4.5M, Tris/HCl 50mM pH 9.1 + 0.05% β-octyl-glucoside
Drop	12 mg/ml protein concentration Sodium Chloride 2.25M, Tris/HCl 25mM pH 9.1 + 0.025% β-octyl-glucoside (seeding)
Temperature	16°C
Cell dimensions (a, b, c) (Å)	a=b=101.76 c=72.72
α, β, γ (°)	90, 90, 120
Space group	R3
Molecules per asymmetric unit	1
Data collection & processing	
Temperature	100°K
Station (Synchrotron)	EMBL-Hamburg X31
Image plate	Mar18
Wavelength (Å)	1.09
Distance (mm)	146.17
Oscillation range (degrees)	1.2
No of images (degrees)	92 (110.4)
No of observations	109525
No of unique reflections	18890
Maximum resolution (Å)	2.0
*Completeness (outermost shell) (%)	98.0 (92.7)
✧R_{sym} (I) (outermost shell) (%)	10.2 (37.0)
$< I / \sigma (I) >$ (outermost shell)	7.53 (3.19)
Outermost shell (Å)	2.07 – 2.00

*Completeness in the range ∞-resmax, where resmax is the maximum resolution to which data were collected. ✧$R_{sym}(I)=\Sigma_{hkl}\Sigma_i| I_{hkl,i}-I_{average,hkl}| / |\Sigma_{hkl}\Sigma_i| I_{hkl,i}|$ x100.

Table 3.5 Distribution of reflections for W67L/W91H rRBP variant data

Shell		*I / Sigma in resolution shells*								
Lower	*Upper*	*No. of reflections with I / Sigma less than*								
limit	*limit*	*0*	*1*	*2*	*3*	*5*	*10*	*20*	*>20*	*total*
∝	4.31	9	23	40	68	103	415	1696	111	1807
4.31	3.42	13	28	60	85	140	481	1713	170	1883
3.42	2.99	27	62	136	201	321	791	1897	18	1915
2.99	2.71	32	116	212	324	531	1147	1907	9	1916
2.71	2.52	47	134	298	473	763	1435	1924	0	1924
2.52	2.37	75	222	423	601	951	1582	1898	1	1899
2.37	2.25	104	267	475	699	1060	1713	1931	0	1931
2.25	2.15	94	281	530	771	1205	1773	1944	0	1944
2.15	2.07	116	350	647	900	1280	1784	1873	0	1873
2.07	2.0	157	431	750	1051	1437	1759	1798	0	1798
	All hkl	***674***	***1914***	***3571***	***5173***	***7791***	***12880***	***18581***	***309***	***18890***

Table 3.6 Completeness of the W67L/W91H rRBP variant data

Shell		*I / Sigma in resolution shells*								
Lower	*Upper*	*% of reflections with I / Sigma less than*								
limit	*limit*	*0*	*1*	*2*	*3*	*5*	*10*	*20*	*>20*	*total*
40	4.31	0.5	1.2	2.1	3.5	5.4	21.6	88.2	5.8	94
4.31	3.42	0.7	1.4	3.1	4.4	7.2	24.8	88.3	8.8	97.1
3.42	2.99	1.4	3.2	7	10.4	16.5	40.8	97.8	0.9	98.7
2.99	2.71	1.7	6	11	16.8	27.5	59.5	98.9	0.5	99.3
2.71	2.52	2.4	6.9	15.4	24.5	39.5	74.3	99.6	0	99.6
2.52	2.37	3.9	11.6	22.2	31.5	49.9	83	99.6	0.1	99.6
2.37	2.25	5.4	13.8	24.6	36.1	54.8	88.6	99.8	0	99.8
2.25	2.15	4.8	14.4	27.1	39.5	61.7	90.7	99.5	0	99.5
2.15	2.07	6.2	18.6	34.4	47.8	68	94.7	99.5	0	99.5
2.07	2	8.1	22.2	38.7	54.2	74.1	90.7	92.7	0	92.7
	All hkl	*3.5*	*9.9*	*18.5*	*26.8*	*40.4*	*66.8*	*96.4*	*1.6*	*98*

3.2.1.4. Structure determination

rRBP: The structure of the native recombinant protein was determined using the molecular replacement technique with the program *AMoRe* (Navaza, 1994). The initial phases for RBP were obtained by using the human serum retinol binding protein structure at 2.0 Å [PDB code: 1rbp, (Cowan *et al.*, 1990)] as a search model. After superposition of human serum retinol binding protein (hRBP) with rat androgen-dependent epididymal retinoic acid binding protein (E-RABP)

(Newcomer, 1995) residues (14-30), 70, 80, (84-86), (106-115), (135-139), (146-151), (159-160), 165, 172 and 174 are considered to be conserved in the RBP structure hence they were included in the search model. Residues that were part of flexible loops or other disordered regions were excluded from the model and these are residues (1-3), (50-67), (94-105), (140-143) and (175-182).

Table 3.7 After *TRAING* solutions for one molecule

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONTF1	31	27.49	151.5	0.0833	0.0714	0	27.8	54	1
SOLUTIONTF1	70.81	82.16	28.98	0.1944	0.2698	0	29.1	53.1	2
SOLUTIONTF1	7.56	42.13	162	0.0278	0.0437	0	29.4	52.9	3
SOLUTIONTF1	60.53	96.24	104.68	0.1825	0.377	0	28.7	52.8	4
SOLUTIONTF1	18.7	138.32	59.77	0.9921	0.4722	0	27.9	53.1	5
SOLUTIONTF1	58.87	90.75	214.26	0.0397	0.0675	0	28.5	53.4	6
SOLUTIONTF1	70.59	42.78	60.3	0.0079	0.0516	0	28	53.5	7
SOLUTIONTF1	75.74	139.85	145.4	0.5675	0.3849	0	29.4	53.1	8
***									
SOLUTIONTF1	**33.24**	**48.05**	**141.56**	**0.377**	**0.2063**	**0**	**45.5**	**46.6**	**9**
***									
SOLUTIONTF1	17.13	109	44.59	0.0238	0.131	0	27.5	53.4	10
SOLUTIONTF1	108.45	80.42	102.12	0.0754	0.5079	0	28.6	53.3	11
SOLUTIONTF1	95.68	96.87	210.33	0.3294	0.2659	0	27.2	54.1	12
SOLUTIONTF1	46.49	85.03	207.72	0.119	0.0714	0	29	53.4	13
SOLUTIONTF1	98.98	132.03	101.91	0.3016	0.5437	0	29.2	52.7	14
SOLUTIONTF1	87.8	141.54	234.77	0.2619	0.5238	0	27.8	53.4	15
SOLUTIONTF1	114.58	79.87	283.5	0.2976	0.5833	0	28.6	52.8	16

Table 3.8 After *FITING* final solution for one molecule

	α	β	γ	T_x	T_y	T_z	cc	R_f	*no*
SOLUTIONF	30.04	27.52	152.9	0.0818	0.0694	0	35.6	53.1	1
SOLUTIONF	70.74	81.95	28.9	0.1955	0.2704	0	35.3	52.6	2
SOLUTIONF	8.06	42.83	161.77	0.028	0.0439	0	36.4	52.2	3
SOLUTIONF	60.13	96.21	104.39	0.1822	0.3757	0	35.2	52.1	4
SOLUTIONF	19.17	138.54	59.61	0.9938	0.4736	0	34.6	52.5	5
SOLUTIONF	58.88	90.22	214.67	0.0411	0.0686	0	36.5	52.8	6
SOLUTIONF	71.65	42.82	60.11	0.0079	0.0503	0	35.4	52.7	7
SOLUTIONF	76.81	140.18	146.35	0.5673	0.3845	0	36.4	52.4	8
**									
SOLUTIONF	**31.99**	**48.44**	**140.97**	**0.3758**	**0.2058**	**0**	**52.8**	**44.6**	**9**
**									
SOLUTIONF	16.86	108.74	44.65	0.0239	0.1313	0	34.6	52.9	10
SOLUTIONF	109.16	80.81	102.29	0.0745	0.5079	0	35.2	52.7	11
SOLUTIONF	95.26	97.72	210.64	0.3294	0.2654	0	34.3	53.2	12
SOLUTIONF	47.02	84.39	207.14	0.1184	0.0714	0	36.9	52.6	13
SOLUTIONF	98.07	131.9	101.39	0.3015	0.5439	0	35.9	52.2	14
SOLUTIONF	88.29	141.56	235.13	0.2618	0.5248	0	34.7	52.9	15
SOLUTIONF	113.9	80.06	283.01	0.2971	0.5819	0	35.7	52.4	16

Note : Tables 3.7-3.8. The columns α, β, and γ correspond to the orientation Euler angles; T_x, T_y and T_z represent the positional parameters (fractions of the unit cell), cc the correlation coefficient, R_f the *R*-factor and *no-* to the sorting number of the peak when the translation function was calculated. The solution is shown in shaded area and the peak number is in bold face.

rRBP double mutant W67L/W91H: The structure of the double mutant at 2.0 Å resolution was determined using the molecular replacement technique with the program *X-PLOR* (Brünger, 1992b). The initial phases for RBP were obtained by using the rRBP structure at 1.7 Å as a search model. The solution appeared clearly in the rotation function list containing the most probable Euler angles as it is shown in the following table.

Table 3.9 Rotation function list for rRBP variant

index,	α	β	γ	RF-function (ε= 0.25)
********	********	********	********	********
1	**0**	**180**	**0**	**4.7434**
********	********	********	********	********
121	32.143	72.5	327.857	2.3633
182	252.125	2.5	252.125	2.1029
206	340.938	127.5	260.007	2.0374
216	77.483	50	71.483	2.0057
226	123.429	105	123.429	1.9891
236	308.754	142.5	223.42	1.9691
247	63.404	60	328.404	1.95
258	338.615	47.5	269.141	1.929
266	296.17	172.5	183.83	1.9178
267	310.183	130	243.206	1.9177
293	312.795	122.5	198.509	1.8924
316	27.596	2.5	27.596	1.872
319	125.106	177.5	114.894	1.8698
321	17.236	65	310.036	1.8656
326	250.559	57.5	172.298	1.8612
336	274.966	50	262.966	1.8524
344	208.286	115	100.286	1.8477
370	262.109	147.5	196.891	1.8258
377	289.652	5	229.652	1.8182
381	34.422	67.5	343.653	1.8146
388	204.706	42.5	155.294	1.8031
399	313.571	77.5	213.571	1.7902
...	...	...	...	...

Note: The columns α, β, and γ correspond to the orientation Euler angles.

3.2.1.5. Refinement of rRBP and W67L/W91H rRBP variant.

The resultant models of rRBP and W67L/W91H rRBP variant structures (one molecule per a.u.) were first subjected to restrained refinement with maximum likelihood method using the program *REFMAC* (CCP4, 1994) followed by the automated refinement procedure (*ARP*) (Lamzin and Wilson, 1993) and to rigid body refinement using the program *X-PLOR*, respectively. The combination of *REFMAC* and *ARP* refinement protocol applied to rRBP model, significantly improved the quality of the electron density map and enabled the building of the residues that have been initially excluded from the molecular replacement solution, which comprised only conserved residues. Cycles of refinement were performed using the slowcool protocol at moderate temperatures initially with the program *X-PLOR* (Brünger, 1992b) and during the final stages of refinement, with *CNS* (Brünger *et al.*, 1998) for both structures. The progress of refinement was monitored through both free and conventional *R*-factors (Brünger, 1992a). Alternative cycles of manual rebuilding with the graphics program '*O*' (Jones *et al.*, 1991), intertwined with refinement cycles using the standard protocol improved further the quality of the model. In both cases simulated annealing omit maps calculated using either *X-PLOR* or *CNS* indicated large pieces of density in the core of the β-barrel and as well as on the surface of two proteins sufficient to accommodate sometimes more than one glycerol molecules (glycerol used as cryoprotectant during data collection). Six glycerol molecules were incorporated in to the rRBP model (five in the W67L/W91H variant structure) at the very final stages of the refinement. In each case, water molecules were inserted in the model

according to the ($2F_o$-F_c) and (F_o-F_c) electron density maps. The final refinement statistics and model quality are summarised in Table 3.10 and portion of the (2Fo-Fc) electron density map at 1.7 Å resolution is shown in *Figure* 3.7 for the rRBP structure.

Table 3.10 Refinement statistics for rRBP and W67L/W91H rRBP variant

Refinement	**rRBP**	**W67L/W91H rRBP variant**
Resolution range (Å)	40.0 - 1.7	40.0 – 2.0
Reflections	27458	18877
No of protein atoms	1407	1398
No solvent molecules	205	218
glycerol molecules	6	5
$^{\star}R_{free}$ *(%)*	27.0	22.9
$^{\ddagger}R_{conv}$ *(%)*	24.8	20.5
Model quality		
Rms deviations in:		
**bonds (Å)*	0.006	0.006
**angles (°)*	1.468	1.379
Temperature factors (Å^2):		
overall	26.5	22.6
average main-chain	24.2	20.1
average side-chain	24.8	21.6

$^{\star}R_{free}=\Sigma_{hkl\subset T}||F_{obs}-k|F_{calc}||/\Sigma_{hkl\subset T}|F_{obs}|$ x100, where hkl⊂T represents the test set (5-10% of the diffraction data). $^{\ddagger}R_{conv}=\Sigma_{hkl}||F_{obs}-k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$ x100. *Deviations from ideal values (Engh and Huber, 1991)

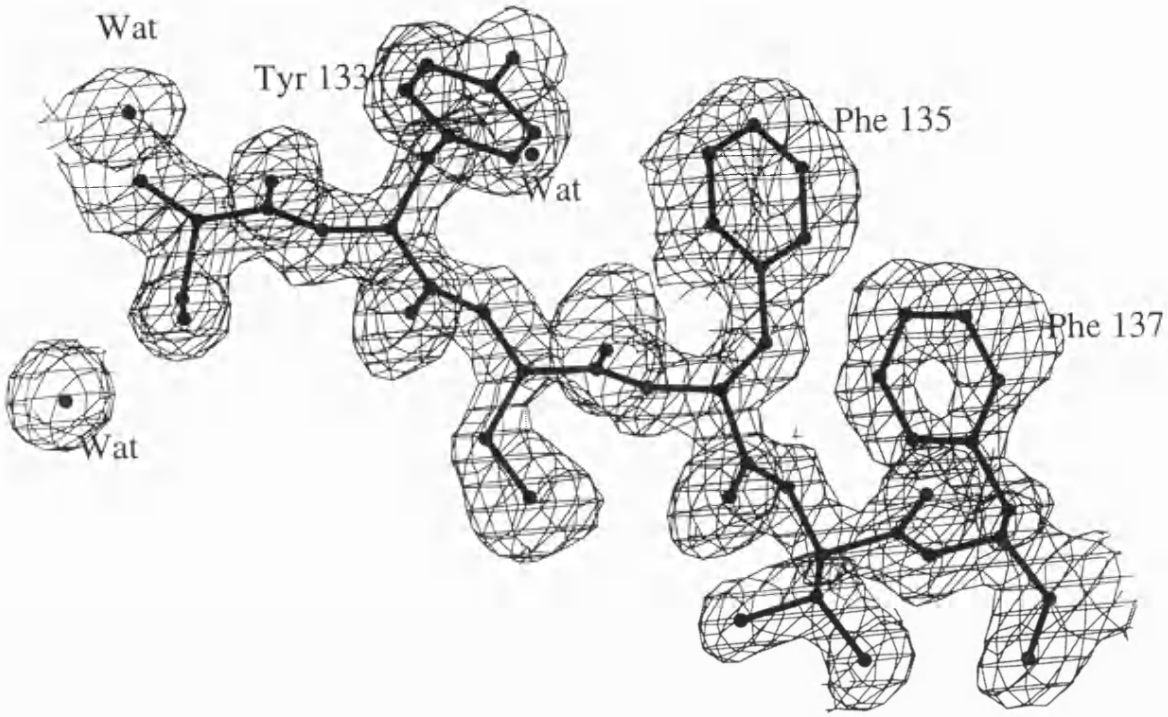


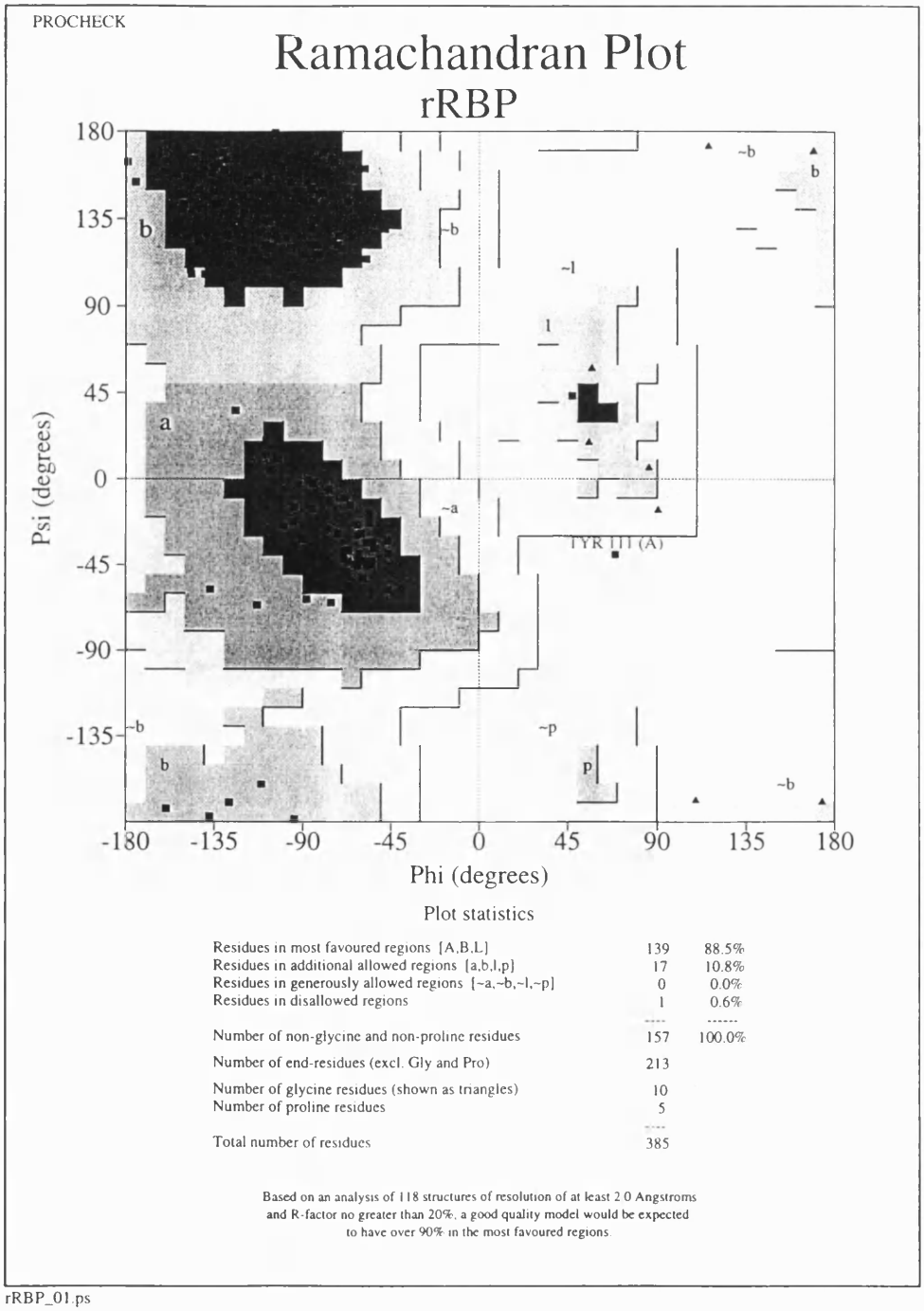
***Figure* 3.7**

Portion of the 2(Fo-Fc) electron density map from rRBP structure at 1.7 Å resolution. The figure was generated using the program BOBSCRIPT (Esnouf, 1997)

3.2.1.6. Structure analysis

The refined crystal structure of rRBP was compared with representative members of the Retinol-binding protein family by superposition with program *MAPS* (Lu, 1998) (Tables 3.12and 3.13). The accessibility of the protein surface was calculated with program *DSSP* (Kabsch and Sanders, 1983) and the program *PROCHECK* (Laskowski *et al.*, 1993) was used to assess the quality of the final structure. Analysis of the Ramachandran (φ - ψ) plot showed that all

residues lie in allowed regions except Tyr 111 in both rRBP and W67L/W91H rRBP variant structures for which the φ and ψ values are $\{\varphi = 69.1°, \psi = -39.8°\}$ and $\{\varphi = 64.6°, \psi = -36.7°\}$, respectively (*Figures* 3.8, 3.9).



rRBP_01.ps

***Figure* 3.8**

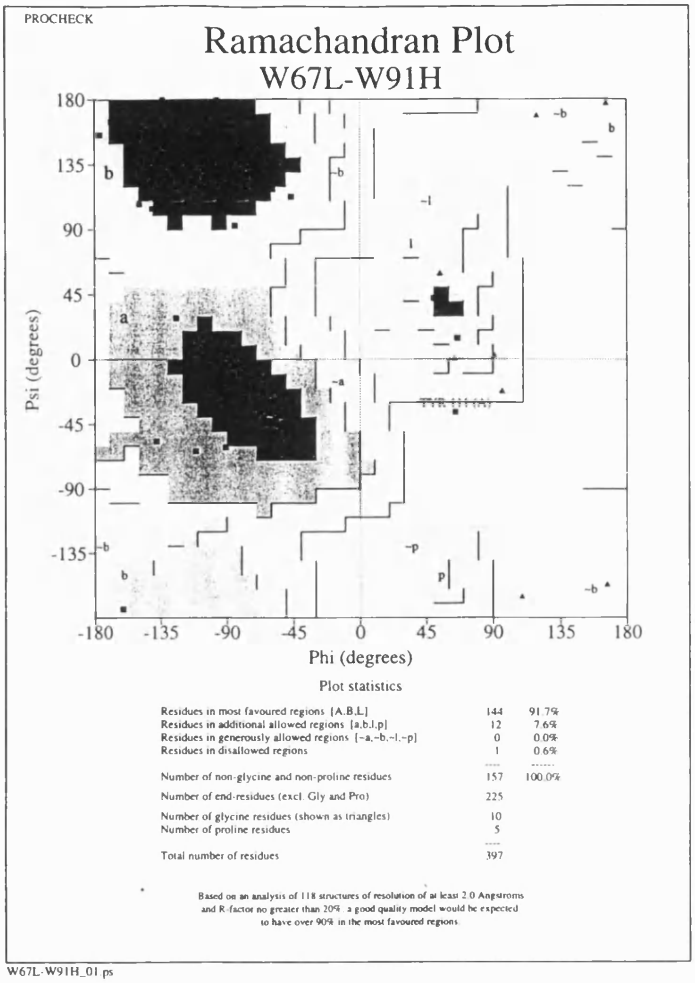


Figure **3.9**

3.2.1.7.. Crystallisation trials on two more rRBP variants: Trp24-Tyr and Trp105-Phe

Crystallisation trials on two more rRBP variants Trp-24-Tyr and Trp105-Phe were performed using the vapour diffusion method. The conditions that were initially tested were based on those that the native recombinant protein and the double mutant that had been crystallised. A

wide range of pH values from acidic to very alkaline was tested, and different salts like sodium chloride, lithium sulphate, cadmium acetate, were also used. Further attempts were also made by dissolving the protein in the reservoir solution with lower salt concentration but none of the above trials was successful. The seeding technique was also attempted with crystals from the double mutant of rRBP (W67L/W91H).

Due to limitations imposed from the available quantities of the two variants the crystal screen I containing a great variety of crystallisation conditions from Hampton Research (Jancarik and Kim, 1991) was not used. The conditions that were studied are summarised in Table 3.11. It can be observed that conditions that ended to turbid drops could be possible targets for optimisation by reducing either the protein concentration (~12mg/ml in the drop), or changing the salt concentration used.

Table 3.11 Summary of crystallisation trials for rRBP variants W24Y and W105F

Trp24-Tyr		
pH matrix	4.5M Sodium Chloride 20mM Sodium Cacodylate, pH 5.4, 6.1, 6.8	Heavy precipitation became lighter with increasing pH.
	4.5M Sodium Chloride 50mM Hepes, pH 7.2	
	4.5M Sodium Chloride 50mM Tris/HCl, pH: 8.1, 9.1	
Addition of detergent	4.5M Sodium Chloride 50mM Tris/HCl, pH 8.6 0.05% β-octyl-glucoside	Turbid
Seeding	4.5M Sodium Chloride 50mM Tris/HCl, pH 8.6 0.05% β-octyl-glucoside seeding from rRBP double variant crystals	Turbid
Trp105-Phe		
pH matrix	4.5M Sodium Chloride 20mM Sodium Cacodylate, pH: 5.4, 6.1, 6.8	Clear drops
	4.5M Sodium Chloride 50mM Hepes, pH 7.2	
	4.5M Sodium Chloride 50mM Tris/HCl, pH: 8.1, 9.1	
More alkaline pH ***(protein dissolved in 3.9M Sodium Chloride)***	4.5M Sodium Chloride 54mM Tris/HCl, pH 9.4	Clear drops
	4.5M Sodium Chloride 50mM Tris/HCl, pH 9.9	
Addition of detergent	4.5M Sodium Chloride 50mM Tris/HCl, pH 8.6 0.05% β-octyl-glucoside	Clear drops
Seeding	4.5M Sodium Chloride 50mM Tris/HCl, pH 8.6 0.05% β-octyl-glucoside seeding from rRBP double variant crystals	Clear drops
Change of precipitant	1.5M Lithium Sulphate 0.1M Hepes, pH 7.5	Clear drops
	1.7, 1.9, 2.1M Lithium Sulphate 0.1 M Hepes, pH 7.5	Drops become more turbid with increasing concentration of Lithium Sulphate
	6mM Cadmium Sulphate 16% 2,4 dimethyl-pentanediol 0.2 M Tris acetate pH 6.8	Clear drops
	6mM Magnesium Acetate 16% 2.4 dimethyl-pentanediol 0.2 M Tris acetate pH 6.8	Clear drops
	2.0M Lithium Sulphate 0.1M Hepes pH 8.8	Turbid
	2.5M Lithium Sulphate 0.1M Hepes, pH 7.5	Turbid

3.2.2. RESULTS AND DISCUSSION

3.2.2.1. Overall structure

Here we report the high-resolution structures of rRBP and a double variant W67L/W91H at 1.7 and 2.0 Å resolution, respectively. In the case of rRBP even from the first cycles of refinement, the high quality of the data was reflected in the electron density map especially around residues that form the β-barrel (*Figure* 3.7). However, the loop region 62-68 was highly disordered hence the incorporation of the initially excluded residues as alanines was possible only after applying the automated refinement procedure (Lamzin and Wilson, 1993). Further refinement considerably improved the phases and mutated residues were replaced according to the sequence apart from those from one flexible loop (residues 62-68) that were introduced to the model only towards the final stages of refinement. The density around residue 67, located in a critical region in the structure was poor even after the last cycle of refinement. Although the complete amino acid sequence of RBP comprises 182 amino acids (Rask *et al.*, 1979), sufficiently clear electron density was visible only up to residue 174 out of 182 residues. As it has been indicated by previous crystallographic studies on RBP the C-terminal of the protein as well as the loop region mentioned above seem to have marked flexibility (Cowan *et al.*, 1990; Zanotti *et al.*, 1993a).

Visual inspection of the rRBP electron density map revealed the presence of small but continuous piece of density in the core of the β-barrel and on the surface of the molecule that could not be accounted for by the addition of water molecules. Simulated annealing omit maps

***Figure* 3.10**

Worm representation of the rRBP (a) and double variant structures (b) showing the location of glycerol molecules bound to the protein.

indicated the presence of six glycerol molecules bound to the protein (which was used as cryo-protectant during data collection), *Figure* 3.10 (a). The final model consists of 205 water molecules. This is considerably higher in comparison with previously reported RBP structures due to better resolution (1.7 Å).

The structure of the double variant (W67L/W91H), determined using the rRBP structure as starting model, had similar features as with the native recombinant protein. The most ordered region of the molecule was formed by residues surrounding the β-barrel while the flexible 62-68 loop, resulted with moderate electron density (better compared to the rRBP structure). Glycerol was used during data collection as cryoprotectant and five molecules were identified in the structure either bound in the core of the barrel or on the surface of the molecule, *Figure* 3.10 (b). Three of these are located in identical positions for both structures of rRBP and variant, in the interior of the β-barrel the corresponding binding site of retinol molecule in the holo-form of RBP. The final model for this variant consists of 218 water molecules.

The overall structures of the two proteins are very similar (r.m.s. deviation 0.32 Å) apart from the flexible loop 62-68 and minor conformational changes of some surface residues. Two loop regions around residues 65 and 93 are involved in crystal packing contacts and the packing arrangement for rRBP is shown in *Figure* 3.11.

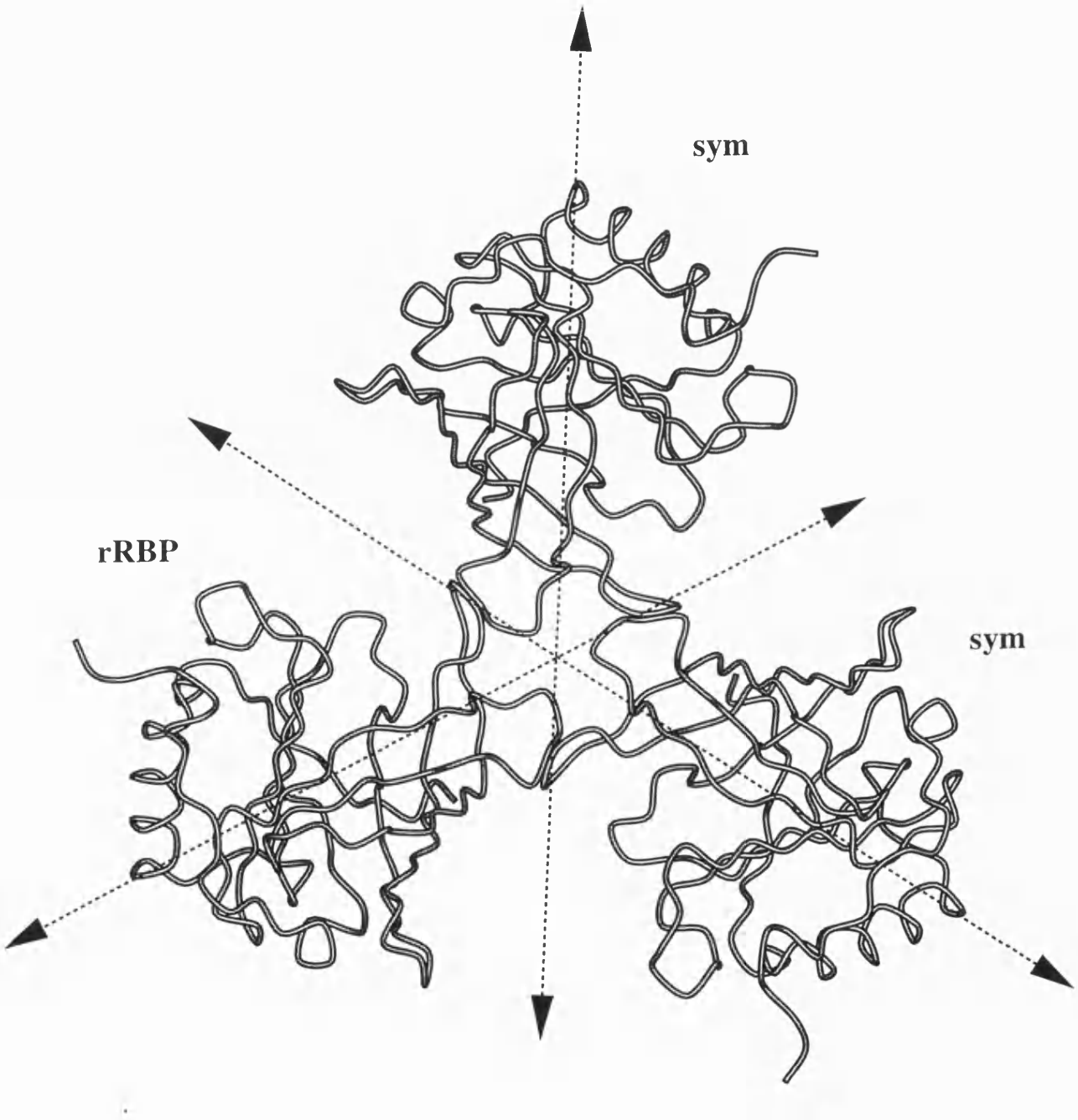


Figure **3.11**

Packing arrangement of rRBP molecules in the crystallographic asymmetric unit.

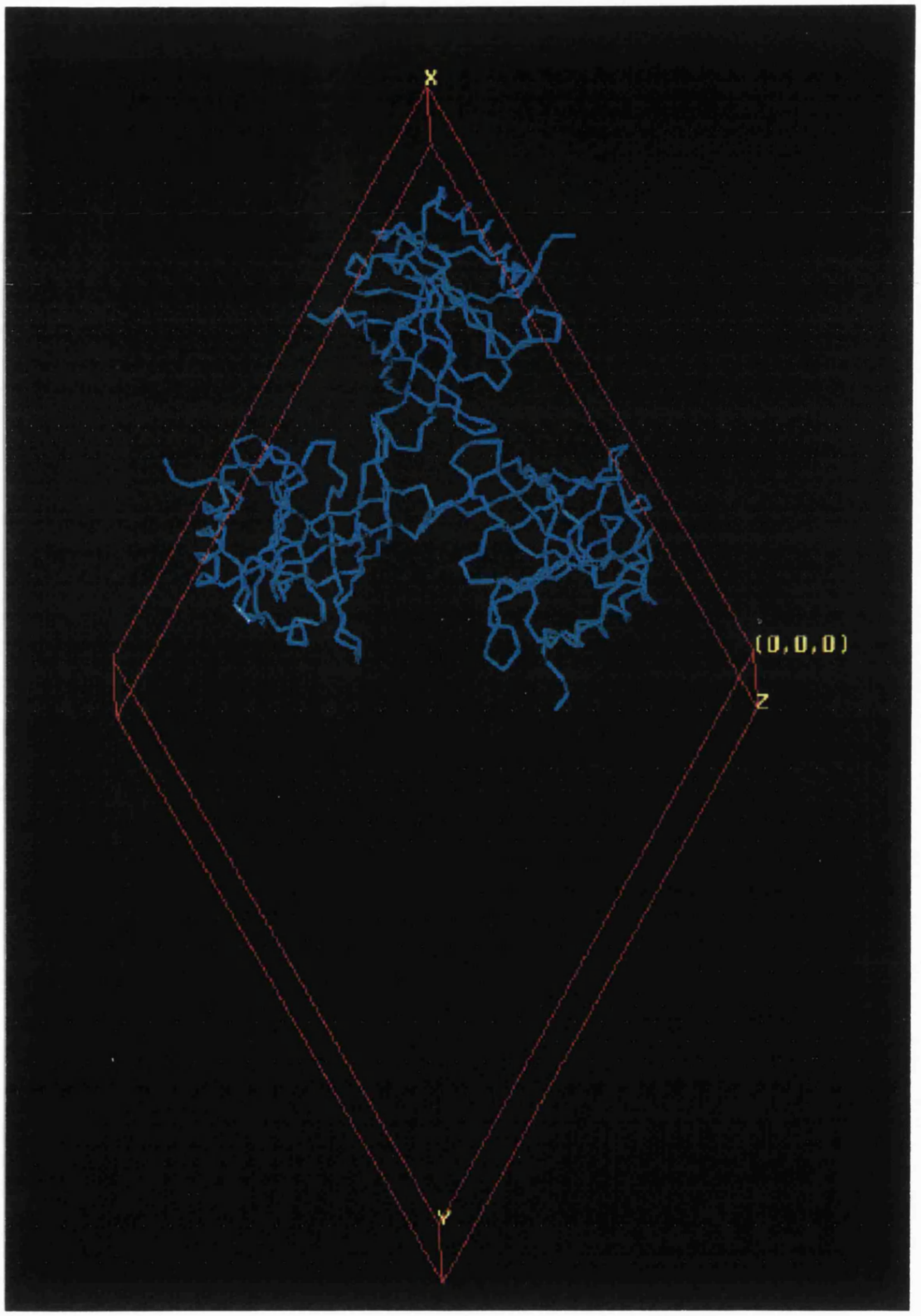


***Figure* 3.11a**

Packing arrangement of rRBP molecules in the crystallographic asymmetric unit.

rRBP and W67L/W91H structures were superimposed with other known homologous structures (members of the superfamily of lipocalins) and the results are shown in Table 3.12. As expected, there is good agreement with other known RBP structures determined previously by Zanotti *et al.*, (1993a) and Cowan *et al.*, (1990). The details of the pdb codes for the structures compared are given in Table 3.13. These comparisons show that all the known structures of this family have retained the 'β-barrel' and the conserved tryptophan residues. However, beyond this structural conservation there appears to be very little amino acid sequence conservation (see *Figure* 3.1).

Note for **Table 3.12. Structural superposition of known RBP structures and its homologues**: *The rms deviations were calculated after superposition using the program MAPS (Lu, 1998)) applied to Cα atoms of residues 1-174 (values in bold face). The number of equivalences/identities is also described (values in normal type face). The total number of residues included in the amino acid sequence of each structure is given (numbers in italics). (The pdb codes for the structures used for the superposition are explained in Table 3.13).*

Table 3.12 Structural superposition of known RBP structures and its homologues

	rrbp	rdm	1bbp	1beb	1brp	1brq	1epa	1epb	1erb	1fel	1fem	1fen	1hbp	1hbq	1mup	1obp	1pbo	1rbp	1aqb	1a3y	*Res*
rrbp	—	**0.32**	**1.53**	**1.40**	**0.50**	**0.53**	**1.56**	**1.66**	**0.64**	**0.69**	**0.65**	**0.67**	**0.64**	**0.66**	**1.68**	**1.44**	**1.46**	**0.59**	**0.62**	**1.53**	*174*
rdm	173/ 171	—	**1.53**	**1.40**	**0.53**	**0.55**	**1.56**	**1.65**	**0.59**	**1.55**	**0.64**	**0.64**	**0.68**	**0.68**	**1.67**	**1.42**	**1.41**	**0.49**	**0.62**	**1.58**	*174*
1bbp	99/ 20	100/ 21	—	**1.69**	**1.62**	**1.68**	**1.47**	**1.44**	**1.68**	**1.46**	**1.69**	**1.64**	**1.52**	**1.67**	**1.67**	**1.65**	**1.67**	**1.58**	**1.49**	**1.57**	*173*
1beb	98/ 23	97/ 23	79/ 15	—	**1.52**	**1.45**	**1.23**	**1.29**	**1.55**	**1.47**	**1.49**	**1.56**	**1.55**	**1.51**	**1.43**	**1.17**	**1.26**	**1.38**	**1.47**	**1.27**	*156*
1brp	166/ 164	166/ 162	99/ 19	99/ 22	—	**0.27**	**1.52**	**1.64**	**0.48**	**0.56**	**0.49**	**0.48**	**0.47**	**0.49**	**1.70**	**1.62**	**1.39**	**0.39**	**0.50**	**1.56**	*174*
1brq	169/ 167	170/ 166	103/ 21	97/ 22	172/ 172	—	**1.56**	**1.65**	**0.51**	**0.59**	**0.53**	**0.53**	**0.52**	**0.51**	**1.77**	**1.50**	**1.45**	**0.45**	**0.53**	**1.54**	*174*
1epa	107/ 25	107/ 25	88/ 21	98/ 27	106/ 26	108/ 26	—	**0.54**	**1.55**	**1.48**	**1.55**	**1.55**	**1.56**	**1.54**	**1.34**	**1.42**	**1.47**	**1.53**	**1.49**	**1.38**	*160*
1epb	108/ 24	108/ 24	89/ 21	97/ 27	109/ 25	108/ 25	159/ 159	—	**1.64**	**1.58**	**1.63**	**1.64**	**1.66**	**1.64**	**1.43**	**1.57**	**1.61**	**1.61**	**1.53**	**1.44**	*160*
1erb	166/ 152	165/ 150	102/ 17	99/ 21	170/ 158	168/ 156	106/ 23	109/ 22	—	**0.23**	**0.26**	**0.19**	**0.15**	**0.18**	**1.77**	**1.44**	**1.61**	**0.36**	**0.36**	**1.66**	*173*
1fel	165/ 151	165/ 150	90/ 16	95/ 21	172/ 160	170/ 158	107/ 23	110/ 23	173/ 173	—	**0.31**	**0.24**	**0.24**	**0.23**	**1.78**	**1.96**	**1.47**	**0.46**	**0.37**	**1.67**	*174*
1fem	166/ 152	167/ 152	97/ 17	97/ 21	171/ 159	169/ 157	105/ 23	107/ 22	173/ 173	173/ 173	—	**0.21**	**0.24**	**0.23**	**1.73**	**1.51**	**1.45**	**1.41**	**0.41**	**1.64**	*176*
1fen	166/ 152	166/ 151	100/ 17	99/ 21	171/ 159	169/ 157	106/ 23	109/ 22	173/ 173	173/ 173	176/ 176	—	**0.18**	**0.14**	**1.78**	**1.51**	**1.60**	**0.38**	**0.38**	**1.70**	*176*
1hbp	166/ 152	167/ 151	93/ 16	98/ 21	171/ 159	171/ 159	107/ 23	110/ 22	173/ 173	174/ 174	174/ 174	174/ 174	—	**0.19**	**1.75**	**1.33**	**1.51**	**0.37**	**0.34**	**1.60**	*174*
1hbq	168/ 155	169/ 155	99/ 17	98/ 21	169/ 158	171/ 160	106/ 23	109/ 22	171/ 170	171/ 170	172/ 171	172/ 171	172/ 171	—	**1.76**	**1.47**	**1.57**	**0.40**	**0.41**	**1.65**	*176*
1mup	91/ 15	91/ 15	77/ 12	111/ 20	87/ 16	90/ 18	108/ 23	109/ 24	90/ 17	96/ 17	96/ 17	90/ 17	90/ 17	90/ 17	—	**1.22**	**1.21**	**1.75**	**1.72**	**1.26**	*157*
1obp	62/ 11	62/ 11	62/ 9	90/ 13	73/ 11	64/ 10	72/ 10	72/ 9	63/ 9	49/ 5	70/ 9	70/ 9	58/ 8	69/ 9	101/ 29	—	**0.56**	**1.53**	**1.48**	**1.03**	*158*
1pbo	56/ 10	55/ 10	62/ 9	88/ 13	63/ 11	58/ 10	73/ 9	78/ 11	70/ 9	57/ 8	68/ 9	69/ 9	66/ 9	68/ 9	99/ 28	142/ 141	—	**1.44**	**1.50**	**1.01**	*157*
1rbp	166/ 164	163/ 160	99/ 18	96/ 21	171/ 171	169/ 169	107/ 26	108/ 25	173/ 161	173/ 161	174/ 162	174/ 162	174/ 162	172/ 161	90/ 16	66/ 10	60/ 9	—	**0.43**	**1.59**	*174*
1aqb	165/ 152	167/ 152	99/ 20	99/ 22	171/ 160	168/ 157	102/ 24	101/ 23	173/ 166	170/ 163	174/ 167	173/ 166	173/ 166	171/ 163	96/ 17	67/ 9	67/ 10	172/ 161	—	**1.69**	*175*
1a3y	89/ 9	88/ 9	57/ 10	110/ 20	87/ 8	85/ 8	102/ 20	103/ 20	93/ 9	87/ 8	89/ 9	90/ 8	88/ 9	90/ 8	128/ 40	107/ 46	104/ 44	89/ 8	87/ 9	—	*149*

Table 3.13 PDB codes for the structures compared

code	Protein	Res. (Å)	Authors
1bbp	Bilin binding protein	2.0	**(Huber *et al.*, 1987)**
1beb	β-lactoglobulin	1.8	**(Brownlow *et al.*, 1997)**
1brp	Retinol transport. Human-RBP (Holo form)	2.5	**(Zanotti *et al.*, 1993c)**
1brq	Retinol transport. Human RBP (apo form)	2.5	**(Zanotti *et al.*, 1993c)**
1epa	Retinoic-acid binding protein	2.1	**(Newcomer, 1993)**
1epb	Retinoic-acid binding protein complexed with retinoic acid	2.2	**(Newcomer, 1993)**
1erb	Retinol transport. RBP complexed with N-ethyl retinamide	1.9	**(Zanotti *et al.*, 1993b)**
1fel	Transport protein. Retinol binding protein complexed with Fenretinide	1.8	**(Zanotti *et al.*, 1994)**
1fem	Retinol binding protein complexed with retinoic acid.	1.9	**(Zanotti *et al.*, 1994)**
1fen	Transport protein. Retinol binding protein complexed with Axerophthene	1.9	**(Zanotti *et al.*, 1994)**
1hbp	Retinol transport. Bovine serum (holo-form)	1.9	**(Zanotti *et al.*, 1993a)**
1hbq	Retinol transport. Bovine serum (apo-form)	1.7	**(Zanotti *et al.*, 1993a)**
1mup	Pheromone-binding. Major urinary protein complex with 2-(sec-butyl). Thiazoline Mouse (Mus Musculus)	2.4	**(Bocskei *et al.*, 1992)**
1obp	Odorant binding protein from bovine Nasal Mucosa	2.0	**(Tegoni *et al.*, 1996)**
1pbo	Complex of bovine odorant binding protein with a selenium containing odorant	2.2	**(Blanchet *et al.*, 1996)**
1rbp	Retinol transport. Retinol binding protein (holo form)	2.0	**(Cowan *et al.*, 1990)**
1aqb	Retinol binding protein from Pig plasma	1.65	**(Zanotti *et al.*, 1998)**
1a3y	Odorant binding protein from nasal mucosa of Pig.	2.25	**(Spinelli *et al.*, 1998)**

Most recently the structure of human RBP with its carrier protein Transthyretin (TTR) was determined by (Naylor and Newcomer, 1999) at 3.2 Å resolution. Superposition of the RBP structure from the complex with TTR on to the rRBP structure and variant indicates that no major conformational change can be detected apart from the flexible loop region 62-67 (r.m.s. deviation 1.05 Å). Due to limited low resolution structure of this complex only gross structural features could be compared with the high resolution structures of rRBP and the variant presented here. However, the complex structure confirms that Trp 67 of RBP is part of the interface with TTR. The most profound differences between the recombinant human serum RBP and RBP-TTR complex are those around residue 62 as well as the C-terminus, both regions appear to be involved in interactions with TTR in the complex. Superposition of RBP from the RBP-TTR complex (Naylor and Newcomer, 1999), RBP in the presence of retinol (Cowan *et al.*, 1990) and rRBP (present structure) is shown in *Figure* 3.12.

3.2.2.2. β-barrel structure

The hydrophobic β-barrel structure is very similar to those from previously reported RBP structures (*Figure* 3.12). The structural similarity of apo-β-lactoglobulin and RBP (Sawyer *et al.*, 1985) also suggests that retinol removal does not necessarily cause a collapse of the β-barrel and both the apo-form of RBP (Zanotti *et al.*, 1993a) and the structures of rRBP and W67L/W91H variant confirm that point.

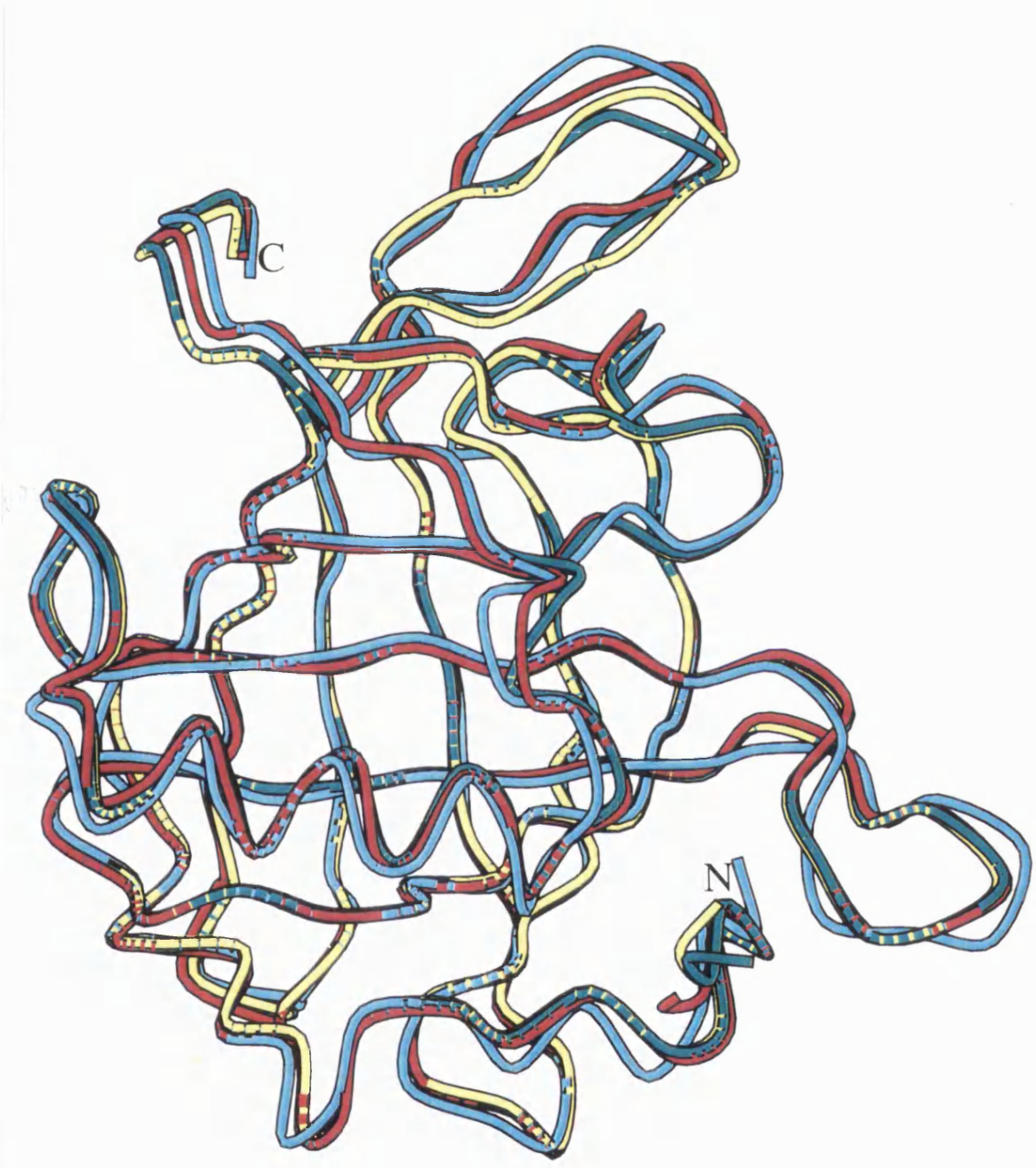


***Figure* 3.12.**

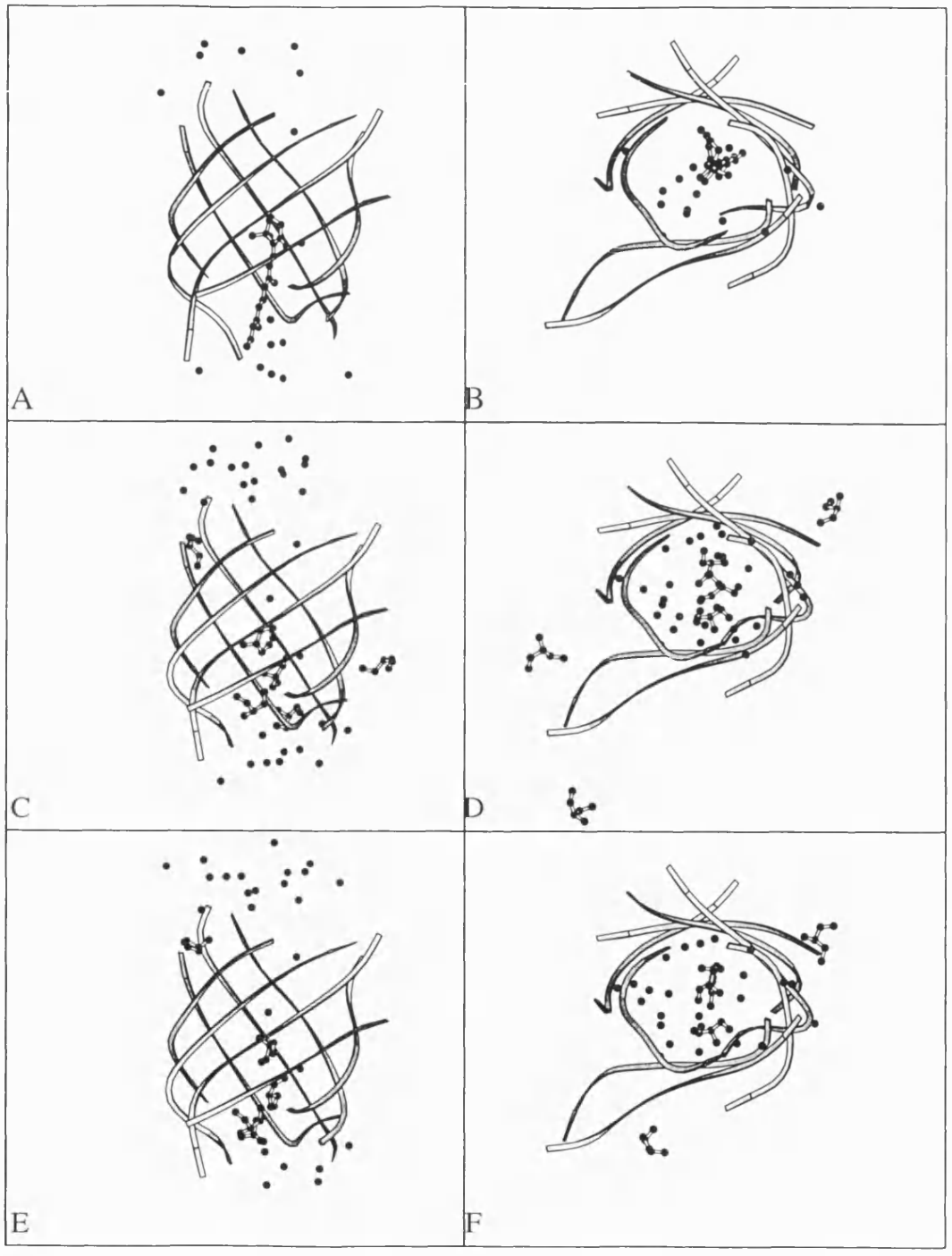
Superposition of the overall structure (only Cα atoms included) of rRBP, W67L/W91H rRBP double variant, human serum RBP with retinol bound (Cowan et al., 1990) and human serum RBP when in complex with TTR (Naylor and Newcomer, 1999), in yellow, green, red and cyan, respectively.

This barrel is known to accommodate retinol molecule as it has already been shown by Cowan *et al.* (1990) (*Figure* 3.13). In the holo-form of RBP, retinol is bound at the interior of the barrel with the β-ionone ring pointing at the bottom of the barrel and the isoprene tail stretching out almost to the surface of the protein. The apo-form of RBP isolated from the human plasma was determined at 2.5 Å resolution by Zanotti *et al.* (1993a) who observed that the internal cavity in the absence of retinol was not empty. Significant portions of (F_o-F_c) electron density were identified but due to the insufficient resolution of the X-ray data it was suggested that the unexplained density could be attributed to solvent molecules.

The high-resolution structure of rRBP at 1.7 Å resolution presented here clarifies the tentative conclusions of Zanotti and co-workers. Analysis of the β-barrel in the present rRBP structure revealed the presence of several water molecules and three glycerol molecules. These were also observed in the double variant structure (*Figure* 3.13). In both cases glycerol (25% v/v) was used as cryoprotectant during data collection under cryogenic conditions and a continuous piece of density was identified after a few cycles of refinement that resulted in significant improvement of the initial phases.

***Figure* 3.13** (overleaf)

Schematic picture of the β-barrel in two orthogonal views: (left)-from an axis parallel to the β-barrel axis pointing from the entrance to the bottom of the barrel (↑) and (right)-from an axis perpendicular to the β-barrel axis with a view in to the interior of the cavity from the entrance of the barrel. The distribution of glycerol and water molecules surrounding the barrel is also shown. (A) and (B) : holo-form of human serum RBP (Cowan et al., 1990); (C) and (D) : rRBP structure; (E) and (F) : rRBP double variant.


A
B
C
D
E
F

The glycerol molecules were incorporated in the model only during the final stages of refinement and are well positioned in the density (*Figure* 3.14) stabilised by hydrogen bonds and van der Waals interactions in the interior of the barrel.

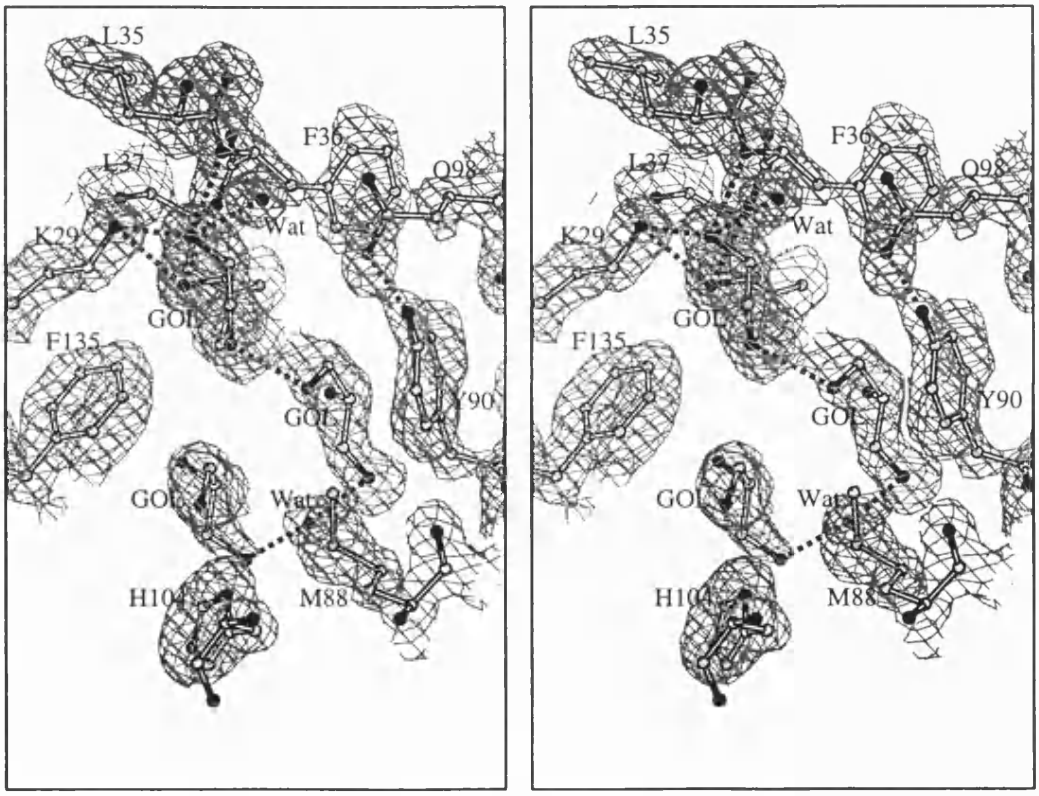


***Figure* 3.14**

Stereo diagram describing the (2Fo-Fc) electron density maps around glycerol molecules (GOL) in the core of the β-barrel. The hydrogen bond interactions are shown in dashes.

The hydrophobic nature of the β-barrel is evident from the presence of a large number of aromatic rings located in the core of the barrel. An important residue Phe36, which is located at the entrance of the β-barrel appears to block the opening of the barrel in the apo-RBP structure (Zanotti *et al.*, 1993a). However, this residue seems to adopt a significantly different conformation by opening the entrance of the barrel when retinol is bound (Cowan *et al.*, 1990; Naylor and Newcomer, 1999). Thus, the side chain of Phe 36 seems act like a 'toll-gate' for retinol binding at the entrance of the barrel. Somewhat similar scenario (as in the case of apo-RBP structure) was observed in the structures of rRBP and the double variant (*Figure* 3.15).

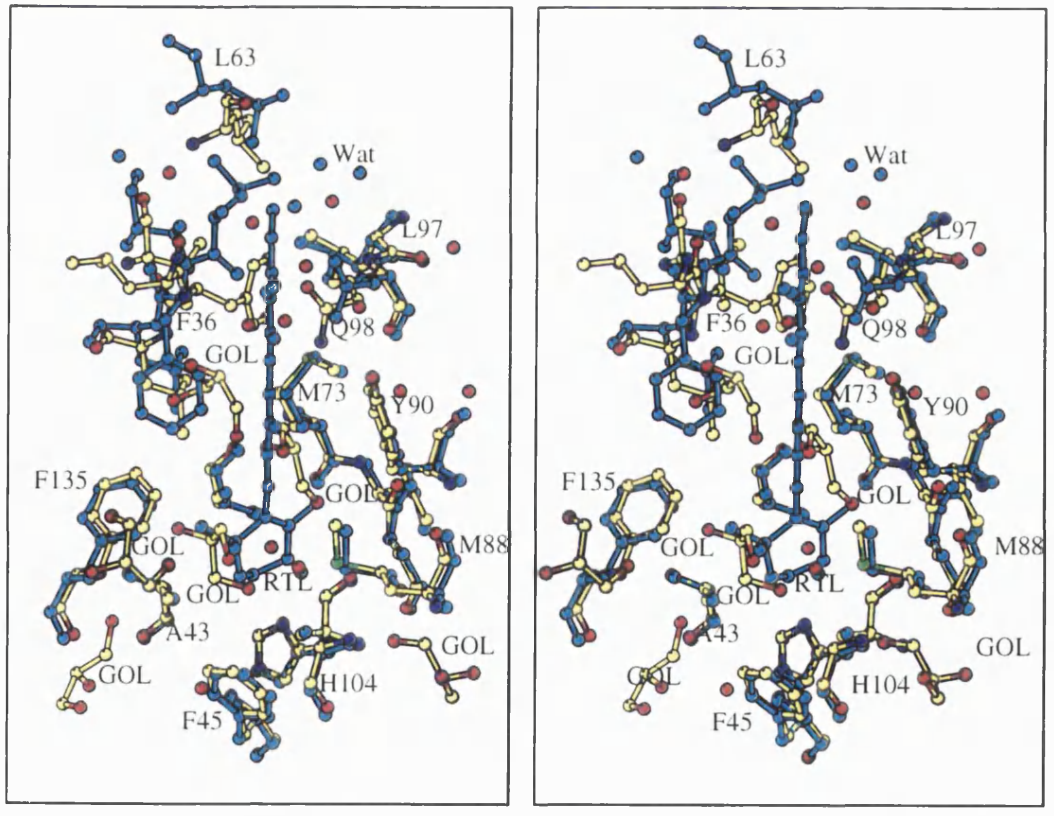


***Figure* 3.15**

Stereo diagram of the superimposed structures of rRBP and retinol bound form of RBP determined by Cowan et al. (1990) showing the conformational change of the side chain of Phe 36 as well as the binding of glycerol molecules in the place of retinol in the interior of the β-barrel

In these structures Phe36 do not seem to prohibit the binding of small molecules such as glycerol in the barrel. Interestingly, the position of two glycerol molecules in the barrel in the rRBP and the variant structures appear to mimic the interactions made by the retinol molecule as defined in the retinol bound form of the structure.

The binding of glycerol molecules as observed in the rRBP and the variant structures seem to provide some clues towards the design of new ligands, which could occupy the 'hydrophobic barrel' of the protein. It is hoped that this knowledge can be used in the design of compounds that can mimic the action of ligands such as retinol.

3.2.2.3. Environment of Tryptophan residues

The environment of the tryptophan residues is highlighted in Tables 3.14-3.16 for rRBP and Tables 3.17-3.19 for the double variant. Out of four tryptophan residues, Trp 24, 91 and 105 are ordered and well defined in the electron density map in the case of rRBP while Trp 24 and 105, which were not substituted in the double variant, fit nicely in the electron density map.

However, Trp67 is disordered in rRBP structure, due to the high flexibility of the loop region 62-68. This loop is involved in crystal packing interactions at the interface of the RBP-TTR complex as shown by Monaco *et al.* (1995) and Naylor and Newcomer (1999) who crystallised the complex comprising both proteins from human source. Both these complex structures indicate that the role of Trp 67 and Trp 91 is rather critical. More detailed examination concerning these particular residues was not possible for any of the above structures

Table 3.14. Hydrogen-bond interactions of Tryptophan residues in rRBP

	Trp24			Trp 91			Trp 105		
atom	***H-bonds distance < 3.3 Å***	***Distance (Å)***	Angle (°)	***H-bonds distance < 3.3 Å***	***Distance (Å)***	Angle (°)	***H-bonds distance < 3.3 Å***	***Distance (Å)***	Angle (°)
N	Ala43 O	2.84	154.9	Val74 O	2.95	157.4	Tyr118 O	2.71	165.5
Nε1	Phe20 O	2.76	145.2	Wat	3.17		Wat	2.87	
O	Ala43 N	2.87	162.3	Val74 N	2.73	160.3	Tyr118 N	2.82	142.0
				Val74 O	3.20	112.9			

Note : Trp 67 is disordered in the present structure of rRBP

Table 3.15 Environment of Tryptophan residues in rRBP

Residue	*van der Waals interactions/no of contacts*	*Accessible protein surface, (Å²)*	*Temperature factor (Å²)*	*Comments*
Trp24	Phe20/15, Thr23/13, Tyr25/12, Ala43/1, Phe45/1, Thr109/1, Tyr111/6, Thr113/1, Tyr114/6, Ala115/5, Phe137/1, Ser138/3, Arg139/16	0.0	19.5	well defined in the electron density map
Trp67	—	—	—	disordered
Trp91	Met73/3, Val74/9, Thr76/2, Lys89/5, Tyr90/18, Gly92/14, Lys99/2	77.0	28.3	well defined in the electron density map
Trp105	Val6/5, Phe9/1, Val11/1, Ala84/2, Lys85/18, His104/17, Ile106/11, Val107/3, Gly117/4, Tyr118/5, Wat/6	21.0	18.6	well defined in the electron density map

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N:3.4 Å. Accessible protein surface per atom for each Tryptophan residue was calculated using DSSP (Kabsch and Sanders, 1983).

Table 3.16 Details of the environment for the three Tryptophan residues in rRBP

	Trp24		Trp 91		Trp 105	
atom	*van der Waals interactions/ no of contacts*	*Accessible protein surface ($Å^2$)*	*van der Waals interactions/ no of contacts*	*Accessible protein surface, ($Å^2$)*	*van der Waals interactions/ no of contacts*	*Accessible protein surface, ($Å^2$)*
N	Thr23/5, Tyr25/1	0.0	Tyr90/7, Gly92/1	0.0	His104/6, Ile106/1, Tyr118/1	0.0
Cα	Thr23/3, Tyr25/1	0.0	Tyr90/3, Gly92/2, Lys99/1	0.0	Ala84/1, His104/3, Ile106/2, Tyr118/1	0.0
Cβ	Thr23/1, Tyr25/1, Phe137/3	0.0	Tyr90/2, Gly92/1, Lys99/1	3.2	His104/1, Ile106/1	0.0
Cγ	Thr23/2	0.0	Tyr90/1, Gly92/1	2.9	Lys85/1, His104/1	0.0
Cδ2	Phe137/1, Arg139/1	0.0	Tyr90/1	2.8	Lys85/3	0.1
Cε2	Phe20/3, Arg139/4	0.0	Val74/2	4.5	Lys85/3, Wat/1	0.7
Cε3	Ala115/1, Phe137/2, Ser138/2, Arg139/1	0.0	Lys89/3, Tyr90/2,	2.3	Val11/1, Lys85/1, Tyr118/1	0.0
Cδ1	Phe20/3, Thr23/1, Phe45/1	0.0	Val74/1, Gly92/3	13.8	Lys85/2, His104/2, Wat/1	0.0
Nε1	Phe20/7, Arg139/1	0.0	Val74/2	11.3	Lys85/3	3.6
Cζ2	Phe20/2, Tyr111/3, Ala115/1, Arg139/4	0.0	Val74/2	21.5	Val6/1, Lys85/3, Wat/1	17.0
Cζ3	Thr113/1, Tyr114/4, Ala115/2, Arg139/2	0.0	Thr76/1, Lys89/2	1.1	Val6/2, Phe9/1, Lys85/1, Wat/1	0.0
Cη2	Thr109/1, Tyr111/3, Tyr114/2, Ala115/1, Arg139/3	0.0	Val74/1, Thr76/1	11.4	Val6/2, Lys85/2, Wat/2	2.6
C	Thr23/1, Tyr25/6, Phe137/1, Ser138/1	0.0	Met73/1, Tyr90/1, Gly92/3	0.0	Ala84/1, His104/4, Ile106/5, Val107/1, Tyr 118/2	0.0
O	Tyr25/3, Ala43/1,	0.0	Met73/2,Val74/1, Tyr90/1 Gly92/3	0.0	Ile106/2, Val107/2, Gly117/4	0.0

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N: 3.4 Å. Accessible protein surface per atom of Tryptophan residue was calculated using X-PLOR (Brünger, 1992).

Table 3.17 Hydrogen-bond interactions of Trp residues in W67L/W91H variant structure

	Trp24			Trp 105		
atom	*H-bonds distance < 3.3 Å*	*Distance (Å)*	Angle (°)	*H-bonds distance < 3.3 Å*	*Distance (Å)*	Angle (°)
N	Ala43-O	2.8	159.5	Tyr 118-O	2.8	163.6
Nε1	Phe20-O	2.9	145.9	Wat-O	2.9	
O	Ala43-N	2.9	171.1	Tyr 118-N	2.9	146.2

Table 3.18 Environment of Tryptophan residues in W67L/W91H variant structure

Residue	*van der Waals interactions/ no of contacts*	*Accessible protein surface, (Å²)*	*Temperature factors (Å²)*	*Comments*
Trp24	Phe20/10 Thr23/13, Tyr25/9 Ala43/1, Phe45/1, Tyr111/4, Thr113/1, Tyr114/4, Ala115/2, Phe137/7, Ser138/4, Arg139/19	0.0	13.1	well defined in the electron density map
Trp105	Val6/2, Val11/1, Ala84/2, Lys85/20, His104/13, Ile106/12, Val107/1, Gly117/2, Tyr118/3, Wat/9	24.0	12.8	well defined in the electron density map

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N:3.4 Å. Accessible protein surface per atom for the Tryptophan residue was calculated using DSSP (Kabsch and Sanders, 1983).

Table 3.19 Details of the environment of the Trp residues in W67L/W91H variant structure

	Trp24		Trp 105	
atom	*van der Waals interactions/ no of contacts*	*Accessible protein surface ($Å^2$)*	*van der Waals interactions/ no of contacts*	*Accessible protein surface, ($Å^2$)*
N	Thr23/5	0.0	His104/5	0.0
Cα	Thr23/3, Tyr25/2	0.0	Ala84/1, His104/3, Ile106/2, Tyr118/1	0.0
Cβ	Thr23/1, Tyr25/1, Phe137/3	0.0	His104/1, Ile106/1, Tyr118/1	0.0
Cγ	Thr23/2	0.0	Lys85/1, His104/1	0.0
Cδ2	Phe137/1, Arg139/1	0.0	Lys85/3	0.1
Cε2	Phe20/2, Arg139/4	0.0	Lys85/3, Wat/1	1.2
Cε3	Phe137/2, Ser138/3, Arg139/3	0.0	Val11/1, Lys85/1, Tyr118/1	0.0
Cδ1	Phe20/2, Thr23/1, Phe45/1	0.0	Lys85/3, His104/2, Wat/1	0.0
Nε1	Phe20/4, Arg139/3	0.0	Lys85/3, Wat/2	3.9
Cζ2	Phe20/2, Tyr111/2, Arg139/4	0.0	Lys85/3, Wat/3	17.6
Cζ3	Thr113/1, Tyr114/2, Ala115/1, Ser 138/1, Arg139/2	0.0	Val6/1, Lys85/1, Wat/1	0.0
Cη2	Tyr111/2, Tyr114/2, Ala115/1, Arg139/2	0.0	Val6/1, Lys85/2, Wat/1	3.6
C	Thr23/1, Tyr25/5, Phe137/1	0.0	Ala84/1, His104/1, Ile106/6, Val107/1	0.0
O	Tyr25/1, Ala43/1,	0.0	Ile106/3, Gly117/2	0.0

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N: 3.4 Å. Accessible protein surface per atom of Tryptophan residue was calculated using X-PLOR (Brünger, 1992).

considering that the resolution of these complex structures was ~3.1 Å.

Thus, only gross structural differences could be identified in the W67L/W91H rRBP variant structure. The effect of sequence substitution at position 67 was not possible to investigate in detail due to poor electron density in this region, but the interactions of histidine at position 91 are listed in Tables 3.20-3.22. The five-membered ring of the tryptophan residue (native) is replaced by a water molecule in the mutant structure and the accessibility of histidine is slightly higher compared to that of tryptophan (*Figure* 3.16)

Table 3.20 Hydrogen-bond interactions of Histidine 91 in W67L/W91H variant structure

	His 91		
Atom	***H-bonds distance < 3.3 Å***	***Distance (Å)***	Angle (°)
N	Val74-O	3.1	159.2
Nδ1	Gly92-O	3.0	113.3
O	Val74-O	3.3	111.1
	Val74-N	2.8	149.1

Table 3.21 Environment of Histidine 91 in W67L/W91H variant structure

Residue	*van der Waals interactions/ no of contacts*	*Accessible protein surface, (Å²)*	*Temperature factors (Å²)*	*Comments*
His 91	Met73/4, Val74/3, Tyr90/11, Gly92/12, Lys99/2, Wat/3	78	23.5	well defined in the electron density map

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N:3.4 Å. Accessible protein surface per atom for the Histidine residue was calculated using DSSP (Kabsch and Sanders, 1983).

Table 3.22 Detailed Environment of Histidine 91 in W67L/W91H variant structure

	His 91	
atom	*van der Waals interactions/ no of contacts*	*Accessible protein surface (Å²)*
N	Val74/1, Tyr90/6, Gly92/1	0.0
Cα	Tyr90/3, Gly92/2, Lys99/1	0.0
Cβ	Tyr90/2, Gly92/1, Lys99/1	8.3
Cγ	Tyr90/1, Gly92/1	4.6
Cδ2	Tyr90/1, Wat/1	18.9
Nδ1	Gly92/3, Wat/1	1.6
Cε1	Val74/1, Wat/1	28.5
Nε2	—	13.4
C	Met73/1, Gly92/4	0.0
O	Met73/3, Val74/1	0.0

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N: 3.4 Å. Accessible protein surface per atom of Tryptophan residue was calculated using X-PLOR (Brünger, 1992).

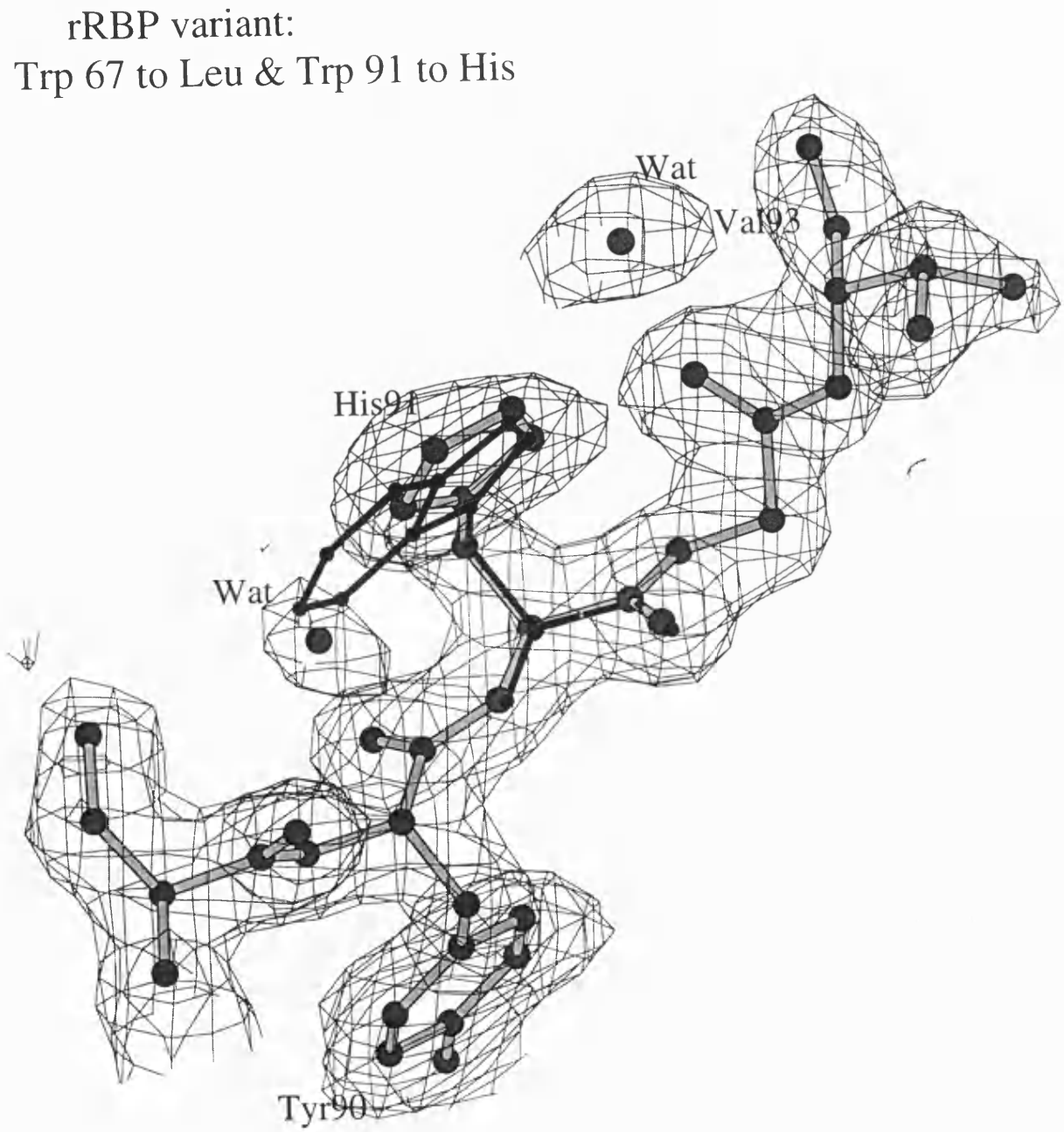


***Figure* 3.16**

Portion of (2Fo-Fc) electron density map at position 91. The figure was prepared using BOBSCRIPT (Esnouf, 1997)

Solvent accessibility calculations for the tryptophan residues in rRBP structure show that Trp24 is buried, Trp91 is partially exposed and Trp105 is exposed to the solvent. Trp24 makes H-bond interactions with Phe20 and Ala43, Trp91 forms H-bonds with Val74 and Trp105 is involved in H-bond interactions with Tyr118 (Table 3.14). Also, all three tryptophan residues make a large number of stacking (van der Waals) interactions as shown in Table 3.19. The interactions involving Trp 105 are presented in *Figure* 3.17.

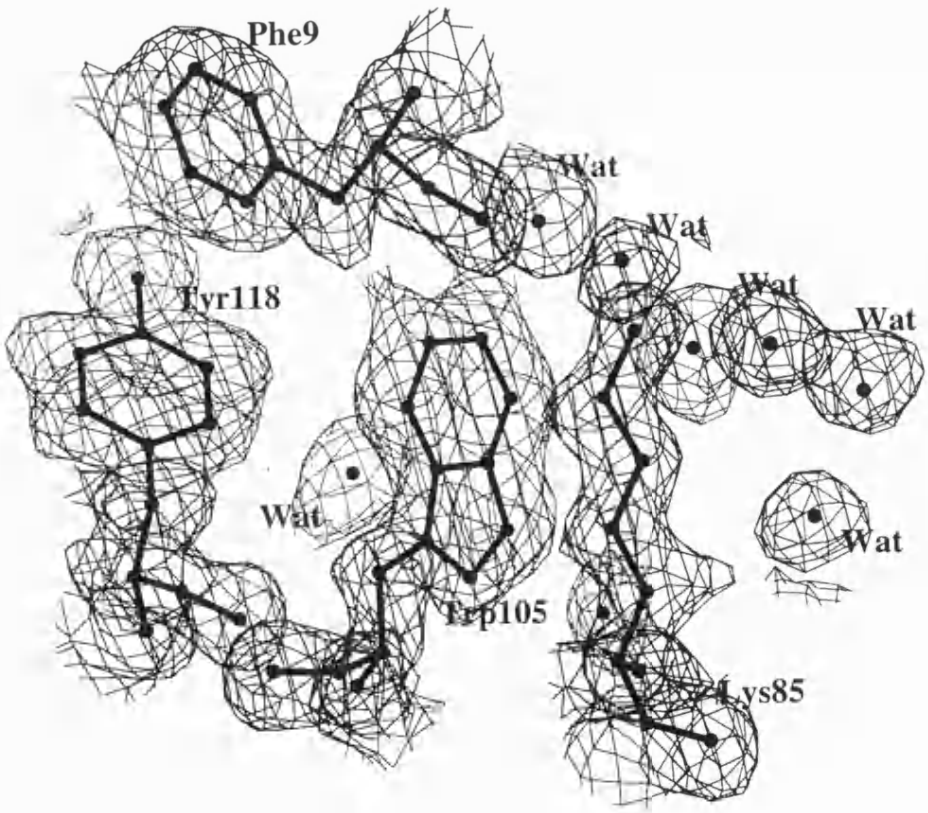


***Figure* 3.17**

Portion of the final 1.7 Å (2Fo-Fc) electron density map for rRBP around Trp105. The map is contoured at 1.0σ level. The stacking interaction between Trp105 and Lys85 is shown. The figure was prepared using the program BOBSCRIPT (Esnouf, 1997).

No significant differences are observed in the double variant structure for tryptophans 24 and 105. Trp 24 is completely buried in the core of the barrel and Trp 105 has become slightly more accessible by the solvent. The H-bond interactions are essentially the same as in the rRBP structure and only minor differences can be observed.

It is worth noting that tyrosine 111, the only residue that lies in a disallowed region in the Ramachandran plot, in both rRBP and the variant structures appears to be in the vicinity of Trp24 and is involved in non-polar interactions. No functional significance has been assigned to Tyr 111 residue. Detailed interactions of Tyr 111 in rRBP and W67L/W91H double variant structure are listed in Tables 3.23-3.28.

Table 3.23 Hydrogen-bond interactions of Tyrosine 111 in rRBP

atom	*H-bonds distance < 3.3 Å*	*Distance (Å)*	*Angle (°)*
N	Thr109 Oγ1	2.9	122.7
	Asp110 Oδ1	3.7	108.5
OH	Asp16 N	2.8	152.9
	Wat O	2.7	0.0
O	Arg139 Nε	2.7	153.7
	Arg39 Nη2	2.9	138.3

Table 3.24 Environment of Tyrosine 111 in rRBP

Residue	*van der Waals interactions/ no of contacts*	*Accessible protein surface, (Å²)*	*Temperature factors (Å²)*	*Comments*
Tyr 111	Lys12/3, Phe15/4, Asp16/5, Arg19/12, Phe20/2, Trp24/6, Thr109/5, Asp110/17, Asp112/11,Thr113/2, Arg139/1, Wat/2	12.0	19.9	well defined in the electron density map

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N:3.4 Å. Accessible protein surface per atom for the Tyrosine residue was calculated using DSSP (Kabsch and Sanders, 1983).

Table 3.25 Details of the environment of Tyrosine 111 in rRBP

atom	*van der Waals interactions / no of contacts*	*Accessible protein surface (Å²)*
N	Thr109/1, Asp110/6, Asp112/1	0.0
Cα	Trp24/2, Thr109/1, Asp110/4, Asp112/2	0.0
Cβ	Arg19/1, Asp110/2, Asp112/1	4.4
Cγ	Arg19/3, Asp110/1	0.0
Cδ1	Arg19/2, Phe20/2, Thr109/1	0.0
Cε1	Phe15/2, Asp16/2, Arg19/1, Thr109/1	0.0
Cδ2	Arg19/3, Asp110/2	1.9
Cε2	Lys12/1, Arg19/1, Wat/1	7.2
Cζ	Lys12/1, Asp16/2, Thr109/1, Wat/1	0.0
OH	Lys12/1, Phe15/2, Asp16/1	1.0
C	Trp24/2, Asp110/2, Asp112/4, Thr113/1	0.4
O	Arg19/1, Trp24/2, Asp112/3, Thr113/1, Arg139/1	1.6

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N: 3.4 Å. Accessible protein surface per atom of Tyrosine residue was calculated using X-PLOR (Brünger, 1992).

Table 3.26 Hydrogen-bond interactions of Tyr 111 in W67L/W91H variant structure

atom	*H-bonds distance < 3.3 Å*	*Distance (Å)*	*Angle (°)*
N	Thr109 Oγ1	2.9	124.2
	Asp110 Oδ1	3.2	107.1
OH	Asp16 N	2.8	147.2
	Wat O	2.4	
O	Arg139 Nε	2.7	151.7
	Arg39 Nη2	2.9	139.6

Table 3.27 Environment of Tyrosine 111 in W67L/W91H variant structure

Residue	*van der Waals interactions/ no of contacts*	*Accessible protein surface, (Å²)*	*Temperature factors (Å²)*	*Comments*
Tyr 111	Lys12/1, Phe15/4, Asp16/2, Arg19/10, Phe20/3, Trp24/4, Thr109/3, Asp110/16, Asp112/10, Thr113/1, Arg139/1, Wat/2	12.0	14.9	well defined in the electron density map

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N:3.4 Å. Accessible protein surface per atom for the Tyrosine residue was calculated using DSSP (Kabsch and Sanders, 1983).

Table 3.28 Details of the environment of Tyr 111 in W67L/W91H variant structure

atom	*van der Waals interactions / no of contacts*	*Accessible protein surface (Å²)*
N	Thr109/1, Asp110/6, Asp112/1	0.0
Cα	Thr109/1, Asp110/4, Asp112/2	0.0
Cβ	Arg19/1, Asp110/2, Asp112/1	4.0
Cγ	Arg19/2, Asp110/1	0.0
Cδ1	Arg19/2, Phe20/3	0.0
Cε1	Phe15/2, Asp16/4, Arg19/1, Thr109/1	0.0
Cδ2	Arg19/4, Asp110/1	2.2
Cε2	Lys12/1, Arg19/1, Wat/1	7.0
Cζ	Asp16/2, Wat/1	0.0
OH	Phe15/2, Asp16/1	1.5
C	Trp24/2, Asp110/2, Asp112/4, Thr113/1	0.3
O	Trp24/2, Asp112/2, Arg139/1	1.9

Note : van der Waals interactions were calculated using CCP4 (CCP4, 1994). van der Waals distances are the maximum allowed values of C-C : 4.1 Å, C-N : 3.8 Å, C-O: 3.7 Å, O-O: 3.3 Å, O-N: 3.4Å and N-N: 3.4 Å. Accessible protein surface per atom of Tyrosine residue was calculated using X-PLOR (Brünger, 1992).

The crystal structures of rRBP and its double variant reveal the sequence substitutions at position 67 and 91 in rRBP molecule do not seem to disrupt the overall native structure of the protein. These results are in agreement with those found by near- and far-UV circular dichroism (Greene, 1998).

The high mobility the loop around residue 67 exhibits, prevents

the more detailed analysis of that critical region for the formation of the RBP-TTR complex. However, it seems that the loop 62-67 adopts a different conformation in the variant structure (*Figure* 3.18).

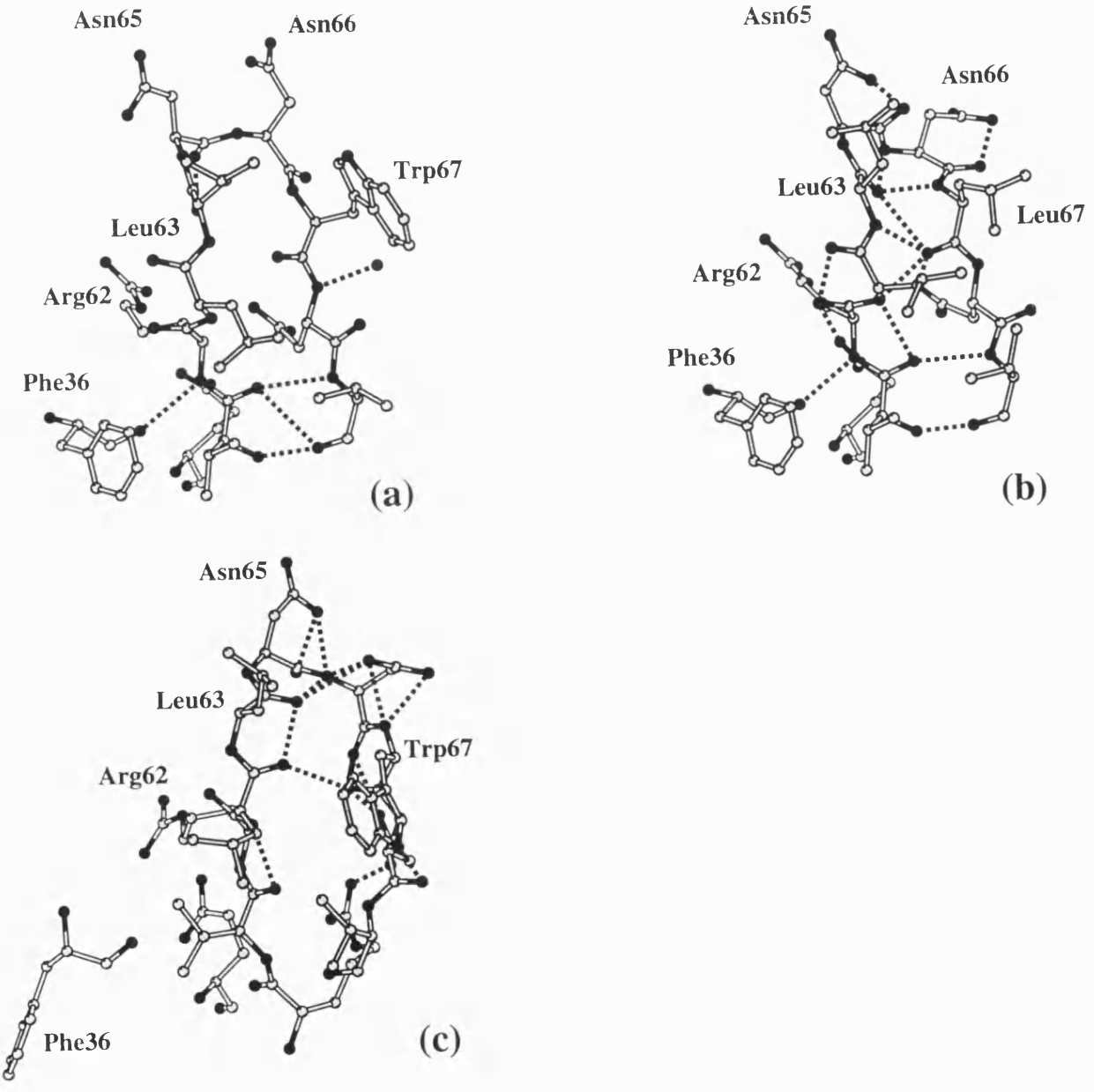


***Figure* 3.18**

Loop region around position 67 for the structures of (a) rRBP-1.7 Å resolution, (b) W67L/W91H-2.0 Å resolution and (c) RBP-TTR complex-3.1 Å resolution.

Further experimental work on RBP should include the optimisation of the crystallisation conditions of the two single variants Trp24-Tyr and Trp105-Phe with the aim to assess the role that conserved residues within the RBP family of molecules (either within the lipocalin superfamily or the RBP family only) serve in the RBP molecule by giving structural evidence that could support a relationship between conservation and the molten globule stability.

4

REFERENCES

4. REFERENCES

Acharya, K.R., Ren, J.S., Stuart, D.I., Phillips, D.C. and Fenna, R.E. (1991) Crystal -structure of human alpha-lactalbumin at 1.7-Å resolution. *J. Mol. Biol.*, **221**, 571-581.

Acharya, K.R., Stuart, D.I., Phillips, D.C., McKenzie, H.A. and Teahan, C.G. (1994) Models of the 3-dimensional structures of echidna, horse, and pigeon lysozymes - calcium-binding lysozymes and their relationship with alpha-lactalbumins. *J. Prot. Chem.*, **13**, 569-584.

Acharya, K.R., Stuart, D.I., Walker, N.P.C., Lewis, M. and Phillips, D.C. (1989) Refined structure of baboon alpha-lactalbumin at 1.7-Å resolution-comparison with c-type lysozyme. *J. Mol. Biol.*, **208**, 99-127.

Adams, P.D., Pannu, N.S., Read, R.J. and Brunger, A.T. (1997) Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl. Acad. Sci.(USA)*, **94**, 5018-5023.

Alexandrescu, A.T., Broadhurst, R.W., Wormald, C., Chyan, C.L., Baum, J. and Dobson, C.M. (1992) H-1-NMR assignments and local environments of aromatic residues in bovine, human and guinea-pig variants of alpha-lactalbumin. *Eur. J. Biochem.*, **210**, 699-709.

Alexandrescu, A.T., Evans, P.A., Pitkeathly, M., Baum, J. and Dobson, C.M. (1993) Structure and dynamics of the acid-denatured molten globule state of alpha-Lactalbumin-a 2-dimensional NMR-study. *Biochemistry*, **32**, 1707-1718.

Anderson, P.J., Brooks, C.L. and Berliner, L.J. (1997) Functional identification of calcium binding residues in bovine alpha-lactalbumin. *Biochemistry*, **36**, 11648-11654.

Aschaffenburg, R. and Drewry, J. (1957) *Biochem. J.*, **65**, 273-277.

Baum, J., Dobson, C.M., Evans, P.A. and Hanley, C. (1989) Characterization of a partly folded protein By NMR methods - studies on the molten globule state of guinea-pig alpha-lactalbumin. *Biochemistry*, **28**, 7-13.

Bavik, C., Ward, S.J. and Ong, D.E. (1997) Identification of a mechanism to localize generation of retinoic acid in rat embryos. *Mechanisms of Development*, **69**, 155-167.

Bell, J.E., Bever, T.A. and Hill, R.L. (1976) The kinetic mechanism of bovine milk galactosyltransferase. *J. Biol. Chem.*, **251**, 3003-3013.

Berliner, L.J., Ellis, P.D. and Murakami, K. (1983) Manganese (II) electron-spin resonance and Cd-113 nuclear magnetic- resonance evidence for the nature of the calcium-binding site in alpha-lactalbumins. *Biochemistry*, **22**, 5061-5063.

Bhutta, Z.A., Bang, P., Karlsson, E., Hagenas, L., Nizami, S.Q. and Soder, O. (1999) Insulin-like growth factor I response during nutritional rehabilitation of persistent diarrhoea. *Archives of Disease in Childhood*, **80**, 438-442.

Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press.

Brew, K., Richardson, R.H. and Sinha, S.K. (1979) Structural basis of the regulation of galactosyltransferase. In Russell, T.R., Brew, K., Faber, H. and Schultz, J. (eds.), *In From Gene to Protein: Information transfer in normal and abnormal cells*. Academic Press, New York, Miami Winter Symposia, Vol. 16, pp. 433-447.

Brew, K., Shaper, J.H., Olsen, K.W., Trayer, I.P. and Hill, R.L. (1975) Cross-linking of the components of lactose synthetase with dimethylpimelimidate. *J. Biol. Chem.*, **250**, 1434-1444.

Brownlow, S., Cabral, J.H.M., Cooper, R., Flower, D.R., Yewdall, S.J., Polikarpov, I., North, A.C.T. and Sawyer, L. (1997) Bovine beta-lactoglobulin at 1.8 angstrom resolution - Still an enigmatic lipocalin. *Structure*, **5**, 481-495.

Brünger, A.T. (1992a) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.

Brünger, A.T. (1992b) *X-PLOR Version 3.1 Manual: A system for X-ray Crystallography & NMR*. Yale University Press, New Haven.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., GrosseKunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905-921.

Brünger, A.T. and Rice, L.M. (1997) Crystallographic refinement by simulated annealing: Methods and applications. *Methods in Enzymol.*, **277**, 243-269.

Calderone, V., Giuffrida, M.G., Viterbo, D., Napolitano, L., Fortunato, D., Conti, A. and Acharya, K.R. (1996) Amino acid sequence and crystal structure of buffalo alpha- lactalbumin. *FEBS lett.*, **394**, 91-95.

CCP4. (1994) The CCP4 suite : Programs for protein crystallography. *Acta Crystallogr.*, **D 50**, 760-763.

Chandra, N., Brew, K. and Acharya, K.R. (1998) Structural evidence for the presence of a secondary calcium binding site in human alpha-lactalbumin . *Biochemistry*, **37**, 4767-4772.

Chaudhuri, T.K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T.P., Uchiyama, H., Ikura, T., Tsumoto, K., Kataoka, H., Matsushima, M., Kuwajima, K. and Kumagai, I. (1999) Effect of the extra N-terminal methionine residue on the stability and folding of recombinant alpha-lactalbumin expressed in *Escherichia coli. J. Mol. Biol.*, **285**, 1179-1194.

Cogan, U., Kopelman, M., Mokady, S. and Shinitzky, M. (1976) Binding affinities of retinol and related compounds to retinol binding proteins. *Eur. J. Biochem.*, **65**, 71-78.

Cowan, S.W., Newcomer, M.E. and Jones, T.A. (1990) Crystallographic refinement of human serum retinol binding-protein at 2 Å Resolution. *Proteins*, **8**, 44-61.

Dauter, Z. (1997) Data collection strategy. *Methods in Enzymol.*, **276**, 326-344.

Dillon, S.C., Taylor, G.M. and Shah, V. (1998) Diagnostic value of urinary retinol-binding protein in childhood nephrotic syndrome. *Pediatric Nephrology*, **12**, 643-647.

Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray structure refinement. *Acta Crystallogr.*, **A47**, 100-119.

Ewbank, J.J. and Creighton, T.E. (1991) The molten globule protein conformation probed by disulfide bonds. *Nature*, **350**, 518-520.

Feher, G. (1986) Mechanisms of nucleation and growth of protein crystals. *Journal of Crystal Growth*, **76**, 545-546.

Fex, G., Albertsson, P.A. and Hansson, B. (1979) Interaction between prealbumin and retinol-binding protein studied by affinity chromatography, gel filtration and two-phase partition. *Eur. J. Biochem.*, **99**, 353-360.

Flower, D.R. (1996) The lipocalin protein family: structure and function. *Biochem. J.*, **318**, 1-14.

Forge, V., Wijesinha, R.T., Balbach, J., Brew, K., Robinson, C.V., Redfield, C. and Dobson, C.M. (1999) Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin. *J. Mol. Biol.*, **288**, 673-688.

Garman, E.F. and Schneider, T.R. (1997) Macromolecular cryocrystallography. *J. Appl. Cryst.*, **30**, 211-237.

Gastinel, L.N., Cambillau, C. and Bourne, Y. (1999) Crystal structures of the bovine beta 4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. *EMBO J.*, **18**, 3546-3557.

Gerken, T.A. (1984) Amino group environments and metal binding properties of carbon-13 reductively methylated bovine alpha-lactalbumin. *Biochemistry*, **23**, 4688-4697.

Greene, D., Ingram, V. and Perutz, M.F. (1954) *Proc. Roy. Soc.*, **A225**, 287.

Greene, L.H. (1998) Investigation into the relationship between sequence conservation, stability, and folding in a model lipocalin: human serum retinol-binding protein. . University of Miami, Coral Gables, Florida.

Greene, L.H., Grobler, J.A., Malinovskii, V.A., Tian, J., Acharya, K.R. and Brew, K. (1999) Stability, activity and flexibility in alpha-lactalbumin. *Protein Engineering*, **12**, 581-587.

Grobler, J.A. and Brew, K. (1992) *alpha-Lactalbumin*. Elsevier Applied Science, London.

Grobler, J.A., Rao, K.R., Pervaiz, S. and Brew, K. (1994a) Sequences of 2 highly divergent canine type-C lysozymes-implications for the evolutionary origins of the lysozyme alpha-lactalbumin superfamily. *Archives of Biochemistry and Biophysics*, **313**, 360-366.

Grobler, J.A., Wang, M., Pike, A.C.W. and Brew, K. (1994b) Study by mutagenesis of the roles of 2 aromatic clusters of aplha-lactalbumin in aspects of its action in the lactose synthase system. *J. Biol. Chem.*, **269**, 5106-5114.

Harata, K. and Muraki, M. (1992) X-Ray structural evidence for a local helix-loop transition in alpha-lactalbumin. *J. Biol. Chem.*, **267**, 1419-1421.

Hase, J., Kobashi, K., Nakai, N. and Onosaka, S. (1976) Binding of retinol-binding protein obtained from human urine with vitamin A derivatives and terpenoids. *J. Biochem.*, **79**, 373-380.

Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51-58.

Hill, R.L. and Brew, K. (1975) Lactose synthetase. *Adv. Enzymol. Rel. Areas. Mol. Biol.*, **43**, 411-489.

Hodam, J.R. and Creek, K.E. (1998) Comparison of the metabolism of retinol delivered to human keratinocytes either bound to serum retinol-binding protein or added directly to the culture medium. *Exp. Cell Res.*, **238**, 257-264.

Hong, C.Y. and Chia, K.S. (1998) Markers of diabetic nephropathy. *Journal of Diabetes and Its Complications*, **12**, 43-60.

Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H. and Kayser, H. (1987) Molecular-structure of the bilin binding-protein (BBP) from pieris-brassicae after refinement at 2.0-Å resolution. *J. Mol. Biol.*, **198**, 499-513.

Inaka, K., Kuroki, R., Kikuchi, M. and Matsushima, M. (1991) Crystal-structures of the apomutant and holomutant human lysozymes with an introduced Ca^{2+} binding-site. *J. Biol. Chem.*, **266**, 20666-20671.

Iyer, L.K. and Qasba, P.K. (1999) Molecular dynamics simulation of alpha-lactalbumin and calcium binding c-type lysozyme. *Protein Engineering*, **12**, 129-139.

Jancarik, J. and Kim, S.H. (1991) Sparse-matrix sampling - a screening method for crystallization of proteins. *J. Appl. Crystallogr.*, **24**, 409-411.

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building models in electron density maps & the location of errors in these models. *Acta Crystallogr.*, **A 47**, 110-119.

Kabsch, W. (1988) Automatic-indexing of rotation diffraction patterns. *J. Appl. Crystallogr.*, **21**, 67-71.

Kabsch, W. and Sanders, C. (1983) Dictionary of the protein secondary structure : pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, **22**, 2577-2637.

Kam, Z., Shore, H.B. and G.Feher. (1978) On the crystallisation of proteins. *J. Mol. Biol.*, **123**, 539-555.

Ke, H.M. (1997) Overview of isomorphous replacement phasing. *Methods in Enzymol.*, **276**, 448-461.

Khatra, B.S., Herries, D.G. and Brew, K. (1974) Some kinetic properies of human milk galactosyltransferase. *Eur. J. Biochem.*, **44**, 537-560.

Kronman, M.J. (1989) Metal-ion binding and the molecular conformational properties of alpha-lactalbumin. *Critical Reviews in Biochemistry and Molecular Biology*, **24**, 565-667.

Kronman, M.J., Sinha, S.K. and Brew, K. (1981) Characteristics of the binding of Ca^{2+} and other divalent metal-ions to bovine alpha-lactalbumin. *J. Biol. Chem.*, **256**, 8582-8587.

Kuroki, R., Kawakita, S., Nakamura, H. and Yutani, K. (1992) Entropic stabilization of a mutant human lysozyme induced by calcium-binding. *PNAS (USA)*, **89**, 6803-6807.

Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins*, **6**, 87-103.

Kuwajima, K., Harushima, Y. and Sugai, S. (1986) Influence of Ca^{2+} binding on the structure and stability of bovine alpha-lactalbumin studied by circular dichroism and nuclear magnetic resonance spectra. *Int. J. Peptide Protein Res.*, **27**, 18-27.

Kuwajima, K., Mitani, M. and Sugai, S. (1989) Characterization of the critical state in protein folding-effects of guanidine-hydrochloride and specific Ca^{2+} binding on the folding and specific Ca^{2+} binding on the folding kinetics of alpha-lactalbumin. *J. Mol. Biol.*, **206**, 547-561.

Lamzin, V.S. and Wilson, K.S. (1993) Automated refinement of protein models. *Acta Crystallogr,*. **D49**, 129-147.

Lamzin, V.S. and Wilson, K.S. (1997) Automated refinement for protein crystallography. *Methods in Enzymol.*, **277**, 269-305.

Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) Procheck - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283-291.

Lindberg, L.A., Sinkkonen, H., Poso, A.R., Tesfa, A.T. and Schroder, J. (1999) Production of monoclonal antibodies and enzyme immunoassay to bovine retinol-binding protein and determination

of retinol-binding protein serum levels and retinol concentrations in serum and liver in dairy cows before and after parturition. *Research in Veterinary Science*, **66**, 259-263.

Linzell, J.L. and Peaker, M. (1971) Mechanism of milk secretion. *Physiol. Rev.*, **51**, 564-597.

Lu, G. (1998) An approach for multiple alignment of protein structures. *in preparation.*

Malinovskii, V.A., Tian, J., Grobler, J.A. and Brew, K. (1996) Functional site in alpha-lactalbumin encompasses a region corresponding to a subsite in lysozyme and parts of two adjacent flexible substructures. *Biochemistry*, **35**, 9710-9715.

McPherson, A. (1982) *The preparation and Analysis of Protein Crystals*. John Wiley & Sons, New York.

McRee, D. (1993) *Practical Protein Crystallography*. Academic Press, London.

Mitra, A.K., Alvarez, J.O., GuayWoodford, L., Fuchs, G.J., Wahed, M.A. and Stephensen, C.B. (1998) Urinary retinol excretion and kidney function in children with shigellosis. *American Journal of Clinical Nutrition*, **68**, 1095-1103.

Monaco, H.L., Rizzi, M. and Coda, A. (1995) Structure of a complex of 2 plasma-proteins-transthyretin and retinol-binding protein. *Science*, **268**, 1039-1041.

Morrison, J.F. and Ebner, K.E. (1971a) Studies on galactosyltranferase. Kinetic investigations with N-acetylglucosamine as the galactosyl group acceptor. *J. Biol. Chem.*, **246**, 3977-3984.

Morrison, J.F. and Ebner, K.E. (1971b) Studies on galactosyltransferase. Kinetic investigations with glucose as the galactosyl group acceptor. *J. Biol. Chem.*, **246**, 3985-3991.

Musci, G. and Berliner, L.J. (1985a) Physiological roles of zinc and calcium-binding to alpha-lactalbumin in lactose biosynthesis. *Biochemistry*, **24**, 6945-6948.

Musci, G. and Berliner, L.J. (1985b) Probing different conformational states of bovine alpha-lactalbumin - fluorescence studies with 4,4'-bis[1-(phenylamino)-8- naphthalenesulfonate]. *Biochemistry*, **24**, 3852-3856.

Musci, G. and Berliner, L.J. (1986) Intramolecular distance measurements in alpha-lactalbumin. *Biochemistry*, **25**, 4887-4891.

Navaza, J. (1994) AMoRe : an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157-163.

Navaza, J. and Saludjian, P. (1997) AMoRe: An automated molecular replacement program package. *Methods in Enzymol.*, **276**, 581-594.

Naylor, H.M. and Newcomer, M.E. (1999) The structure of human retinol-binding protein (RBP) with its carrier protein transthyretin reveals an interaction with the carboxy terminus of RBP. *Biochemistry*, **38**, 2647-2653.

Newcomer, M.E. (1995) Retinoid-binding proteins - structural determinants important for function. *FASEB Jl*, **9**, 229-239.

Newcomer, M.E., Jones, T.A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L. and Peterson, P.A. (1984) The 3-dimensional structure of retinol-binding protein. *EMBO Jl*, **3**, 1451-1454.

Ottonello, S., Maraini, G., Mammi, M., Monaco, H.L., Spadon, P. and Zanotti, G. (1983) Crystallisation and preliminary X-ray data of human-plasma retinol-binding protein. *J. Mol. Biol.*, **163**, 679-681.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In Carter, C.W.J. and Sweet, R.M. (eds.), *Methods Enzymol.* Academic Press, New York, Vol. 276, pp. 307-326.

Permyakov, E.A., Grishchenko, V.M., Kalinichenko, L.P., Orlov, N.Y., Kuwajima, K. and Sugai, S. (1991) Calcium-regulated interactions of human alpha-lactalbumin with bee venom melittin. *Biophys. Chem.*, **39**, 111-117.

Permyakov, E.A., Ostrovsky, A.V. and Kalinichenko, L.P. (1988) Kinetics of dissociation of the alpha-lactalbumin complexes with Ca^{2+} and Mg^{2+} Ions. *Biophysica*, **33**, 413-416.

Permyakov, E.A., Yarmolenko, V.V., Kalinichenko, L.P., Morozova, L.A. and Burstein, E.A. (1981) Calcium-Binding to Alpha-Lactalbumin - Structural Rearrangement and Association Constant Evaluation By Means of Intrinsic Protein Fluorescence Changes. *Biochem. Biophys. Res. Commun.*, **100**, 191-197.

Pike, A.C.W. (1995) Structure-function relationships of alpha-lactalbumin. PhD. thesis, University of Bath, Bath.

Pike, A.C.W., Brew, K. and Acharya, K.R. (1996) Crystal structures of guinea-pig, goat and bovine alpha-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure*, **4**, 691-703.

Powell, J.T. and Brew, K. (1976) A comparison of the interactions of galactosyl transferase with a glycoprotein substrate (ovalbumin) and with alpha-Lactalbumin. *J. Biol. Chem.*, **18**, 1771-1776.

Preels, J.-P., Bell, J.E., Schindler, M., Castellino, F.J. and Hill, R.L. (1979) Involvement of histidine-32 in the biological activity of alpha-lactalbumin. *American Chemical Society*, **18**, 1771-1776.

Quadro, L., Blaner, W.S., Salchow, D.J., Vogel, S., Piantedosi, R., Gouras, P., Freeman, S., Cosma, M.P., Colantuoni, V. and Gottesman, M.E. (1999) Impaired retinal function and vitamin A availability in mice lacking retinol-binding protein. *EMBO Jl*, **18**, 4633-4644.

Rao, K.R. and Brew, K. (1989) Calcium regulates folding and disulfide-bond formation in alpha-lactalbumin. *Biochem. Biophys. Res. Commun.*, **163**, 1390-1396.

Rask, L., Anundi, H. and Peterson, P.A. (1979) The primary structure of the human retinol-binding protein. *FEBS lett.*, **104**, 55-58.

Redinbo, M.R. and Yeates, T.O. (1993) Structure determination of plastocyanin from a specimen with a hemihedral twinning fraction of one-half. *Acta Crystallogr.* **D49**, 375-380.

Ren, J.S., Stuart, D.I. and Acharya, K.R. (1993) Alpha-lactalbumin possesses a distinct zinc-binding site. *J. Biol. Chem.*, **268**, 19292-19298.

Rhodes, G. (1993) *Crystallography made crystal clear*. Academic Press, London.

Rodgers, D.W. (1994) Cryocrystallography. *Structure*, **2**, 1135-1140.

Ronne, H., Ocklind, C., Wiman, K., Rask, L., Obrink, B. and Peterson, P.A. (1983) Ligand-dependent regulation of intracellular protein-transport-effect of vitamin-A on the secretion of the retinol-binding protein. *J. Cell Biol.*, **96**, 907-910.

Rosales, F.J., Jang, J.T., Pinero, D.J., Erikson, K.M., Beard, J.L. and Ross, A.C. (1999) Iron deficiency in young rats alters the distribution of vitamin A between plasma and liver and between hepatic retinol and retinyl esters. *Journal of Nutrition*, **129**, 1223-1228.

Rossmann, M.G. (ed.) (1972) *The Molecular Replacement Method.* Gordon and Breach, New York.

Sarkar, G. and Sommer, S.S. (1990) The Megaprimer method of site-directed mutagenesis. *Biotechniques*, **8**, 404-407.

Sawyer, L., Papiz, M.Z., North, A.C.T. and Eliopoulos, E.E. (1985) Structure and function of bovine beta-lactoglobulin. *Biochem. Soc. Trans.*, **13**, 265-266.

Schaer, J.J., Milos, M. and Cox, J.A. (1985) Thermodynamics of the binding of calcium and strontium to bovine alpha-lactalbumin. *FEBS lett.*, **190**, 77-81.

Schindler, M., Sharon, N. and Preels, J.P. (1976) Reversible inactivation of lactose synthase by the modification of His 32 in human alpha-Lactalbumin. *Biochem. Biophys. Res. Commun.*, **69**, 167-173.

Shechter, Y., Patchornik, A. and Burstein, Y. (1973) *Biochemistry*, **12**, 3407-3413.

Shewale, J.G., Sinha, S.K. and Brew, K. (1984) Evolution of alpha-lactalbumins the complete amino-acid-sequence of the alpha-lactalbumin from a marsupial (Macropus-Rufogriseus) and corrections to regions of sequence in bovine and goat alpha-lactalbumins. *J. Biol. Chem.*, **259**, 4947-4956.

Smith, L.J., Alexandrescu, A.T., Pitkeathly, M. and Dobson, C.M. (1994) Solution structure of a peptide fragment of human alpha-lactalbumin in trifluoroethanol - a model for local-structure in the molten globule. *Structure*, **2**, 703-712.

Smith, S.G., Lewis, M., Aschaffenburg, R., Fenna, R.E., Wilson, I.A., Sundaralingam, M., Stuart, D.I. and Phillips, D.C. (1987) Crystallographic analysis of the 3-dimensional structure of baboon alpha-Lactalbumin at low resolution-homology with lysozyme. *Biochem. J.*, **242**, 353-360.

Song, H., Inaka, K., Maenaka, K. and Matsushima, M. (1994) Structural changes of active site cleft and different saccharide binding models in human lysozyme co-crystallised with hexa-N-acetyl-chitohexaose at pH 4.0. *J. Mol. Biol.*, **244**, 522-540.

Soprano, D.R. and Blaner, W.S. (1994) *The retinoids*. Raven Press, New York.

Steinrauf, L.K. (1998) Structures of monoclinic lysozyme iodide at 1.6 angstrom and of triclinic lysozyme nitrate at 1.1 angstrom . *Acta Crystallogr.*, **D54**, 767-779.

Stuart, D.I., Acharya, K.R., Walker, N.P.C., Smith, S.G., Lewis, M. and Phillips, D.C. (1986) Alpha-lactalbumin possesses a novel calcium-binding loop. *Nature*, **324**, 84-87.

Stuart, D.I., Levine, M., Muirhead, H. and Stammers, D.K. (1979) The catalytic structure of cat pyruvate kinase at a resolution of 2.6 A. *J. Mol. Biol.*, **134**, 109-142.

Sundaram, M., Sivaprasadarao, A., DeSousa, M.M. and Findlay, J.B.C. (1998) The transfer of retinol from serum retinol-binding protein to cellular retinol-binding protein is mediated by a membrane receptor. *J. Biol. Chem.*, **273**, 3336-3342.

Tsuge, H., Ago, H., Noma, M., Nitta, K., Sugai, S. and Miyano, M. (1992) Crystallographic studies of a calcium-binding lysozyme from equine milk At 2.5 Å resolution. *J. Biochem.*, **111**, 141-143.

Vanderheeren, G. and Hanssens, I. (1994) Thermal unfolding of bovine alpha-lactalbumin-comparison of circular-dichroism with hydrophobicity measurements. *J. Biol. Chem.*, **269**, 7090-7094.

Veillon, C. and Vallee, B.V. (1978) Atomic spectroscopy in metal analysis of enzymes and other biological material. *Methods in Enzymol.*, **54**, 446-484.

Vicente, C.P., Fortuna, V.A., Margis, R., Trugo, L. and Borojevic, R. (1998) Retinol uptake and metabolism, and cellular retinol binding protein expression in an in vitro model of hepatic stellate cells. *Molecular and Cellular Biochemistry*, **187**, 11-21.

Wang, M., Scott, W.A., Rao, K.R., Udey, J., Conner, G.E. and Brew, K. (1989) Recombinant bovine alpha-lactalbumin obtained by limited

proteolysis of a fusion protein expressed at high levels in Escherichia coli. *J. Biol. Chem.*, **264**, 21116-21121.

Yeates, T.O. (1997) Detecting and overcoming crystal twinning. *Methods in Enzymol.*, **276**, 344-358.

Zanotti, G., Berni, R. and Monaco, H.L. (1993) Crystal-structure of liganded and unliganded forms of bovine plasma retinol-binding protein. *J. Biol. Chem.*, **268**, 10728-10738.